

A BOTTOM-UP GEOSPATIAL DATA UPDATE MECHANISM FOR SPATIAL DATA INFRASTRUCTURE UPDATING

TianWenwen ^{a,*}, ZhuXinyan ^a, LiuYi ^b

^a State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University
129 Luoyu Road, Wuhan, Hubei, 430079, China - geotww@foxmail.com

^b School of Geodesy and Geomatics, Wuhan University
129 Luoyu Road, Wuhan, Hubei, 430079, China

KEY WORDS: Geography, Vector, Integration, Matching, Model

ABSTRACT:

Currently, the top-down spatial data update mechanism has made a big progress and it is widely applied in many SDI (spatial data infrastructure). However, this mechanism still has some issues. For example, the update schedule is limited by the professional department's project, usually which is too long for the end-user; the data form collection to public cost too much time and energy for professional department; the details of geospatial information does not provide sufficient attribute, etc. Thus, how to deal with the problems has become the effective shortcut. Emerging Internet technology, 3S technique and geographic information knowledge which is popular in the public promote the booming development of geoscience in volunteered geospatial information. Volunteered geospatial information is the current "hotspot", which attracts many researchers to study its data quality and credibility, accuracy, sustainability, social benefit, application and so on. In addition to this, a few scholars also pay attention to the value of VGI to support the SDI updating. And on that basis, this paper presents a bottom-up update mechanism from VGI to SDI, which includes the processes of match homonymous elements between VGI and SDI vector data, change data detection, SDI spatial database update and new data product publication to end-users. Then, the proposed updating cycle is deeply discussed about the feasibility of which can detect the changed elements in time and shorten the update period, provide more accurate geometry and attribute data for spatial data infrastructure and support update propagation.

1. INTRODUCTION

Over the past decades, the majority of geospatial information has been captured, processed, managed and updated by the public institutions, such as spatial data infrastructure. SDI as public data producer usually adopts a top-to-bottom mechanism to collect and update geospatial information, then share with the regional and local institutions. The mechanism is limited to the institutions to design the updating schedule for keeping the consistency between geo-spatial information and real world. However, more and more remote sensing image and aerial photos have been acquired by advanced technology, how to process the mass data, detect the changes, update the out-time data in time to reflect the real world to users has become an intractable issue.

With the development of web 2.0, new solutions has emerged to create and share information on the internet, such as RSS, Tag, Trackbook, Blog, Wiki, SNS and so on. The new Web 2.0 is a two-way collaboration, the users interactive with Web sites, not only provide information for the site, but also provide information service for others. Wikipedia is a typical example, a lot of the individual user can provide any content of articles, the content is managed through the management and administrator group consisted by a relatively small amount member. It provides a broad, loosely organization metadata, users can access all history of any articles, including the creation of the first version and various change. An emerging trend in the wider information community is the growing use of open portals to collect and share information, both spatial and non-spatial (McDougall, 2009). Google earth, Openstreetmap, Wikimapia etc. all provide the web2.0 platform and service for public users to create and share the geo-spatial information as

volunteered geographic information (Goodchild, 2007). The most successful example is Openstreetmap, which is founded by Steve Coast in 2004. Currently, the project has over 300,000 users, the size of geospatial vector data has reached 200G, not only cover England, but also all over the world. Because the information of OSM is updated very quickly and free for use, many companies and researchers adopt the data of OSM. The most outstanding example is the application of Hanti earthquake, (Zook and Graham et al., 2010)

There are a growing consensus to recognize the role of VGI in support of SDI updating, many literature have studied the topic and discuss many related issues, such as new rules and standards and the applicability of VGI in SDI (Coleman and Georgiadou et al., 2009), the possibility of citizen volunteered spatial information for building SDI (McDougall, 2009), linking contributor motivations to program effectiveness (Coleman and Sabone et al., 2010), the opportunities and constraints of volunteered geographic information in spatial data infrastructure (Coleman and Eng, 2010), whether volunteered geographic information can be a participant in eEnvironment and SDI (Mooney and Corcoran, 2011) etc. Based on the above researches, more and more person realize how to find a way or design strategies and systems from VGI to SDI has become an opportunity and challenge in the GIS research. There are so many problems need to deal with, such as how to collect the data from crowd-sources on the VGI platform, how to filter and clean the data, how to get the information from the multi-sources data, how to match homonymous elements between VGI and SDI, how to detect the changes and alert the institutions to verify the credibility of the information, how to update SDI database timely and so on.

