

Exploring the potentials of volunteered geographic information as a source for spatial data acquisition

Izyana Ariffin, Badariah Solemon, Rina Md. Anwar, Marina Md. Din, Nor Nashrah Azmi

College of Information Technology, Universiti Tenaga Nasional, Kajang, MALAYSIA

E-mail: izyana@uniten.edu.my, badariah@uniten.edu.my, mrina@uniten.edu.my, marina@uniten.edu.my, nashrah@uniten.edu.my

Abstract. The advancement of technologies nowadays enables participation by non-professionals, known as volunteers to participate in producing, sharing and consuming geographic information. Such information, termed as *volunteered geographic information* (VGI) has created a new approach of gathering geographic information. This paper discusses the traditional way of acquiring geographic information and potentials of VGI as an information source in GIS applications. We also review four commonly cited applications which rely on volunteers for their geographic information based on five criteria; the geometry type available in the applications, availability of user profile, average number of attributes used in the applications, data type of the information (raster or vector) and the domain the application belongs to. This review serves as a preliminary study in designing a GIS application used for asset management which aims at exploiting volunteers to produce geographic information related to assets.

1. Introduction

Technological revolution in the field of Information Technology nowadays has facilitated the rapid development of Geographic Information System (GIS). GIS can be defined as one type of information system that stores, displays, manipulates and analyzes spatial data particularly map-based information [1, 2]. The usage of GIS spreads across many application domains, namely transportation, travel, health surveillance, inventory management, military and many others [3]. However, despite its usefulness, GIS is very difficult to use; “there is little hope that community residents with no computing experience will be able to use the software”, even with a proper training being provided [4, 5]. As technology advances, many improvements have been done by the software providers especially on the graphical user interface [3]. These improvements, with the help of embedded global positioning system (GPS) in mobile devices, have inadvertently enabled GIS to be accessible and consumed by a much wider range of users, especially the Internet users.

In the early days, mapping information in GIS were done only with military interest. In the 20th century, mapping information started to be used by civilian, mostly by National Mapping and Cadaster Agencies (NMCAs) [6]. However, the updating is seriously lagging and very little efforts were taken by the authorities to improve the level of details of existing maps [7]. Only in recent years that public has started to engage with the production as well as consumption of mapping information, as the result of Web 2.0 technologies. Successful implementation of OpenStreetMap (OSM), Google Earth and



HDTraffic proves that public can be a reliable source for geographic information. In this paper, we aim at reviewing features available in commonly cited GIS applications that have successfully gained active participation by volunteers. This review serves as a preliminary study to a research that investigates the effectiveness of using VGI in distributed asset management. We discuss the common terminologies used to describe VGI, and explain how VGI can be harnessed as an alternative source of geographic information in GIS. We then explain the results of our review on the common features that are available in those successful applications. The review is used as a basis for designing an application which targets at utilizing volunteers to produce geographic information related to assets. We conclude our paper by discussing some issues in VGI and the future works that we find useful for our current research project.

2. Volunteered geographic information

The emergence of Web 2.0 technologies, miniaturization of GPS devices and advancement of broad band communication links have empowered the Internet users to be involved in producing geographic information. These Internet users, referred to many as *citizens* or *sensor* [7], who voluntarily produce, share and use geographic information, were termed by Goodchild as VGI [7]. Previously, Turner [8] described similar act of sharing location information with friends and visitors as *neogeography* while some others refer to similar nature of producing information as *crowdsourcing* geo-information. *Crowdsourcing* was first coined by Howe [9], in which he was inspired from the fact that “under the right circumstances, groups are remarkably intelligent, and often smarter than the smartest people in them”. This was referred as “wisdom of crowds” by Surowiecki [5]. Apart from these, *participatory GIS* (PGIS), *public participants GIS* (PPGIS) and *community-integrated GIS* are among other terms that are commonly used to illustrate the approaches that are “context-and issue-driven rather than technology-led and seek to emphasize community development involvement in the production and/or use of geographical information” [10].

