Challenges in VGI for scientific projects

Volunteered Geographic Information (VGI) is a recent trend that has been successfully used in order to collect and share geographic information. This method is of interest for scientists who are in need of data and who want to get people involved in their cause. In this paper we discuss the challenges and opportunities that scientists may face when using the concept. An initial challenge is to find users who are willing to contribute. Second, scientist must get these users to interact with the application and with each other. The final goal is to end up with high quality data that can be used for scientific research.

CHALLENGES IN VGI FOR SCIENTIFIC PROJECTS

Jens Ingensand[1], Sarah Composto[1], Marion Nappez[1], Timothée Produit[1], Olivier Ertz[2], Mathias Oberson[2], Daniel Rappo[2]

[1] University of Applied Sciences Western Switzerland, G2C Institute, Department EC+G

[2] University of Applied Sciences Western Switzerland, MEI Institute, Department Comem+

ABSTRACT

Volunteered Geographic Information (VGI) is a recent trend that has been successfully used in order to collect and share geographic information. This method is of interest for scientists who are in need of data and who want to get people involved in their cause. In this paper we discuss the challenges and opportunities that scientists may face when using the concept. An initial challenge is to find users who are willing to contribute. Second scientist must get these users to interact with the application and with each other. The final goal is to end up with high quality data that can be used for scientific research.

INTRODUCTION

VGI allows citizens to gather geographic data, making data creation cheaper and access to geographic information easier (Goodchild M. , 2007a). The success of GPS-equipped smartphones and tablet computers has made it easy for any citizen to access and share geographic information anywhere on the planet.

It is in this context that the scientific community has started using the concept to gather geographic information that otherwise a) would have been very expensive and b) would have taken much time to get. VGI as a concept however relies on the fact that citizens

1. Find an application / a website among a vast choice
2. See an interest in spending time sharing geographic information and
3. Use the tool to collect and share information (Rotman, et al., 2012).

These three points are crucial for the success of a project involving VGI. Especially within the context of scientific projects these points are even more difficult to address since scientist often have very specific prerequisites regarding data quality (e.g. accuracy, completeness, etc). Moreover scientific projects are often limited in time and the development of VGI applications (concept, implementation, testing) is time-consuming.

In this paper we want to address the challenges for using VGI within scientific projects. This paper is structured as follows: first we address the question how scientist can get users to be aware of their projects and to discover an application or website. Second we discuss how users can be convinced to not only use an application one time, but frequently. The third point concerns data quality and the question how high quality data can be collected. Another important issue are legal aspects such as the identification of the data owner and how data which has been collected by several users can be reutilized. Finally we will describe ideas for further developments and present our conclusions.

FINDING USERS

Obviously VGI without volunteers is difficult. An initial challenge is therefore to get people to see an interest in using a VGI platform.

First of all advertising allows users to notice the project and to understand the project's intentions (Engels, 2015). A vast choice of advertising possibilities from traditional media such as newspapers, radio and television or more recent media such as social media can be used. The question that needs to be raised in this context is "Who are the users?", "In which context do we situate the intended users (work, free time, stress, etc.)?" or "During which time of day would advertising make sense (e.g. while the intended users travel to work, read newspapers, watch television)?"

Media campaigns may stimulate the users' participation (Schoberth, Heinzl, & Preece, 2006), however such campaigns can be expensive and time-consuming. According to (Porter, 2008) good ways to find users is through word of mouth, blogs, reviews, links, and so on. Moreover if we focus on a specific community, it is possible to connect as early as possible with users. Citizen must be engaged as soon as possible in the project process because it will affect the application method such as how data are gathering (Engels, 2015).

According to Rogers (Rogers, 2003) there are four main components regarding diffusion: the innovation, the communication channels, the time and the social system. Indeed "diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system" (Rogers, 2003). For this purpose social networks are regularly used by citizens and thus provide a good way to find new users (Bartoschek & Kessler, 2012).

Within the context of scientific projects starting a new VGI project may not be as easy, as the scientific context and thereby also the outcome and the benefits may be more difficult to explain.