* Corresponding author.

In this paper, we try to build a mechanism or a cycle to collect real-time data form VGI to SDI, in turn SDI public the verified data to VGI. The mechanism or cycle includes collecting VGI data, preprocessing data, matching the homonymous elements between VGI and SDI vector data, detecting changes, verifying the credibility of the changes, updating SDI database, updating the products to end-users. The paper is organized as follows, section 2 briefly surveys the issues associated with the utilisation of volunteered spatial information in SDI; section 3 looks at the difficulties posed by the above issues in section 1 and give a framework of the mechanism or cycle, describes he mechanism or cycle itself in details, section 4 take the mapping departments of china as an example ,discusses the problem of the mechanism or cycle adopting in SDI; section 5 gives the conclusions and future directions of research.

2. PROBLEMS AND CHALLENGE OF VGI

There are a lot of issues that may exists in the mechanism of VGI in support of SDI updating, such as data quality, data structure, data differentiation, data copyright, data confidential and privacy and so on. Firstly , data quality depends on the way of collecting data in VGI , the users use mobile phone or on-board GPS to record the GPS track, draw the outlines of geographic elements on remote sensing image or aerial photos to acquire vector data, add attribute information according to their own experiences, upload audio and pictures to contribution data. There are many problems in the contributed data quality. The civil GPS do not have enough accuracy compared to professional GPS equipments; the feature drew by users without professional training may not fully meet the requirements of the professional departments, which is difficult to directly used by the professional surveying and mapping database; professional surveying and mapping data pay attentions on the standardization of attribute data, such as word length and data type, conversely in VGI operating environment which only has established some standard to limit the modification behaviour; the semantic ambiguity brought by the different opinions of cognitive limitations from users to collect attribute information. If using VGI data to in the support of updating professional data in surveying and mapping departments, the new rules and standards should be built for data preprocessing, filter and conversion to meet the standards of professional surveying and mapping departments. Secondly, in VGI platform the same features edited by many users may cause geometry objects with mass points, which brought more difficulties in a wide range of data acquisition and inquires. For instance, if the inquiries or requests over 0.25 square degrees of data by users will be refused server in OpenStreetMap. The massive geometry objects not only set up a barrier to the inquiry and request, but also obstruct the updating and modification, so how to build new, flexible and efficient data types to satisfy the professional departments become a big issue. Thirdly, for the differences of informationization level in different areas and the distribution of VGI users, the differentiation of VGI information becomes a problem. Usually, in densely populated areas VGI data is intensive, wherever the data in the most inaccessible areas is sparse. Beyond that, there is also data copyright, data security and privacy etc. This paper doesn't discuss in this section.

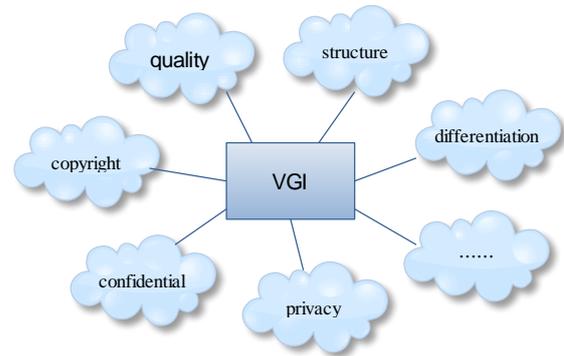


Figure 1. The problems of VGI data

3. FRAMEWORK OF THE MECHANISM

Based on the above discussion in section 2, aiming at settling the problem that how to utilise VGI data to update professional surveying and mapping department database timely and in turn professional department publics the enhanced data to VGI platform, this paper present a mechanism as organic cycle to support the requirement. This mechanism emphasizes the thought that “data from users, to users”, the users provide more real-time data by various ways, such as GPS, drawing the features of geographic objects etc. to SDI, SDI verify the validation of data and release the verified data to the user, which promote each other and help to improve the efficiency of data updating. Firstly, we give the current methods for VGI data collection and preprocessing; secondly according to the characteristics of VGI data, discuss the matching algorithms for point, line, polygon between VGI and SDI vector data to detect the changes; thirdly, provide automatic SDI database linkage updating rules; forth, release the real-time and verified data to VGI.

