Utilising volunteered geographic information to assess resident’s flood evacuation shelters. Case study: Jakarta

A.N.L. Kusumo a, b, *, D. Reckien b, J. Verplanke b

a Jakarta City Planning Department, Jakarta Capital City Government, Jakarta, Indonesia
b Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT

Research on disaster response frequently uses volunteered geographic information (VGI), due to its capability to provide near real-time information during and after a disaster. It is much less commonly used in spatial planning related to disaster management. However, VGI appears to have considerable potential for use in spatial planning and offers some advantages over traditional methods. For example, VGI can capture residents’ preferences in a much faster, more timely, and more comprehensive fashion than is possible with, for example, questionnaires and surveys. This research investigates the usefulness of VGI for planning flood evacuation shelters. Using Jakarta, Indonesia, as a case study, we use VGI to capture the locations of flood evacuation shelters based on residents’ preferences during flood periods in 2013-2014 and 2014-2015 and compare these with the locations of official shelters. Floods frequently affect Jakarta and the city administration uses VGI in flood emergency responses. Moreover, Jakarta has been identified as having the largest number of active Twitter users among cities worldwide. Thus, Jakarta is an appropriate place to study the use of VGI for planning evacuation shelters. VGI generated by Twitter users was used to identify the shelter locations preferred by Jakarta residents, and more precisely the flood evacuees. Of 171,046 tweets using keywords relating to flood evacuation, the content of 306 tweets indicated that they had been sent from inside or near evacuation shelters. The spatial pattern showed that those tweets were sent from 215 locations, mostly located near flooded areas. The analysis further showed that 35.6% of these shelter locations preferred by residents intersected with the locations of official evacuation shelters. As a general conclusion, our study demonstrates the advantages of using VGI for spatial planning, which mainly relates to the ease of capturing community preferences over a large area.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Background

Social media have been acknowledged as key communication channels during situations of crises and disasters. Authorities and emergency response agencies use social media as a valuable source of information as well as a useful platform for the rapid delivery of information (Kreiner & Neubauer, 2012). It aids disaster responses and management before and during an event, e.g., by sending alerts, identifying critical needs, and focusing responses (Carley, Malik, Landwehr, Pfeffer, & Kowalchuck, 2015). Residents use social media to request help during a crisis, to share views and experiences on current topics, and to criticise responses by government agencies and other organisations (Takahashi, Tandoc, & Carmichael, 2015). Social media thereby allows residents to take part in the disaster response and management, as well as in other participatory processes (Goodchild, 2007).

Using social media, residents often voluntarily provide data on their own locations—known as Volunteered Geographic Information (VGI). VGI is made accessible by harnessing tools to assemble and disseminate these geographic data (Goodchild, 2007). VGI can aid disaster response and management (Takahashi, Tandoc, & Carmichael, 2015) by increasing the speed of interaction between victims and relief organisations. Some social media applications provide VGI, including Twitter, Flickr and Open Street Map which...
have a geolocation feature (Schade et al., 2011). All these are used in situations requiring near real-time disaster response and management (Carley et al., 2015; Kreiner & Neubauer, 2012; Takahashi et al., 2015) before and during a disaster. However, VGI is rarely used in addressing spatial planning problems arising after an extreme weather event or disaster, relating to future disaster mitigation and/or climate change adaptation. VGI has considerable potential for such applications and appears to offer advantages over traditional methods. For example, VGI could be used for the planning of evacuation shelters, enabling residents’ knowledge and preferences for shelters to be captured in a much faster, more timely, and comprehensive fashion than is possible with, for example, questionnaires and surveys.

The advantages of VGI are made use of in other case studies related to urban planning (Brabham, 2009), for instance with regard to people’s participation in validating land use/cover in urban areas. Similarly, the endorsement of VGI by the government is also seen in many cases. Governments use VGI as a platform to accommodate reporting from the community. However, the use of VGI also has challenges (Johnson & Sieber, 2013), e.g. such as how to assure the validity of the data provided by the community.

This research investigates the advantages and disadvantages of using VGI for capturing residents’ preferences for (official and unofficial) flood evacuation shelters and explores the usefulness of this information for urban planning. The research is motivated by the observation that official evacuation shelters provided by the authorities are not frequently used by the residents—a fact that is notified within the planning department of Jakarta for a long time (one of the authors is working there) and recognized by research since at least 30 years (Perry, 1979). Different factors determine people’s use of official and non-official evacuation shelters (Stein, Duenas-Osorio, & Subramanian, 2010). According to Rahman, Mallick, Mondat, and Rahman (2014) these comprise drainage capacity, soils, technical feasibility, capacities, basic facilities, environmental impact, accessibility, land availability, and maintenance.