KEEPING AND INVOLVING USERS

Once a user has found a website or installed an app the goal is to involve the user in the project. According to (Preece & Shneiderman, 2009) social participation can be divided in successive levels. At first users just watch into an application. However in a second time some of these users decide to return to the application and to participate only a little in order to avoid exposing themselves. Finally a few of these users return again to the application and participate more actively. Thus more we climb a level of social participation, less users there are. Those successive levels are named reading, contributing, collaborating and leading.
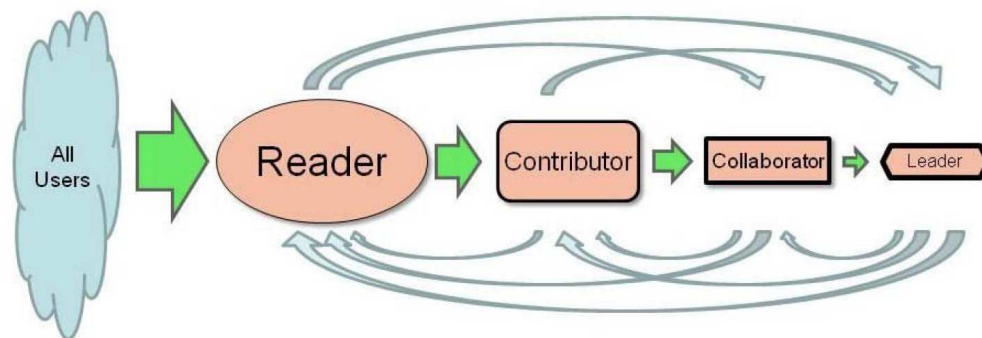
The Reader-to-Leader Framework (Preece & Shneiderman, 2009)

The difference between the contributing and collaborating levels are the users' initial intentions. Contribution means that a user adds e.g. a picture or a comment although initially he does not have the intention to make any contribution. If a user regularly contributes through an application, he becomes a collaborator (Preece & Shneiderman, 2009). According to Riddell (Riddell, 2014) different reasons could motivate users to participate more such as health or development concerns, empowerment, environmental protection, learning new skills, making social connections and having fun. On the other hand some people tend to resist to new things and thus require strong encouragement from a trusted friend or respected authority (Fogg, 2002). This fact can be illustrated by an example of a Peruvian village (Rogers, 2003): Most residents of this village do not boil water despite of sanitary recommendations. In fact, the majority of people who do boil their water were persuaded during a health worker's regular visits.

The contribution of volunteers can be categorized in terms of the complexity of the task as well. Haklay (Haklay M. , 2013) suggests a scale from citizens serving as a simple sensor (level 1 – crowdsourcing), as basic interpreters (level 2 – distributed intelligence), as participators in easy tasks (level 3 – participatory science) or as participators in more complicated tasks (level 4).

Contribution levels in VGI (Haklay M. , 2013)

Bartoschek and Kessler (Bartoschek & Kessler, 2012) state that a key for keeping users is usability. A website's or an application's interface must be adapted to the people who will use it (Preece & Shneiderman, 2009). For instance if the intended audience are children, it is important that the interface contains pictures as children do not have an extensive vocabulary and do not like to read a lot. (Alburo, et al., 2005).

In any case the application's interface must be kept as simple as possible or as Porter (Porter, 2008) points out - a complicated interface suggests a complicated service. Therefore a complicated interface can demotivate people using the application. In order to show people an application's ease of use, it is also possible to make a "How it Works" namely a short graphic which explains the main points of the application. Another issue regarding VGI applications is the time for a contribution. Schmidt et al. (Schmidt, Klettner, & Steinmann, 2013) for instance have identified time as the main barrier for collaborating to the OpenStreetMap project.

Furthermore an application's success also depends on the vitality of the subject and the size of the community (Schoberth, Heinzl, & Preece, 2006). According to a study the growing size of a community does not only have positive effects (Whittaker, Terveen, Hill, & Cherny, 1998). Indeed this tendency can increase information overload and it can get more difficult to find common ground which finally leads to users ceasing collaboration. Bartoschek and Kessler (Bartoschek & Kessler, 2012) suggest that a mixed form of VGI and social networks could keep users to collaborate. Moreover quick feedback through an application can incite citizens in continuing using an application.

