

CROWDSOURCING IS RADICALLY CHANGING THE GEODATA LANDSCAPE: CASE STUDY OF OPENSTREETMAP

Steve Chilton
Chair of the Society of Cartographers
Middlesex University, Fenella Building
The Burroughs, Hendon, London NW4 4BT, UK
steve8@mdx.ac.uk

Abstract

Our social world is changing rapidly, and cartography is no exception to this. This paper expands on this by examining the effect it has had on data collection for mapping, using OpenStreetMap's crowdsourcing of geodata as a case study.

There are parallels to what is happening with data collection in other aspects of modern life. Wikipedia has changed the whole knowledge landscape, whilst applications like Flickr and You Tube have similarly allowed the easy storage and dissemination of photographs and videos respectively.

In today's society there is a need for instant information, particularly in crisis situations. Hurricane Katrina, which hit Louisiana with such effect in 2005, showed the problems that inadequate map data can have for aid and emergency work. Similarly there has been a pressing need for up-to-date map data to analyse and understand modern conflicts like the one recently in Gaza.

Various developments have created an environment that has allowed the collection of geodata to move from National Mapping Agencies and major commercial data providers to what is now sometimes called User Generated Content providers – or crowdsourced data collectors. Some of these changes have been: the unscrambling of satellite signals, cheaper domestic GPS units, the availability of satellite imagery in the public domain, and the availability and ease of use of open source development tools.

The OpenStreetMap project is the leading global example of the effectiveness of crowdsourcing of geodata. It has been so successful that it is challenging both traditional GIS systems and map data providers, by providing an easy to use, up-to-date and readily available data and map service to the GI community. OpenStreetMap has over 140,000 contributors, and in many parts of the world already has compiled more detailed mapping than traditional mapping suppliers. As an example, the database includes for Germany alone over 34,000 car parks, 25,000 WLAN hotspots, and 9,500 restaurants, all accurately positioned and displayed via the web map interface. OpenStreetMap was the first online mapping service to accurately map and display the

new London Heathrow Terminal 5 and it's new access routes by road and rail – on the day of it's official opening.

It is no coincidence that major commercial data providers now have community data collection and/or feedback systems in place, e.g. Google (Map Maker), Tele Atlas (Map Insight), and NAVTEK (Map Reporter). Recently OpenStreetMap was one of 5 mapping-related websites listed in the Guardian newspaper's 100 top sites to watch in 2009. Just some of the uses already being made of OpenStreetMap data are: the award-winning OpenCycleMap.org; it's use by aid workers in Baghdad; in online routing services at OpenRouteService.org and in specially formatted output for downloading to Garmin devices. Two companies have been formed, in Germany and USA, to add value to the output by providing customised OpenStreetMap datasets.

Other commentators have labelled these changes, and more importantly the innovative mapping outputs that can be achieved using the data, as being part of a Neogeography, and even suggested that it has been the Renaissance of Geographic Information.

Introduction - what is crowdsourcing?

One definition of crowdsourcing is that from Wikipedia, which is “a distributed problem-solving and production model. Problems are broadcast to an unknown group of solvers in the form of an open call for solutions. Users--also known as the crowd--typically form into online communities, and the crowd submits solutions. The crowd also sorts through the solutions, finding the best ones.” Another which comes closer to describing what happens in the data arena is “Using the general public to do research or other work, which may or may not be paid The term may also refer to user-generated content.” (from <http://dictionary.zdnet.com/definition/Crowdsourcing.html>, accessed 27th July 2009).

In some instances (for instance Amazon's Mechanical Turk¹) a system of “payments” is involved, and may be the driver behind people's involvement. In others the getting and sharing of knowledge is the driver (a good example being the Guardian newspaper's investigation into the MP Expense Scandal in the UK, The newspaper created a system to allow the public to search methodically through 700,000 expense claim documents, and over 20,000 people participated). In the field of geodata collection the ‘crowd’ is very often very narrow and even quite specific. For instance the suppliers of geodata to Satnav sellers will target the existing user base to try to get updates and changes from them.

OpenStreetMap – a classic crowdsourcing success

Wikipedia is often held up as the best known and most successful example of crowdsourcing on a global scale (incidentally the term itself is derided by Wikipedia co-founder Jimmy Wales), and has certainly changed the landscape of people's approaches

to what might loosely be termed 'fact finding'. I would like to suggest that the OpenStreetMap project² is going to have a similar impact in changing the geodata landscape, particularly in areas like the UK. As will be seen later the coverage that has been achieved, it's accuracy, it's availability and global impact are all changing they way individuals and organisations are thinking about the collection, purchase and use of geodata.

