Applied Geography 79 (2017) 115–126

Contents lists available at ScienceDirect

Applied Geography

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

An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas

Noam Levin [a, c, *], Alex Mark Lechner [b, d], Greg Brown [c, e]

[a] Department of Geography, The Hebrew University of Jerusalem, Mt Scopus, Jerusalem, 91905, Israel
[b] Regional Water and Land Resources, Sustainable Minerals Institute, The University of Queensland, St Lucia, Brisbane, Queensland, 4072, Australia
[c] School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, Queensland, 4072, Australia
[d] School of Environmental and Geographical Sciences, University of Nottingham Malaysia Campus, Semenyih, Selangor, 43500, Malaysia
[e] Natural Resources Management & Environmental Sciences Department, California Polytechnic State University, San Luis Obispo, California, 93407, USA

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form
11 October 2016
Accepted 22 December 2016
Available online 9 January 2017

Keywords:
Flickr
OpenStreetMap
Place values
Public participation GIS (PPGIS)
Volunteered geographic information (VGI)
Wikipedia

ABSTRACT

Parks and protected areas provide a wide range of benefits, but methods to evaluate their importance to society are often ad hoc and limited. In this study, the quality of crowdsourced information from Public Participation GIS (PPGIS) and Volunteered Geographic Information (VGI) sources (Flickr, OpenStreetMap (OSM), and Wikipedia) was compared with visitor counts that are presumed to reflect social importance. Using the state of Victoria, Australia as a case study, secondary crowdsourced VGI data, primary crowdsourced (PPGIS data) and visitor statistics were examined for their correspondence and differences, and to identify spatial patterns in park popularity. Data completeness—the percent of protected areas with data—varied between sources, being highest for OSM (90%), followed by Flickr (41%), PPGIS (24%), visitation counts (5%), and Wikipedia articles (4%). Statistically significant correlations were found between all five measures of popularity for protected areas. Using stepwise multiple linear regression, the explained variability in visitor numbers was greater than 70%, with PPGIS, Flickr and OSM having the largest standardized coefficients. The social importance of protected areas varied as a function of accessibility and the types of values (direct or indirect use) expressed for the areas. Crowdsourced data may provide an alternative to visitor counts for assessing protected area social importance and spatial variability of visitation. However, crowdsourced data appears to be an unreliable proxy for the full range of values and importance of protected areas, especially for non-use values such as biological conservation.

1. Introduction

The advent and combination of several technologies, namely the global positioning system (GPS), the internet, Google Maps, smartphones and social media, have transformed the ways we generate, consume and interact with geographic information (Goodchild, 2007; Haklay, Singleton, & Parker, 2008). In the world of Web 2.0, where citizens can be seen as sensors (as termed by Goodchild, 2007), and where people voluntarily contribute geographic data, either knowingly or unknowingly, new avenues open for spatial research. Whereas in the past, most geographic information and maps were generated by governments, non-governmental organizations, and major companies, in the recent decade the sharing of volunteered geographic information (VGI; Fast & Rinner, 2014) is rising, and concerns are being raised about its credibility (e.g., Flanagin & Metzger, 2008; Spielman, 2014). The quality of user-generated or crowdsourced data is an obvious challenge when using big (geo) data, which is characterized by volume, velocity, and variety (Goodchild, 2013).

Volunteered geographic information (Goodchild, 2007) is a form of *crowdsourcing* (Howe, 2006) where ideas or content are solicited for a certain project, from a large group of people, especially from an online community, with enabling technical infrastructure including hardware, software, and the Geoweb – online location-based services (Fast & Rinner, 2014). *Public participation GIS* (PPGIS) is a term that originated in 1996 to describe how GIS technology could support public participation processes (NCGIA 1996a, 1996b). Although the term *crowdsourcing* did not originate

* Corresponding author. Department of Geography, The Hebrew University of Jerusalem, Mt Scopus, Jerusalem, 91905, Israel.
E-mail address: noamlevin@mscc.huji.ac.il (N. Levin).

http://dx.doi.org/10.1016/j.apgeog.2016.12.009
0143-6228/

Table 1
Sources used for examining the popularity and values associated with protected areas.

Source	Description
Visitation	Official visitation statistics, available for 342 parks from Parks Victoria (2011).
PPGIS	Study participants (n = 1905) were sampled, recruited, and requested to identify place values using markers on a Google Map of Victoria, Australia. The place values were a typology of values (biological/conservation, heritage/cultural, intrinsic/existence, learning/education/research, recreation, scenic/aesthetic, therapeutic/health, wilderness/pristine (see Brown, Weber and De Bie, 2014 for the value definitions). Participants could place as many of the different types of markers as needed to identify places values. Place values (also called *landscape values*) are a type of "relationship" or "transactive" value that bridges both "held" values (what is personally important) and "assigned" values (the relative importance of external objects) (Brown & Weber, 2012).The number and type of place values were counted within each protected area.
Flickr	The total number of georeferenced Flickr photos (between 2004 and 2012) within each protected area (further details on this dataset are provided in Levin et al., 2015). Flickr is a photo-sharing website, and Flickr photos have been shown to be well-correlated with official visitation statistics (Levin et al., 2015; Wood et al., 2013).
Wikipedia	The total number of Wikipedia edits which were found for Wikipedia articles on protected areas. The numbers of Wikipedia edits were counted in October 2015. Most protected areas did not have a Wikipedia article.
OpenStreetMap	The total number of vertices within each protected areas as found in the OpenStreetMap roads, tracks, and trails layer provided by Geofabrik (http://www.geofabrik.de/data/download.html). OpenStreetMap started in 2004 and provides a free digital map of the world created collaboratively by internet users (Haklay, 2010). The density of vertices within protected areas is expected to reflect the physical infrastructure, as well as the interest and familiarity in a protected area by the public. Parks with more facilities for visitors (e.g., trails, tracks, roads) and which are planned to attract more visitors, will likely have more features which can be mapped using OSM. Thus, parks with more OSM vertices are an indication of more mapped features, but also of greater familiarity and interest by the public in mapping it. In addition, the density of vertices can indicate the accuracy (i.e. scale) in which the mapping was done. The number of vertices used to digitize a certain feature depends both on the geometric complexity of that feature, and also on the levels of generalization and simplification used by the person who digitized that feature. Thus, more vertices can also indicate more care and details in the mapping process.