Using VGI for scientific projects may be more complicated, as scientific project often have very specific intentions that not necessarily meet a citizens' personal concerns. Furthermore there are not necessarily existing communities that can help adopting a VGI platform. The challenges for scientist (in addition to aforementioned elements) are therefore to: a) explain the project's context in an understandable manner and to find incitements to convince users to contribute. b) define the user's level of contribution (e.g. ranging from simple digitization to more complex tasks) and c) to find and offer ways to make users interact (among each other, but also with the scientists).

Engels (Engels, 2015) suggest a framework how scientists can be involved in citizen science project. In this framework scientists interact with the users of a platform at several stages: scientists can a) express a need for data, b) give recognition to volunteers, c) train volunteers and d) comment and validate data that has been added. Engels suggests that a deep involvement of the scientist can stimulate the circle of involving users, using a system, sharing data and validating data.

Goodchild (Goodchild M. F., 2007b) argues that data that is produced by local observers has the potential to be more accurate than data that is collected through other techniques since these citizens are familiar with the area. Data quality however depends on a series of different factors.

Quality is often defined as "fitness for use" (Chrisman, 1983). According to the International Standards Organization (ISO) quality means "the totality of characteristics of a product that bear on its ability to satisfy stated and implied needs". Van Oort (van Oort, 2006) cites different aspects that define spatial data quality such as

- Lineage: history of a dataset
- Positional accuracy (horizontal and vertical): accuracy of coordinate values of a dataset relates to the reality on the ground
- Attribute accuracy: accuracy of all attributes other than the positional and temporal attributes of a dataset
- Logical consistency: internal consistency of a dataset (fidelity of relationships that are encoded in a dataset)
- Completeness: measure of the lack of data (errors of omission) and the presence of excess data (errors of commission)
- Semantic accuracy: link with the way in which the object is captured and represented in the database to its meaning and the way in which it should be interpreted
- Usage, purpose and constraints: help for a potential user of a dataset
- Temporal quality: validity of changes in the database in relation to real changes
- Variation in quality: homogeneity within a database
- Meta-quality: information on the quality of the quality description
- Resolution

Spatial data quality is not a recent issue and has been important for all citizen science projects in general (Engels, 2015). Van Oort (van Oort, 2006) for instance identifies some work about spatial data quality dating back to the late 1960s. With the emergence of VGI the data quality issue is growing – especially if the data collected is used for decision making (van Oort, 2006). Any errors and uncertainties in spatial data may have practical, financial and even legal implications (Huisman & De By, 2001). For several disciplines high quality data is crucial, such as public health, urban planning, mobility or environmental protection.

Citizens are not necessarily experts regarding the data that the creators of VGI applications want them to gather and to share. According to Cho (Cho, 2014) data quality depends on the background and training of the users. The problem of geographic accuracy for instance can be an important issue if the data is collected using smartphones / tablet computers in urban areas due to GPS reception problems. Furthermore it is also possible that some users willingly share inaccurate data. Data may contain different degrees of error and according to Huisman and De By (Huisman & De By, 2001) there are three main categories namely gross errors (mistakes), variations in the data measurement (imprecisions) and classification errors.

In order to ensure data quality, researchers have identified several possibilities: One of these possibilities is the users' authentication. If a user creates an account, it allows him to interact with other users, e.g. by commenting on others users' contributions (Castelein, Grus, Crompvoets, & Bregt, 2010). According to the Hickling Arthurs Low Corporation (Hickling Arthurs Low Corporation, 2012) "login authentication ensures that error reporting and change submission are only performed by authorized users". Moreover registration allows for user classification (Maué, 2007). A newly registered user can be restricted with some application features (restrictions on participation) such as for instance modifying others users' contributions. In order to access all features, users must gain a high reputation by collaborating through the application. However according to a study made by Zhou (Zhou, 2014) authentication discourages new user from reporting.

