Can Volunteered Geographic Information Be a Participant in eEnvironment and SDI?

Peter Mooney, Padraig Corcoran

Can Volunteered Geographic Information be a participant in eEnvironment and SDI?

Peter Mooney and Padraig Corcoran

Department of Computer Science,
National University of Ireland Maynooth (NUIM),
Maynooth, Co. Kildare. Ireland
{peter.mooney,
padraig.corcoran}@nuim.ie

Abstract. We investigate the potential role Volunteered Geographic Information (VGI) can play in eEvironment and various Spatial Data Infrastructures (SDI) on a local, regional, and national level. eEnvironment is the use and promotion of ICT for the purposes of environmental assessment and protection, spatial planning, and the sustainable use of natural resources. An SDI provides an institutionally sanctioned, automated means for posting, discovering, evaluating, and exchanging geospatial information by participating information producers and users. A key common theme shared by both definitions is public participation and user-centric services. We pose the research question: is VGI (an example of public participation and collaboration) is ready to participate in eEnvironment and SDI?

Keywords: VGI, OpenStreetMap, SDI, Spatial data

1 Introduction

On first glance VGI appears to have all of the required ICT ingredients to provide a dynamic picture of the environment. VGI's ability to leverage large numbers of dedicated "citizen sensors" [5] is unprecedented. As a consequence the amount of VGI available on the Internet today has grown enormously in the past few years. Initiatives such as Wikimapia, Google Mapmaker, OpenStreetMap (OSM), geotagging in Flickr, geolocation in Twitter, Geonames, etc have seen VGI become a "hot topic in GIS research" [15] and is now one of the rapid growth areas of GIS. Recently VGI (such as OpenStreetMap) has begun to provide an interesting and feasible alternative to traditional authoritative spatial information from National Mapping Agencies and corporations. However the fact that VGI is a spatial form of the user-generated content in Web 2.0 has raised serious concerns and reservations within the GIS, Geomatics and Environmental Science communities [15, 14] about its quality, accuracy, sustainability, and fitness for use/purpose. Kessler et al [11] points out that compared to other projects building on user contributed content, such as Wikipedia, VGI remains on the GIS periphery and consequently has been restricted to web-based mapping applications [8] and not considered for involvement in "serious geomatics applications" [17]. De Longueville et al [13] (also in [12]) comments that work-flows have been implemented to create, validate, and distribute VGI datasets for various thematic domains but its exploitation in real-time and its integration into existing concepts of Digital Earth, such as SDI, still needs to be further addressed". SDI are created for specialists and experts with the goal of making diverse and heterogeneous data available and accessible. Gouveia et al [6] stress that the development of SDI, throughout the world, has facilitated improved public access to environmental information because of its inherently spatial characteristics. As GPS and web-enabled mobile devices have become ubiquitous we feel that it is important to leverage these new information sources and work towards stronger integration capabilities. Our paper attempts to make a case for VGI as a participant in eEnvironment and SDI.