OpenStreetMap – the crowd acts

OpenStreetMap (OSM) was started in the UK by Steve Coast in August 2004, because of his frustration with the tight copyright that pertains to the Ordnance Survey (UK's National Mapping Agency) maps and data. In his view this was restricting him from producing a local map, and at a macro level stifling innovation by not making the map data available to users without excessive royalty payments. So with a consumer grade GPS he started collecting tracks around his local area of central London, and writing some reasonably unsophisticated software to display this data. He soon realised, after talking to people of a similar frame of mind (he was a research student at UCL [University College London] at the time) that there could be much more to this than he had perhaps originally thought. Shortly after presenting his idea at the EuroFOO conference he started getting others interested and contributing data in like mode. The mushrooming interest in the feasibility of actually mapping a significant geographical area (now spreading rapidly from it's UK source) is shown by the fact that within 16 months there were 1,000 registered OSMers. By now coders galore were being attracted to this novel project and editors, map display tools and other significant refinements to allow scaling of the work were taking place. Exactly three years in, August 2007, there were 10,000 registered participants. In March 2009 another milestone was reached when the figure of 100,000 participants was reached – an amazing increase in uptake.

Similar data can be shown as evidence of the exponential increase in the amount of data being recorded³. There are now people mapping all around the globe. Not surprisingly the amount of coverage does vary from country to country, and within countries. However, there are many areas – particularly in Europe – where the basic map coverage is pretty well complete, and people are using this data in some fantastically innovative ways. For the other aspects that really makes OpenStreetMap, and its superb geodata, stand out from other data sources are firstly that it is completely free of charge, and secondly that it is released under a licence⁴ which allows you to do pretty much what you like, as long as you mention the original creator and the licence, and that anyone else can do the same with anything you produce.

The varied extent of data coverage within OpenStreetMap could be a stumbling block. However, this is rapidly changing and is certainly very good in developed areas. Gorman (2008) noted an analysis that had been done by several OSM project members of the coverage of the world's capital cities – with the results mapped via Geocommons's Maker! website - which concluded that "OSM is slightly ahead of

Google/TeleAtlas worldwide and in Africa and Asia. In Europe, OSM is well ahead. Google is slightly ahead in Oceania, and well ahead in North and especially South America.”

The quality of the data produced by masses of volunteers with no professional training might well be challenged. However, work by Haklay (2008) has proved this an unfounded suggestion, and in fact in certain aspects (eg currency and accuracy of data) OSM was found in fact to be superior to that of comparable commercial data suppliers – in this case OS Meridian and Mastermap data. Haklay concluded that “OSM is better than Meridian 2 in terms of positional accuracy, and less accurate than Mastermap”. The availability, accuracy and price (free) of the OpenStreetMap data has led some local authorities in the UK to question the need to have total reliance on being locked into a contract for their geodata with the National Mapping Agency (OS). The release of UKMap’s large scale data⁵ has further challenged the former monopolistic situation in the UK.

The effect of availability of data

Crowdsourcing geodata, in the way that OpenStreetMap does, gives the possibility of really current data being available to users. What effects can this have? A simplistic answer is illustrated by the fact that if you were navigating to the new London Heathrow Terminal 5 on the day it opened in March 2007 and were using ANY of the proprietary map services then you wouldn’t have had a complete and accurate map to help you. With OpenStreetMap there was a complete and accurate map of the intricate new service and link roads available for that day⁶. Other services may have a large lead time for getting the data from survey to map output. At a slightly more serious level Maron (2007) noted that in the aftermath of Hurricane Katrina in 2005 aid agencies such as Red Cross were relying on Google maps for data, and without local knowledge were not aware of the demise of the US Route 90 bridge. As Maron points out “the issue is with the data providers, Navteq and TeleAtlas, whose business processes insert huge delays between reality and its representation catching up”. Crowdsourcing (and particularly local knowledge) could have avoided this, and various initiatives such as the Global Connection Project⁷ are working on improving this situation.

More recently aid work during and following the Israeli/Gaza conflict was hampered by the lack of up-to-date map data. The OpenStreetMap community stepped up to map Gaza. With GPS surveying out of the question during the height of the crisis, volunteers digitized geographic features from Yahoo! Maps aerial imagery over southern Gaza. In northern Gaza, OSM raised funds to purchase recent imagery from Digital Globe. Coordination for the effort largely took place through the OSM wiki. Through rapid internet based, collaborative research (or crowd-sourcing), a comprehensive catalog of Gaza data sources was compiled, made consistent, and applied to OSM. Once that the crisis had subsided, volunteers entered Gaza and worked with locals to complete the map. This is now part of a larger project called the Humanitarian OSM Team⁸.

There are many examples of the OpenStreetMap crowdsourced data being taken in interesting and innovative directions – often due to its very availability to researchers, hackers, government institutions and commercial companies. Just some of them will suffice to illustrate the point. There are specialist maps for cycling⁹, routing applications^{10 11}, skiing¹², topography¹³ and for maritime use¹⁴. There are many and varied instances of the OpenStreetMap map tiles being used as part of notable websites. They don't come much more notable than the US White House¹⁵ and the German Supreme Court¹⁶.

