Coerced Geographic Information: The Not-so-voluntary Side of User-generated Geo-content

Grant McKenzie, Krzysztof Janowicz

Department of Geography, The University of California, Santa Barbara, USA
Email: {grant.mckenzie; jano}@geog.ucsb.edu

1 User-generated Geo-content

It has been seven years since Michael Goodchild published his seminal paper describing citizens as sensors and defining user-generated content of a geographic nature as *Volunteered Geographic Information (VGI)* (Goodchild, 2007). Two years prior to this publication, the term *User-generated Content (UGC)* began to appear in the media alongside such terms as *Web 2.0*, *social Web*, and *New Media*. In the past seven years, the world of user-contributed content has changed substantially. While individuals and groups still actively contribute geographic information to open platforms such as *OpenStreetMap* and *Wikimapia*, an increasing trend has been to contribute data, a lot of which contains geographic attributes, to private data silos. In many cases the motivation for contributing to these private data sources is completely unrelated, or at least secondary, to the actual data being captured. Unbeknownst to many users of these systems, their data are being used for purposes other than how it was originally intended, both by commercial entities as well as the academic research community. Furthermore, the ability to access data contributed to these systems deviates significantly from the open-access nature of traditional VGI systems.

In this short paper we explore the differences between what has historically been labeled VGI and what we propose should lie under the title of *Coerced Geographic Information (CoGI)*. We give examples of datasets that are generated through different means and outline a set of five criteria that may be used to define the differences between VGI and CoGI. Lastly, we present a *VGI to CoGI Scale* that is used to rate current platforms that collect user-generated geo-content.

The value of geospatial data has not gone unnoticed in the recent rise of online social networking (OSN). Virtually every social media platform established in the last five years either began with, or has come to incorporate geographic data in their applications. Platforms such as *Foursquare*, *Yelp* and *Jiepang* were founded in geospatial data, offering users the ability to *check-in* to places or share their mobile device location with friends. More popular applications such as *Facebook*, *Twitter* and even *Sina Weibo* began without the ability to geotag content and later added this feature as, *a)* geolocation technology became ubiquitous and, *b)* the companies realized the power in knowing the locations of their clients. While the ability to geotag photos or check-in to a specific location has been sold as a *feature* to users of these platforms, it has come at the cost of location privacy, a concept that is not fully explained by the platforms, nor understood by the average user contributing to these applications.

Take for example a user uploading photos to *Yahoo's* photo sharing application *Flickr*. The primary purpose of uploading photos, to most users, is to share moments with their friends, family or even the public. The fact that many mobile image capturing devices (e.g., mobile phones, digital cameras) include a location tag in the image header is either not known to the original publisher or seen as a secondary *feature* of arguably little importance. What is truthfully not understood by the vast majority of contributors is the value of this "secondary" geodata. Previous work (Girardin et al., 2007; Toyama et al., 2003) has shown just how rich this geodata is through the construction of gazetteers and travelogues. The *Flickr* API even offers the ability for developers to extract the location information directly from a photograph's exchange image file format (Exif).[1] If the academic community is able to construct such robust data models with a minimal amount of data accessed through public-facing APIs, one can only imagine what is possible given the full set of data.

[1] https://www.flickr.com/services/api/flickr.photos.getExif.html

A second example is found in the gamification aspects of online social networks. Applications such as *Foursquare* offer users the ability to check-in to places in order to gain points and receive badges. The more places you check-in the more points you receive. From a social perspective this is quite appealing, I can compete with my friends for points and mayorships. The *only* cost of this entertainment is sharing my location. While the game-play and enjoyment of users is the motivating factor, the benefit to Foursquare is enormous. The company has built an entire business model around geospatial data contributed by their users. At last count, Foursquare lists more than 50 million[2] *venues* (Points of Interest) contributed by more than 50 million users. These sensitive personal data are used for location-targeted advertising, business registration and is even sold to third parties (Van Grove, 2013).

OSN contributions aside, location data is being gathered from a great range of sources. Credit Card companies have access to location-specific transaction records and reward programs keep track of where people buy fuel and groceries. Opening the *Privacy Settings* on our location-enabled mobile device will present the user with all applications that have access to location data. Applications such as a *Brightest Flashlight* or *Angry Birds* have no motivation for collecting location information other than resale (Hong, 2012). One of the primary question of interest is if users are aware of what is being collected and done with their data.

From a research community perspective it is important to realize that many individuals contributing data to platforms such as *Twitter* are unaware that their data is being used for research purposes. A limited few may be aware that their location data is used for direct marketing and advertising, but the vast majority of contributors are unaware that existing research employs psychological profiling techniques as well as terrorist threat detection models (Mahmood, 2013), for example, on their seemingly private data and that their tweets may appear as examples in scientific papers.