We address the objective mentioned above using Jakarta, Indonesia, as a case study. Flooding affects Jakarta on an almost annual basis. Moreover, authorities in Jakarta are forerunners in the use of social media, using Twitter to coordinate flood emergency responses. Jakarta itself has been identified as the city with the largest number of active Twitter users worldwide in a case study undertaken by Semiocast (2012). And Indonesia is recognized as the country with the most Twitter users, judged by the number of posted tweets in June 2012. A tweet is any message posted to Twitter and comprises five administrative cities on the mainland and one administrative coastal region, covering the marine area and islands to the north of the mainland. Only the five administrative cities with their 267 sub-districts are considered in the research.

Flooding has been an issue in Jakarta since the colonial era. Based on historical records, major floods occurred in 1654, 1872, 1909, and 1918 (Team Mirah Sakethi, 2010). Currently, floods happen nearly every year. In 2002 and 2007 Jakarta was severely affected by two ‘50-year’ floods (i.e., floods with a statistical probability of occurring once every 50 years). According to Firman, Surbakti, Idroes, and Simarmata (2011), the 2002 flood covered about one-fifth of Jakarta’s total area. Hundreds of thousands of people were made homeless, 68 persons were killed, 190,000 people suffered from flood-related illnesses, and about 422,300 had to be evacuated. Flood losses were estimated at nine trillion Indonesian Rupiahs (USD 998 million) (Akmalah & Grigg, 2011).

The government of Jakarta province has incorporated the use of VGI in flood emergency responses, through an online resource known as “Peta Jakarta” provided by the Jakarta Government’s Regional Disaster Management Agency (BPBD), in collaboration with the SMART Infrastructure Facilities, and Twitter. Peta Jakarta (@petajkt) is a system that utilises social media to gather, sort, and display information about flood events in Jakarta in real time (BPBD Jakarta, 2015). Jakarta’s residents can also use the platform to report conditions in their neighbourhood. Residents’ reports may include information on flood events, evacuation processes, traffic jams, and other flood-related problems.

One of the reasons behind the development of the Peta Jakarta was the enormous volume of VGI being generated by residents of Jakarta through their use of social media. Based on research by Semiocast (2012), Jakarta holds the first place as the most active city using Twitter among all cities worldwide, based on the number of posted tweets in June 2012. A tweet is any message posted to Twitter which may contain photos, videos, links and up to 140 characters of text (see http://www.twitter.com). Semiocast analysed a sample of 10.6 billion public tweets posted by 517 million Twitter users. More than 2% of posted tweets came from Jakarta (Fig. 2).

2. Data and methods

Our study adopts the following procedure to determine the usefulness of VGI data for planning evacuation shelters. First, we determine the location of Twitter users in or near evacuation shelters; second, the spatial pattern of users in or near evacuation shelters, and third, their preferences regarding the use of these shelters. To this end, the study employs secondary as well as primary data, which is analysed using various methods (Fig. 3). In the following sections, we describe data retrieval, processing and analysis in more detail, following the three steps listed above.

2.1. Data retrieval and processing

This study used both primary and secondary data. Primary data was collected to determine the preferences of individuals (Twitter users) regarding shelters. However, due to a low response rate of the questionnaire aimed to determine information on location preferences also secondary data were used for preference elicitation. Secondary data comprise information on Twitter use, land use categories, and other spatially explicit GIS data from the statistical offices.

2.1.1. Twitter data

Twitter data was retrieved from the “Digital On-line Life and You” (DOLLY) archive—the massive database of geolocated tweets developed by “The Floating Sheep Team” that allows for
real-time research and analysis. Building on top of existing open source technology, the Floating Sheep team has created a back-end that ingests all geotagged tweets (~8 million a day), does basic analysis, indexing and geocoding to allow real-time search throughout the entire database (3 billion tweets since Dec 2011 (Zook, Graham, Shelton, Stephens, & Poorthuis, 2016)) using the Twitter Application Program Interface (API). API is a set of routines, protocols, and tools for building software applications. According to Durahim and Coşkun (2015), Twitter API is the method most commonly used to gather data from Twitter. We requested geotagged information and received a random 1% of tweets of all geotagged tweets with and without keywords in Jakarta. Spatially, the study includes data within the bounding box of Jakarta, i.e., between 5.20166°S and 6.37248°S and 106.390266°E and 106.974274°E. However, this bounding box does not exactly represent the administrative boundary of Jakarta Province. Therefore, the data was clipped (Fig. 6.).

Temporally, the study restricts itself to tweets that were sent

---

**Fig. 1.** Map of Jakarta Province, which located in 5°19’12”–6°23’54” S, 106°22’42”–106°58’18” E. Sources: Google (2015), Jakarta Capital City Government (2014).

