United Nations E/CONF.101/CRP16

ECONOMIC AND SOCIAL COUNCIL 19 July 2012

Tenth United Nations Conference on the Standardization of Geographical Names
New York, 31 July – 9 August 2012
Item 8(a) of the provisional agenda*
National standardization:
Field collection of names

Crowd-Sourcing Geospatial Information for Government Gazetteers

Submitted by Australia**

* E/CONF.101/1.

** Prepared by: Laura Kostanski (Australia), Secretary Committee for Geographical Names of Australasia

Summary

This paper outlines the results of a six-week Churchill Fellowship undertaken by Laura Kostanski to study methods for integrating crowd-sourced toponyms into government gazetteers. During the research trip in early 2012 Laura Kostanski visited with key government departments, agencies and organisations which are involved in pilot studying or researching methods for crowd-sourcing geospatial information. In addition to meeting with academics at University College London, McGill University, Nottingham Geospatial Institute, Carnegie Mellon and Stanford Universities, she met with representatives and staff of national and international mapping and gazetteer agencies including Lantmäteriet (Sweden), Ordnance Survey (UK), Permanent Committee on Geographic Names (UK), Kadaster (Netherlands), Canadian Geographic Names Board, United Nations Group of Experts on Geographic Names, United States Geological Survey and the United States Department of State.

Key Findings

Across the range of programs Laura Kostanski was exposed to during the Fellowship there were key themes which can be summarised as guidelines for any agency considering developing a crowd-sourcing application for geospatial information:

1. Focus on end-users
2. Don't rely solely on digital technology
3. Don't assume crowd-sourced data will fill all the gaps
4. Work with special interest groups
5. Create a new app, but also reach out to existing platforms
6. Develop strong internal-agency policies for quality assuring crowd-sourced data
7. Consider appointing 'knowledgeable-notifiers'
8. Create mechanisms for tracking uploads, changes and deletions
9. Give and you shall receive
10. Consider redefining data interests

Conclusions

Far from the potentially negative scenario anticipated by those unexposed to crowd-sourcing data techniques, the Fellowship learnings have demonstrated that Volunteered Geographic Information (VGI) has inherent strengths and weaknesses - but none so disastrous as to negate the potential success of integrating crowd-sourcing practices within existing government processes. This is the most important concept acquired during the Fellowship- that crowd-sourcing offers an opportunity for governments to redefinetheir geospatial data collection principles- to become more aligned with increasingly extensive end-user requirements.

Crowd-sourcing can provide data previously uncollectable by most government agencies. However, it is not a perfect and final solution. There are inherent biases and gaps in crowd-sourcing (similar to those which exist in current data collection methods) and because of these any crowd-sourced application should be developed to complement, not compete with, existing programs. Government being able to enter into the crowd-sourcing domain is predicated on the ability of agencies to freely distribute their data, or components of it, to participants and end-users. Open Data agreements are thus a necessity, to varying degrees, in this sphere.

The findings outlined in this paper are intended as introductory guides to some of the programs being piloted or introduced by naming and mapping agencies in various corners of the globe. While

not a definitive analysis, the paper does highlight some of the key trends and experiences which have common application across a range of data collection, maintenance and distribution processes for toponymic gazetteers.

Introduction

The Changing Role of Geospatial Information

This period in history could be defined as one of the greatest dreams of cartographers – for it is a time where we are witnessing the transformation of the map from an object which was purpose-built and limited to specific-use cases, to an application which in the digital age can underpin the majority of work in the government planning, emergency response, statistical analysis, automated navigation and business delivery domains. While digital map productsincluding Google Maps, Bing, Map Quest and applications such as Open Street Map (OSM) have brought the potentials of cartography to the forefront of most people's work or personal lives there has been a flow-on of effects into the world of cartography and the realm of toponymy.