2 Overview of Related Literature

Citizens, experts and non-experts alike, are increasingly participating in the process of generating continuous spatial information and collaborating with others in problem-solving tasks. This highlights the transition of the role of users from just mere data consumers to active participants and providers [2]. Traditionally, SDI building follows a top-down approach. This scenario leads to the provider-consumer paradigm, where only official providers like National Mapping Agencies (NMAs) and other environmental agencies, centrally, manage and deploy resources according to institutional policies. In this approach end-users can only be consumers [2]. However, VGI has changed this. There has been a transition in the role of users from just mere "data consumers to active participants and providers" [2]. Budhathoki et al [1] argue that SDI and VGI are not separate entities but are complementary phenomena. Budhathoki et al believe that these phenomena can be brought within a "single framework where the role of the user of SDI is re conceptualized to *produser* (producer and consumer of spatial data [2]) and VGI is included in the SDI-related processes. To enhance consumption of spatial data from SDI Omran and van Etten [16] suggest using a social network approach to spatial data sharing as a means of improving spatial data exchange in SDI. However, in a social network model (just as in VGI) there needs to be "a redefining of the rules about spatial data sharing and transferring more responsibility to more individuals in organizations [16]. Most SDI typically comprise of participants such as National Mapping Agencies (NMA), government agencies, private organizations, etc who have traditional or commercial roles in producing spatial information. Ho and Rajabifard [9] believe that this view "leaves a large part of society (community groups, concerned citizens) with none or nominal roles in SDI and are excluded or disengaged" where VGI is potentially left on the fringes. Research on SDIs and other spatial data sharing structures has not specifically considered the challenges facing grassroots data users. This is addressed by Elwood [3] who emphasizes the need for local data integration and accessibility to local users. SDIs are predicated on an assumption of openness to data sharing and exchange, conceptualizing data as a public good and assuming institutional and individual openness to sharing. In all likelihood, no single approach is wholly sufficient, given the social, political and technological complexity of spatial data sharing. Ho and Rajabifand [9] argue that as a visible representation of citizens' thoughts, observations, collected spatial and environmental data, VGI can be a "potential barometer for people's environmental concerns and attitudes and potentially lead to better citizen 'buy-in' to SDIs". This could also help inform the SDI managers what grassroots users and citizens currently require from an SDI.

In Section 1 we mentioned that VGI was not used in "serious applications" but there are some examples. Pultar et al [18] show applications to wildfire evacuation modeling and travel scenarios of urban environments. Over et al [17] develop prototype 3-D models using German OSM data. The "extensive produser (producer and consumer of spatial data [2]) base" in VGI referred to by Budhathoki et al [1] is a now large enough to be considered by SDI-related initiatives and eEnvironment. Some authors outline the problems in SDI development which are actually positives in VGI. Thellufsen et al [19] argue that the effective development of SDI is often a "fragmented" activity requiring inter-organization collaboration. Unfortunately many of the stakeholders in this collaboration resist "data sharing across organizational boundaries due to loss of control, power and independence. This is an area where VGI is very strong through the collaborative nature and ethos of the community based upon an inherent understanding and willingness to *share* data. Budhathoki et al [1] argue that VGI has harnessed a "large number of participants, without being coordinated by any formal organization, and without the lure of monetary or personal gain". Thellufsen et al [19] conclude that currently too many organizations, with spatial data useful for SDIs, are data "silo-minded" precisely at a time when they ought to be outreaching and co-operative. The authors suggest that these organizations should focus on building motivation awareness before actual solutions. In the next section we provide some results of analysis of OSM which provides some open questions for the inclusion of VGI in SDIs.

3 Using VGI in SDI: The OpenStreetMap case study

In this section we outline some results from some experimental analysis of OSM in Europe. To give an overview of some of the problems that VGI must tackle before becoming an active player in VGI we use the methodology of Grus et al [7] as well as some examples of the characteristics of OSM. Grus et al [7] developed a goal-oriented assessment view approach for assessing the realization of SDI goals and is demonstrated by its implementation in the Dutch SDI. As concluded by Giff and Crompvoets [4] SDIs must "not only to justify expenditure on their implementation but also to determine whether or not they are achieving their objectives". For this case-study we have analyzed the OSM databases for UK and Ireland, France, Germany, Austria, and Estonia. We show sample results under a number of headings: metadata, the nature of collaborative contributions, and data scale issues.