**Fig. 2.** Top 20 cities by the number of posted tweets. Source: Semiocast (2012).
during two periods of severe flooding, categorised by BPBD Jakarta (2015) as an event demanding emergency response. The flood periods that were analysed include December 2013 to March 2014 and from December 2014 to March 2015 (Fig. 4.). The retrieved data contains one or more relevant hashtags and keywords, i.e. #banjir, #banjirjkt, #evakuasi, #logistik, #relawan, pengungsi, korban, @petajkt. These hashtags and keywords were determined by the authors (the first author is a native Indonesian living in Jakarta) in cooperation with Mrs Fitria Sudirman from Peta Jakarta (http://www.petajakarta.org). Peta Jakarta was the official operator consultant of DKI Jakarta who managed Twitter reports from the community. The retrieved data, therefore, represents locations of Twitter users in Jakarta Province talking about the flood during the flood.

2.1.2. Data of community preferences on evacuation shelters

This research generated primary data using questionnaires sent to select Twitter users to capture residents’ preferences regarding shelter locations. Those selected were people who sent information through Twitter related to evacuation shelter locations and who were identified as evacuees (see Fig. 5). The questionnaire was designed in Survey Monkey. The link to the Survey Monkey questionnaire was sent to the respondents through their Twitter account. The questionnaire was a mix of open and closed questions regarding respondents’ use of evacuation shelters during previous flood events (see Appendix B).

2.1.3. Secondary data related to evacuation shelters

Secondary data related to the distribution of official evacuation shelters was collected from BPBD and Jakarta Spatial Planning Department (DPK). The data was used to compare the community preferences for shelter locations with the locations of official evacuation shelters. Table 1 lists all secondary data types and sources.

2.2. Data analysis and methods

2.2.1. The location of those Twitter users in or near evacuation shelters

We employed a two-fold content analysis. First, we conducted a validity or relevance check of downloaded geolocated Twitter data; second, we undertook a content classification analysis of relevant tweets (Fig. 5.).

The validity check determined the relevance of tweets to the topic of evacuation shelters using Atlas.ti Software. The aim of this check was to filter the tweets that were contextually relevant to the

Fig. 3. Analytical framework. Source: Authors’ elaboration.

Fig. 4. Twitter impressions during the flood period of 2014/2015 (Holderness & Turpin, 2015).
Fig. 5. Flowchart of the analysis of Twitter data. Source: Authors’ elaboration.

Fig. 6. Source locations of 60,517 tweets related to the floods in 2013/2014 and 2014/2015 in Jakarta Province. Source: Authors’ elaboration.

Table 1
Types and sources of secondary data used. Source: Authors’ elaboration.

| Data                        | Year           | Source                                                   |
|-----------------------------|----------------|----------------------------------------------------------|
| Road                        | 2014           | Jakarta City Planning Department                         |
| Provincial boundary         | 2014           | Jakarta City Planning Department                         |
| Flood area (flooded to a depth of more than 10 cm) | 2002, 2007, 2014/2015 | Jakarta Disaster Management Agency, Jakarta Planning Board |
| Land use category           | 2014           | Jakarta City Planning Department                         |
flood evacuation shelters in Jakarta. We used a deductive approach, which starts with predefined keywords regarded as relevant by an expert (in this case the researcher). For instance, #banjir (flood) and #evakuasi (evacuation), and related words derived from the same root (Holderness & Turpin, 2015). Atlas.ti was also used in the identification of the locations mentioned by the twitter user (Appendix E).

Approximately 135,885 tweets were posted between December 2013 to March 2014 and 35,160 tweets from December 2014 to March 2015. Data clipping yielded 60,517 tweets that were sent within the administrative boundary of Jakarta Province (Fig. 6.).

2.2.2. The spatial pattern of users in or near evacuation shelters

People who send tweets with geolocation data indirectly allow their location to be disclosed, as it can be collected from the geo-tagged tweet. However, this occurs with a degree of inaccuracy. The Twitter location does not necessarily match the actual location of the person tweeting. To deal with this inaccuracy tweets can be grouped into spatial units, with the degree of accuracy required for data analysis.

The location of the Twitter data is obtained as a point feature, which can be ascribed to administrative boundaries or any other spatial unit (Pootrhuis et al., 2014). The selection of the appropriate type of spatial unit is highly dependent on the purpose of the research, which in this case required using the smallest spatial unit available. Other scholars have used buildings or land use categories when representing shelter sites (Chang & Liao, 2014; Gall, 2004; Kar & Hodgson, 2008).

It is necessary to consider the positional accuracy in order to select the most appropriate spatial unit. Many studies have investigated the accuracy of VGI (Goodchild & Li, 2012). For example, Haklay (2010) compared the data of Open Street Map with survey data, which showed an average deviation between the geolocation and actual location of 6 m. In this study, the accuracy assessment was conducted using the content of tweets as control data (Comber et al., 2013). Using purposive sampling, i.e., tweets that clearly mentioned the location of the person in the text, we tested the distance between the geolocation and the actual location. The mean distance was used as a basis for choosing the spatial unit.

