What Volunteered Geographic Information is (good for) – designing a methodology for comparative analysis of existing applications to classify VGI and its uses

Nicole Ostlaender, Robin S. Smith, Bertrand De Longueville, Paul Smits
{nicole.ostlaender | robin.smith | bertrand.de-longueville | paul.smits}@jrc.ec.europa.eu
Joint Research Center (JRC), European Commission

Geographical Information Science (GI-Sc) has a new rising star and buzz-word: Volunteered Geographic Information (VGI). VGI describes a phenomenon that transfers the principles found in the Internet-based user-driven content of Web 2.0 to the particular case of geospatial data and information. As a research topic, this highly diverse data-source sees the convergence of a number of themes in GI-Sc. These include the adoption of new technological paradigms; utilising non-technical popular knowledge as a source of data creation and validation; and the democratisation of geospatial technology. Amongst all these competing perspectives, VGI can exist in many guises that make it difficult to understand its characteristics, properties and potential roles. Thus, what is currently lacking is a systematic framework that classifies VGI products and sources, and the problem domains they are used to address. Drawing together the various perspectives on VGI in the GI-Sc field, the framework should provide guidance on data and workflow characteristics to help produce new applications. If carefully defined, such a framework would support informed decisions on how to use VGI in GI applications and, thereby, facilitate the transition of VGI from a mere 'hot topic' to a pillar of tomorrow's GI application design. In order to create such a framework, we suggest a bottom-up approach: a comparative analysis of the large pool of existing applications to harness the merits (and identify the limitations) of existing efforts. As an initial step, which we present in this paper, we developed a methodology based on business process modeling that allows the description of these applications and prepares the results for the subsequent comparative analysis.

Background and related work

First discussed by Goodchild (2007), Fritz et al. (2009) identified a number of terms relating to user-generated geospatial content and VGI (including crowdsourcing, web mapping 2.0 and neogeography, amongst others), where related applications help (predominantly) citizens to "…voluntarily collect, organize and/or disseminate geographic information and data in such a manner that the information can be used by many others" (Tulloch, 2008; 161).

Although the term has been recently coined, notions associated with citizen-based data collection, validation and generation have been long-standing, for example, annual bird-counts in the 1900s in the United States (Gouveia et al., 2004; after Lee, 1994), British land-use surveys in the 1930s (Stamp, 1937) and 1960s (Coleman, 1961). More recently, wider ICT adoption has led to something of a renaissance of

such projects in environmental monitoring (e.g. Gouveia et al., 2004; Oscarson and Calhoun, 2007; Monk et al., 2008), alongside useful theoretical discussions linking VGI and the socio-technical considerations of Public Participation GIS (Tulloch, 2008). Increasingly, the diffusion of Internet-enabled mobile devices is allowing those with technical and socio-political access (Smith and Craglia, 2003) to use VGI applications for many purposes: from finding new bike routes[1] to reporting violent acts during political crises[2] (Barhee, 2008); from monitoring forest fires (De Longueville et al., 2009) to avoiding traffic jams[3]; and from validating land-use maps (Fritz et al., 2009) to assessing earthquake intensity (De Rubeis et al., 2009). However, detailed methodologies to explore VGI content and applications are yet to emerge, notable as concerns can be readily raised about the credibility, reliability and overall value of VGI (Flanagin and Metzger, 2008) and citizen-sourced data, in general (Gouveia et al. 2004).

Describing VGI and VGI applications using business processes

The VGI applications we aim to describe and compare range from complex, multi-purpose software products such as Google maps[4] by Google™, to singular workflows where humans use or produce VGI through one or more software products. An example of the latter is the application 'Just Landed' by Jer Thorp (2009) that, inspired by looking into the spread of epidemics, harvests and analyses Twitter messages to map airplane landings. The methodology we develop should readily describe these applications, so that the characteristics of VGI can be captured alongside the role it plays within the application. Therefore, the methodology has to be platform-, product- and workflow-independent. As the outcomes of the comparative analysis shall eventually lead to a framework that could guide application design, the methodology should also foresee the collection of characteristics for future applications' requirements analysis.

In order to meet all these specifications, we abstain from describing applications as a whole but, instead, apply a goal-driven use case analysis. As we deal with workflows, that possibly involve not only interaction with more than one software product but also human-to-human interaction, we use 'business use cases' or 'business processes' instead of the more commonly known system use cases (cf. Cockburn, 2001). Business process modeling (BPM) is the activity of representing the processes of an entity under consideration, e.g. a business enterprise, so that the current processes may be analyzed and improved in future. In the software design domain, business process modeling is an artifact of the business modeling discipline of the IBM Rational Unified Process (RUP®) (see e.g. Kruchten, 2000) that lays the foundation for the analysis of functional and non-functional requirements.