The rapid uptake of digital mapping products by an ever expanding user base has witnessed an exponential rise in the demand for geospatial data. Where national mapping agencies were once the arbiters of officially-sanctioned geospatial information, the increasing demand for points-of-interest and open data have seen the proliferation of private companies entering into the geospatial data arena and the development of crowd-sourcing or Volunteered Geographic Information (VGI) systems. Where digital mapping companies were once purchasing data from national mapping agencies at regular intervals, some have now developed their own systems for harvesting data from the web and VGI enthusiasts.

Many reasons exist for these moves away from relying on officially sanctioned geospatial information from government sources, to seeking the crowd's assistance in keeping maps updated. Chief among the motives is the fact that the processes employed to collect, approve, maintain and distribute data by government agencies are to various degrees time-consuming and costly. There is a necessity for the laborious nature of government geospatial data governance arrangements, mainly being that the final product is considered highly-reliable and an accurate reflection of the community and/or government's perspective of the landscape. However, the consequence of these processes is that from the time of collection through to final publication the data is relativelyout-dated compared to crowd-sourced information.

In addition to this conundrum of information timeliness is the issue of gazetteer coverage. As mentioned in the *Four Faces of Toponymic Gazetteers* paper submitted by Australia for UNCSGN (refer tohttp://unstats.un.org/unsd/geoinfo/ungegn/docs/10th-uncsgn-docs/econf/E_CONF.101_57_The%20Four%20Faces%20of%20Toponymic%20Gazetteers.pdf) official gazetteers have tended towards indexing toponyms for classic-topographic features, such as mountains, waterways and buildings. Whereas, the major focus of online and mobile map applications is directed towards promotion of, and way-finding between, points of interest (such as businesses and service points). Therefore, the traditional government gazettes in most cases are not covering the wide range of feature types required for the latest mapping systems and community user-case requirements. The somewhat limited information coverage and issue of timeliness can be considered the major underlying factors predicating the development and use of crowd-sourced

geospatial data (in addition to the demands for government data to be open for use by the general public).

Investigating the latest VGI developments

Having researched the area of cultural toponymy for the past decade, and worked in applied toponymy for the Victorian Office of Geographic Names, Australia, since 2007, I was familiar with the processes involved in assigning official names to features and roads. My experiences in responding to, and attempting to resolve, the issues associated with emergency responders not being able to locate particular incident sites, or other service providers having their programs fail to reach key stakeholders, also meant that I developed a strong understanding of the gaps present in official geospatial datasets maintained by most government departments and agencies. These failures were due to many reasons- amongst which the use of unofficial, colloquial or slang toponyms by the community was a common theme.

I began to realise that for traditional data collection, quality assurance, maintenance and distribution systems to meet the real-world requirements of most end-users and stakeholders, they would need to be augmented with new processes for collating and distributing localised information. However, most government datasets are governed by legislation, regulations, policies and principles which emphasise processes for making data 'official'. These governance arrangements are critical to ensuring the quality of the datasets endorsed by government agencies and distributed for a multitude of internal and external stakeholder requirements. Thus, the consequences of potentially incorporating 'unofficial' information into 'official' government datasets can be profound.

Ultimately, I knew from experience at the Office of Geographic Names that the general community believed that when they marked-up changes to Google Maps (or similar applications such as Bing, Map Quest and OSM) that this information would be fed-back to government agencies (especially emergency services) and all associated mapping products consequently updated. However, as most naming authorities would recognise, this is not the case as most are private companies (with the exception of OSM) and there are currently no requirements for them to provide data change-request information to Australian government agencies.

I was of the opinion then, and remain so today, that to develop and provide relevant geospatial products to the community, government mapping agencies must provide opportunities for unofficial information to be incorporated into official datasets and there must be platforms for citizens and organisations to notify of changes required to mapping information, particularly toponyms and addresses.With this issue in mind I applied for, and in 2011 I was awarded, a Churchill Fellowship from the Winston Churchill Memorial Trust of Australia to study methods for crowd-sourcing unofficial toponyms (place names) and incorporating them into gazetteers (geospatial datasets) for improved knowledge transfer during emergencies or for ongoing planning and analysis purposes.