There are several possibilities to analyze data once collected and to rule out false data. These methods can be categorized into automated methods and user-controlled methods. (Fonte, et al., 2015)

Regarding automatic methods it is first of all important that the possibilities of users to make errors must be minimized: in many cases attributes depend on other attributes and if a user has filled in one attribute, other attributes can be changed or deactivated automatically. When it comes to geographic information it is often

possible to rule out areas where the information that a user can share is impossible (e.g. sharing a flower in the lake or sharing a car on to of a mountain).

The community itself can be used to improve data quality through the implementation of options that allow users to confirm or correct data. (Goodchild & Li, 2012). This idea has been successfully implemented in several projects. The "Notes" function for example is a tool that allows users to report errors or to make annotations in the OpenStreetMap data. This concept has another positive side-effect: users get feedback on their data and thereby may get even more interested in using an application.

Another way to insure data quality is through the utilization of standards, specifications, protocols and metadata formats. In the case of some applications such as the GLOBE (Global Learning and Research to Benefit the Environment) project – which aims at improving the earth's environment – protocols are established in order to ensure high quality data (Goodchild M. F., 2007b). Furthermore protocols can be adjusted to prevent mistakes by observing users during data acquisition in order to understand what kind of mistakes they make (Engels, 2015).

Communication is also a crucial point in the field of VGI. Indeed project managers must communicate with users in order to train, support and help them (Engels, 2015). As an example error messages can be sent to users in order to ask them to check data they have just digitized.

LEGAL ISSUES

If a user shares information, he does work. Even if the user is not paid for his work, in some cases he can stay the owner of his data and/or have the right to delete or modify his data. Another issue is the reutilization of already collected data for different purpose. Furthermore geographic data can be collected anywhere on the planet – if the data collected is located on private ground the owner of the ground may have the right to ask the creators of a VGI platform to delete or modify the data.

Cho (Cho, 2014) identifies different legal concerns: a) Data quality and credibility, b) Ownership and intellectual property rights: e.g. there could be difficulties in tracing the data source due to the number of contributors. In other words it could be difficult identify the person who digitized data. One solution to this problem is the Creative Commons license. c) Liability: Paid professionals are held responsible for their data. However it is difficult to do the same with volunteers, especially if a group of users has contributed to the same set of data. It is very difficult to determine who is liable if there is a mix of responsibilities between the various volunteers and also between the professional controllers. (Janssen, 2011)

According to Janssen (Janssen, 2011) another issue is privacy, which is defined by "any information related to an identified or identifiable person". Janssen quotes the example of "citizens as sensors"; where volunteers are localized and tracked. The location can also relate to the location of other people. The main question is to know what the reasonable expectation of privacy is. One possible solution to this problem is to inform volunteers about the data they share.

The legal situation for these issues is important and in many cases not clearly specified (it can even change from country to country). It is therefore important that the creators of a VGI platform clearly specify how the collected data will be used and if the data may or may not be used within other contexts. Moreover according to Cho (Cho, 2014) a way to solve these different issues is to develop standards and to address the different issues in licenses and documentation.

CONCLUSIONS

VGI is an efficient possibility for scientists to get data and to make people aware of specific problems. This way of collecting georeferenced information has some advantages such as the fact that it is cheap and that theoretically enormous amounts of data can be collected. Moreover VGI can reach regions where no scientist will or can put their foot such as for instance private ground. However VGI is not just about quickly developing a working webpage or mobile application that lets users share information and hoping that everybody will use it.

Although most citizens have a device in their pocket that lets them quickly consume and share information, there is a vast choice of applications, webpages, social networks etc. around that the creators of new VGI platforms either need to compete with or use in order to get citizens interested in their application. We therefore suggest some important points that scientist need to be aware of before creating a VGI platform:

First scientist need to figure out how their subject can be explained to citizens in a way that he can understand. This point has two important consequences. By understanding the importance of a subject, citizens can become motivated in helping and sharing information. Furthermore motivated citizens can motivate other citizens to do the same.

Second, the creators of VGI platforms need to figure out how to get users to find an application or a web-platform. Methods such as newspaper ads, radio transmissions, social networks and the word of mouth can be used. Each method has its advantages and disadvantages in terms of cost or the people that can be reached.

