Current status and future trends in crowd-sourcing geographic information

Giles M. Foody[1], Peter Mooney[2], Linda See[3], Norman Kerle[4],
Ana-Maria Olteanu-Raimond[5] and Cidalia C. Fonte[6]

1. School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK
2. Department of Computer Science, Maynooth University, Maynooth, Ireland
3. Ecosystems Services and Management Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
4. Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands
5. IGN France, COGIT Laboratory, 73 Avenue de Paris, 94160 Saint-Mandé, France
6. Department of Mathematics, University of Coimbra / INESC Coimbra, Coimbra, Portugal

June 2015

**EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY**

The last foresight report was written at a time of dramatic change. The report highlighted key issues and directions for the geographic information community. Over its medium term horizon, it saw geography as changing, notably with new players entering and shaping the geographic information sector, the increase in location-awareness, the growing trend for information to be available to the public and increasingly open approaches adopted. One particular area highlighted in the foresight report was the anticipated growth of crowd-sourcing geographic information.

Since the publication of the first foresight report in 2010 there have been considerable developments in the general topic that it outlined as crowd-sourcing. This activity is to some extent masked by the considerable variety of expressions used to describe the inputs of citizens to the geographic information sector. A wide variety of terms have been used, including crowdsourcing, neogeography, user generated content, and volunteered geographic information (VGI). These various terms are often used to help to differentiate between activity that is passive or active or perhaps truly volunteered as opposed to information provided for a modest, and possibly non-financial, reward. Here, there is no particular desire to distinguish between the different approaches, although the detail can be important, and the focus is simply on citizen-derived geographical data. The citizens may potentially be anyone, they could be children or adults, they may be amateurs or experts, they may have differing motivations and may even be contributing without knowing so.

With the proliferation of location aware devices and opportunities of web 2.0, it is now possible for citizens to easily acquire, share and use geographical information. This has had a revolutionary effect across a broad spectrum of activity from routine daily life applications through retailing to science. Resources such as Google Earth, Bing maps and even citizen generated maps through activities such as OpenStreetMap (OSM) are now widely and routinely used by diverse amateur and professional communities. While this trend is set to continue there is likely to be further growth, much linked to the provision of free or at least inexpensive data associated with the launch of new Earth Observation satellites and access to official government resources. These tremendous opportunities do, of course, come with challenges. The latter include dealing with a variety of concerns with the data. This includes the data deluge (e.g. the Sentinel 2 satellite due for launch will produce ~1 TB of data per-day and is just

one of over 350 Earth observing satellites are to be launched by some 40 different countries by 2023) but also problems linked to data emerging from variable sources, in inconsistent formats and often without reference to any form of standards. Moreover, the data generated may be poorly described and associated with little if any metadata. To realise the full potential of citizen sensing there is a need to establish good practices and perhaps even protocols for some activities. This will be a challenging task, not least due to issues such as the diversity of data sets generated, devices used and sensitivities to error and uncertainty. There is also clearly a strong desire to 'not kill the golden goose' by laying down rules and procedures that end up make volunteering an onerous task that ultimately deters the provision of citizen data. A variety of priorities to address have been identified including issues such as standardisation and interoperability (Brown *et al*., 2013), especially in relation to issues such as the INSPIRE directive, and groups are working on defining good practices to encourage mapping related applications. In particular, COST Action TD1202 is working on the identification of good practices and, where appropriate, protocols for the acquisition, description, storage, dissemination and use of citizen derived data in relation to common mapping applications. Before looking to likely future trends it may be helpful to first focus on some of the main aspects of current citizen sensing activity.

The field of citizen mapping is currently dominated by OSM but geographical information is acquired in a range of projects which may ultimately be mapped. The most-established citizen science projects that acquire geographical information are in the general area of ecology and conservation but the range of application areas is expanding rapidly, facilitated by recent technological advances. But even in these relatively long-established areas of activity there are strongly contrasting approaches and priorities. For example, the free tagging of OSM and lack of protocols contrast sharply to the rigorous protocols often found in ecologically-orientated citizen science projects, perhaps reflecting differences in the original purpose of the projects and the usability of the data for other applications. Geotagged photographs are also widely used as a source of geographical information. The number of repositories for geotagged photographs is rising rapidly, including popular social media, and these photographs may be used for a range of applications. For example, geotagged photographs may be interpreted to indicate the land cover at the location and used as reference data in the validation of land cover maps. The potential of such resources is, however, greatly limited by concerns such as the spatial distribution and nature of the data acquired. There are so many potential data sets and applications that may make use of citizen data it may be helpful to focus on some of the benefits and limitations in one growing area as an example: crowd-sourcing to aid disaster risk management.

