Dynamic Surveying Adjustments for Crowd-sourced Data Observations

Sergiusz Pawlowicz¹, Didier Leibovici¹, Roy Haines-Young² Richard Saull³, Mike Jackson¹

Abstract

Advances in positioning, imaging, location-based services capabilities, and broadband connectivity enable public participation in environmental monitoring and decision making in a manner previously only possible for professional scientists. Data collected by volunteers has long been an important factor in environmental programmes, but the difficulties in applying quality control measures and in ensuring an appropriate sample size of observations for a given area limit their scientific value. This paper addresses this challenge in describing a surveying architecture, allowing efficient in-field data collection from a GPS enabled ubiquitous device. The originality of this architecture comes from a real-time analysis of surveyed responses in order to "drive" the survey for optimised precision and validity. Spatial awareness from the surveying engine allows one to modify the sets of questions for each surveyor volunteer as the survey goes along, and even to try to recruit volunteers in the area. As a result this architecture is able to support the implementation of a dynamic and directed approach to in-field data collection with real-time quality control driven by an adaptive surveying modelling technique insuring optimised data collection and personalised feedback to users. A plug-in architecture gives the possibility to extend using RSS, Twitter and Mappiness data sources, with other real time feeds, then allowing the adaptive surveying engine to use multilevel multi-sourced sets of information.

The paper describes this conceptual architecture and a technical solution for its implementation, independent of the mobile hardware producer, tablets, smart phones, netbooks, laptops, in order to allow the widest public participation opportunity possible. The implementation was tested on Google Android and Apple iPhone devices with a use case coming from the Tranquillity Report of the Campaign to Protect Rural England (CPRE).

1. Introduction

Unprecedented growth and popularity of ubiquitous devices in last few years, their location based services capabilities, broadband connectivity and ease of use, allow extending public participation (PP) in environmental decisions. Environmental datasets collected under rigorous scientific construction, funded by governments or private sector, can be now tested and improved by the public not considered scientifically as `experts' (Seeger 2008). Automatic monitoring stations, although staying more sensitive, could not perceive a natural environment as holistically as humans (Chong and Kumar 2003). Modern devices – smart phones, equipped with Global Positioning System (GPS) sensors – are becoming widely available with Internet connectivity, and could be used for self-driven surveys even by non-qualified persons.

Ways in which people perceive their environments through culture, morality and social interaction (Raento, Oulasvirta, and Eagle 2009) can be fed into the expert system containing previously collected

¹ Centre for Geospatial Science, University of Nottingham, U.K.

email: lgxsp7@nottingham.ac.uk, Internet: http://www.nottingham.ac.uk/cgs

² School of Geography, University of Nottingham, U.K

³ SciSys U.K. Ltd

information, analysed in real-time and brought back through broadband internet to in-field surveyors, multiplying the outcome. The data itself may be photographic records, sound recordings or text input.

Currently environmental datasets are usually collected under a rigorous scientific process, funded by governments or the private sector. Because such survey often demands the collection of data in the field by scientists or trained staff the costs can be high and timescales for data collection long (Carver et al., 1995). This in turn often leads to a spatial resolution, spread of time period for the collection and frequency of sampling that is less than ideal especially where an area is experiencing rapid change. Statistical techniques can help in optimizing the nature of the collection, but they are often not sufficient (Conroy and Gordon 2004). Costs, time, and both the spatial and temporal resolution associated with traditional means of environmental data collection using professional surveyors/scientists, can be overcome by using volunteered geographic information (VGI).

One of the potential problems of volunteered geographic information (VGI) is ensuring its quality. Innocent mistakes and intentional falsehoods can reduce not only the quality of the information, but also people's confidence in VGI as a legitimate source of data (Mummidi and Krumm 2008). Poor data quality leads to three major consequences: user dissatisfaction, increased operational costs and employee dissatisfaction (Wang, Strong, and Guarascio 1996). Accuracy of data collected during surveys is an important parameter, as on average, about half of what informants report is probably incorrect in some way (Bernard et al., 1984).

We propose an architecture solution for location-aware self-driven surveys for environmental monitoring. A knowledge base system, located on a remote server, capable of storing large datasets and powerful enough to interpret partial results to give statistically and qualitatively refined questions, will enable a feedback process which can be repeated several times during the same survey. If there are more surveyors in the same area, the expert system can simultaneously parse and mix collected information enabling the better validation of collected data. This architecture has been tested on a use case based on the "tranquillity data" (see section 5), which looks at disadvantageous changes of the natural environment (MacFarlane et al., 2004). The previously proposed GIS model used in tranquillity mapping is extended and re-surveyed as a validation and updating process. The GIS model used in this research is compatible with Open Geospatial Consortium standards.