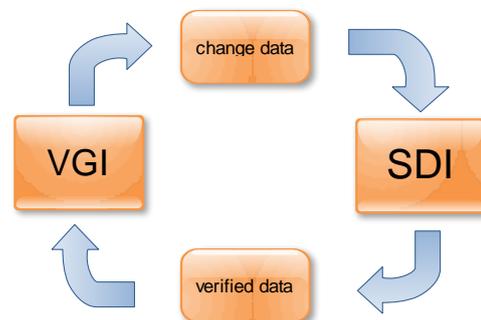


Figure 2. Framework of the mechanism

3.1 The methods of data collection and preprocessing

Nowadays, we make a general survey of current VGI websites, VGI data is mainly achieved by the following kinds: GPS track or point of interest recorded by handheld or on-board GPS; the outlines of geometry object for points, lines and polygons drew on online high resolution remote sensing image or aerial photos; attribute information based on the user's cognition; the texts, images, audios and videos which contain position information etc.. According to the nature of data, it can be divided into graphics data and semantic data, graphics data includes GPS track, POI and the outlines of geometry objects; semantic data includes attribute information and related text, images and video etc. Because there is a great gap between VGI data

acquisition methods and professional surveying and mapping methods, VGI data has some characteristics: the uncertainty of data quality; the differentiation of data distribution; the outlines of geometry objects drew by public not accord with the standards of professional surveying and mapping department, the complexity of big geometric objects edited by multi-users etc. So, if professional surveying and mapping department wants to harness VGI data, the process of data filter, cleaning, pretreatment and conversion would be necessary to meet the standards of professional surveying and mapping. The next part in this section will discuss the proposed problems.

At present, the mainstream VGI web sites such as Openstreetmap in order to get high quality data, design some rules to restrict the source and quality of data. They use user permissions, user contribution and the level of user professional to judging the certainty of data source, provide certain data format and fixed form for uploading data, adopt the multi-user and third-party to verify the validity of data. We think that two layer should be established for data filtering and cleaning, the first layer for getting reliable data source, the second layer for achieving reliable data. The verified data should be processed by format conversion, transformation of global coordinates, topological relationship building, space entity index, the consistent of graphics entity expression to meet the requirements of matching homonymous elements.

3.2 The algorithms of homonymous elements

The more important aspect of VGI data is the homogenous elements recognition and matching, which is fundament for detecting the changes and updating. The earliest case is a project of United States census bureau, in 1983 the United States census bureau merged its existing data map with USGS data map which was new, using the rubber band technology to adjust and iterative the node on the map constantly, for getting the matching relations between homogenous elements(Saalfeld, 1988). Since then, many scholars began to discuss the theoretical and practical problems in homonymous elements matching. The current data matching algorithm, mainly chose geometry, semantic and topological relations as matching indexes to find homogenous elements. According to the characteristics of VGI data, such as without scale concept, we provide an algorithm to extract multi-level tree data structure globally; choose the feature points as center points to match on each layer; combining multiple matching indexes, match homogenous elements of point, line and polygon in a given threshold area. For the data cannot be determined, professional department need to design update schedule, adopting other method to collect the change data, such as field collection.

