

Response to: CALL FOR PARTICIPATION: SPECIALIST MEETING ON  
VOLUNTEERED GEOGRAPHIC INFORMATION

Ben Lewis, Harvard University

I am interested in the phenomenon of the collaborative development of public goods, and am particularly interested in those which are GIS related. I have worked professionally in the GIS field since the early 90's and since 2000 have followed closely the development of the open source GIS software movement and also the evolution of open GIS protocols. In 2001 I started a website [opensourcegis.org](http://opensourcegis.org) to help keep track of the growing number of projects in this area. Traffic on the site has grown steadily since it began. In 2002 I developed an open source peer to peer web service sharing tool called ROMap which I demonstrated with the OGC at the United Nations Summit in South Africa. I have recently developed an application called Geonomy which is an experiment in collaborative dataset development built on open source software. Recently I have taken a position at Harvard's Center for Geographic Analysis.

I believe that the open source world provides guidance for the development of volunteered geographic information in proven strategies for balancing a distributed contributor base with the need for quality control. Wikipedia also provides lessons in this area, and there are others. For a number of reasons I believe that open source GIS components are a particularly good way to build tools to enable the development of geospatial datasets. Such tools will need to evolve rapidly with input from the best, most motivated programmers from around the world and open source provides proven mechanisms to enable that. Open source applications naturally gravitate toward open standards and the more recent ones tend to be highly web compatible. Open source also lowers the barrier to deployment of applications which enable the development of specialized spatial datasets.

As an experiment in making masses of public GIS data available in an easy to use way using current technologies and existing data, I developed the application [geonomy.com](http://geonomy.com) together with a friend. In this application we do several things which I think pertain to this meeting. First we geocoded and tagged the English language version Wikipedia. We performed a first cut by parsing looking for lat/long signatures. As it turns out there are more than 30 ways in which people geocode pages in Wikipedia. This allowed us to map some 60,000 wikipedia pages.

( <http://www.geonomy.com/geonomy/viewHome.do?lat=39.124&lon=-94.591&zoom=2&tagName=Wikipedia> ) That is a large collaboratively built GIS dataset. It is quite likely the fastest growing such dataset, apparently doubling in size every 6 months. Of course there are other ways to Geocode unstructured data as well a next step for Geonomy.

The Geonomy platform provides a number of other features which make it a potential platform for volunteer geographic data development. Users can add their own point features to the system, associating a location with a description, tags and URL. Users can add tags to their own features and to existing features. The system supports the insertion

of remote web services so that virtually any dataset can be displayed as a backdrop. The system also supports moderation so that an administrator can decide whether to publish data which has been submitted or not. Data which has been thus collaboratively assembled can in turn be made available as a service or KML document for others to use in their own system, closing the loop and making for a service oriented architecture.

Geonomy is itself built from open source components such as Mapserver and PostGIS. The client will be reworked to be basemap agnostic so that it can support others in addition to Google Maps and WMS. The entire Geonomy platform is currently in the process of being turned into an Open Source software project, soon making it possible for anyone to download and install (and improve) a Geonomy instance.

I think it will be partly through applications such as Geonomy that we will be able to accelerate the development of collaboratively built datasets which are useful for humanity. These systems will evolve based on what works in the real world to include appropriate mechanisms for ingesting data and for finding the right balance of automated collaborative quality control (where there is no individual doing a final check), versus mixed quality control where a steward or stewards are involved in checking and verifying data before it is published. Different datasets will require different levels of control. Determining the level of expert involvement should be based on factors such as the type of data being checked, the way that the data will be used, the cost of the final product, and other factors. An open source platform will allow groups to design systems that best fit their needs with a low barrier to entry.

An interesting question: Why do people volunteer? Many people who contribute to open source applications and to Wikipedia are not volunteers, but are paid by a company to create or edit content/code. Content which is contributed is made available to all under an open source style license. I would argue that license definition is a key component in the design of a strategy for volunteered data development.

Why do people contribute to projects which give their creation away? As Eric Raymond famously pointed out in *The Cathedral and the Bazaar*, when it comes to contributing to open source software projects, that people are often motivated to “scratch their own itch”. Because the itch is scratched in a way that can be replicated for free over a global network, at no cost the scratched itch becomes part of something bigger, that is, a software application or an encyclopedia. Because the network also allows anyone to cheaply check the person’s work and improve it, (again scratching their own itch), we end up with a model which supports the development of important public goods of reasonably goods quality by harnessing the power of individuals acting in their own self interest.