2. VGI surveying versus traditional spatial surveys

Understanding and exploiting data and services emerging from online communities is one of the recent challenges of Geographic Information Science. Using the emerging technologies of the social web, GI user roles switched from being data consumers to becoming data producers, then challenging the usability of this generated GI data (Bishr and Kuhn 2007). In most European countries geodata is provided either by public or commercial institutions. On the one hand this procedure is of high quality in the sense of accuracy and homogenous integrity but on the other hand it tends to be very expensive and sometimes out of date so that is not always practical or suitable. Using GPS enabled smart mobile phone technology as a data collection tool deployed by the general public or amateur is another means of data collection often referred to as "crowd-sourcing" or volunteered geographic information (VGI).

Traditional geospatial surveying techniques								
Strengths							Weaknesses	
	•	datasets	collected	under	rigorous	scientific	• time consuming	

construction	 expensive long path from start to end not enough data
Opportunities	Threats
• may be useful to audit the validity on several samples gathered using our dynamic techniques	 probably rare in the future as economy shrinks and results are needed quickly

Table 1 Traditional geospatial surveying techniques

Previous publications have reported on the pros and cons of such data as compared to authoritative data collected from recognised survey organisations (Bernard et al., 1984, Jackson et al., 2010). Table 1 summarises the strengths and weaknesses of traditional surveying techniques. Crowd-sourced surveying is not new in environmental studies, but up to now this method of collection carried a lot of uncertainties as described in Table 2.

3. A dynamic surveying methodology

In order to alleviate the weaknesses of VGI we propose a dynamic adaptive surveying approach, which proceeds by real-time analysis of the data observations collected along with a knowledge base system containing past data, rules and models. The latter can then adapt the survey for each current volunteering surveyor in a spatially-aware way as the knowledge base engine uses all the current surveyors along with the spatial information of the surrounding environment. The detailed properties and potential problems with this approach are listed in Table 3.

Crowd-sourced surveying techniques				
Strengths	Weaknesses			
 volunteered information from people motivated in the subject, not only specialists massive amount of samples 	 innocent mistakes intentional falsehoods like all computer driven self-surveys, it is harder to achieve uniform quality in the results 			
Opportunities	Threats			
 data could be produced by the "ordinary people" without the barriers of the rigorous traditional survey construction quick path from start to results interpretation, allowing regular updates 	 stability of the computer system fed with unknown amount of data at the same time user interface limits of smart-phones, user interface incompatibility, excluding part of society from participation 			



The dynamic approach improves the VGI surveying process by making it both adaptive and modelbased. Instead of being just collected in-field and stored for further analysis, data are injected into a model (knowledge base), where new observations are compared in real-time with historical data and stochastic expectations enabling validations, corrections and suggestions, allowing a quality controlled survey in order to eliminate errors and optimise quality of outcomes and coverage.

The knowledge base system can trigger further data collection in a time-loop, where a remote server validates, identifies possible errors or even removes obvious errors.



Figure 1: Dynamic surveying principle

3. Geospatial knowledge base system

An overview of the geospatial knowledge base system is presented on Figure 1. It is based on open



Figure 2: Technical solution

standards (Lee and Percivall 2008). The whole platform assumes constant Internet connectivity and is based on a client-server architecture.

A relatively generic implementation of the Data model of surveying system as shown in of Figure 2 consists of:

• Geoserver - a Java-based software server that allows users to view and edit geospatial data. Using open standards defined by the Open Geospatial Consortium (OGC), GeoServer allows flexibility in map creation and data sharing.

• Map base layer – geographical context and administration boundaries are based on Openstreetmap (OSM) and Ordnance Survey (OS) raster layers.

• PostGIS RDBMS, which adds support for geographic objects to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS). This two-way engine is the basis for collecting survey results and allows storing and processing data for post-survey analysis.

• Surveying smart-phone, Android- or Apple Iphone based devices, with 2- and 3-G Internet connectivity, GPS, a touch-screen and the built-in HTML5 web browser, which enables a direct GPS to browser interface.

