Position Paper for the Workshop on Volunteered Geographic Information, 13-14 December, 2007

Chris Rewerts
Research Scientist
US Army Corps of Engineers
Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
Champaign, Illinois, USA

The issues surrounding volunteered geographic information (VGI) are widely varied. In preparing this position paper for the December 2007 VGI workshop, it is assumed the reader is has previously read Michael Goodchild's introductory paper, "People as Sensors: The World of Volunteered Geography".

The Army has an extensive and intensive strategy to expand use and capabilities of geospatial data and information. It may seem that the formal and institutionalized types of development and uses of geospatial data preclude that VGI would be a topic of import, but I don't think that is the case. Some of the specific areas particular to research that I have been involved with recently are in the domains of ecology and the battle space mapping.

In the case of ecological research and management there are a many types of data that are not literally "volunteered", but they are being developed and submitted by persons whose expertise is not data development, sharing, archival, or documentation of geospatial metadata. We have been able to observe this in the participation with a Department of Defense funded long term ecological research program involving many researchers with projects focused on the area of an Army installation for over ten years. As part of the overall program part of our participation was the creation and management of a data and knowledge repository. Researchers were directed to submit data, metadata, results, reports, articles, and any other relevant products of their projects to the repository. The repository was then envisioned to serve not only as an archive, but as a resource for the installation and other researchers.

For example, a field researcher may collect data with a geospatial component as part of their research. These data will be analyzed, perhaps combined with other data and processes, with the typical outcome being center on publishing a report or journal article. Since most data are now in digital form, it is a potential waste that not only the new data, but the analysis portion of what was done to the data were reposited in such a way as to preserve the provenance of the final results reported.

In the process of managing the ecological data repository a number of lessons learned are relevant to VGI. One of the most likely avenues for technology to begin to address this problem seems to be more automated means for data and data products to be self-documenting and self-describing. Firstly, we found researchers were unknowledgeable or

unwilling to provide proper metadata. This is understandable since geospatial metadata standards are rather onerous to those not familiar with them, as well as the difficulty that some of the data being provided were not typical geospatial products, for which standards are even less routine. What seems to be needed, and what also seems to be evolving, are more automated documentation and integration of metadata generation by the tools used to collect and store data. For example, cameras integrated with GPS (Global Positioning System) that records time and location with the other photography parameters as part of the digital image. Next, data are rarely in such cases by themselves, but processed with statistical software, combined with other data, analyzed with GIS (Geographical Information System), modeling, or other software, then visualized in some manner to be presented as part of the new knowledge or results reported or published as part of the end products of the research. Now that so much of data and processes of research being digital, there may be much to be gained by focusing not on the resulting journal articles of research, but also on the provenance of the results. Thus, not only should the original metadata be automatically collected and managed, as suggested previously, but all the processing of the data to its final state as presented in the articles. The rudiments of this are present in some cases, such as a GIS that records information in new data layers about their sources and analysis process that created them.

Since data collection protocols are often not standardized, especially for individual research projects, it can often be difficult to combine data on similar subjects collected in different places with different method. Thus, one of the challenges of VGI is the ability to sensibly aggregate data from differing sources.

Another source of Army examples of challenges similar to VGI are data collected by military operations in the battlespace. In some of the examples of data sources of VGI with "people as sources", there are many types of information being collected by soldiers and sensors that could be more widely useful if better and more automated means of providing means to assess quality, to self-document and describe, and aggregate with other sources and existing data.