In February 2012 I set out to visit governments, agencies and organisations which I understood were experimenting with or applying crowd-sourcing solutions to geospatial data collection, maintenance and distribution problems. Over the course of 6 weeks in 6 countries I met with 28 groups and over 60 people to discuss their pilot studies and methods and took away lessons on the key positives, negatives, issues and potentials associated with crowd-sourcing of geospatial information. I learnt that this is a nascent area, populated with passionate toponymists, cartographers and geospatial

experts- whose work has only begun to understand all the permutations and variations involved in developing robust mechanisms and processes for engaging communities with geospatial information.

The final fellowship report is published on the Trust's website (www.churchilltrust.com.au) and I provide herein a snapshot of what could be considered the most pertinent messages absorbed during the journey. A note of caution in reading this report- the unique aspect of the Churchill Fellowship was the opportunity to discuss key developments, most of which are currently unpublished, with leading practitioners and hear their anecdotes (which don't usually make it into published formats), share stories and query the finer details of their strategies, successes and failures- to develop a comprehensive report on the 'current state of things', rather than a polished appraisal of the most recent academic literature in this domain.I therefore provide for the reader an unadulterated summary of the key thoughts, ideas, concepts and discussion items I encountered during my 45 day trip across the more northerly regions of the globe- with a promise for future peer-reviewed academic articles based on the lessons outlined herein.[i]

Key Questions & Lessons Learnt

Are government agencies considering using crowd-sourcing as a method for collecting geospatial data? Why? Are there key perceived strengths and weaknesses of this method?

All of the government agencies visited during the course of the Fellowship are considering and pilot studying approaches for crowd-sourcing geographic information:

- In Sweden Lantmäteriet have run a successful pilot study for crowd-sourcing unofficial names
- Kadaster in the Netherlands have been working with specialised interest groups to crowd-source change notifications for features
- Ordnance Survey in the UK have engaged with walking groups to update special interest geospatial information
- In Canada the Geographic Names Board are working with colleagues in Natural Resources to investigate approaches for extending traditional field-collection methods
- At USGS there is a pilot study being developed into an ongoing crowd-sourcing program which will potentially allow updates to be made to the national gazetteer

Within each of these projects there are two common patterns which have underscored their initiation and development. These can be summarised as follows:

1. *Impetus for developing crowd-sourcing platforms*

There is an increasing awareness that with the rise of GIS there has been an exponential increase in end-user demands for up to date and reliable geospatial information to be provided by government agencies. Where these data needs are not being met through government, private mapping companies are filling the gaps by sourcing data from other agencies. This creates an ongoing set of challenges for government mapping departments, particularly where there is a perception in the public that marking-up changes to private, online mapping services such as Google Maps, will necessarily result in all government mapping products being updated (particularly those used by Emergency Service Organisations (ESOs)).

In addition, traditional field-collection methods are either being removed (as is the case in the Netherlands, where remote-sensing is now the main data collection method) or existing datasets have been shown to be out of date, contain inaccurate information, or limited in scope to such a degree that key stakeholders such as ESOs cannot rely solely on the government datasets for information (ie policies usually dictate that only official toponyms be included in gazetteers, which excludes names

for commercial properties or general points of interest). Based on these reasons, interested government agencies are seeking new solutions for repositioning their geospatial information to supply the public with the data they are requesting.

2. *Cultural Change*

All government agencies visited during the Fellowship acknowledged that a significant hurdle to overcome was the acknowledgement of 'information gaps' in their current geospatial datasets. Because government agencies are usually tasked with providing accurate and official geospatial information, there is a common perception that recognizing data gaps could open a Pandora's box of problems- including making redundant all government geospatial data development programs and their responsible agencies.

