Finding the Quality in Quantity:
Establishing Trust For Volunteered Geographic Information

Jeremy Severinsen[1], Femke Reitsma[2]

[1]Department of Geography
University of Canterbury. Christchurch, New Zealand
Land Information New Zealand
Phone: +64 3 374-3856
Email: jeremy.severinsen@pg.canterbury.ac.nz

[2]Department of Geography
University of Canterbury. Christchurch, New Zealand
Phone: +64 3 364-2987 Ext. 4180
Email: femke.reitsma@canterbury.ac.nz

Presented at SIRC NZ 2013 – GIS and Remote Sensing Research Conference
University of Otago, Dunedin, New Zealand
August 29th – 30th 2013

***Keywords and phrases*:** VGI, crowdsourcing, trust, quality, geographic data, quality assurance

1.0 INTRODUCTION

In the modern environment of Web 2.0, the increasingly connected nature and online presence of societies and communities across the world is drastically and continually altering the role of the 'user' of spatial information. The proliferation of new, mobile, and spatially aware technologies has resulted in the ability for ordinary citizens to not only consume, but also produce data and contribute local expertise to spatial datasets in volumes that are increasing exponentially (Seeger 2008). Where this crowdsourced data explicitly or implicitly captures a location and associated detail, it is now widely referred to as 'Volunteered Geographic Information' (VGI) (Goodchild 2007), and is eliciting great interest from the scientific community. This data is a rich source of near real time information, and marks a significant paradigm shift in the creation and dissemination of geographic information, from a top down authoritative approach to a democratic model with the end user also acting as a data producer (Budhathoki et al 2008, Goodchild 2008), or a phenomenon now referred to as 'Neogeography' (Goodchild 2007). With all of these factors sparking interest in this new tsunami of information, questions have emerged surrounding issues of integration with authoritative datasets and spatial data infrastructures (SDIs), hinged on how the quality of VGI can be accurately determined.

The large volumes of VGI already found on the internet have been created by a heterogeneous and often disparate group of authors often with no reference to, and little understanding of, relevance and coordinated quality standards and protocols. For this reason, VGI has become accepted as somewhat of a "Pandora's Box", with re-users often reluctant to incorporate this rich data source into any scientific analysis for fear of these perceived shortcomings, a fear which could be considered justified for many, but unfounded for an equal number of volunteered geographic features.

The aim of this research is to remove this ambiguity surrounding VGI by quantifying the level to which a particular VG feature can be deemed 'trustworthy', and enable this knowledge to be translated into a rich and ongoing source of information, to be put to productive re-use under the same principles that guide interoperable data sharing inherent in any Spatial Data Infrastructure (SDI). By ascribing a numerical value to this trust rating this research seeks to close the feedback loop for VGI and facilitate a means by which user-created content is not only relevant for other user groups, but also enable reuse of this data by actors traditionally seen as authoritative producers. In this

way, crowdsourced information could be used to augment a variety of fundamental datasets, and thus cement the place of VGI within the geographic data value chain.

Furthermore this research presents a data structure that will capture all of the information required to make a accurate assessment of quality. By capturing this against an individual feature, this information will maintain transitivity with that feature, and enable a consumer of that data to accurately extract VGI of a known quality from the vast quantity that exists in the world of Web 2.0.

2.0 PRIMARY QUALITY INDICATORS

Three primary factors have been identified for use in determining trust in a volunteered geographic feature, and can be summarised as author reputation, spatial precision and temporal quality.

Of the three broadly identified components to quantify trust for VGI, author reputation is probably the most widely discussed across a range of academic disciplines, as well as having more deeply established roots in both social and semantic network analysis. The terms "credibility", "reputation" and "reliability" are often used interchangeably when assessing the source of a particular piece of crowdsourced data, and a large body of academic research has focussed on the idea that the credibility of a source of GI can be used as a proxy for the inherent quality of that data (Flanagin and Metzger 2008).