2.2.3. The preferences of using these shelters among Twitter users

Analysis of residents’ preferences of evacuation shelters was conducted using a qualitative, comparative method that considered the following data:

- Land use categories
- Evacuation shelters classified as official shelters by the planning boards of Jakarta
- Results of the questionnaire survey

In this study we used land use categories to categorize and characterize the location and preferences of evacuation shelters, because this is a categorization that is useful for urban planning and management. However, other studies might characterize evacuation shelters based on other features, such as height or age of the building, which may also be important characteristics. Therefore, to supplement the survey information, we determined the number of people tweeting from shelter sites located in each land use category, and how many of them are at or near official evacuation shelters. This way, we could determine patterns of use of unofficial shelters (hereafter residents’ evacuation shelters’) and official evacuation shelters.

3. Results

3.1. The location of those twitterers users in or near evacuation shelters

Our analysis shows that 306 tweets could be recognized as coming from evacuation shelter locations in 2013–2014 and 2014–2015. By overlaying the tweet data on the flood maps of 2013–2014 and 2014–2015, we could analyse the spatial distribution of tweet locations.

The locations of tweets from 2013–2014 were clustered in the central area of Jakarta which was most severely affected by the flood event (Fig. 7a), i.e., the Kampung Pulo neighbourhood in Jatinegara District. Some areas in Kampung Pulo are located in the floodplain of the biggest river in Jakarta, the Ciliwung River. For many years the floodplain has been occupied by slums inhabited by low-income residents. Like many other slums (Kit, Lüdeke, & Reckien, 2011), Kampung Pulo occupies an area where the risk of flooding is high (Khomarudin, Suwarsono, Ambarwati, & Prabowo, 2014). In response to the higher risk of Kampung Pulo in comparison with other locations in Jakarta, voluntary organisations have donated aid to set up evacuation shelters.

In comparison with the 2013–2014 flood event, the 2014–2015 flood area was smaller. Similarly, the impact of flood event in 2014–2015 was also less than 2013–2014, which is most likely the reason why there are fewer tweets during the 2014–2015 event. Moreover, according to BPBD Jakarta (2015), the time of inundation per month from December 2013 to March 2014 was 4, 20, 20 and 8 days, respectively; whereas from December 2014 to March 2015 the days of inundation per month comprised 5, 2, 7, and 4 days, respectively. Consequently, the flood in 2013–2014 was larger than in 2014–2015, which is probably another reason for the smaller number of tweets referring to evacuation shelters in 2014–2015.

There were only 48 tweets mentioning evacuation shelters in 2014–2015, and 258 tweets in 2013–2014. The tweets were dispersed in several locations throughout the city, rather than being concentrated in the flood area, as was the case in 2013–2014 (Fig. 7b.).

3.2. The spatial pattern of users in or near evacuation shelters

Of the 306 tweets related to evacuation shelters, 86 mentioned the detailed location of the shelter. These 86 tweets were used as a sample to calculate the mean distance between the geolocation and the actual location based on what was mentioned in the tweet. The average distance between the geolocation and the actual location was 188.28 m, ranging from 0 to 5405 m (Appendix C).

Since the distances between the geolocation and the actual location were quite large, each tweet point was converted so that its location was recorded as being within one of the spatial units selected for further analysis. This procedure helped to group tweets from different geolocations that mentioned the same location in the text. There are several types of spatial units which could accommodate the mean distance of 188.28 m. Buffers, hexagons and land use zones are some of them.

The first spatial unit considered was land use zones (Fig. 8.), i.e., polygons ranging in size from 3 m² to 3 km². However, if we directly converted the point of the tweets into land use zone locations using spatial joining without buffering the result could be misleading when the evacuation shelter lays close to a border between different land use categories. Tweets that are actually posted close to but not from within evacuation shelters (or have been posted from within but are recorded as outside due to transmission speed) could already belong to a different land use zone. The right-hand
Hexagon tessellations are commonly used to simplify point data (Raposo, 2013). According to Birch, Oom, and Beecham (2007), hexagon tessellations have several advantages, e.g., over regular square grids. The width of hexagons has conformity, meaning that using hexagons allows covering an area without overlapping neighbourhoods. The nearest neighbourhood is more symmetrical in a hexagonal tessellation compared to a rectangular grid since the length of each line of the hexagon contour is equal. Data can also be visualised more clearly.

In this study, the hexagon used had equal length sides of 200 m, based on the 188.28 m mean distance between geolocations and actual locations. Following conversion, the 306 tweets from residents’ evacuation shelters (points) were located in 215 hexagons representing residents’ shelter sites (Fig. 9).

3.3. The preferences of using these shelters among Twitter users

Twitter users’ preferences for evacuation shelters were analysed
using results from the questionnaire combined with the spatial pattern of the hexagons. The responses to the questionnaire were generated from people identified as evacuees—in contrast to e.g. volunteers—from Twitter content. In total 269 relevant tweets sent from 184 Twitter accounts were identified as having been posted by evacuees.

Fig. 9b shows a tweet sample of an evacuee. In some cases, evacuee and volunteer could not be differentiated, for example several tweets mentioned only “I am at evacuation shelter”. This type of respondents and other people who only gave information about the evacuation shelters have been classified as “Other people”.