In addition there are several companies now building on the crowdsourced data by supplying OpenStreetMap based services, support and consulting. These include Cloudmade¹⁷, geofabrik¹⁸, Mapnik Consulting¹⁹, and Itoworld²⁰. This shows a serious belief in the value of the data and the business that can be built around it.

Other geodata examples

The realisation that crowdsourced data could have commercial benefits for geodata providers wasn't long in coming. The main global players have all instigated some form of user feedback/input system. Google has Map Maker²¹, which “allows you to contribute, share and edit map information for certain regions around the world”. TeleAtlas allows users to feedback changes through Map Insight²², and NAVTEK allows users to report new data through its Map Reporter²³ interface. However, in all cases there is a fundamental difference between contributing data through these portals and to contributing data to OpenStreetMap. That is that the terms and conditions all transfer the data to the company rather than remaining with the individual contributor. With OSM you can contribute AND do other interesting things with the data, with the examples mentioned you cannot.

Other examples of crowdsourcing of (or with) geodata abound in this ever changing world. One that is used a lot is that of radio stations augmenting their traffic updates with data supplied by people sending in SMS messages as they experience jams or roadwork delays. Another interesting case was the search for the crashed plane of noted aviator and explorer Steve Fossett in 2007. On September 8, the first of a series of new high-resolution imagery from Digital Globe was made available via Mechanical Turk. By September 11, up to 50,000 people had joined the effort, scrutinizing more than 300,000 278-foot-square squares of the imagery – but to no avail.

One could argue that one of the first examples of successful crowdsourcing was the Longitude Prize, which was a reward offered by the British government through an Act of Parliament in 1714 for a simple and practical method for the precise determination of a ship's longitude. Although many know the story of John Harrison gaining a significant reward for his chronometer solution, there were also several other awards to people from the crowd attracted by the task, who variously improved the chronometers, developed a lunar distance method and constructed a superior dividing engine.

Conclusion

We have seen that commercial companies are encompassing crowdsourcing and also that new wave companies are monetising crowdsourced data that is User Generated Content (UGC). But, what is the longer term situation and are there likely to be winners and losers. Mike Dobson (2009) asks “will the winners be a) established commercial companies that capitalise on UGC to augment their data? Or b) new competitors that commercialise UGC and augment these data to compete with established commercial systems?” It is actually likely to be the consumers of data that win, in that they are presented with a situation where large amounts of accurate and current geodata is available, and the proprietary suppliers will be forced by the changes illustrated in this paper to reconsider their policies on derived data and their pricing models.

Notes

- 1 <https://www.mturk.com/mturk/welcome>
- 2 http://wiki.openstreetmap.org/wiki/Main_Page
- 3 <http://wiki.openstreetmap.org/wiki/Stats>
- 4 Currently OpenStreetMap data is released under a Creative Commons Attribution-ShareAlike 2.0 licence, but this is very likely to change to an ODBI licence very shortly. For licence details see: <http://wiki.openstreetmap.org/wiki/Licence>
- 5 <http://www.theukmap.co.uk/>
- 6 <http://www.opengeodata.org/?m=200803>
- 7 <http://www.cs.cmu.edu/~globalconn/index.html>
- 8 http://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team
- 9 <http://www.opencyclemap.org/>
- 10 <http://www.openrouteservice.org/>
- 11 <http://www.yournavigation.org/>
- 12 <http://openpistemap.org/>
- 13 <http://toposm.com/ma/>
- 14 <http://www.openseamap.org/>
- 15 <http://www.whitehouse.gov/change/>
- 16 <http://www.bundesverfassungsgericht.de/organisation/anfahrt.html>
- 17 <http://www.cloudmade.com/>
- 18 <http://www.geofabrik.de/en/index.html>
- 19 <http://www.mapnik-consulting.com/>
- 20 <http://www.itoworld.com/>
- 21 <http://www.google.com/mapmaker>
- 22 <http://mapinsight.teleatlas.com/mapfeedback/index.php>
- 23 <http://mapreporter.navteq.com/dur-web-external/>

References

- Dobson, M., 2008, Data quality and neogeography, available to download from http://www.wun.ac.uk/ggisa/seminars/archive/autumn08_program/index.html, accessed 30 July 2009
- Gorman, S, 2008, OpenStreetMap vs. Google/TeleAtlas Street Coverage, <http://blog.fortiusone.com/2008/12/12/openstreetmap-vs-googleteleatlas-street-coverage/>, accessed 30 July 2009
- Haklay, M., 2008, How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England, submitted to Environment and Planning B.
- Maron, M., 2007, U.S. Route 90 .. Katrina destroyed that bridge, but it's still in Google maps! <http://brainoff.com/weblog/2007/04/25/1246>, accessed 30 July 2009