

Crowdsourcing the Public for the Public Sector

By Raphael Sussman, Coordinator, Land Information Ontario, Ministry of Natural Resources

For years, Conservation Authorities throughout Ontario have asked the public to help them track rare bird sightings, water levels and other environmental indicators. Long before Google, Grand River Conservation Authority gave out maps to volunteers and asked them to mark and track Bald Eagle nests as they appear and disappear. Crowdsourcing



People are always looking.

is not new; using it to acquire and maintain geospatial data is actually an old idea. What has changed is that today's Internet-savvy people are both able and willing to interact on a massive scale and with much greater ease.

Crowdsourcing is when an organization makes an open request for a group of people with an interest in a particular subject area to voluntarily complete a task typically performed by a paid individual. The Crowd can be anybody and everybody.

Geospatial data and crowdsourcing seem to go hand in hand. Anyone in the world with a browser and some free time can look around, collect, and share information about their surroundings. Web-based systems such as Google Maps, Bing Maps, as well as Global Positioning System (GPS) service suppliers, now accept information from their clients to improve the quality of their data products and applications. People are making the geospatial data that they use in their daily lives better, together with other users and their data supply companies.

The traditional model for any data owner, in either the public or the private sector, is that they are responsible for creating and managing their data to a known standard of quality (accuracy, currency, and completeness). The data owner is also responsible for verifying and validating any changes to data that they distribute or use themselves. Any suggestions for changes to data content would normally spawn an examination of the geography in question by the owner, followed by a modification of the data, if deemed necessary by the owner. Many organizations see crowdsourcing exclusively as filling this change detection function and use the crowd-based feedback as a means of focusing their ongoing data improvement efforts, rather than systematically reviewing all of their data holdings looking for possible problems.

The modern ideas that led to crowdsourcing have percolated up through the Internet culture where information and functionality is expected to be open, free, shared, and created by

anyone; it reinforces the concept of community. Taking advantage of an environment where individuals are connected, communicating and seeking information all of the time creates a platform ripe for exchange and ready for true interaction and working together.

If you want to build and maintain a data set through crowdsourcing, it is essential to create an

environment that includes a critical mass of participating individuals in order to be successful. Internet culture is diverse, helpful and accessible but most importantly, open and available to *anyone*. The crowd can range from new and inexperienced participants to experts and professionals, with the quality of data contributed varying accordingly. Both expert and amateur alike need to be targeted in any campaign, but knowing specifically who is in the crowd goes a long way to providing a known level of confidence in the way in which the data should be applied.

There is an expectation from the contributing individual, often unspoken, that their time and information has value and that they will get something in return. Whether it is the thrill of contributing or improving a service that they find valuable, acknowledging their contribution is important. Success requires that there is a perceived benefit to sharing knowledge.

How crowdsourced data enters the system and is subsequently acknowledged before being distributed is possibly the most important question for an organization interested in maintaining their data through crowdsourcing. Processes and methods for validating the results require consideration before embarking on any project, but are much more significant when the sources of the data differ dramatically. To complicate matters, there is often an expectation by contributors that contributions will be accepted or rejected by the crowdsourcing community at large. When considering the idea of using the crowd to manage your data, some of the questions in your analysis could include:

Who validates data changes?

- Crowd validation
- Subject matter experts or retained staff
- No one
- Algorithm based on voting or ranking

Who is in the crowd?

- Knowing and ranking individuals or contributing

organizations can help score contributions and give some contributions more authority

How are submitted changes incorporated?