In order to get information on the location preferences of evacuees we sent the link to a questionnaire survey to the 184 accounts of evacuees. Several challenges were encountered in getting the feedback from the respondents. First, people tended to ignore the questionnaire. In this case, we sent several reminders over the course of 5 days. After six reminders, only three accounts provided feedback on their preferences. The low number of returned questionnaires lends support to one of our initial arguments made above, i.e. that obtaining information about preferences of shelter use is difficult to obtain with traditional data collection methods (Appendix D). Another challenge was the limited number of characters (140) in Twitter, which restricted the information we could provide to introduce the research.

The distance of evacuation shelters from the flood area is one of the criteria determining people’s preferences in selecting an evacuation shelter (ARC (American Red Cross), 2002; FEMA, 2015; Kar & Hodgson, 2008). Based on the mean distance of each resident’s evacuation shelter site to the nearest flood-prone area, we found that the shelter sites in Jakarta were mostly located within flood-prone areas. About 60% of residents’ evacuation shelter sites (hexagons) in Jakarta were within an area flooded in 2013–2014 and/or 2014–2015 (Fig. 10.). One possible reason is that people look for a safe location near their home. For example, some evacuees take refuge on the second floor of a neighbour’s house.

Kongsomsaksakul, Yang, and Chen (2005) mention that the ideal location for an evacuation shelter is outside the flooded area, but within 1 km distance of it. In the case of Jakarta, about 31% of residents’ shelter sites (hexagon) fulfilled this criterion—a fact which was also confirmed by answers to the questionnaire (Appendix D)—indicating that the evacuation shelters used by respondents were located very close to the flooded area (i.e., between 200 m and 1 km distance).

Respondents mentioned that the main reason for their choice of evacuation shelters was accessibility, safety from flood and proximity to their home. Accessibility is clearly an important factor for people considering where to go when they are evacuating (CCCMCluster, 2014; Tai, Lee, & Lin, 2010). One of the respondents added that the proximity of the evacuation shelter to their house allowed them to monitor conditions at their home at any time (Appendix D). The average distance from the shelter to the respondent’s houses was between 200 and 300 m. However, one respondent mentioned that the shelter was located 2 km away...
from his house. He added that this shelter, provided by a religious organisation, was the closest one he could reach.

All the respondents stated that they reached the evacuation shelter on foot (Appendix D)—none of them used a car, motorbike or public transportation. This result is consistent with research by Chang and Liao (2014), who found that people chose to walk rather than drive to evacuation shelters. On the other hand, Kar and Hodgson (2008) assumed that people usually travel by car to the shelters. The mode of transport used to reach the evacuation shelter is also dependent on the type and impact of the flood. For example, one of the tweets in this study said that the author had used a rescue boat to reach the evacuation shelter (Appendix E).

According to our findings, however, the walking distance from the shelter location should be the main factor considered in planning evacuation shelters in Jakarta.

Additionally, we identified the land use types where residents’ evacuation shelters were located on the land use map. It was not always possible to locate the shelter precisely on a particular land use type since the spatial unit of the shelter was a hexagon, and one hexagon could contain several types of land use. Thus our analysis could only provide a general overview of the types of land use preferred for the location of the evacuation shelters.

Our results show that the residents’ evacuation shelters were mostly located in open/green spaces. This is also the land use type most frequently used by official evacuation shelters provided by the government. Some of these shelters were tents, rather than permanent buildings, e.g. the central evacuation shelter of Jakarta. This shelter was located in one of the largest open/green spaces in the city, where it also served as a logistics and coordination centre. Another tent shelter was located at the train station erected in a park that is part of the buffer zone alongside the railway. The second most common land use type for the location of shelters was residential land. People found a shelter near their house, often provided by neighbours or families. Several tweets sent by volunteers indicated that they had provided their house as a temporary shelter for their neighbours. Offices were the third most common type of land use used for shelters, at sites mainly chosen by residents. Based on the tweets, several shelters were located in the basements of office buildings. Fig. 11 shows the distribution of land use types used for residents’ evacuation shelters.

3.4. Comparison of official and residents’ evacuation shelters

By comparing residents’ shelter sites and official evacuation shelters, we could obtain an overview of people’s use of official evacuation shelters. Based on the spatial plan, there were 2645 official evacuation shelters at the time our study was conducted. The spatial unit that shows the location of official evacuation shelters is based on land use zones, which is different from the hexagons that identify the site of residents’ evacuation shelters. Hence, it is difficult to determine whether or not the residents used the official evacuation shelters.

To deal with this issue, we analysed the spatial join between official and residents’ shelter sites, assuming that the intersection of an official evacuation shelter with the hexagon of a residents’ evacuation shelter site points to the use of an official evacuation shelter. Overall, 35.6% of residents’ shelter sites intersected with official evacuation shelters (Fig. 12.).