3. VGI as an alternative spatial data collection method

The data available in GIS is categorized into spatial and non-spatial information. Spatial information can be grouped into two (2) categories; raster and vector while non-spatial information, commonly known as attributes, are the physical information associated with the spatial information. The data sources for raster often come from digital remote sensing images and aerial photographs, while the data sources for vector come from global positioning system (GPS) and survey measurements. One of the commonly used method for collecting spatial data for GIS is known as field data collection [11]. In the past decades, this method was solely authoritarian; professional experts produce, dissemination is radial, and amateurs consume [12]. The method has been widely known to be very costly [4], in turn causes many organizations to avoid adopting GIS in their daily operational activities despite its capability in enhancing decision making process through the provision of basic data and complex analyses [1]. It relied fully on professionals due to the belief that the acquisition of geographic information requires extensive training, thus to be beyond the abilities of amateurs [12]. Thus, the accuracy of the data is considered of high quality due to the positional accuracy of the GPS device and techniques used by the professional land surveyor in capturing the geographic references. Additionally, due to the limitation of technology in the past, public only had local reference for geographic information, i.e. the local governments and agencies, who are able to provide detailed geographic information that can be verified and integrated at the national level [12].

The concept of utilizing public as source of geographic information was far from reality until the availability of Web 2.0 technologies. The advancement of spatial technologies, fostered by the proliferation of information supplied by the Internet users has resulted significant changes in the way organizations acquire geographic information to be used in GIS. VGI is now globally accepted as a reasonable means to collect and update geographic knowledge of local experts [4]. With VGI, not only the production and update of geographic information can be improved [13], organizations can also exploit volunteers, resulting in reduced operating cost [2]. Although the quality of VGI is still a major

concern, volunteers are often more interested to map their immediate surroundings, which may contribute to surprisingly good quality of information [6].

Our main research objective is to look at the effectiveness of using volunteers to collect information related to assets which are geographically scattered. This research was originally driven by Coleman's questions; if an organization were to employ VGI for its source of information, how it is going to attract new volunteers, and how does it keep existing volunteers engaged [14]? Therefore, to design an application which targets volunteers to actively produce the geographic information, it is critical that we review the elements that exists in successful VGI projects, and consider the elements for our to-be-developed application. The following section discusses successful VGI projects that are commonly cited in literatures.

3.1 Successful existing VGI projects

The plethora of GPS-embedded mobile phone devices available in the market nowadays catalyzes many VGI-applications to come into usage. OpenStreetMap (OSM) (<http://www.openstreetmap.org>) project for example, has been very successful in collecting geographic information in most part of the world. Volunteers actively provide and consume geographic information in OSM, resulting the project to become the backbone for free and commercial printed or digital maps, navigation devices, location-based services or research initiatives [4]. Additionally, Waze mobile application (www.waze.com), has been widely used by drivers to get to destination by using least congested routes. Waze uses drivers' nearest telecommunication towers and GPS location to detect traffic jams and provide alternative routes for drivers to avoid heavy traffic roads. Drivers can also share information with nearby drivers if there are accidents or road maintenance at nearby location. In disaster and crisis management, Scipionus (<http://gregstoll.dyndns.org/scipionus>) and Ushahidi (<http://www.ushahidi.com>) become the information platform for disaster relief. The Internet users are able to follow the progress of crises, in real time through the eyes of those directly involved in the disaster [15, 16]. Wikimapia (<http://www.wikimapia.com>) is also one example of successful applications which relies on volunteers for its information source. Users can mark rectangles and polygons on a map to indicate places of interest, as well as update those information entered by other users. In addition to these, there have been numbers of VGI-applications actively used by the Internet users which include finding new bike routes (<http://www.mapmyride.com>), tracking infectious disease (<http://healthmap.org>), monitoring forest fires [17], locating cell phone coverage and Wi-Fi hotspot (<http://opensignal.com>), tracking local weather (<http://weathersignal.com>) and many others.