There has been significant research into the concepts of trustworthiness and expertise as components of credibility (Elwood 2008, Heipke et al 2010, Golbeck et al 2003, Golbeck & Hendler 2004, Coleman et al 2009, Van Exel et al 2008, de Longueville et al 2010). Expertise is generally accepted to be the result of one or both of two factors – some form of formal qualification, or experience. Furthermore, Goodchild (2008) explored the concept of expertise as derived from experience, described as a contributor's "activity space". In simple terms, this concept argues that local knowledge surpasses any formal qualification for the contribution of quality VG features. Consequently, if a feature is contributed by a local about their local environment, then it can be expected that this feature could be given a quality rating that would equal or surpass that ascribed to a contributor with formal geographic training or experience.

Furthermore, as existing facilitated VGI solutions have matured, so too have comparisons between the data collected through portals such as OpenStreetMap and more traditionally authoritative sources (Du et al 2012, Haklay et al 2010, Craglia 2007, Osterman & Spinsanti 2011, Heipke 2010, Haklay 2010). Without an authoritative reference dataset, however, the spatial characteristics of quality are the most difficult to incorporate into a trust calculation for VGI, without extending this research into a study of the semantic web and ontological considerations, which are outside of the scope of this work. Given these restrictions a more appropriate means to measure spatial precision is through map scale. The more familiar an author is with a created feature's real world counterpart, the more likely that contributor will be to create the feature at a smaller scale, with a higher level of detail and spatial precision. Therefore by collecting scale information at capture, a subsequent user of that feature could infer judgement on the author's knowledge and expertise, and thus the extent upon which the spatial accuracy and precision of the feature could be relied (De Longueville et al 2010). This approach is proposed in this case.

Finally, in many respects the real value added by VGI is its currency. VGI has been employed in several coordinated responses to natural disasters due to the speed at which up to date information can be collected and disseminated (Zook et al 2010, Poser et al 2010), therefore reinforcing the information relevance, and augmenting traditional authoritative sources. It is essential to not only capture information about a feature's creation and most recent edition, but to also place this into the context of general change in its immediate area – due to a lack of change to a real world object, a feature with an aging creation date may still accurately reflect that object, and therefore remain current and relevant. A feature may also be one of many, although slightly separated, instances of the same real world object, created by a number of users. Parker et al (2012) state that VGI is likely to be most relevant to the user when a geographic feature is dynamic rather than static in nature. Any quality assessment of the temporality of VGI must therefore account for the context of its surrounding features and general surrounding activity, through an assessment of a feature's currency and provenance (Ye et al 2011).

3.0 PROPOSED VGI-TRUST ALGORITHM

Based on the factors identified as having a measurable impact on the quality of geographic information, this research presents an algorithm that assesses these factors at an appropriate weighting. At a high level, this algorithm will include information about the data author, aspects of its spatial precision, and information on its provenance and currency, or temporal quality, each of which are ascribed a weighting within the formula. The model will be tested with volunteer contributions in post-earthquake Christchurch, where contributors will be asked to augment the Topo50 map series produced by Land Information New Zealand with user information on the changed state of the city.

One of the primary drivers of this research is to establish an algorithm that produces a trust rating for VGI that is easily interpreted, and can inform any re-user of these data as to the inherent quality of a given feature, irrespective of that user's expertise and training in the spatial sciences. As such, the "trust rating" is a measure between 0 and 10, with 0 marking a feature that cannot be relied upon in any way, and 10 being a feature that can be considered of exceptional quality. By establishing this rating system, quality percentiles can be inferred, as can degrees of trust, and these can be applied to fitness for purpose assessments for data reliance.

4.0 CONCLUSIONS

Crowdsourced geographic information is now prolific in the online world, and the issue with this increase in VGI is the question of trust. How can an end user of VGI quickly and effectively determine the quality of that data? This research is a starting point for practical quality analysis of VGI, through an analysis of a series of factors that will be collected during a facilitated case study, in order to produce an algorithm that could provide an estimate of quality for the vast spectrum of volunteered features. This research can logically be extended to include the examination of the semantic web and volunteer ontologies, both to further enhance the assessment of spatial precision and as a content quality indicator in its own right. These considerations are at this time outside of the scope of this research.