3.1 Metadata

Grus et al [7] suggest two metadata related indicators for SDI: *Metadata-standard applied in the national SDI geoportal* and *Metadata are produced for a significant fraction of spatial datasets*. Metadata can be inserted into the OSM database through the use of the accepted tagging structures. Tag keys such as "source","source:ref", "source:url" can be used as metadata to document the source of imported bulk data, tracing from aerial imagery, etc. Tags such as "note" and "attribution" can make specific statements about mapping techniques, errors, or other information useful to other mappers or users of the data. Tags offer OSM contributors with an opportunity to document their contributions. In the global OSM database the "source" tag is widely used of the 73 million objects using the source tag 51% are points, 47% are polygon, and 40% are relations. However our analysis shows this is dominated by bulk imports such as the French import of the EEA Corine Land Cover dataset. The "note" tag is used on five million objects: 5.8% of these are points, 1.74% of these are polygons, and 5.77% are relations. The tagging of OSM can rapidly degenerate into a folksonomy by leaving tagging up to individual contributors. But this can be seen as a flexibility which can grow to accommodate other metadata needs for example. Best practices, at the very minimum for bulk import or tracing of aerial imagery using OSM editors, can and should be implemented. All edits are tracked in OpenStreetMap so there is a clear trail of "who did what". While there are well known advantages in spatial data interoperability in using Dublin Core and ISO 19115 metadata schemes Kalantari et al [10]) argue that the folksonomy approach of VGI could have unpredicted advantages by leading to "more user-generated metadata potentially initiating richer metadata and better possibillty of user/creator updating".

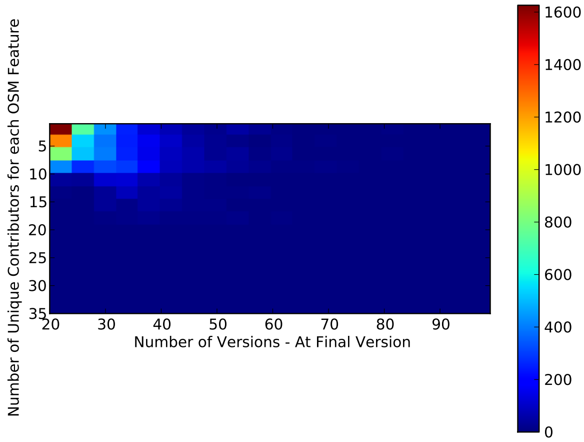

Fig. 1. A 2-D heat map histogram showing the distribution of the number of versions (x axis) against the number of unique contributors (y)

3.2 Collaborative Contributions

In Figure 1 we show a 2-D heat map histogram of the distribution of the number of versions (x axis) of OSM polygons and polylines against the number of unique contributors (y) to each feature. $10,000$ "high edit" (more than 20 edits from OSM contributors) polygons and polylines from the UK and Ireland OSM databases were analyzed. Most of these features have a small number of contributors (≤ 5) while the number of versions (subsequent edits) of these features have a mean of 35 versions. We analyzed the complete historical record for these $10,000$ objects. Table 1 shows a summary of the time between edits of consecutive versions of the same object. Almost 42% of consecutive edits are separated by an editing time of 1 week to 1 month. Almost 38% of consecutive edits have 1 hour and 24 hours between them. The results in Table 1 gives an indication of the rate of change of the OSM database for these $10,000$ objects. To integrate the most up-to-date versions of these features into an SDI one must consider the time between edits. While almost half of edits (updates) happened with a period of 24 hours or more one must consider if the OSM database is changing too quickly for integration into an SDI.

3.3 Spatial Scale Issues

As Mooney and Corcoran [14] outline differences between countries in Europe to how spatial data is represented in the corresponding OSM database particularly

Table 1. Distribution of time between consecutive edits for 10,000

No. Edits	**% of Edits**	**Time Between Edits**
3,866	3.30%	$\leq$ 5 minutes
11,478	9.80%	5mins $\leq$ 30mins
2,400	2.05%	30mins $\leq$ 1hour
12,183	10.40%	1hr $\leq$ 2hr
21,318	18.20%	2hr $\leq$ 12hr
11,608	9.91%	12hr $\leq$ 24hr
3,391	2.90%	24hr $\leq$ 1week
49,084	41.91%	1week $\leq$ 1month
1,784	1.52%	> 1month

