Beyond data collection in the VGI research field: a systematic mapping study on intended use and analysis methods

Carlos Granell¹, Frank Ostermann²

¹ Institute of New Imaging Technologies, Universitat Jaume I of Castellón (Spain)
Email: carlos.granell@uji.es

² Department of Geo-Information Processing, Faculty of Geo-Information Science and Earth Observation
University of Twente (The Netherlands)
Email: f.o.ostermann@utwente.nl

Although it has been seven years, the essential meaning of VGI nowadays remains pretty much the same: Users and citizens are producing (and consuming) large volumes of data tied with a location or geographical reference. These bits of geospatial data are gaining recently importance due to improvements on pervasive technology (e.g. automated GPS traces), handheld devices and above all to the proliferation of location-based social network sites and services along with enabling analytics technology that have permitted users to easily create and share location-aware social data. The bottom-line idea is that one’s location is a central aspect. No matter what people and/or services are sensing (e.g. noise, air quality), spatio-temporal context is a must to help us to scientifically best understand and interpret our surrounding environment.

GIScience is an open and multidisciplinary field with tight connections to other scientific fields and disciplines (Sui, 2014; Blaschke and Merschdorf, 2014). The same holds for the VGI research topic. As GIScientists, we have a great opportunity to leverage location and geospatial technology at the forefront of the mainstream technology in general and applied to social media data and VGI in particular. Altogether, it might lead to a plethora of new and emerging patterns on the utilization of VGI to permit advancing the VGI research field beyond its predominant focus in data collection.

Recent works in the literature, e.g. Neis and Zielstra (2014), examine the nature of VGI data itself (e.g. quality, precision), or focus on the role of the user (e.g. gender, motivating factors to contribute) or centre on the capabilities of VGI sources (reliability, documentation, easy-to-use). Even though these facets are of vital interest to enhance the role of VGI in scientific contexts, equally important is to identify what analyses methods have been used and for what purpose. We argue that VGI research is not at an early stage, but rather at a stage where a body of knowledge and best practices should emerge. This is a difficult task because research on VGI is so technology-driven and changes at a fast pace. Still, some consolidation is happening, and we want to contribute by identifying important analytical trends and use patterns on the utilization of VGI, in order to shape future research and applications. To do so, we look at VGI with the following research questions:

- What are the most investigated intended uses and how have these changed over time?
- What are the most frequently VGI data sources, and in what context/application?
- What are the most frequently (spatial) data analyses methods, and in what context/application?

Rather than conducting a systematic literature review, we take a complementary, less time-consuming approach named systematic mapping study (Peterson et al., 2008; Casteleyn et al., 2014).
Engström and Runeson (2011) summarized this type of study as a useful tool that it is performed “at a higher granularity level with the aim to identify research gaps and clusters of evidence in order to direct future research”. It essentially aims to find and classify the primary studies in a specific topic area. In this work, we perform a systematic mapping study to derive a consolidated overview on intended use in conjunction with analysis methods based on relevant studies that have been published on location-aware social data analysis and VGI over the past few years. The aim of the systematic map study is then to come up with a visual overview, the map, which connects the application coverage and types of analysis methods employed, and their evolution over time.

Preliminary results suggest that the potential value of VGI has been demonstrated and partly realized through the wide variety of application domains in which location is or might be a determining factor. We realize though that most of the work in VGI analysis goes to early phases of data analysis such as data preparation, i.e. merging, finding missing values, cleaning, annotating, filtering and so forth in order to make several data sets ready for the “true” analysis. Despite the enormous importance in science of defining the right hypothesis, framing the research question(s) and the analyses strategies needed to empower the validity, scope and impact of the analytical results, we detect that it is a pending issue that must still be properly addressed in most studies. In addition, special attention is required to identify sources and types of VGI that best fit with the analysis needs. The more variety in the data sources, the greater the level of confidence in the results at the cost most likely of increasing considerably the complexity associated to data integration and contextualization.

Some scholars even envision an emerging scientific field that would combine data science, data analysis, GIScience and social sciences skills under the label of “location intelligent” (Wachowicz, 2013), with the ability to shape standardized methodologies to facilitate data collection, to carry out data analysis, and to compare the results beyond the irruption of popular technology and short-term research objectives. In this way, the VGI systematic map may be an ideal tool for early and experience researchers alike as it provides an understanding of the existing VGI literature on specific facets (intended use, analytic methods) but also for end users, as main contributors to the VGI field, who may realize of the type of applications that come out as a consequence of their volunteered data contributions.

References
Sui, D., 2014. Opportunities and Impediments for Open GIS. Transactions in GIS 18(1), 1-14