Barbara Poore

**Mapping the Unmappable: Is it possible, ethical, or even desirable to incorporate volunteered geographic information into scientific projects?**

> We should be mapping information that in some ways has been historically unmappable because it is 1) not valued or is 2) actively seen as threatening or is 3) simply too hard to map using traditional tools.—Anselm Hook, WhereCamp 2009, quoted in http://www.ugotrade.com/2009/06/02/location-becomes-oxygen-at-where-20-wherecamp/

> Participatory approaches did not originate as a methodology for research, but as a process by which communities can work towards change.—Pain and Francis 2003: 46

Is "volunteered geographic information (VGI)" a good label for a crowdsourced phenomenon such as OpenStreetMap? This question, which has been asked by a number of researchers and bloggers (Haklay 2008, Gorman 2010), could and should be debated at this workshop as it exposes a number of potential social, ethical, and cognitive issues in the use of so-called volunteered geographic data in pursuit of science.

The enthusiasm over volunteered geographic information (VGI) among academic geographers has been little short of remarkable, manifested in numerous journal articles, paper sessions at conferences, and a specialist meeting sponsored by the National Center for Geographic Information and Analysis (NCGIA). Recently, several National Mapping Agencies have debated whether VGI might be incorporated into national map products and spatial data infrastructures (SDI). My agency, the U.S. Geological Survey (USGS), is currently undertaking an experimental project on this topic which will be described at the workshop. The term "VGI" has likely caught on within the academic community not only because it originated from an authoritative source (Goodchild 2007) but also because it rolls off the tongue with ease. In many respects, however it is misleading because it does not distinguish between very different types of user-generated content (UGC).

Not only do many researchers gloss over these distinctions, but pay scant attention to the appropriateness of the term, nor to the implications of UGC for science. The following taxonomy of different types of UGC and methods of mediation, with examples, was inspired by remarks by Don Cooke of TeleAtlas at the NCGIA meeting (Glennon 2008).

| Type | example | mediation |
|---|---|---|
| street data | sensor in car (consensual) | raw data |
| opinions | push-pin maps; Foursquare | software algorithms |
| citizen science | Audubon bird count | experts |
| involuntary | tag analysis (no consent) | experts |
| crowdsourced sdis | OpenStreetMap | software mapping platform |

This table is a work in progress, it does not capture everything about user types, nor does it make complete sense; further research is necessary. For example, there are categorical issues with the "type" column. Street data is a particular form of data while "involuntary" describes the state of knowingness of the user/generator. "Crowdsourced sdi's," a term I invented due to structural resemblances between some crowdsourced online mapping applications and spatial data infrastructures, would likely be resisted within the neogeography community, and so forth.

For exploring the questions of whether and how UGC can be used by scientists, important issues lie along three axes: social, ethical, and cognitive. On the social axis, one has to consider the differences among autonomous individuals with car sensors mapping streets, possibly for a vendor;

versus individuals collecting data under some direction from a scientist for the good of science; versus online communities "mapping the unmappable," as Anselm Hook (quoted above) reminds us.

The ethical issues involve questions such as: Who is doing the mapping? Do they know they are providing information? Do they know about/agree with what they are contributing to? What do they get back? In their article on participatory research Pain and Francis (above) note that motivations for community mapping projects are not directed at scientific research but toward change in the community.

The use of Flickr, the online photo sharing application, in academic research provides an example how all user-generated data are not created equal. In one instance, the Library of Congress uploaded a number of photographs from their American Memory project and invited users to contribute tags identifying the photos. Many of the tags were geographic, and the project was deemed a success (Springer et al. 2008). In other instances, researchers have made use of an aggregation algorithm that allows them to create polygons (called Alpha shapes) from location tags that users associate with uploaded photos. The resulting shapes, when mapped, fairly approximate the boundaries of the geographic phenomena the tags refer to (Cope 2008). Unlike the Library of Congress experiment, this use of VGI would be termed involuntary VGI since the taggers were not necessarily aware that their tags were to be used in such a way. One would have to carefully study the copyright agreements to which users assent, but I suspect most users just click "accept" without reading. In this case, there is little harm done because the specific identity of individual users is not disclosed. But more and more social media applications are linking location with identifiable users, and more and more cameras and user devices have embedded GPS. The temptations for researchers and perhaps the unscrupulous are mighty.

The final question is cognitive, concerning the difference between local knowledge and scientific knowledge. How is community knowing channeled into scientific data? What is gained, what is lost? These questions are not new, but citizen science is a deliberate attempt to blur established boundaries. Some types of geographic knowledge, such as those associated with extremely local manifestations of global climate change like backyard budburst, fall into Hook's category "too hard to map." They can only be collected by volunteers, but these data are extremely desirable from a scientific point of view. What sort of system, training, or mentoring will be required to move these types of data into databases? Will this change the way citizens think of science? The other side of this cognitive issue is what changes will be wrought in the scientific community and the scientific method with the widespread introduction of user-generated data? The changes might be as profound as those predicted from data-driven cyberinfrastructures (Timmer 2010). Philosophers of science will surely address these questions, but practitioners ought to be aware of the implications as well.

References

Cope, A. 2008. The Shape of Alpha. http://code.flickr.com/blog/2008/10/30/the-shape-of-alpha/ Accessed May 19, 2010.

Glennon, A. 2007. Research Issues on Volunteered Geographic Information, Group III. In *Workshop on Volunteered Geographic Information*. Santa Barbara, CA: National Center for Geographic Information and Analysis. http://www.ncgia.ucsb.edu/projects/vgi/docs/present/GroupIII_Glennon.pdf

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

Gorman, S. 2010. Why VGI is the Wrong Acronym. *Off the Map*. April 15, http://blog.fortiusone.com/2010/04/15/why-vgi-is-the-wrong-acronym/

Haklay, M. 2008. How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. *Environment and Planning B* (in press).

Pain, R. and Francis, P**.** 2003: Reflections on participatory research. *Area* 35, 46–54.

Springer, M., B. Dulabahn, P. Michel, B. Natanson, D. Reser, D. Woodward, and H. Zinkham. 2008. For the Common Good: The Library of Congress Flickr Pilot Project. Washington, D.C.: Library of Congress.

Timmer, J. 2010. Keeping computers from ending science's reproducibility. In *Ars Technica*. http://arstechnica.com/science/news/2010/01/keeping-computers-from-ending-sciences-reproducibility.ars