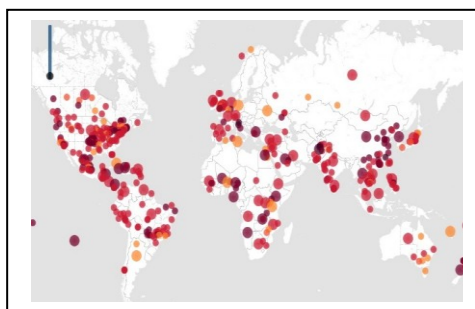


Figure 1. A snapshot of HealthMap application.
Source: <http://healthmap.org>

3.2 Reviewed projects

As mentioned earlier, this paper serves as a preliminary study for creating a VGI application for asset management. The application is targeting volunteers to supply geographic information associated to assets of an organization. To ensure the application is able to attract huge number of volunteers to participate, we selected four commonly cited projects which rely on public for the production of geographic information. We intended to evaluate applications used for asset management but very

limited literatures were found. Many asset management applications only use crowdsourced information which is not associated with geographic references. As the application we are developing uses geographic location to map the asset location, it is essential to review similar nature of projects although the domain is different. The projects we selected are OSM, Waze, HealthMap and Wikimapia. The evaluation criteria used in the review are shown in table 1.

Table 1. Evaluation criteria.

Evaluation Criteria	Description
Geometry types	The geometry types of the information; point, polyline or polygon
User profile	The availability of login or user profiling function in the application
Average number of attributes	Associated physical attributes of the data for example name, size, lot number etc.
Type of data	Type of data produced by volunteers; raster or vector
Domain	The domain of the application for example health, navigation, utility, transportation, mapping etc.

3.3 Evaluation criteria

The following sections explain the details of each criteria and its relevance to our study.

3.3.1 Geometry types

GIS has long been recognized for its difficulty to use [3-5], and volunteers are normally lacking of technical knowledge on the tools to use as well as the domain knowledge required in order to produce geographic information. Hence, creating a GIS application which is easier to be used by volunteers is always the highest priority. Nielsen suggested that in order to increase active participation, users should find the application “easy to contribute, without technical, logistic, legal or intellectual barriers” [18]. In GIS, point geometry is always easier to digitize as compared to polyline or polygon. Additionally, different users have different views of the information they want to see in the application, and different views would require different level of details, of geometry, imagery, attributes (semantics) etc. [19].

3.3.2 User profile

Many studies venture into finding what motivates volunteers to contribute information and what contributes to these volunteers’ stickiness to the application. Although they do not gain financial reward, volunteers do provide information to enhance their personal reputation and own sense of worth in the online community [13]. Thus, volunteer’s profile is required in order for him to be recognized among the online community members. Additionally, it has been suggested that volunteers need to be rewarded in order to obtain active participation [18]. To reward, being able to track who is the volunteer and assess the quality level of the information he contributes are mandatory. It is indeed one of the features that an application designer needs to look at when designing an interface for crowdsourcing application [20].

3.3.3 Number of attributes

Number of physical data and the complexity level of the attributes can deter volunteers from entering information. Complex attributes which require advanced technical knowledge may lead to users’ frustration when describing the geographic information. The more associated information is required, the longer the time it takes for volunteers to key in information. With the rapid development of the interface used in crowdsourcing applications, it is expected that the data models will become more sophisticated [6]. Users normally prefer a simple and less complex application to use [20]; therefore we need to see how simple should a GIS application be in order to obtain huge participation by the volunteers.

3.3.4 Type of data

Using volunteers to map raster information requires a considerable skill in the use of photogrammetric techniques, expensive equipment for observation and analysis and substantial investment in large-format printing [5]. However, certain applications have made the interface simple that enables volunteers to consume information related to raster. We intend to review the type of data entered by volunteers in each of the application in order to measure the minimum level of expertise required by such volunteers.