ACKNOWLEDGEMENTS

The support of Land Information New Zealand in this research is gratefully acknowledged.

REFERENCES

Budhathoki, N., B. Bruce & Z. Nedovic-Budic (2008) Reconceptualising the Role of the User of Spatial Data Infrastructure, *Geojournal,* Vol. 72, pp. 149-160.

Coleman, D., Y. Georgiadou & J. Labonte, (2009) Volunteered Geographic Information: The Nature and Motivation of Produsers, *International Journal of Spatial Data Infrastructures Research,* Vol. 4, No. 1, pp. 332-358.

Craglia, M. (2007) Volunteered Geographic Information and Spatial Data Infrastructures: When do Parallel Lines Converge? *Position Paper for the VGI Specialist Meeting*, Santa Barbara 13 – 14 December.

De Longueville, B., N. Ostlander & C. Keskitalo, (2010) Addressing Vagueness in Volunteered Geographic Information (VGI) – A Case Study, *International Journal of Spatial Data Infrastructures Research,* Vol. 5.

Du, H., N. Alechina, G. Hart, M. Jackson, S. Anand, J. Morley, D. Leibovici, M. Ware, (2012) Geospatial Information Integration for Authoritative and Crowd Sourced Road Vector Data, *Transactions in GIS,* Vol. 16 No. 4, pp. 455-476.

Elwood, S. (2008) Volunteered Geographic Information: Future Research Directions Motivated by Critical, Participatory, and Feminist GIS, *Geojournal,* Vol. 72, pp. 173-183.

Flanagin, A. & M. Metzger, (2008) The Credibility of Volunteered Geographic Information, *Geojournal,* Vol. 72, No. 3-4, pp. 137-148.

Golbeck, J., J. Hendler, (2004) Inferring Reputation on the Semantic Web, *Proceedings of the 13th International World Wide Web Conference,* May, Vol. 316.

Golbeck, J., B. Parsia, & J. Hendler, (2003) Trust Networks on the Semantic Web, Springer Berlin Heidelberg pp. 238-249

Goodchild, M. (2007) Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0, *International Journal of Spatial Data Infrastructures Research,* Vol. 2, pp. 24-32.

Goodchild, M. (2008) Spatial Accuracy 2.0, *Proceedings of the Eighth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences,* Vol. 1, pp. 1-7.

Haklay, M. (2010) How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets, *Environment and Planning B: Planning and Design,* volume 37, pp 682 – 703.

Haklay, M., S. Basiouka, V. Antoniou, A. Ather, (2010) How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information, *The Cartographic Journal,* Vol. 47 No. 4 pp. 315-322.

Heipke, C. (2010) Crowdsourcing Geospatial Data, *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 65, pp. 550-557.

Osterman, F. & L. Spinsanti, (2011) A Conceptual Workflow for Automatically Assessing the Quality of Volunteered Geographic Information for Crisis Management, *Proceedings of AGILE,* April.

Parker, C., A. May & V. Mitchell (2012) Understanding Design With VGI Using an Information Relevance Framework, *Transactions in GIS,* Vol. 16, No. 4, pp. 545-560.

Poser, K. & D. Dransch, (2010) Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation, *Geomatica,* Vol. 64, No. 1, pp. 89-98.

Seeger, C. (2008) The Role of Facilitated Volunteered Geographic Information in the Landscape Planning and Site Design Process, *Geojournal,* Vol. 72, No. 3-4, pp. 199-213.

Van Exel, M., E. Dias, & S. Fruijtier, (2010) The Impact of Crowdsourcing on Spatial Data Quality Indicators, *Proceedings of the 6th GIScience International Conference on Geographic Information Science,* September p. 213.

Ye, M., K. Janowicz, C. Mulligann & W. Lee (2011) What You Are is When You Are: The Temporal Dimension of Feature Types in Location-Based Social Networks, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* November, pp. 102-111.

Zook, M., M. Graham, T. Shelton, & S. Gorman, (2010) Volunteered Geographic Information and Crowdsourcing Distster Relief: A Case Study of the Haitian Earthquake, *World Medical Health Policy 2,* No. 2, pp. 7-33.