2 Criteria and Scale

Given the examples listed above it is clear that there should be some criteria that can be applied to any given user-contribution platform in order to place it on a VGI to CoGI scale. To start the discussion, we introduce five criteria in this section each of which can be used to rate a platform on a 3 point scale from 0 (highly CoGI) to 2 (highly VGI).

(I) **Equivalent bi-directional data access**. One of the truly limiting aspects of CoGI and one that explicitly differentiates it from traditional VGI is the accessibility of the contributed data. The value of traditional VGI platforms is that it is as easy to access the data as it is to contribute. *OpenStreetMap*, for example allow users to download entire *planet* files while platforms such as *Twitter* restrict consumption of data through limited APIs and application such as *Strut* offer no ability to consume the contributed data, i.e., they are data one-way streets.

(II) **Limited terms of use**. This criteria relates to the restrictions on how the data can be used once consumed. Conventional VGI systems, e.g., Wikimapia, have very open (often *free to all*) licenses, while more commercially oriented systems often claim the rights to contributed data and restrict the terms of use. Still other platforms act as data silos, allowing no external reuse of the data.

(III) **Awareness of Contributed Data**. An important question to ask of any platform is whether the users are aware of the data that they are contributing. For example users may not be aware that their IP address is being recorded or that their location may be inferred through Wi-Fi positioning. CoGI platforms range in their level of transparency regarding what is being collected.

(IV) **Awareness of Data Usage**. Transparency on how contributed data is used is an important criteria. Sites like *OpenStreetMap* offer a reasonable level of transparency when it comes to data usage. Most contributors realize that the data they contribute can be used for almost any purpose. In contrast it is less likely that users of the *Foursquare* application are aware that their *check-ins* are being used to target advertising at their friends.

(V) **Active User Involvement**. Lastly, *active* vs. *passive* user involvement is of interest. The act of downloading an application and contributing data to it indicates active involvement in the user-generated content process. Alternatively, geosensors such as bluetooth, RFID tags or, CCTV cameras do not offer users the option of contributing data, but rather collect content generated by users. Existing research in this area has focused on related concepts of *opt-in* versus *opt-out*

[2] These are speculative numbers as Foursquare keeps this information private

provisions as they relate to crowd-sourced data (Harvey, 2013). A particular malicious example were trash bins in the city of London that would spy on the MAC addresses of mobile phones to determine the location and movement pattern of any citizens[3].

Given the above criteria it is possible to describe user-generated geo-content platforms in terms of their VGI-CoGI*ness*. Table 1 shows a sample of six such platforms along with a 0-2 scalar rating for each criterion. The total is calculated across all criteria with high values, maximum 10, indicating a high tendency towards VGI and low values, minimum 0, depicting applications leaning towards CoGIS.

Table 1. Scalar of sample platforms based on five criteria

Platform	**I**	**II**	**III**	**IV**	**V**	**Total**
OpenStreetMap	2	2	1	2	2	**9**
Flickr	1	1	1	1	2	**6**
WikiMapia	2	2	1	2	2	**9**
Foursquare	1	1	1	0	1	**4**
Google Map Maker	1	1	1	1	2	**6**
Brightest Flashlight App	0	0	0	0	1	**1**

3 Conclusions

The recent increase in user-generated geo-content has given rise to a vast number of platforms eager to provide tools for data contribution as well as data consumption. While methods for generating truly *Volunteered Geographic Information* have continued to thrive in this environment, a new division of user-generated geo-content has emerged, one that is not-so-volunteered. As commercial entities have realized the power of geographic information, so have they come to develop tools to ascertain this information. This paper outlines a number of ways in which VGI and CoGI differ and presents important distinctions of which users of user-generated geodata must be aware.

References

F. Girardin, et al. (2007). 'Understanding of tourist dynamics from explicitly disclosed location information'. In *Symposium on LBS and Telecartography*. Citeseer.

M. F. Goodchild (2007). 'Citizens as sensors: the world of volunteered geography'. *GeoJournal* **69**(4):211–221.

F. Harvey (2013). 'To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information'. In *Crowdsourcing Geographic Knowledge*, pp. 31–42. Springer.

J. Hong (2012). 'Analysis of Most Unexpected Permissions for Android Apps'. Online. http://confabulator.blogspot.com/2012/11/analysis-of-top-10-most-unexpected.html.

S. Mahmood (2013). 'Online Social Networks and Terrorism: Threats and Defenses'. In *Security and Privacy Preserving in Social Networks*, pp. 73–94. Springer.

K. Toyama, et al. (2003). 'Geographic location tags on digital images'. In *Proceedings of the eleventh ACM international conference on Multimedia*, pp. 156–166. ACM.

J. Van Grove (2013). 'Foursquare partners with Gnip to sell check-in stream'. Online. http://www.cnet.com/news/foursquare-partners-with-gnip-to-sell-check-in-stream.

[3] http://www.bloomberg.com/news/2013-08-12/snooping-garbage-bins-in-city-of-london-ordered-to-be-disabled.html