3.3.5 Domain

Most of VGI applications that exists are designed for mapping the world. However, the information that comes from volunteers might be influenced by their interests or hobbies; volunteers tend to contribute information to community which they have something in common, and inadvertently form many social linked groups. These groups have totally different needs and different ways to keep them to be always motivated to volunteer [6]. It is good to gain an insight on the groups that exists to identify the most dominant domain that volunteers are in.

4. Review analysis

We visited all four projects mentioned earlier and analyzed the criteria we depicted in table 1. Table 2 summarizes the findings.

Table 2. Summary of review findings.

<i>Project</i>	<i>Geometry types</i>	<i>User profile</i>	<i>Average number of attributes</i>	<i>Type of data</i>	<i>Domain</i>
<i>OpenStreetMap</i>	<i>Point</i> <i>Polyline</i> <i>Polygon</i>	<i>Required</i>	<i>5-8</i>	<i>Vector</i>	<i>Mapping</i>
<i>Waze</i>	<i>Point</i> <i>Polyline</i> <i>Polygon</i>	<i>Optional</i>	<i>6</i>	<i>Vector</i>	<i>Navigation</i>
<i>Wikimapia</i>	<i>Polygon</i>	<i>Optional</i>	<i>7</i>	<i>Vector</i>	<i>Mapping</i>
<i>HealthMap</i>	<i>None</i>	<i>Required</i>	<i>7</i>	<i>None</i>	<i>Health</i>

4.1 Geometry types

OSM supports many types of geometry for mapping real world objects (RWOs) such as supermarket, kiosk, hospital, vending machine, post office, hotel and many others. To make it easier for users to find the objects to map, categories are used i.e. amenities, shopping and transport. In Waze, users are allowed to update information about traffic by adding point geometry to indicate hazard, accident or police appearance on a map. It also allows user to add newly constructed road by using ‘*Pave road*’ function, which will track the starting point, the route and the end point of the new road (polylines). Besides updating through the mobile application, users are also encouraged to update the map used in Waze. Similar to OSM, Waze has many RWOs which are grouped into categories for easy finding. These include hospital, clinic, police station, university and many others. However, the list of RWO in Waze is more extensive when compared to OSM’s. Unlike Waze, as Wikimapia was meant to encourage people to describe the world, the information provided by volunteers is linked to polygon geometry only. However, experienced users are allowed to add or update information related to polyline objects such as river, railroad and roads. The experience is measured based on the ratings on user profiles. On the other hand, HealthMap works totally different from OSM, Waze and Wikimapia. Since the application targets at getting volunteers to supply information regarding outbreaks, users describe the place by typing the place name to mark the location of the disease. Again, since the category of information is already specific (disease outbreak), volunteers do not have to select the category for the RWO, in this case, the disease location.

4.2 User profile

Before a volunteer can map inside OSM, login is required. However, only minimum number of information is required from a '*mapper*' when creating a login account, perhaps to increase the convenience of using such application. Unlike OSM, if one intends to Waze as a source for traffic information, login is not required. Only if a '*wazer*' plans to update the map used in Waze that login is made mandatory. This is perhaps due to the need for quality checking of the information provided by the '*wazer*'. Wikimapia operates just like Waze. The application allows users to provide attribute information related to the place i.e. name, description and categories without logging in. However, if a user wishes to provide information with related to polyline, login and experience are required as new users or low rating users are not allowed to contribute information related to polyline objects. Like in OSM, login is required if a volunteer wants to provide information regarding disease outbreak in HealthMap. Having a login function for the application that spreads outbreak information is seen very relevant, as any false information appears in the application may create panic situation among the citizens.

4.3 Average number of attributes

In OSM, the number of attributes available for users to key in is not consistent; it varies from one object to another, but on average, each object requires between 5-8 attributes associated with the object. However, in contrast, number of attributes required for updating map information in Waze is consistent for all objects. This definitely improves the usability of the application, as consistency is crucial in creating an effective user interface [21]. In Wikimpia and HealthMap, users are allowed to key in 7 associated attributes to the object they created. However, in HealthMap, only 1 object is available for user to map, which is the disease. Therefore the number of attributes can be considered consistent as in Waze.