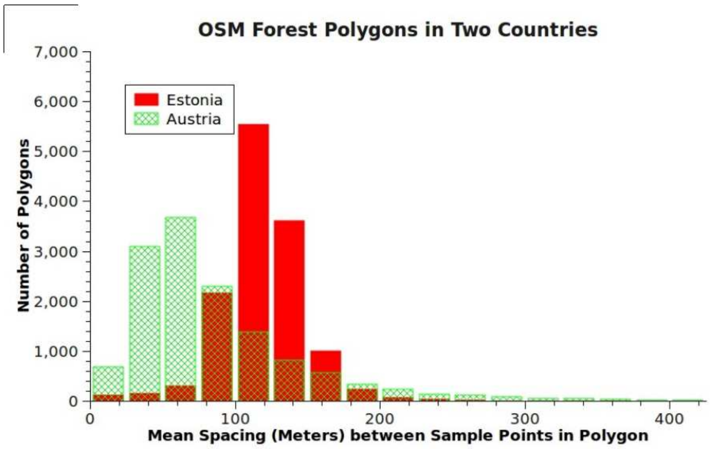

Fig. 2. Spacing between points in polygons in countries with government generated data.

in the number of nodes used to represent features. Differences in representation is an artefact of the different surveying and sampling methodologies employed by contributors to OSM. Figure 2 shows a plot of the mean spacing between nodes in landuse polygons in OSM for Estonia and Austria. Both Estonia and Austria have benefited from the donation of spatial data from Corine and Government sources respectively. It is evident from figure 2 that the scales of the two datasets are different. Grus et al [7] cite the inclusion of harmonized datasets within an SDI as an important indicator of a successful SDI.

4 Conclusions

VGI is a rapidly evolving user-generated content movement. It's sustainability going forward into the long-term is uncertain. VGI will need to introduce improved management of contributions and contributors to prevent the spatial data moving indefinitely between a status of good and bad quality. Leveraging those dynamic updates (indicated in Table 1) is a key step to making the most of VGI. It is necessary that the VGI community can demonstrate that issues such as data quality (see Mooney and Corcoran [15]), scale and harmonization problems (Figure 2), etc can be detected effectively and efficiently. We agree with Budhathoki et al [1] that is "unlikely that VGI will completely replace SDIs". However, we believe that there is adequate scope and motivation for VGI to become a key stakeholder (both as spatial data producer and consumer) in SDIs. Diaz et al [2] remark that VGI is forcing the expert producers to rethink their traditional approaches of spatial production. Many open research questions remain. One of the key questions in relation to GIS is how VGI will interact with this community? As Budhathoki et al [1] (and Mooney and Corcoran [15]) remark "VGI is unlikely to satisfy the vast majority of institutional and professional GI producers whose requirements in terms of data quality, timeliness, and completeness are very strict". Can a suitable middleground between professional GI producers and VGI can be found? Wiemann and Bernard [20] indicates that "the full integration of VGI (such as OSM) within SDI is not yet possible". However we believe that the generation of knowledge from a variety of sources of spatial information can play a decisive role in building knowledge-based structures made accessible through the vehicle of SDI. Giff and Crompvoets [4] conclude that SDIs must "engage its users to clearly think through the processes involved in the provision of spatial information products and services". Engagement of users/communities is one of the most impressive characteristics of the VGI phenonema.

Acknowledgements

Dr. Peter Mooney is a research fellow at the Department of Computer Science NUIM and is funded by the Irish Environmental Protection Agency STRIVE programme (grant 2008-FS-DM-14- S4). Dr. Padraig Corcoran is a lecturer and post-doctoral researcher also at the Department of Computer Science NUIM. Dr. Corcoran is part of STRAT-AG which is a Strategic Research Cluster grant (07/SRC/I1168) funded by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support.

References

1. Budhathoki, N., Bruce, B., and Nedovic-Budic, Z. Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal 72* (2008), 149–160.
2. Daz, L., Granell, C., Gould, M., and Huerta, J. Managing user-generated information in geospatial cyberinfrastructures. *Future Generation Computer Systems 27*, 3 (2011), 304 – 314.

