Access to data in 'data-poor' settings: Can VGI help?

Position paper for the workshop on the Role of Volunteered Geographic Information in Advancing Science: GIScience 2012, Columbus, OH

Jonathan Cinnamon, PhD Candidate, Department of Geography, Simon Fraser University
Nadine Schuurman, Professor, Department of Geography, Simon Fraser University

Volunteered geographic information (VGI) is broadly framed as a citizen-initiated bottom-up alternative to traditional centralized top-down geospatial data collection (Goodchild, 2007). As commonly conceived, VGI is created by distributed citizens interacting with geospatial technologies on a voluntary basis without oversight or the use of formalized data collection frameworks (Haklay *et al.*, 2010). This popular image of VGI is indeed exciting, and certainly, growing attention to it from the GIScience community might be attributed to the appeal of more open and equitable data. VGI created under these conditions is likely to have value for some uses, including for the development of crowdsourced alternatives (e.g. OpenStreetMap and WikiMapia) to government and corporate products. Despite this value however, this purely grassroots 'gold standard' ideal for VGI may restrict its use for scientific inquiry or decision-making purposes due to concerns regarding the quality and value of amateur 'asserted' data compared to official 'authoritative' data sources. To make citizen-volunteered data valuable for scientific purposes, we must envision new ways to enhance data quality; otherwise, VGI will remain an interesting, but ultimately minor phenomenon. With some ingenuity and flexibility, however, VGI could become a major source of useful data for science and decision-making.

The promise of vastly increased quantities of spatial data due to the VGI phenomenon holds particular promise for settings that are 'data-poor', especially those in the world's low- and middle-income countries (LMIC). Many LMIC suffer from very poor access to appropriate, up-to-date sources of data regarding numerous national development-related concerns – e.g. environmental, social, health, and economic issues – which means decisions are often made without evidence based on data. Even vital registration data – including births and deaths – are not collected in many LMIC, especially in sub-Saharan Africa (London *et al.*, 2002). Without data, evidence-based decision-making (EBDM) cannot occur in these settings and national development ultimately suffers. While many phenomena bypass low-resource settings, there may be unique enablers in the case of VGI that encourage its use. For instance, the increasingly widespread use of mobile technologies replete with their expanding geographic coverage (James, 2010), and the growing push to improve the availability of broadband Internet in these settings (Kim *et al.*, 2010) could seed the popularity and use of VGI in LMIC. The main challenge to harnessing VGI for science and decision-making in LMIC is the same as in high-income settings: the issue of data quality.

We argue that a broader conceptualization of VGI that includes data creation with an appropriate degree of oversight may serve to enhance the potential of VGI for science and decision-making. Although this may reduce some of the appeal of VGI as a singularly grassroots initiative, the increasing availability of quality data for LMIC and other data-poor settings may outweigh this drawback, given the societal benefit the accumulated data would create. Although there have been some VGI quality assessment efforts that use algorithmic and 'wisdom of the crowds' techniques (e.g. Haklay *et al.*, 2010; Mummidi & Krumm, 2008), modified approaches to VGI itself may also be a solution. For instance, Seeger (2008) describes 'facilitated VGI' (fVGI), an assisted voluntary data collection model in which a participant group is asked to volunteer spatial information according to predefined questions or criteria, using either face-to-face or distributed approaches. Ostensibly, this method could improve data quality, especially if the participants are chosen due their knowledge of the specific phenomenon. Further, the data created should be suitable for its intended purpose, given that it is volunteered according to specific criteria.

Our ongoing research project on injury data collection and injury prevention in Cape Town, South Africa has begun to use fVGI as a source of injury location information. As this project strives to identify and characterize injury hotspot (high-incident) locations to allow for decision-making on injury prevention, we require current and reliable data on where injuries occur. This information was unavailable in Cape Town and our prior efforts had limited success in collecting injury location data. We decided to harness VGI to acquire this information, however, we needed quality data (credible, specific) to serve our purpose. Thus, our fVGI approach utilized a knowledgeable participant group: emergency medical services (EMS) paramedics. EMS workers regularly attend the scene of injury events and thus they have knowledge of high-incident locations in the city. Participants were asked to interact with a map interface, and volunteer their opinion of injury hotspots by clicking on the map. This face-to-face interaction is likely to have promoted honesty and integrity of the submissions, and thereby also likely improved the quality of the data created. We believe that the data created are valuable and of sufficient quality to complement the larger injury data collection and prevention project.

Our data collection approach points to the potential value of *hybrid* approaches to data collection. Two hybridities are identified by Goodchild (2009) that potentially could improve the quality and value of volunteered data: 1) the merging of authoritative and asserted data sources; and 2) the coming together of neogeographers and those with formal training in geospatial technologies. Combining the successful elements of authoritative (e.g. formal coordination, oversight) and asserted (local knowledge) data collection, and the proficiencies of amateur and professional geographers may be an exciting prospect for fostering access to quality data in traditionally data-poor settings. Perhaps, taking a pragmatism-inspired 'if it works it has value' perspective may serve to increase the utility of citizen-volunteered data for science and decision making; that is to say, we should be seeking ways to make volunteered data more useful and beneficial for society, even if this means straying somewhat from its original framing as a purely grassroots phenomenon. Recent literature has noted this potential, including Girres & Touya (2010) who observed the need to strike a balance between user freedom and the importance of following specifications for contributors to OpenStreetMap. Although this is a fruitful area for inquiry for VGI – and spatial data more broadly – it must be noted that hybrid approaches will come with their own set of challenges; notably, questions of access and ownership of data created by multiple parties.

Harnessing VGI for science and decision-making in LMIC could help to close the divide caused by data access limitations. So far, most VGI is of indeterminate value because it is created without quality assurance mechanisms. Finding ways to improve the quality of VGI should be a major focus, including through exploring the potential of approaches that combine citizen efforts with a degree of 'expert' oversight. These approaches are especially relevant in LMIC where data are traditionally scarce.

References

Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS, 14*(4), 435-459.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211-221.

Goodchild, M. F. (2009). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services, 3*(2), 82-96.

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *Cartographic Journal, 47*(4), 315-322.

James, J. (2010). Penetration and growth rates of mobile phones in developing countries: An analytical classification. *Social Indicators Research, 99*(1), 135-145.

Kim, Y., Kelly, T., & Raja, S. (2010). *Building Broadband: Strategies and Policies for the Developing World*. Washington, D.C: World Bank.

London, J., Mock, C., Abantanga, F., Quansah, R., & Boateng, K. (2002). Using mortuary statistics in the development of an injury surveillance system in Ghana. *Bulletin of the World Health Organization, 80*(5), 357 - 364.

Mummidi, L., & Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal, 72*(3), 215-227.

Seeger, C. (2008). The role of facilitated volunteered geographic information in the landscape planning and site design process. *GeoJournal, 72*(3), 199-213.