until a decade later, PPGIS methods that collect spatial data from a large group of people for public participation are considered a type of crowdsourcing (Brown, 2015). The related fields of VGI and PPGIS, as forms of crowdsourcing for geographic knowledge production, usually differ in their sampling and mapping methods, and both have experienced significant growth as evidenced in the number of real-world applications and academic publications (Brown & Kyttä, 2014; Mukherjee, 2015; Sui, Elwood, & Goodchild, 2012).

While the majority of the examples of VGI mapping have been for land cover mapping (e.g., roads, land use), crowdsourced data can also offer original insights as preferences for geographic locations (Brown, 2015; Lechner et al., 2014). Locational information can be manually entered (Gao, Li, Li, Janowicz, & Zhang, 2014) or extracted from devices such as smartphones (Birenboim & Shoval, 2016; Shoval, 2007). Associations with places can be identified through geotagged photos (e.g., Li, Crandall, & Huttenlocher, 2009), along with methods that analyze and classify the images (e.g., Yanai, Yaegashi, & Qiu, 2009). Wikipedia articles offer a wide array of options for investigating cultural values associated with sites using quantitative text analysis (as done by Michel et al., 2011 for millions of scanned books, or by Yasseri, Spoerri, Graham, & Kertész, 2014 for millions of Wikipedia articles), as well as by examining the edits to Wikipedia articles (Graham, Hogan, Straumann, & Medhat, 2014; Hardy, Frew, & Goodchild, 2012). The density of OpenStreetMap (OSM) edits has been shown to be highly correlated with socioeconomic status of inhabitants (Haklay, 2010).

Addressing uncertainty associated with crowdsourced data quality such as completeness and accuracy is critical to ensure the reliability of geographic analyses (Comber et al., 2013; Devillers & Jeansoulin, 2010; Shi, 2010). For VGI data, mutually reinforcing observations, independently derived from different user-generated sources, can serve as a method of data triangulation (Denzin, 2006) to test the correspondence between and completeness of crowdsourcing information. Indeed, integrating multi-sourced volunteered geographic data has been suggested as an approach for enriching the information available and overcoming some of the limitations associated with a single source of user generated data (Sester, Arsanjani, Klammer, Burghardt, & Haunert, 2014).

In this paper, our aim is to compare user-generated information from several sources with official observations to determine the degree to which different Web 2.0 sources correspond and complement each other. The specific theme pursued here is the importance of protected areas as perceived by the public. A protected area is a clearly defined geographical space, recognised, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values (Dudley, 2008). While protected areas may be highly important for their ecological or biological features, of which the public may not be fully aware, our aim is to examine a range of social values associated with protected areas. Because park visitation is a volitional human behaviour, there is clearly personal and social importance (from benefits) associated with visitation activity or people would not visit protected areas. Further, protected area funding (an indirect measure of importance) is often linked to visitation. Thus, the social value of protected areas can be assessed in multiple ways ranging from visitation, a type of popularity measure, to the assessment of specific place values (e.g., scenery, recreation) that people associate with protected areas from crowdsourced methods. In this study we have assessed social value using PPGIS data which accounts for both direct and indirect, and use and non-use values.

Official statistics concerning visitors' numbers can represent a real world reference for the popularity of protected areas; however, visitation statistics are missing for most protected areas globally (Eagles et al., 2002), and even when visitation statistics are available, additional information on what people actually do during their visit is often lacking (Buckley, Robinson, Carmody, & King, 2008). Recently there have been studies whose aim was to predict visitation rates using statistical modeling to estimate the economic value generated by protected areas (Balmford et al., 2015), to quantify human presence in protected areas using crowdsourced and remote-sensing data (Levin, Kark, & Crandall, 2015), and to assess the potential attractiveness of protected areas based on the species which are present in them using the Flickr photo-sharing website (Willemen, Cottam, Drakou, & Burgess, 2015). PPGIS and VGI methods have been used to identify both use and non-use values in protected areas in multiple global settings including the U.S. (Brown & Alessa, 2005; Brown, Kelly, & Whitall, 2014; Van Riper and Kyle, 2014), Canada (Beverly, Uto, Wilkes, & Bothwell, 2008), and Australia (Brown, Weber and De Bie, 2014; Van Riper