From the information gathered during the Fellowship, it would appear that the key to overcoming these perceived issues is an acknowledgement akin to that from crowd-sourcing providers such as Open Street Map (OSM). In discussions OSM noted that the level of granularity and quality assurance provided by government agencies far exceeds that available from a crowd-sourcing platform. Therefore it can be strongly argued that in the high-quality micro-geospatial world there will always be a role for government datasets. Once this important fact regarding granularity and quality is acknowledged within institutions, from observations during the Fellowship, then it would appear that government agencies are successful in embarking on crowd-sourcing pilot studies and programs.

What lessons can be learnt from government and non-government agencies? And what are the limitations of crowd-sourcing data?

There are certain lessons which were common across the range of pilot studies and real-world applications observed during the Fellowship. These ten lessons and experiences could be summarised as guidelines for any agency considering developing a crowd-sourcing application for geospatial information.

1. ***Focus on end-users***

Research at McGill University, University College London (UCL), Carnegie Mellon University and Nottingham Geospatial Institute have all shown, and this is corroborated by most of the government agencies contacted, that when a crowd-sourcing platform is developed without consulting end-users the result is a highly-technical program which won't necessarily appeal nor be accessible to the vast majority of the population. Therefore, it is necessary to not only spend considerable development time consulting with IT programmers, but to also get out in the community and consult them on what they need from any geospatial platform.

2. ***Don't rely solely on digital technology***

Interestingly, with the rapid uptake of technologies such as laptops, smart phones and tablets, it could be assumed the latest and greatest developments in crowd-sourcing would be in the digital domain. However, most practitioners have realised that technologies can actually hamper the interaction of communities with maps, especially when they are asked to collaborate and mark-up information or changes. Limited access to high-speed and reliable internet connections, in addition to the training required to interact with most online GIS applications, means that *the common paper map and some stickers or pens are usually the most reliable and people-friendly tools a crowd-sourcer can use…..* A good example of paper maps came from Todd Huffman at OSM- *Field*

Paper Notes. In addition, examples provided by MukiHaklay at UCL, of community mapping programs, confirmed the idea that technology developments are critical, but it is important to supplement these digital ideals with real-world traditional tools.

3. Don't assume crowd-sourced data will fill all the gaps

As noted in recent research publications on VGI, crowd-sourced geospatial data is not the panacea to all the world's geospatial data problems. While traditional field-collection methods can be criticised for their often limited scopes, cultural hegemonic practices and timeliness, it would appear that crowd-sourced geospatial data can be criticised for demographic biases and information gaps. Of particular interest were the research results from the team at UCL, which demonstrated that

- Over a period of 6 months 80% of data collected for OSM GB was contributed by only 100 people (1 person contributed 8% of the data)
- In 2009, of 117,000 registered users for OSM, 71% never mapped anything and only 29% (33,452 people) did. Of people who contributed 44% only contributed once.
- 96% of OSM contributors were male, 49% had a university level of education and their average age was 20-40 years.

While work being undertaken at McGill University has demonstrated that women have a strong interest in developing and working with community crowd-sourcing platforms, and work at CrowdFlower has demonstrated that online micro-tasking can complement NGO literacy programs, there is still a bias present in the most popular online crowd-sourcing geospatial programs.

These biases mean that crowd-sourced geospatial data is often lacking in areas where the population have lower education levels and employment opportunities. While this can obviously be tackled over time through the development of specifically-targeted training programs (which in turn can provide many opportunities for under resourced communities), these solutions will take time to implement and reap the benefits of. Thus, in the meantime, a certain degree of caution needs to be exercised when contemplating the use of crowd-sourced data- it should be considered a complementary data collection program rather than a replacement one.

4. Work with special interest groups

One option for overcoming issues associated with user-fatigue and filling in data gaps is to focus on developing projects targeted at user-groups. Many organisations are working on pilot studies with special interest groups (Ordnance Survey, Kadaster, SWAP, USGS, GNBC). The rationale behind this is that groups such as bird watchers and hikers have an inherent interest in utilising maps which include details of their particular geospatial requirements- for example, walking tracks. This interest translates to enthusiasm to participate in crowd-sourcing programs which will result in a net benefit to the group or user.