Dynamic geospatial surveying techniques						
	Strengths	We	aknesses			
 added sy user possible required GSM cc more a improve a dyna data col 	patio-temporal dimension transparently to the e real-time adaptive surveying when it is l and technically possible (e.g. effective overage) appealing involvement from VGI users, ed 'public spirit' of submitted data mic, directed, iterative approach to in-field lection with real-time quality control	• potential technology limits as weak GPS signal, low transfer or even no internet availability outside urbanized areas				
	Opportunities	'hreats				
 live usage data professi sma automat to the ge it is tagged is to prese even n dynamic part unlimite 	e dataset refining, possible AI-like system aset may be crowd-sourced or fully onal, depends on the scenario art-phone can be connected with other ic sensors and collected data sent altogether eospatial knowledge base to improve quality s possible to include multimedia data (geo- images or video recordings) with the dataset ent it after collection for e.g. illustrative or nachine-processing purposes (image/audio c recognition) tial results can be mixed with almost ed layers of live information, e.g. twitter or	internet network depe harder to understand the dynamic content more expensive then s	ndent for an average user because of static crowd-sourced surveys			

RSS feeds
• results can be aggregated and presented live for
ad-hoc interpretation
• can share technology engine, and applied to
different research areas, can be useful in both: social
and natural sciences

Table 3 Potential of dynamic geospatial surveying techniques

The presentation layer is built on OpenLayers Javascript library. OpenLayers makes it easy to put a dynamic map in any web page. It can display map tiles and markers loaded from any source. OpenLayers is a pure JavaScript library for displaying map data in most modern web browsers, with no server-side dependencies. OpenLayers implements a JavaScript API for building rich web-based geographic applications, similar to the Google Maps and MSN Virtual Earth APIs, OpenLayers is Free Software, developed for and by the Open Source software community. OpenLayers is fully supported by the Geoserver spatial engine.



Figure 3: Information flow

The knowledge cascading within the knowledge base is orchestrated using an Open Source stack as illustrated on Figure 3.

.4. Interoperability aspects

Recent advances in Internet technologies, coupled with wide adoption of the web services paradigm and interoperability standards, makes the World Wide Web a popular vehicle for geospatial information distribution and online geoprocessing (Vatsavai et al., 2006). Associated with this shift is a new emphasis on context-aware computing (Townsend 2006), Neo-cartography spans ubiquitous cartography, user participation and considerations for geomedia techniques. This new expansion of multimedia and internet cartography combines the latest Web developments with traditional cartography and imagery research (Gartner and Rehrl 2009). New technical possibilities in the field of ubiquitous computing based on a growing wireless network coverage opens a wide spectrum of scenarios available for environmental surveys using public participation (PP).

As an alternative to HTML5 style applications and for some "system-driven" surveys, full-featured applications for particular smart-phone platform (Rogers 2010), also exist, e.g. Sypiens Survey1⁴ or mQuest Survey⁵ (available on Android platform), or SurveyPocket⁶ (for Iphone mobiles).

The received data is analysed in real-time in a cascading manner as illustrated in Figure 3: quality control, predictive model using combined data sources. Where insufficient data exists for the required precision of prediction the system will use the ability to locate and contact (via SMS, email, web pop-up) other volunteers in the field in the required locality and prompt them to collect further data samples.

5. The tranquillity use case

The "tranquillity" project (CPRE 2006) defined the concept and a measure of disadvantageous changes of the natural environment. "Tranquillity" is a nature preservation campaign from CPRE: "Campaign to Protect Rural England wants a beautiful, tranquil and diverse countryside that everyone can value and enjoy". Disadvantageous changes or threats to natural tranquillity are for example: new roads, more planes and runways, increased light pollution, new buildings and infrastructure. The initial knowledge base we used for the use case includes findings from the previous work: Tranquillity Mapping 2004, carried out by CESA (Centre for Environmental and Spatial Analysis) and PEANuT (Participatory Evaluation and Appraisal in Newcastle upon Tyne) at Northumbria University (MacFarlane et al., 2004).

The already proposed GIS model for tranquillity mapping is extended with a community layer, resulting from public participation (PP and VGI). Jackson, et al., 2008, have built Tranquillity maps from layers of positive and negative attributes of landscape – based on a nationwide survey to test what tranquillity means to people and the different factors which make up 'tranquillity' (see Table 4).