4.4 Type of data

As expected, the data type produced by volunteers in OSM, Waze and Wikimapia is vector based. However, OSM does provide a function for volunteers to get the information available in OSM in raster format, and this has been made simple so that volunteers do not need to possess deep knowledge in photometry or remote sensing image processing. In HealthMap, as user does not draw any geometry in the application, hence, there is neither vector nor raster data being generated in the application.

4.5 Domain

OSM is clearly being exploited for collecting geographic information used for mapping activities. The information available in OSM is used to help map agencies fill in the information gaps in their existing maps. However, Waze is different. Although Waze also provides a platform for volunteers to update its maps, the main function of the application focuses at updating drivers with the latest nearby traffic information. The information is consumed for navigation purposes. Like OSM, Wikimapia is used mainly for mapping as the information supplied by volunteers is used to add more information to the existing maps. However, Wikimapia focuses more on describing the place rather than providing very accurate mapping information especially for streets. HealthMap, as the name suggests, is used for collecting information related to health domain.

4 Discussion & future works

The result of the review suggests that volunteers prefer to contribute geographic information pertaining to mapping. All applications use vector data by the volunteers, except for HealthMap which uses description to describe geographic location as opposed to geometry. To increase data quality, all applications require users to login before they can produce geographic information. However, this review is merely performed on only four commonly cited applications in literatures, and there are many other applications which rely on volunteers that were not included. In addition, as the review was performed on the breadth of the application rather than depth, further studies need to be conducted to

evaluate each criteria in detail. This includes finding number of RWOs created by volunteers based on geometry and its type. A similar study has been done, but it focused on heavily edited objects only at few selected areas [22]. By knowing such details, we can identify popular geometry produced by the volunteers and perhaps further investigation can research into finding why such geometry is in favour. Although geometry should represents RWO actual characteristics - for example land parcel is in polygon and road is in plotline, this knowledge may not be possessed by the volunteers. In GoogleMaps, many tourist attraction places and shops are drawn as point instead of polygon, indicating that volunteers are clueless about RWO's suitable geometry. However, as Nielsen suggested, to increase participation, volunteers are encouraged to edit, rather than create, therefore the application owner or organization should provide templates for jobs to be done [18], in this context, the geometry of the object should be pre-defined.

Besides, finding details about the attributes can be further explored by looking at the complexity of the attributes. We can examine whether complexity of attributes plays an important role in increasing participation from volunteers. In the reviewed applications, the associated information to the geographic information is considered minimum and can be easily understood by non-professionals. However, in the future, mapping may evolve into three-dimensional, and the information produced by volunteers might have to be spatially analyzed, based on organizations' specific needs. Thus, there might be time when volunteers also need to contribute information that requires technical background. Additionally, as organizations start relying heavily on volunteers to produce information, the demand for level of information completeness may increase-organizations may start requesting volunteers to produce information in a way that it can be easily digested for decision making, without any further information processing done in between the information producers and consumers. Limitation of using VGI as a source for data acquisition also remains arguable by many. To what extent can we rely on these volunteers? How reliable is the information produced and why do they want to produce? This is what our research focuses to study. We intend to examine whether VGI can be applied in asset management domain, the type of geometry volunteers prefer to use, the motivation, and the elements required to sustain such volunteers. Although quality of the information produced by these unprofessional volunteers also is still a major concern, under certain circumstances, having poor quality data is always better than none. Until today, researchers pushed for more studies to be done in the area of VGI before it can be used as the primary source of information in any GIS application [23]. Majority still regard VGI as a complementary way of acquiring data instead of a substitution.