Third, once a citizen has been convinced to try out a website or a mobile application, he must be encouraged to not only consume the information once, but to return several times in order to contribute and to collaborate. Techniques that can be used to achieve this goal are on one hand interface techniques such as user experience engineering, but also methods that make the user understand that his contribution matters. Another important technique is that users get something in return for their work (e.g. a map or points in a ranking of contributors)

The quality of data and legal issues are other important aspects of VGI. Data quality can be verified and corrected using automated methods (e.g. using validated reference data) or through the community itself (e.g. by verifying and commenting on data). Furthermore it is important to verify and announce legal aspects before the application is released.

REFERENCES

Alburo, J., Komlodi, A., Preece, J., Druin, A., Elkiss, A., & Resnik, P. (2005). *Evaluating a cross-cultural children's online book community: Sociability, usability and cultural exchange.*

Bartoschek, T., & Kessler, C. (2012). *VGI in Education - from K-12 to Graduate Studies.*

Castelein, W., Grus, L., Crompvoets, J., & Bregt, A. (2010). A characterization of Volunteered Geographic Information. *13th AGILE International Conference on Geographic Information Science - Portugal.*

Cho, G. (2014). Some legal concerns with the use of crowd-sourced Geospatial Information. *7th IGRSM International Remote Sensing & GIS Conference and Exhibition.*

Chrisman, N. (1983). The role of quality information in the long-term functioning of a GIS. *Proceedings of AUTOCARTO6.*

Engels, B. (2015). Citizen science: An overview of the current state, the possibilities and challenges and the opportunities for the future. *WaLTER.*

Fogg, B. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do.* San Francisco: Morgan Kaufmann.

Fonte, C. C., Bastin, L., Foody, G., Kellenberger, T., Kerle, N., Mooney, P., . . . See, L. (2015). VGI Quality Control. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* La Grande Motte, France.

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

Goodchild, M. F. (2007b). Citizens as sensors: the world of volunteered geography. *GeoJournal*.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 110-120.

Haklay, M. (2008). How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey Datasets.

Haklay, M.(2013). Citizen science and volunteered geographic information overview and typology of participation.

Hickling Arthurs Low Corporation. (2012). *Volunteered Geographic Information (VGI) Primer*.

Huisman, O., & De By, R. (2001). *Principles of Geographic Information Systems*. Enschede: ITC The International Institute for Geo-Information Science and Earth Observation.

Janssen, K. (2011). Spatial data contributed by citizens: legal challenges. *INSPIRE Conference*.

Li, C., & Bernoff, J. (2008). *Groundswell: Winning in a World Transformed by Social Technologies*. Boston: Harvard Business Review.

Maué, P. (2007). Reputation as tool to ensure validity of VGI. *Workshop on volunteered geographic information*.

Porter, J. (2008). *Designing for the Social Web*. Berkeley: New Riders.

Preece, J., & Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction*, pp. 13-32.

Priedhorsky, R. (2007). Specialist Meeting on Volunteered Geographic Information Position Paper. *Workshop on Volunteered Geographic Information*. Santa Barbara.

Riddell, K. (2014). *Community based citizen science: Learning from success in the United States and Canada.*

Rogers, E. (2003). *Diffusion of Innovations, 5th Edition*. New York: Free Press.

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Jacobs, D. (2012). Dynamic Changes in Motivation in Collaborative Citizen-Science Projects. *CSCW '12 Computer Supported Cooperative Work*. Seattle.

Schmidt, M., Klettner, S., & Steinmann, R. (2013). Barriers for Contributing to VGI Projects. *International Cartographic Association*.

Schoberth, T., Heinzl, A., & Preece, J. (2006). Exploring Communication Activities in Online Communities: A Longitudinal Analysis in the Financial Services Industry. *Proceedings of the 36th Hawaii International Conference on System Sciences*.

van Oort, P. (2006). *Spatial data quality: from description to application*. PhD Thesis, Wageningen.

Whittaker, S., Terveen, L., Hill, W., & Cherny, L. (1998). The dynamics of mass interaction. *Proceedings CSCW '98*.

Zhou, Y. (2014). *Volunteered Geographical Information: an Alternative Solution for Overcoming the Chasm Between Stormwater Management and Community Participation.*