---

[1] http://www.mapmyride.com/
[2] http://www.ushahidi.com/
[3] http://www.waze.com
[4] http://maps.google.com

The bottom-up comparative analysis, which this methodology enables, acts as the base of a systematic framework that classifies VGI products and sources and provides guidance on data and workflow characteristics for future VGI application design. To achieve this, we base the methodology on the Eriksson-Penker Business Extension for UML (Eriksson & Penker, 2000), and characterize a business process as an activity that has a given number of inputs and outputs and is carried out by an actor to achieve a certain goal. As we need to be able to compare the various applications, we extend these elements with a list of characteristics that should be described for each process. These characteristics were captured iteratively by describing various processes for a range of existing applications. A particular focus was put on identifying issues related to input or output of VGI, such as data quality and coverage of the captured information. The relationships between existing VGI applications (bottom-up-approach), method development and comparative analysis are shown in Figure 1. This methodology is now being applied to describe existing VGI applications. Patterns that we aim to identify from our analysis should, for example, allow us to determine which particular types of VGI can be used to achieve particular goals.

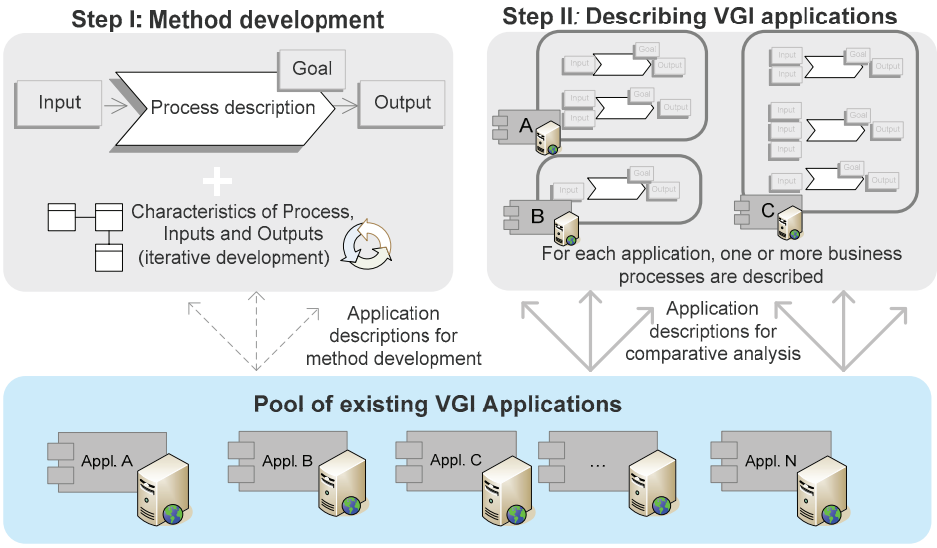

Figure 1: Relationships between existing VGI applications, method development and comparative analysis

Conclusions and further work

Within the ambit of VGI, the rapid production of massive quantities of online georeferenced content has led to a range of issues that the GI Sciences are aiming to address. Such activities rest within a long tradition of citizens gathering and sharing data about their local environment but ICTs are offering an increased potential to rapidly generate and share such content. Alongside theoretical considerations from the PPGIS field, more detailed investigation is needed from a technological perspective, to better understand VGI's content, associated applications and the potential for its reuse and adaptation, respectively.

As presented in this paper, one approach involves understanding what VGI is already good for, made evident by the current range of applications containing a variety of processes, roles and goals. Through

business process modeling, a strategic analysis can be brought to bear on VGI applications; offering not only a structured and replicable approach but also the means to specify details for comparative analysis, the logical next step in the presented approach.

References

Cockburn A. (2001) Writing effective use cases. Addison-Wesley, New York.

Coleman, A. (1961) The second land-use survey: Progress and prospect. *Geographical Journal* 127, 68–186.

De Longueville, B., Smith, R.S. and Luraschi, G. (2009) "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. *Proceedings of the 2009 International Workshop on Location Based Social Networks*. November 3, 2009, Seattle, Washington.

Flanagin, A.J. and Metzger, M.J. (2008) The Credibility of Volunteered Geographic Information. *GeoJournal* 72 (3-4) 137-148.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. and Obersteiner, M. (2009) Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing* 1, 345-354.

Goodchild, M. J. (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal* 69 (4) 211-221.

Gouveia, C., Fonseca, Câmara, A. and Ferreira, F. (2004) Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management* 71, 135–154.

Kruchten P. (2000) *The Rational Unified Process. An Introduction*. Addison -Wesley, Reading MA.

Monk, J., Ierodiaconou, D., Bellgrove, A. and Laurenson, L. (2008) Using community-based monitoring with GIS to create habitat maps for a marine protected area in Australia. *Journal of the Marine Biological Association of the United Kingdom* 88 (5) 865-871.

Oscarson, D.B. and Calhoun, A.J.K. (2007) Developing vernal pool conservation plans at the local level using citizen-scientists. *Wetlands* 27 (1) 80-95.

Smith, R.S. and Craglia, M (2003) Digital Participation and Access to Geographic Information: A Case Study of Local Government in the United Kingdom. URISA Journal 15 (APA II) 49-54.

Stamp, L.D. (1937), The Land of Britain: the report of the land utilization survey of Britain.

Tulloch, D.L. (2008) Is VGI participation? From vernal pools to video games *GeoJournal* 72 (3-4) 161-171.

Thorp, J. (2009) Just Landed: Processing, Twitter, MetaCarta & Hidden Data. last access: 2009-12-15 at: http://blog.blprnt.com/blog/blprnt/just-landed-processing-twitter-metacarta-hidden-data.