5. Create a new app, but also reach out to existing platforms

There are an overwhelming amount of online resources, programs and apps out there. Because of the crowded apps market, one problem many organisations experienced was the promotion of their crowd-sourcing programs. In particular, the team at Lantmäteriet acknowledge that while it was important to develop their own crowd-sourcing program within their existing systems, their calls out to the public for information were often 'lost in the haze' of other competing programs and resources. As one person in the UK mentioned to me 'why would members of the general public come to a government mapping website to download one specific app which only focuses on notifying of data errors?' While this approach might appeal to geospatial data-philes, the appeal to the general populace is limited.

Thus, Lantmäteriet propose to work with popular online mapping programs- to exchange data supply for feedback received by these platforms from the general public. Something similar should be considered a worthwhile option for any agency who would like to extend their reach into the general community.

6. Develop strong internal agency policies for quality assuring crowd-sourced data

Yes, it is possible to quality-assure crowd-sourced data. Just as each contributor will have their own motivations for supplying information, so too agencies will have their own agendas, policies and legislation requirements to fulfil when processing the data. To establish how quality assurance processes might be performed it is essential to run a pilot-study not only in collaboration with contributors and end-users, but with the assistance and input from agency staff. It is the agency staff who will be able to readily identify key issues with the processes and develop policies for issue resolution to be applied across the crowd-sourcing program in the future.

7. Consider appointing 'knowledgeable-notifiers'

This is a process applied by USGS and OS within the government domain, and CrowdFlower in the open community realm. It acknowledges that certain members of the community can be considered trustworthy of supplying or quality-checking data uploads. Thus, rather than just relying on general information being submitted by random members of the community, there is an overarching system applied for appointing volunteers to positions of authority for scrutinising information and, in some cases, being able to approve it for submission to government agencies for their consideration.

Systems for this process can be automated, as that used by CrowdFlower, or based on participation status (ie for a national parks mapping program the Chief Ranger of a park could be considered to hold a position of authority for verifying information submitted by the public). However, as noted by Jeannie Stamberger at Carnegie Mellon University, there needs to be caution applied in making assumptions about people's knowledge or authority over particular domains- and the usual biases in approving or overlooking information are also evident in this method.

8. Create mechanisms for tracking uploads, changes and deletions

A good way for quality assuring information is to develop in-built mechanisms for tracking uploads, changes and deletions to data. Within OSM this mechanism is devolved to the crowd who self-regulate the maintenance and accuracy of data. Other systems, such as that piloted by USGS, provide nightly reports of changes and highlight instances where one participant has uploaded or changed feature information more than 'x' times. Systems developed at McGill University included requiring participants to provide their email details and notifying private business owners of crowd-sourced changes to their details (so that they could verify the information updates).

9. Give and you shall receive

Time and again the most common message received during the Fellowship was- *it is essential to give something back to participants*. The 'something' can be tangible objects such as data downloads, points, cash, rewards; or, intangible items such as communication (ie responding to social media commentary), building community cohesion, improved information for end-user work requirements (such as ESOs) and personal or group satisfaction. No matter what it is, the feedback

must be something valued by the target contributor community to ensure their initial or ongoing engagement with the crowd-sourcing program.

10. Consider redefining data interests

One of the key reasons for government agencies seeking to implement crowd-sourcing opportunities, is the acknowledgement that existing geospatial datasets have gaps. These gaps have developed over time for multiple reasons, one of which should be the impetus for a wholesale analysis of existing data capabilities- both the structure of data systems and scope of information collected.

Crowd-sourced data allows the opportunity for government agencies to reconsider how they store information (ie can gazetteers now include unofficial information alongside official information, or is a separate gazetteer of crowd-sourced data required?), what data is collected (ie given the capabilities of digital mapping systems, can polygons and lines be collected rather than just x,y coordinates?), and how information is referenced (ie can addresses be assigned and points geocoded?). All of these considerations should be addressed when an agency considers developing a crowd-sourcing application.

How can communities be motivated to participate in crowd-sourced mapping projects?