Positive attributes	Negative attributes		
 Seeing a natural landscape Hearing birdsong Hearing peace and quiet Seeing natural looking woodland Seeing the stars at night Seeing streams Seeing the sea Hearing natural sounds Hearing wildlife Hearing running water 	 Hearing constant noise from cars, lorries and/or motorbikes Seeing lots of people Seeing urban development Seeing overhead light pollution Hearing lots of people Seeing low flying aircraft Hearing low flying aircraft Seeing power lines Seeing towns and cities Seeing roads 		

Table 4. Positive, negative and combined attributes of tranquillity mapping from the Easter Bank

http://www.androidzoom.com/android_applications/productivity/sypiens-survey_kqwc.html

⁵ http://www.androidzoom.com/android applications/productivity/mquest-survey-demo mpue.html

⁶ Software available on Apple iTunes marketplace.

Holiday weekend, April 2006 survey

The original tranquillity dataset resolution is low, as the basic grid size is only 500m x500m (Figure 4). The first aim of our experiment was to increase the resolution and validate at this new resolution, the



Figure 4: A fragment of the map showing the range of tranquillity in Derbyshire area. Reproduced courtesy of the CPRE, 2007. Mapped from left: positive attributes, negative attributes, composite weight of both.

previous findings of the ordered negative and positive key layers.

A first experiment with six participants, aiming at validating the tranquillity but also as a proof of concept of our architecture, was conducted in October 2010. Participants were given a smart-phone, which had pre-cached fragments of the CPRE tranquillity database content for the selected area close to Derby. The survey was user-driven; the participants were free to use the application any time they wanted. The application was dynamically serving two of the most negative, two of the most positive and two random attributes based on position (GPS) to the participant enabling them to provide their personal opinion. It was also possible to answer a simple yes/no question, to express positive or negative feelings related to the location. Details of the answers and behaviour recordings during this validation experiment is available online (see Figure 5).

Off-line, using a local smart-phone database and local-on-line; using a laptop to serve (via WiFi connection) the dynamic survey in an active mode, has been tested. This was necessary in this rural area where the GSM internet signal is weak or unavailable.



Figure 5: Initial densifying and validating survey results, online version is available at: <u>http://testsurvey.pawlowicz.name/</u>. Map base layer: CC-By-SA by Openstreemap.org

6. Discussion and conclusion

People, as source and recipient of VGI, build up a common sense environmental knowledge-base through collaboration with a large distributed community (Singh 2002). This spatial social web as citizen science presents a unique opportunity to achieve the goal of reducing the cost of collection, maintenance and update of geospatial data (Bishr and Kuhn 2007). The INSPIRE European Union directive, encourages European citizens "to support the understanding of the complexity and interactions between human activities and environmental pressures and impacts" (INSPIRE 53:44). Our approach allows one to go down the line of developing the capability of achieving the above statements through public participation.

Addressing the quality of the data and the interoperability issues when approaching the general public, we proposed an open framework and an architecture allowing seamless outreach to wide communities and enabling the survey to be adaptive to the user-context. The survey driven in real-time by the knowledge based system becomes optimised according to the rules implemented, e.g., spatial awareness, quality control. Describing -a surveyed questionnaire being the same whatever the situation - as of interaction level 0, our architecture can be though of enabling an interaction level 1- where already known information can be used to drive the survey (e.g., existing geographic feature, previously estimated distribution of answers)- an interaction level 2 - dynamically integrating current information from other surveyed participants- and an interaction level 3 – dynamically integrating information from other sources (e.g., Tweeter).

For the proof of concept and using the "tranquillity data" project, the knowledge base system was

limited to dynamic verification of previously identified attributes and mapping of results, but more complex workflows generating the updated questionnaire, triggering for more data or more surveyors can be developed without slowing the efficiency of the system.

The system can be made available on-line so that everyone and particularly communities could perform their own 'densifying and validating' surveying experiment for tranquillity data or any other environmental assessment. It is nonetheless believed that at this stage the knowledge-based engine has to be first made more user friendly. Future research will focus on more transparent caching of data in the information cloud, creation of a more stable and intuitive user interface, with API extension. For semantic interoperability purposes this future work will need improved standardisation of geospatial ontologies in order to, for example, develop multidisciplinary knowledge based systems, which is the backbone of this dynamic surveying system.

Besides applying the dynamic surveying system to "tranquillity" we are also doing experiments in the health-related field (Adams et al., 2011). Aggregated live Twitter feeds are planned to be used for information support during spatially enabled in-field surveys in both contexts.