The disaster domain has turned out to be an attractive field for citizen sensing, with the vast majority of projects focusing on the post-disaster response and management phase. There are good reasons for this development: disasters are exceptional, highly visible events that often generate tremendous compassion and generosity, with the provision of VGI offering an easy and non-monetary way for members of the general public to help. The aftermath of disastrous events, at least when measured by media attention, also tends to be of limited duration, hence VGI projects can be quickly established and long-term continuity challenges are not an issue. However, the disaster domain offers its own set of challenges. Ways for volunteers to contribute are manifold, as is the number of volunteer types. On one hand there have been many successes: voluntary mapping platforms such as OSM scored some of their most visible moments when disaster-torn places such as Port-au-Prince, Haiti, where comprehensively mapped in a matter of a few weeks following the 2010 earthquake. The data set this provided to the many disaster responders were immensely useful. Similarly, the platform Ushahidi provided an effective vehicle for people located in the disaster-affected area to report on the situation, be it on the state of roads, bridges or other critical infrastructure, or to file requests for specific assistance. The above examples highlight what volunteers are very good at: base mapping and reporting of local knowledge, with OSM combining both very effectively. However, VGI has also been used in several disaster damage assessment projects where remotely located volunteers mapped features such as structural damage, landslides, or temporary shelters

using remote sensing imagery. Some campaigns only made use of professional volunteers with a remote sensing background (e.g., the GEO-CAN campaign following the Haiti earthquake), while others allow anyone to participate (the most visible platform being Tomnod). A key issue of concern with these data sets is their quality. Research on the value of such contributions, and on the ideal approach to harness the assistance of volunteers, is ongoing, including in the COST Action TD1202. The main problems faced by volunteer-based damage mapping are how to identify suitable volunteers, how to instruct them, how to monitor (and when needed influence) their mapping, and how to integrate the contributions from many volunteers at a time marked by both urgency and a frequent lack of validation data. Research is ongoing on the modelling of different volunteers types (e.g., the able and well-meaning versus those that are challenged by the task, that are indifferent to the accuracy of the results, or even those that aim at sabotaging the campaign), but also on how to make optimal use of multiple damage labels for a given structure that individually may be of questionable accuracy.

Over the coming years more work will be required to maximise the utility of VGI in other phases of disaster risk management, such as for hazard assessment, monitoring of potentially hazardous situations, and early warning. The potential of volunteered information, especially when coupled with physical sensors, is enormous, but more work is needed to establish proper conceptual frameworks to generate meaningful and long-term contributions, much of this also requiring protocols, and clear guidance on how to engage and train volunteers. The disaster domain is clearly also a sensitive one, leading to legal and ethical concerns. These include ethical concerns related to post-disaster reporting of damage and potentially of victims. Volunteered information can also lead to the realisation of certain risks affecting a given area, which can influence property prices or insurance premiums. The above-mentioned Tomnod damage mapping platform highlights yet another ethical challenge: VGI is increasingly seen as valuable. Tomnod was recently bought by a large satellite operator, meaning now that the generous volunteered contributions by people trying to help in a disaster situation are at risk of being commercially exploited. For such situations clear transparency rules are needed.

Finally before highlighting anticipated future developments it should be noted that citizen data, although typically arising from amateurs, has potential impact on authoritative mapping bodies. While the activities of the various VGI communities has not substantially changed the way bodies such as national mapping agencies (NMAs) produce data, change in the future is anticipated. In particular, the economic models of NMAs are changing and need to be adapted to the new reality in which VGI is abundant by, for example, proposing paid services based on geographical data and not only data or 'win to win services'. To-date only a few NMAs are significantly engaged with VGI and typically using it only for change detection and error reports. More NMAs are likely to exploit the substantial potential of VGI when current barriers to its use, such as concerns on VGI quality and heterogeneity, legal and ethical issues, and crowd motivation and sustainability, are broken down.

Over the next five years it is anticipated that citizen derived data will grow considerably and be used in increasingly diverse ways. Given that the amount of spatial data available is increasing exponentially (Craglia and Shanley, 2015) and diversity of data sources and types is also increasing, one key issue will be the assessment of the fitness for use, which is intimately related to data quality and uncertainty. Data harmonization may play an important role in the era of big data, since it may enable data comparison, allowing the application of the law of big numbers (Kuhn, 2007) and contribute to an automated and fast preliminary data quality assessment and even data conflation. When multiple sources of data are available that may potentially be useful, methodologies also need to be developed to assist the users with the selection of a data set, or combination of data sets, for use in a specific application. Decisions such as these will be aided by the provision of information about the data and hence meta-data is likely to become increasingly important with citizen derived data sets. It is anticipated that there will be considerable emphasis placed on addressing the various concerns that exist with the quality of VGI. Projects may

generate their own quality assurance approaches to meet their specific needs. Similarly, bodies such as the NMAs may develop their own processes and methods for using citizen generated geographic data. Given the huge amount of data it is likely that there will be a focus on the development of automated approaches for the assessment of VGI quality. This will be challenging given the greatly varied nature of the data, which can be unstructured and heterogeneous, but essential for many uses.