As mentioned previously, communities can be motivated to participate in various ways. Some suggested options include:

- Reach out to existing popular programs, such as Google Earth, OSM or similar online mapping platforms, and establish collaborative data exchange agreements
- Target specialist communities, such as walking groups or national parks users, who have an inherent interest in using the resultant data for their activities
- Engage in feedback-mechanisms. The adage that 'those who give shall receive' applies in the crowd-sourcing domain. Feedback can be tangible or intangible.
- Provide training programs, and don't hesitate to supplement digital programs with tried-and-tested methods of paper, stickers and pencils.

Can crowd-sourced data be quality assured?

Yes.

The most common perception of crowd-sourced data is that it is unreliable and therefore cannot be implemented into official government data collection, maintenance and distribution programs. However, all of the programs investigated during the course of the Fellowship had implemented forms of data quality assurance. These methods varied between three main themes:

1. Let the crowd self-regulate information (OSM model)
2. Appoint 'knowledgeable-notifiers' (An aspect of semi-manual USGS and automated CrowdFlower models)
3. Develop in-house data checking methods which rely on existing staff to qualify information before it is transferred into permanent datasets and redistributed (OS, Kadaster, GNBC)

Each of these methods has inherent strengths and weaknesses which are comparable to those extant within already established quality assurance models of government geospatial agencies.

What are the implications of developing crowd-sourcing platforms for government geospatial data?

Far from the potentially negative scenario anticipated by those unexposed to crowd-sourcing data techniques, the Fellowship learnings have demonstrated that VG) has inherent strengths and

weaknesses - but none so disastrous as to negate the potential success of integrating crowd-sourcing within existing government processes. This is the most important concept learnt during the Fellowship- that crowd-sourcing offers an opportunity for government to redefine their data collection purpose to become more aligned with end-user requirements. Crowd-sourcing can provide data previously uncollectable by most agencies. However, it is not a perfect and final solution. There are inherent biases and gaps in crowd-sourcing (similar to those which exist in current data collection methods) and because of these any crowd-sourced application should be developed to complement, not compete with, existing programs.

Conclusions

A vast array of information has been collected during the period of the Fellowship- ranging from theories to projects to policies- from key leaders in the domain of crowd-sourcing geospatial information. Of all the learnings listed, the key messages can be distilled to the following conclusions:

1. Crowd-sourcing is an excellent method for complementing/supplementing existing government geospatial data collection, maintenance and distribution processes.
2. Quality assurance of crowd-sourced data can be provided through different pre-determined mechanisms.
3. Crowd-sourced data is fallible and imbued with issues similar to those which already exist in the 'official' government geospatial data domain.
4. Governments need to consider the requirements of end-users and begin to increase the scope of their data collection and provisioning to incorporate those items which are most required by the community.
5. Government being able to enter into the crowd-sourcing domain is predicated on the capacity of agencies to distribute their data to participants and end-users. Open Data agreements are a necessity, to varying degrees, in this sphere.
6. GIS capabilities should be captured by government agencies, and geospatial datasets updated to reflect digital mapping requirements (ie provide addresses, include lines and polygons, assign geocodes).

Basic elements of crowd-sourcing acquired during the Fellowship have been discerned and distilled into the diagram provided below. The basic concept is that there is a distinction to be made between collecting geospatial data for specific emergency situations, and for running ongoing geospatial data collection and maintenance programs. Within both of these domains there are specialised aspects of running crowd-sourcing platforms, which in the perfect scenario would be combined to allow any government agency the opportunity to generate and receive as much information from as wide a community base as possible.