Based on the analysis above, we could determine the land use type of official evacuation shelters that was most used by residents. The results show that the intersections between residents’ evacuation shelters and formal shelters occurred most frequently on the sites of educational centres: 53.5% of all cases where formal and residents’ shelters intersected were educational centres. Green space was the second most common land use type where formal and informal shelters intersected, accounting for 29.6% of the total. The remainder of cases were religious, health and sports centres.

Education centres are used by the people in their daily lives. Therefore, people are aware of their locations. Moreover, there are education centres in every neighbourhood in Jakarta, including about 2700 public schools and 4100 private schools of all levels (DKI Jakarta Province Government, 2015).

Our results suggest that the sparse usage of official evacuation shelters relate to people’s lack of awareness. In response to the questionnaire (Appendix D), the respondents mentioned that they never used the official evacuation shelters because they did not know they existed. One of them mentioned that official evacuation shelters are only set up when the flood event occurs. Moreover, if
there were no evacuation shelters provided by the government, respondents indicated to prefer to go to the house of a neighbour or family member that was safer than their house. Two of the respondents mentioned that they often use the sites of their daily activities. One of these shelters was a religious centre and the other the home of family members. Location preferences were strongly influenced by the familiarity that evacuees felt with the shelter site. Therefore, our analysis also shows that social networks in a community are very important during disaster situations. Using VGI data, our analysis provides important insights into the distribution and patterns of the use of informal residents’ evacuation shelters and official government shelters.

4. Discussion

Our analysis has shown that using VGI for disaster planning and management has more potential applications than just capturing real-time information (Erskine & Gregg, 2012). In this study, the focus was on information provided on evacuation shelters during flood events. As a result, the location of evacuation shelters used by evacuees during floods could be identified. We showed that the general pattern of the evacuation shelter locations captured using VGI could provide important inputs for evacuation shelter planning. Moreover, this research covers an area of 662 km². Residents’ preferences regarding evacuation shelters in this large area could be determined in a relatively short time. By only using secondary data of VGI, we were also able to map the distribution of evacuation shelters without conducting a field survey. Obtaining the same information using traditional data collection methods would be time-consuming and costly (e.g., Mooney, Sun, & Yan, 2011). Our study shows that VGI has potential as a cost-effective substitute for traditional data collection methods.

Another benefit of using VGI, and specifically Twitter datasets, is the easiness in selecting and accessing the information. The analysis of Twitter content uncovered various types of information...
related to the evacuation shelters. Firstly, VGI analysis could identify different types of users based on the content of the tweets. In this research, Twitter users were classified as volunteers, representatives of government agencies and NGOs, and evacuees. Evacuees were then selected for a questionnaire survey on preferences regarding evacuation shelters. Thus VGI was also a platform for identifying survey respondents as a sample of the population.

Moreover, through content analysis, we were also able to identify the time frame when people had tweeted based, for example, on the use of present verb tenses in the sentences. Thus, we could identify who was in an evacuation shelter at the time they sent the tweet. Using this type of information, we could pinpoint the location of evacuation shelters more accurately.

However, content analysis of tweets can be time-consuming and potentially reduce the advantage of time-efficiency. Choosing proper keywords is therefore key, as are adjustments in keywords in the iterative process of content analysis. Several factors need to be considered. First, in filtering the content of the tweets, synonyms of each keyword should be considered. Some people use another word with the same meaning. Slang words should also be included, especially when the users are young people. Moreover, the adjectives, verbs, and nouns formed from the same root word all need to be considered. Another factor to consider is the use of metaphors. Keywords are influenced by the characteristics of each language, and the same keyword could have many connotations. All these aspects were taken into account in the content analysis of tweets for our study. However, despite the care taken to guarantee consistent outputs in the content analysis, unorthodox use of language and particularly the use of metaphors could potentially lead to the inclusion of tweets whose content is not relevant to the topic under consideration—a possibility that also in our case cannot be ruled out completely.

There are other drawbacks of using VGI. The problem of geographic inaccuracy is currently a concern of many researchers. It was also an issue in this study. The accuracy assessment found a considerable deviation between the geolocation supplied with the Twitter data and the actual location mentioned in the content of the tweets. There are many possible explanations for this. One possible reason is that people tweet while moving; this would affect the geolocation, particularly if a time interval elapsed before the tweet was sent. It is also possible that people send tweets about their experiences in the shelter after moving away from the location.

Poorhuis, Zook, Shelton, Graham, and Stephens (2014) argue that different technologies of geotagging have issues with accuracy. The various types of GPS and Wi-Fi also influence the accuracy of geotagging. However, based on this analysis, the advantages of using VGI data outweigh potential disadvantages.