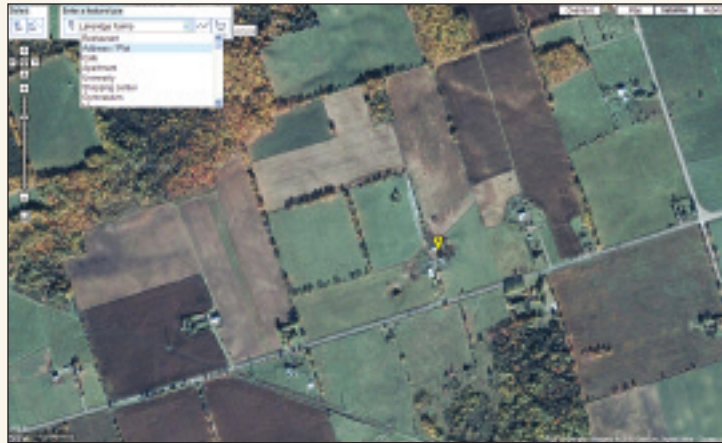
- Automated and/or manual processes
- How often is the content re-entered, transformed, or re-formatted?

Benefits and Challenges

There are many benefits to engaging a crowd to create, update and upgrade a geographic data set to be used and/or distributed by a public organization. First and foremost is the potential of reduced cost through the expenditure of less resources to achieve equal or better results. The budgets of most public organizations are being reduced, but without any changes in public expectation around providing the same or better service. Crowdsourcing is one potential way of balancing fiscal restraints and user expectations.

Distrust by the public often plagues public sector organizations. When the public contribute voluntarily they are actively participating in solutions, and more involved in their government's decisions. Removing the "Us-versus-Them" attitude by giving the public an actual role in making policy by contributing what they know and having it accepted as needed by their government, is a huge potential change.

It is logical that data reviewed by more people more often will lead to data that is more current, accurate and complete. This is especially true for base data. It may lead to fewer, and better, data sets rather than competing and conflicting data. Building on the idea that users are more likely to use their own crowdsourced common base data, is the compounded benefit of more easily integrated, reliable thematic data.



Screen capture of Google Map's edit feature.

Crowdsourcing is not an easy ride. A public organization faces the increased liability of decisions that were influenced by data collected in such a non-traditional method. There may also be an incorrect public perception. For example, the distinction between producing and coordinating is often not obvious. The public has a high expectation that data and information produced by government is authoritative, and they may not distinguish between data

maintenance and coordinated data maintenance. Public engagement can also be a double-edged sword. The public may feel upset when changes they propose are not incorporated immediately, or at all.

The current model for any organization, especially the public sector, of creating and maintaining all of the geographic data the public consumes, is not sustainable. Changes happen too quickly for any one organization to be

aware. Public sector budgets will continue to be reduced, resulting in fewer resources, and priorities will be on action rather than research and data collection.

After all is said and done, crowdsourcing can work. Geographic data can be made more current, more accurate, more complete, and at a lower cost. But before crowdsourcing can become reality, it is necessary that there be a fundamental change in the relationship between the public and the public sector.

Raphael Sussman is a professional Geographic Information Systems (GIS) Manager, Urban and Regional Planner, Landscape Architect, Project Manager, and Ontario Land Surveyor. He worked for many years in municipal government, specializing in Enterprise Architecture and GIS, and now coordinates Land Information Ontario. Raphael can be reached by email at: raphael.sussman@ontario.ca.

Calendar of Events

July 12 to 16, 2010

ESRI International User Conference

Geography – Opening the World to Everyone
San Diego, California

www.esri.com/events/user-conference/index.html

July 26 to 30, 2010

GeoWeb 2010

Going Real Time
Vancouver, British Columbia
<http://geowebconference.org>

September 14 to 17, 2010

GIS in the Rockies 2010

Loveland, Colorado
www.gisintherockies.org

September 28 to October 1, 2010

GIS-Pro 2010

URISA's 48th Annual Conference
Orlando, Florida
www.gis-pro.org

October 28 to 29, 2010

Geomatics Atlantic 2010

"Surveying the Past, Mapping our Future"
Fredericton, New Brunswick
www.geomaticsatlantic.com

November 17, 2010

GIS Day

"Discovering the World Through GIS"
www.gisday.com