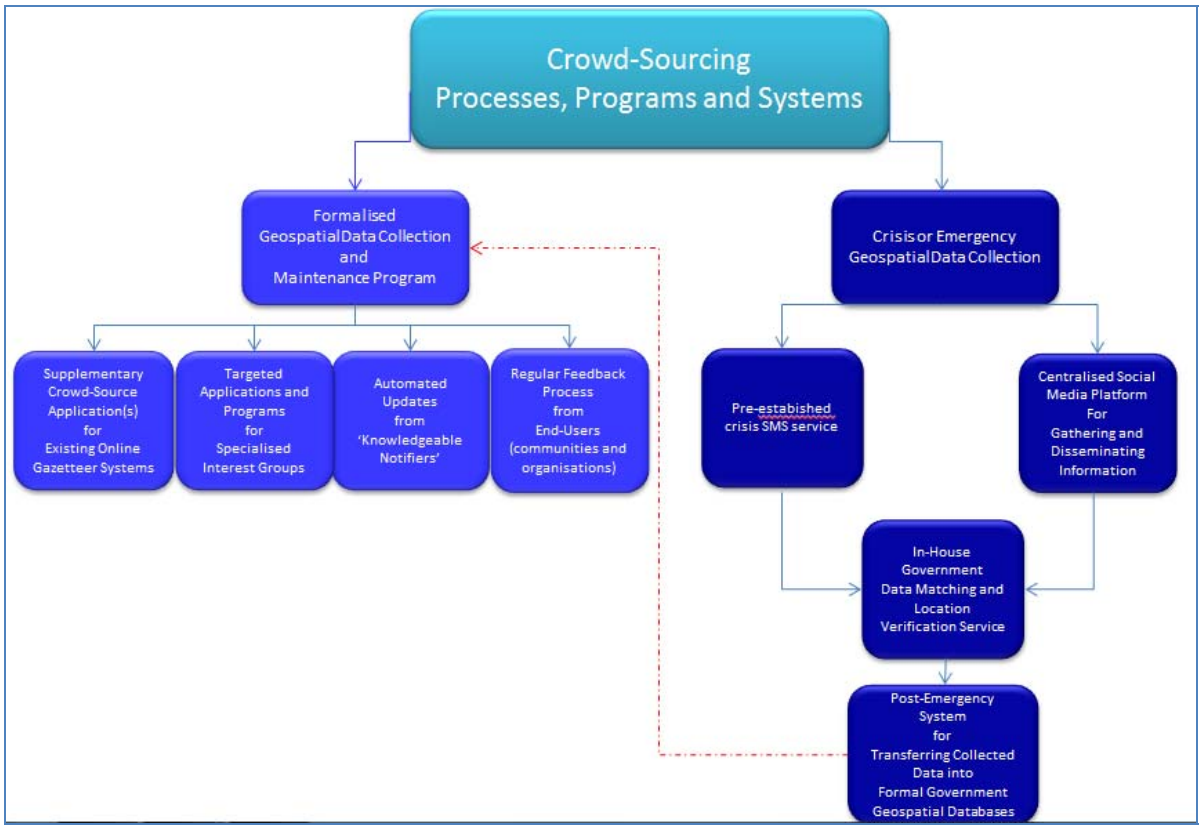

Figure Error! No text of specified style in document.**-1 Summary of methods for crowd-sourcing geospatial information**

Crowd sourcing offers an opportunity to keep up to date on real-world changes as they happen. It provides insights not available to traditional geospatial data collection methods. However, given gaps in crowd-sourcing information, there will always be a requirement for government agencies to produce and distribute 'official' data, which has a sound quality assurance process applied. There will always be a role for traditional data collection methods- but they will need to reconfigured or reengineered to suit modern technological developments.

In conclusion, I would like to state that this is an emerging field of technology, and I feel privileged to have had the opportunity granted through the Churchill Trust to travel and explore the developments at this nascent stage.

[i] A note on Biases and Errors- the information presented herein is my interpretation of information discussed with people who were often speaking openly and frankly about key developments and issues. I have chosen to not indicate the direct sources of information in the main body of the document as the details cannot be said to officially represent the viewpoint of any one person or organisation. I apologise in advance and accept responsibility for any unintentional misreporting. Additionally, I openly acknowledge that the Fellowship focused primarily on western countries, and aim in coming years to explore developments in other cultures.