5. Conclusion

It has been observed that the number of projects that utilizes VGI as a new method for collecting data is rapidly growing. In this paper, we performed review on four commonly cited applications that make use of VGI and explored into five criteria that we considered important in designing our future application. It can be concluded that in order to attract volunteers to keep producing information, an application should consider deploying simple geometry to describe the RWOs, using minimum and preferably consistent number of attributes across all RWOs, implementing user profile functionality besides encouraging use of vector data as opposed to raster. However, as the extensiveness of this review is on the breadth rather than depth, more studies are needed to really identify the criteria of an application which can lead to increased volunteers participation.

References

1. Campbell, H. and I. Masser, *GIS In Organizations: How Effective Are GIS In Practice?* 1995: CRC Press.
2. Longley, P., *Geographic information systems and science*. 2005: John Wiley & Sons.
3. Goodchild, M.F., *Geographic information systems and science: today and tomorrow*. *Annals of GIS*, 2009. **15**(1): p. 3-9.
4. Goodchild, M.F., *Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0*. *International Journal of Spatial Data Infrastructures Research*, 2007. **2**: p. 24-32.

5. Goodchild, M., *NeoGeography and the nature of geographic expertise*. Journal of Location Based Services, 2009. **3**(2): p. 82-96.
6. Heipke, C., *Crowdsourcing geospatial data*. ISPRS Journal of Photogrammetry and Remote Sensing, 2010. **65**(6): p. 550-557.
7. Goodchild, M.F., *Citizens as sensors: the world of volunteered geography*. GeoJournal, 2007. **69**(4): p. 211-221.
8. Turner, A., *Introduction to neogeography*. 2006, O'Reilly Media Inc, Sebastopol CA.
9. Howe, J. (2006) *The Rise of Crowdsourcing*. **14**, 1-4.
10. Dunn, C.E., *Participatory GIS—a people's GIS?* Progress in human geography, 2007. **31**(5): p. 616-637.
11. Lwin, K.K. and Y. Murayama, *Web-Based GIS System for Real-Time Field Data Collection Using a Personal Mobile Phone*. Journal of Geographic Information System, 2011. **3**(4).
12. Goodchild, M.F., *in the World of Web 2.0*. International Journal, 2007. **2**: p. 24-32.
13. Coleman, D.J., Y. Georgiadou, and J. Labonte, *Volunteered geographic information: The nature and motivation of producers*. International Journal of Spatial Data Infrastructures Research, 2009. **4**(1): p. 332-358.
14. Coleman, D.J., *Potential contributions and challenges of VGI for conventional topographic base-mapping programs*, in *Crowdsourcing Geographic Knowledge*. 2013, Springer. p. 245-263.
15. Roche, S., E. Propeck-Zimmermann, and B. Mericskay, *GeoWeb and crisis management: issues and perspectives of volunteered geographic information*. GeoJournal, 2013. **78**(1): p. 21-40.
16. Zook, M., et al., *Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake*. World Medical & Health Policy, 2010. **2**(2): p. 7-33.
17. De Longueville, B., R.S. Smith, and G. Luraschi. *OMG, from here, I can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires*. in *Proceedings of the 2009 International Workshop on Location Based Social Networks*. 2009: ACM.
18. Nielsen, J., *Participation inequality: Encouraging more users to contribute*. Jakob Nielsen's alertbox, 2006. **9**: p. 2006.
19. Craglia, M., et al., *Next-Generation Digital Earth*. International Journal, 2008. **3**: p. 146-167.
20. Nakatsu, R. and E. Grossman, *Designing effective user interfaces for crowdsourcing: an exploratory study*, in *Human Interface and the Management of Information. Information and Interaction Design*. 2013, Springer. p. 221-229.
21. Johnson, J., *Designing with the mind in mind: Simple guide to understanding user interface design rules*. 2010: Morgan Kaufmann.
22. Mooney, P. and P. Corcoran, *Characteristics of heavily edited objects in OpenStreetMap*. Future Internet, 2012. **4**(1): p. 285-305.
23. Wilson, N., et al., *Interpreting Google flu trends data for pandemic H1N1 influenza: the New Zealand experience*. Euro surveillance: bulletin européen sur les maladies transmissibles= European communicable disease bulletin, 2009. **14**(44).