Future developments in citizen sensing will also require greater consideration of the citizen as well as the end use of the data generated. A greater understanding of the citizen sensors is required as is a two-way dialogue with those using the VGI, especially as the citizens may be the source of useful ideas. Feedback to citizen contributors is likely to become important, especially in developing the citizen's skill and maintaining motivation. Real-world benefits and motivating reasons for citizens to participate in the acquisition of VGI need to be developed, ranging from calls to altruistic spirit and helping achieve a common good to gamification. There also need to be developments in relation to a set of legal and ethical issues. Some concerns are already evident, such as those mentioned above in relation to crowd-sourcing to aid disaster risk management. The legal and ethical issues may, perhaps, be particularly apparent when VGI is used by a legally mandated organisation such as an NMA. A series of important questions arise and need to be answered in the near-term. For example, in relation to the fundamental issue of legal responsibility, is this a matter for the citizen or the NMA, or indeed for both? Cho (2014:10) argues that there must be legal protection for volunteers in VGI data collection and projects. Otherwise "the ensuing litigation may destroy the VGI model before it reaches its full potential". The nature of the exact VGI information or data used and which use-case it is applied to may help to determine which legal, ethical and privacy issues are most prominent. When information about individual citizens is transferred and presented within a geographical context the resulting profile information could be both "highly revelatory and involuntary" (Scassa 2012:p5) and this can raise important ethical issues that need to be addressed. It is anticipated that VGI will increasingly be harvested from sources as diverse as social-media and wearable devices which while potentially yielding vast amounts of useful VGI, including human movement, it comes with a suite of concerns ranging from privacy to the legal and ethical issues touched on earlier. These are complex issues with, for example, privacy legislation appearing to lag behind technological advance and differing between countries.

There are also serious concerns with the re-use of VGI. In many instances, especially when VGI is mined from open resources, it may be used for applications the original provider is uncomfortable with. As the ability to integrate and fuse together greater numbers of complex and disparate data sets increases it is of crucial importance that the issue of data re-use is addressed. Data re-use also links to legal concerns. For example, if the VGI was acquired by digitising from a map or image without relevant permission, what are the implications to those that re-use the VGI?

One critical issue related to the diversity and quality of spatial data is the need to develop good practices. Here, there is a tension between the desire to encourage volunteers without constraints on their activities and the desire to acquire highly useful data. The latter could be aided by the specification of best practices or even protocols but if these become too onerous they may actually act to deter volunteers. So, for example, much current VGI is derived from geo-tagged photographs. The latter vary greatly in their value as a source of geographical information and the adoption of some basic good practices could greatly expand the value of the photographs while following best practices could help meet demands of some communities that have demanding data requirements. Thus, for example, the value of photographs to some applications could be enhanced by simple actions such as the encouragement of acquisitions from multiple directions to convey information on the homogeneity of the landscape as well as the provision of basic meta-data on the location from which the photographs were taken and their date of acquisition. It is likely that the cameras used by citizen sensors will help provide better photographs for derivation of VGI in the future. For example, trends in the photographic capability of mobile phones suggest that the

accuracy of geolocation will be enhanced and the development of 3D systems may provide a step change in the useability of content in geotagged photographs. These developments combined with advances in image analysis and processing, including enhancement of automatic classification algorithms, as well as increasing access to hardware such as high quality unmanned aerial vehicles (UAVs) should greatly help exploit the potential of crowd-sourcing geographic information to support an increasing array of applications.

References

Brown M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., Forrest, D., Haklay, M. and Jackson, M. (2013). Usability of Geographic Information; Current Challenges and Future Directions, *Applied Ergonomics*, 44, 855–865.

Cho, G. (2014). Some legal concerns with the use of crowd-sourced Geospatial Information. *IOP Conference Series: Earth and Environmental Science*, 20(1), 012040. http://doi.org/10.1088/1755-1315/20/1/012040

Craglia, M., Shanley, L. (2015): Data democracy - increased supply of geospatial information and expanded participatory processes in the production of data, *International Journal of Digital Earth*, DOI: 10.1080/17538947.2015.1008214

Kuhn, W. (2007). Volunteered Geographic Information and GIScience. *Position Paper for the NCGIA and Vespucci Workshop on Volunteered Geographic Information*; Santa Barbara, CA, December 13-14.

Scassa, T. (2013). Legal issues with volunteered geographic information. *The Canadian Geographer / Le Géographe Canadien*, 57(1), 1–10. http://doi.org/10.1111/j.1541-0064.2012.00444.x

Additional resources

There is a vast array of material on the subject in the academic and popular literature as well as available on social media or web sites. Interested readers may wish to use the following resources as entry-points to the wide variety of resources available:

TED talks:
    http://www.ted.com/talks/james_surowiecki_on_the_turning_point_for_social_media?language=en
    http://www.ted.com/talks/clay_shirky_on_institutions_versus_collaboration?language=en)

Example of popular VGI initiatives:
    http://www.openstreetmap.org

Example of key academic context:
    Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography, *GeoJournal*, 69, 211-221.