3. Elwood, S. Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science 22*, 1 (2008), 71–90.
4. Giff, G. A., and Crompvoets, J. Performance indicators a tool to support spatial data infrastructure assessment. *Computers, Environment and Urban Systems 32*, 5 (2008), 365 – 376.
5. Goodchild, M. F. Neogeography and the nature of geographic expertise. *Journal of Location Based Services 3*, 2 (2009), 82–96.
6. Gouveia, C., Fonseca, A., Cmara, A., and Ferreira, F. Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management 71*, 2 (2004), 135 – 154.
7. Grus, L., Castelein, W., Crompvoets, J., Overduin, T., van Loenen, B., van Groenestijn, A., Rajabifard, A., and Bregt, A. K. An assessment view to evaluate whether spatial data infrastructures meet their goals. *Computers, Environment and Urban Systems In Press, Corrected Proof* (2010), –.
8. Haklay, M. A comparative study of openstreetmap and ordnance survey datasets for london and the rest of england. *Environment and Planning B: Planning and Design 37*, 4 (2010), 628–703.
9. Ho, S., and Rajabifard, A. Learning from the crowd: The role of volunteered geographic information in realising a spatially enabled society. In *Proceedings of GSDI 12 World Conference - Realising Spatially Enabled Societies* (Singapore, October 2010).
10. Kalantari, M., Olfat, H., and Rajabifard, A. Automatic spatial metadata enrichment: Reducing metadata creation burden through spatial folksonomies. In *GSDI 12 Proceedings of the 12th Global Spatial Data Infrastructures Conference* (October 2010), GSDI, pp. 773–779.
11. KeBler, C., Janowicz, K., and Bishr, M. An agenda for the next generation gazetteer: geographic information contribution and retrieval. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2009), GIS '09, ACM, pp. 91–100.
12. Longueville, B. D. Community-based geoportals: The next generation? concepts and methods for the geospatial web 2.0. *Computers, Environment and Urban Systems 34*, 4 (2010), 299–308.
13. Longueville, B. D., Ostländer, N., and Keskitalo, C. Addressing vagueness in volunteered geographic information (VGI) – A case study. *International Journal of Spatial Data Infrastructures Research 5* (2009).
14. Mooney, P., and Corcoran, P. A study of data representation of natural features in openstreetmap. In *Proceedings GIScience 2010: The Sixth International Conference on Geographic Information Science*, R. Weibel and S. Fabrikant, Eds. Springer Verlag LNCS, Zurich, Switzerland, 2010.
15. Mooney, P., Corcoran, P., and Winstanley, A. C. Towards quality metrics for openstreetmap. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2010), GIS '10, ACM, pp. 514–517.

16. Omran, E. E., and van Etten, J. Spatial-data sharing: Applying social-network analysis to study individual and collective behaviour. *International Journal of Geographical Information Science 21*, 6 (2007), 699–714.
17. Over, M., Schilling, A., Neubauer, S., and Zipf, A. Generating web-based 3d city models from openstreetmap: The current situation in germany. *Computers, Environment and Urban Systems 34*, 6 (2010), 496 – 507. GeoVisualization and the Digital City - Special issue of the International Cartographic Association Commission on GeoVisualization.
18. Pultar, E., Raubal, M., Cova, T. J., and Goodchild, M. F. Dynamic gis case studies: Wildfire evacuation and volunteered geographic information. *Transactions in GIS 13* (2009), 85–104.
19. Thellufsen, C., Rajabifard, A., Enemark, S., and Williamson, I. Awareness as a foundation for developing effective spatial data infrastructures. *Land Use Policy 26*, 2 (2009), 254 – 261.
20. Wiemann, S., and Bernard, L. A comparative study of proprietary geodata and volunteered geographic information for germany. In *Proceedings AGILE 2010: The 13th AGILE International Conference on Geographic Information Science*, M. Painho, M. Y. Santos, and H. Pundt, Eds. Springer Verlag, Guimarães, Portugal, 2010.