Our study suggests that VGI can be used in the planning of evacuation shelters using the feedback from targeted respondents. Evacuees were asked to answer a questionnaire survey sent to their Twitter account. However, we received only three responses out of a total of 184 accounts to which the questionnaire was sent. When people were confronted with a long list of questions to an issue that has long passed, i.e. the flood crisis, they did not respond. We therefore conclude that Twitter is not a useful platform for getting more in-depth feedback from residents. This result contradicts Brabham (2009), who maintains that VGI has the potential to elicit active public participation in urban planning projects. Our analysis shows that Twitter works best as a source of information provided voluntarily by the community. Requesting people to respond and participate more actively in urban planning initiatives was not successful in this case.

Overall, our analysis was able to capture the residents’ preferences regarding evacuation shelters through VGI and location identification. Some of the technical limitations encountered could be overcome, at least in part, by using VGI in combination with other approaches, such as participatory mapping. Goodchild and Li (2012) see the role of VGI mainly in the initial and hypothesis-generating step of the research, due to technological limitations (e.g. accuracy). VGI is still weak in capturing in-depth preferences. Overall, however, our analysis has provided important insights into how the planning, organisation, and notification of official residents’ evacuation shelters can be improved.

5. Conclusion

This research focussed on using VGI in evacuation shelter planning as one crucial part of emergency response. Our results show that 35.6% of people who sent tweets from evacuation shelters were possibly using a formal evacuation shelter. Furthermore, the most frequently used land use category for residents’ evacuation shelters and official evacuation shelters provided by the government was ‘green/open spaces’. This was followed by the land use type ‘schools and education centres’ as concerns official shelters. We conclude that people only used the official shelters when they knew the locations from daily activities. People’s unfamiliarity can explain the failure by residents to use official shelters near their locations.

Overall, VGI proves to be a useful approach for capturing residents’ preferences regarding evacuation shelters, when used in conjunction with, e.g., land use data. VGI data provides a preliminary overview of the topic of interest, in the form of data of a general nature covering a broad area. However, according to our research, VGI should be combined with other approaches to fully understand residents’ preferences about specific spatial planning problems.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeog.2017.07.002.

References

Akmalah, E., & Grigg, N. S. (2011). Jakarta flooding: Systems study of socio-technical forces. Water International, 36(6), 733–747. http://dx.doi.org/10.1080/02503839.2011.1070729.

ARC (American Red Cross). (2002). Standards for hurricane evacuation shelter selection. Retrieved January 21, 2016, from: http://www.floridadisaster.org/Response/engineers/documents/2007SR/2007SR-AppxC.pdf.

Birch, C. P. D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modeling, 206(3–4), 347–359. http://dx.doi.org/10.1016/j.ecolmodel.2007.03.041.

BPBD Jakarta. (2015). Flood map Jakarta (in Bahasa). Retrieved July 29, 2015, from http://petajakarta.org/banjar/en/.

Brabham, D. C. (2009). Crowdsourcing the public participation process for planning projects. Planning Theory, 8(4), 242–262. http://dx.doi.org/10.1177/1470161609339428.

Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., & Kowalchuck, M. (2015). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. Safety Science, 90, 48–61. http://dx.doi.org/10.1016/j.ssci.2014.06.002.

CCCMCluster. (2014). The mend guide comprehensive guide for planning mass evacuations in natural disasters. Retrieved from http://www.globalccl.com/cluster/files/publications/MEND_download.pdf.

Chang, H.-S., & Liao, C.-H. (2014). Planning emergency shelter locations based on evacuation behavior. Natural Hazards, 76(3), 1531–1571. http://dx.doi.org/10.1007/s11069-014-1557-x.

Comber, A., See, I., Fritz, S., Van der Velde, M., Perger, C., & Foody, G. (2013). Using control data to determine the reliability of volunteered geographic information about land cover. International Journal of Applied Earth Observation and Geoinformation, 23, 37–48. http://dx.doi.org/10.1016/j.jag.2012.11.002.

DKI Jakarta Province Government. (2015). Number of school in Jakarta (in Bahasa). Retrieved January 7, 2016, from http://data.jakarta.go.id/dataset/jumlahsekolah-guru-murid-danruangkelasmenurutjenissekolah/dki jakarta.

Durahim, A. O., & Cojkan, M. (2015). #iamhappybecause: Gross National Happiness through Twitter analysis and big data. Technological Forecasting and Social Change, 95, 92–105. http://dx.doi.org/10.1016/j.techfore.2015.06.035.

Erskine, M. A., & Gregg, D. C. (2012). Utilizing volunteered geographic information...
to develop a real-time disaster mapping tool: A prototype and research framework. In International conference on information resources management (Conf-BRM) (p. 27). Retrieved from: http://www.researchgate.net/publication/272204458_UUtilizing_Volunteered_Geographic_Information_to_Develop_a_Real-Time_Disaster_Mapping_Tool_APrototype_and_Research_Framework.

FEMA. (2015). Safe rooms for Tornadoes and Hurricanes: Guidance for community and residential safe rooms. Washington D.C. Retrieved from http://www.fema.gov/media-library-data/1456226019987-05b13d287f7f65d08e2483f6ed46acdf/fema_p361_jan2016_508.pdf.