Acknowledgements

The research reported here was sponsored by SciSys UK Ltd. and the UK EPSRC.

Bibliography

- Adams, N., Ye, C., Jiang, W., Pawlowicz, S., Anand, S., Leibovici, D.G., Jackson, M. (2011): Participatory Health Surveys Using Ubiquitous Computing: gastrointestinal illnesses application case study. In: Proceedings of the 19th GISRUK conference, Portsmouth 27-29 April 2011, UK.
- Bernard, H.R., Killworth, P.D., Kronenfeld, L. (1984):The Problem of Informant Accuracy: The Validity of Retrospective Data. Annual Review of Anthropology 13 (1) (October): 495-517
- Bishr, M., Kuhn, W. (2007): Geospatial Information Bottom-Up: A Matter of Trust and Semantics. In The European Information Society, 365-387
- Carver, S., Heywood, I., Cornelius, S., Sear, D. (1995): Evaluating field-based GIS for environmental characterization, modelling and decision support. International Journal of Geographical Information Science 9 (4): 475-486
- Chong, C.Y., Kumar, S.P. (2003): Sensor networks: Evolution, opportunities, and challenges." Proceedings of the IEEE 91 (8): 1247-1256
- Conroy, M.M., Gordon, S.I. (2004): Utility of interactive computer-based materials for enhancing public participation. Journal of Environmental Planning and Management 47 (1): 19-33
- CPRE (2006): CPRE leaflet Saving Tranquil Places: East Midlands. CPRE, October 2006
- Danielsen, F., Burgess, N., Balmford, A. (2005): Monitoring Matters: Examining the Potential of Locallybased Approaches. Biodiversity and Conservation 14 (11) (October 8): 2507-2542
- Gartner, G., Rehrl, K., (eds) (2009): Location Based Services and TeleCartography II. Springer, Berlin Heidelberg. http://www.springerlink.com/content/k5207h250812x308/.INSPIRE. 11:53:44. INSPIRE. http://inspire.jrc.ec.europa.eu/
- Jackson, M. J., Rahemtulla, H., Morley. J. (2010). The Synergistic Use of Authenticated and Crowd-Sourced Data for Emergency Response, Proc. 2nd Int. Workshop on Validation of Geo-Information Products for Crisis Management (VALgEO), 11-13 Oct. 2010, Ispra, Italy, pp 91-99. Available online: <u>http://globesec.jrc.ec.europa.eu/workshops/valgeo-2010/proceedings</u>
- Jackson, S., Fuller, D., Dunsford, H. (2008): Tranquillity Mapping: Developing a Robust Methodology for

Planning Support. Online: http://www.cpre.org.uk/resources/countryside/item/download/542

- Lee, C., Percivall, G. (2008): Standards-based computing capabilities for distributed geospatial applications. Computer 41 (11): 50-57
- MacFarlane, R., Haggett, C., Fuller, D., Dunsford, H., Carlisle, B. (2004): Tranquillity Mapping: developing a robust methodology for planning support. Report to the Campaign to Protect Rural England. Countryside Agency, North East Assembly, Northumberland Strategic Partnership, Northumberland National Park Authority and Durham County Council, Centre for Environmental & Spatial Analysis, Northumbria University.
- Mummidi, L.N., Krumm, J. (2008): Discovering points of interest from user's map annotations. GeoJournal 72 (3): 215-227
- Raento, M., Oulasvirta, A., Eagle, N. (2009): Smartphones: An Emerging Tool for Social Scientists. Sociological Methods Research 37 (3) (February 1): 426-454
- Rogers, R. (2010): Developing portable mobile web applications. Linux Journal 2010 (197): 3
- Seeger, C. J. (2008): The role of facilitated volunteered geographic information in the landscape planning and site design process. GeoJournal 72 (3): 199-213
- Singh, P. (2002): The public acquisition of commonsense knowledge. In: Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access.
- Townsend, A. (2006): Locative-Media Artists in the Contested-Aware City. Leonardo 39 (4): 345-347.
- Vatsavai, R., Shekhar, S., Burk, T., Lime, S. (2006): UMN-MapServer: A High-Performance, Interoperable, and Open Source Web Mapping and Geo-spatial Analysis System. In Geographic, Information Science, 400-417
- Wang, R.Y., Strong, D.M., Guarascio, L. (1996): Beyond accuracy: What data quality means to data consumers. Journal of management information systems 12: 5–34