Firman, T., Surbakti, I. M., Idroes, I. C., & Simarmata, H. A. (2011). Potential climate-change related vulnerabilities in Jakarta: Challenges and current status. Habitat International, 35(2), 372–378. http://dx.doi.org/10.1016/j.habitatint.2010.11.001.

Gall, M. (2004). Where to Go? Strategic modelling of access to emergency shelters in Mozambique. Disasters, 28(1), 82–97. http://dx.doi.org/10.1080/00122330410001675675.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. Geographical Reviews, 97(1), 3–22. http://dx.doi.org/10.1111/j.0016-7227.2006.00244.x.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geographic information. Spatial Statistics, 1, 110–120. http://dx.doi.org/10.1016/j.jspa.2012.03.002.

Holderness, T., & Turpin, E. (2015). From social media to GeoSocial Intelligence: Tagging the digitally tagged social data in geographic research. Retrieved from: http://papers.ssrn.com/abstract=2513938.

Haklay, M. (2010). How good is volunteered geographical Information? A comparative study of OpenStreetMap and ordnance survey datasets. Environment and Planning B: Planning and Design, 37(4), 682–703. http://dx.doi.org/10.1068/b35097.

Haklay, M. (2010). How good is volunteered geographical Information? A comparative study of OpenStreetMap and ordnance survey datasets. Environment and Planning B: Planning and Design, 37(4), 682–703. http://dx.doi.org/10.1068/b35097.

J. Haklay (2010). How good is volunteered geographical Information? A comparative study of OpenStreetMap and ordnance survey datasets. Environment and Planning B: Planning and Design, 37(4), 682–703. http://dx.doi.org/10.1068/b35097.

Raposo, P. (2013). Scale-specific automated line simplification by vertex clustering on a hexagonal tessellation. Cartography and Geographic Information Science, 40(5), 427–443. http://dx.doi.org/10.1080/15230406.2013.803707.

Khosrawipour, M., & Shokr, A. M. (2016). Assessing the change-related vulnerabilities in Jakarta: Challenges and current status. Computers in Human Behavior, 50, 392–398. http://dx.doi.org/10.1016/j.chb.2015.04.020.

Laptev, I., Shechtman, E., Bregler, C., & Schmid, C. (2005). Learning to find unusual actions. In CVPR 2005.

Levin, J. D., & Shuler, J. M. (1994). Assessing the change-related vulnerabilities in Jakarta: Challenges and current status. Computers in Human Behavior, 50, 392–398. http://dx.doi.org/10.1016/j.chb.2015.04.020.

Lin, X., Ma, Z., & Zhao, J. (2015). Learning to find unusual actions. In CVPR 2015.

Lukoseviciute, K., & Schmid, C. (2010). Assessing the change-related vulnerabilities in Jakarta: Challenges and current status. Computers in Human Behavior, 50, 392–398. http://dx.doi.org/10.1016/j.chb.2015.04.020.

Mooney, P., Sun, H., & Yan, L. (2011). VGI as a dynamically updating data source in location-based services in urban environments. In Proceedings of the 2nd international workshop on Ubiquitous crowdsourcing - Ubicrowd ‘11 (p. 13). New York, New York, USA: ACM Press. http://dx.doi.org/10.1145/2030010.2030105.

_The_ Tornado, Hurricanes approach? The role of risk, information, and location. Social Science Quarterly, 91(3), 816–834. http://dx.doi.org/10.1111/j.1540-6237.2010.00721.x.

Tao, C.-A., Lee, Y.-L., & Lin, C.-Y. (2010). Urban disaster prevention shelter location and evacuation behavior analysis. Journal of Asian Architecture and Building Engineering, 9(1), 215–220. http://dx.doi.org/10.1016/j.jabe.2010.09.215.

Takahasahi, B., Tandoc, E. C., & Carmichael, C. (2015). Communicating on twitter during a disaster: An analysis of tweets during typhoon Haiyan in the Philippines. Computers in Human Behavior, 50, 392–398. http://dx.doi.org/10.1016/j.chb.2015.04.020.

Team Mirah Sakethi. (2010). Zoon, M., Graham, M., Shelton, T., Graham, M., & Stephens, M. (2014). Using geo-tagged digital social data in geographic research. Retrieved from: http://papers.ssrn.com/abstract=2513938.

Team Mirah Sakethi. (2010). Zoon, M., Graham, M., Shelton, T., Graham, M., & Stephens, M. (2014). Using geo-tagged digital social data in geographic research. Retrieved from: http://papers.ssrn.com/abstract=2513938.

Team Mirah Sakethi. (2010). Zoon, M., Graham, M., Shelton, T., Graham, M., & Stephens, M. (2014). Using geo-tagged digital social data in geographic research. Retrieved from: http://papers.ssrn.com/abstract=2513938.