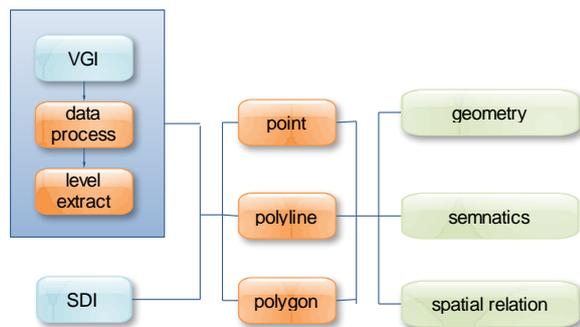


Figure 3. The matching of homonymous elements

3.3 Detect changes and define change type

By matching the homonymous elements between VGI and SDI vector data, the change information could be acquired more easily through the automatic or artificial interactive way, then the change information will be extracted and submitted to the professional surveying and mapping departments to identify and release real-time data. This update method belongs to incremental updating. When changes once happen, they will be detected and verified, and updated the professional spatial database, at the same time the out-of-time data will be archived. Lastly, SDI public the new and verified data products to end-users, which is a loop update mechanism for each other. Here, according to the type of data, the change types could be classified into three types, attribute change, geometry change, geometric and attribute both changing. Then based on the operation of the change, the change types could be divided into add, deletion and modification. We could integrate the first and second classification, the change types will be classified into the following 13 kinds of combination types: attribute add; attribute deletion; attribute modification; geometry add; geometry deletion; geometric modification; attribute add, geometry add; attribute add, geometry modification; attribute deletion, geometry add; attribute deletion, geometry deletion; attribute deletion, geometric modification; attributes modification, geometry add; attributes modification, geometric modification. As the attribute is affiliated with geometry, so when the geometry is deleted, the attribute doesn't exist.

3.4 The confirmation and publication of change for SDI

The data from VGI has so many uncertainty factors, such as possible malicious data. So after detecting the data changes, the changes need to be extracted and submitted to SDI department to verify the validation of data. The validated method mainly can through the way of the combination of the field surveying and mapping and the indoor detection. Field surveying and mapping usage professional equipment and experts on the field survey to confirm the data changes, and collect professional data accord with standard of professional data, the surveying and mapping indoor usage the latest remote sensing image or aerial photographs through professional interpretation to confirm the reliability of data. The two methods also need certain time period, the updating period of field surveying and mapping depends on the professional department update schedule; the updating period of surveying and mapping indoor depends on the collection period of remote sensing image or aerial photos. Generally speaking, the collection period of aerial photos is shorter than that of remote sensing image. After surveying and mapping departments verify data and collect high precision and reliability data, including the data indoor compiling, the various related databases are updated, then, online data products are updated, by releasing to the user.

3.5 The ring of updating

The current update mechanism of VGI is by amateur in VGI platform to contribute data, in contrast, that of professional surveying and mapping departments is by professional expert and equipment to collect and update to realize, there is so great gap between them. But in recent literatures and researches, some began to pay attentions to the topic about that professional departments provide data for VGI platform, such as Bing Maps provide remote sensing image to Openstreetmap. There are also some researches based on VGI data to provide real-time change information for professional departments such as disaster

emergency. Both of them are one-way data collection and updating, how to unify both organically in together, promote each other, this paper is to solve the problem. In VGI platform, the user to provide data, at the same time, they also data users, so some scholars proposed the concept of producer (Coleman and Georgiadou et al., 2009). In this paper the user can be called closing points. In the two levels of understanding, firstly, in data update participants of the ring, mainly there are two participants, one is from VGI platform public data provider, the other is professional surveying and mapping departments, both focusing on different aspects and with different purposes to take data update work in turn; In data flow, data comes from VGI platform acquisition, through a variety of process eventually went back to VGI platform. In this mechanism, we can think of the data is not a pool of stagnant water, but the flow of the source of living water. Both of them complete each other, VGI focus on the change of the data, in the contrast, SDI focus on validation and heavy acquisition. The changes found by VGI can promote SDI to schedule the update plans, meanwhile, SDI release new data to VGI platform which can promote amateurs to provide real-time data more actively. But there are also some problems, such as the users of VGI may only pay attention to and update the region who are living or be believed more interesting, and other remote areas of the update working may need professional departments to complete, and SDI department may not carry out the validation work timely and effectively, which will affect the data update and flow.

4. APPLICATION

In this section, this paper takes one professional surveying and mapping department in china as an example, connecting with data updating workflow and requirements, discusses the proposed mechanism for the possible application prospect and problems. Firstly, in China VGI data source is less than other countries in the world, there are many reasons, such as the public do not contribute data actively and data privacy. For example, Openstreetmap in China has scarce data, only in metropolises such as Beijing and Shanghai can get high quality data. Secondly, in the aspect of the professional surveying and mapping departments actual update situation in China, although at present many cities in china have been built the fundamental geographic information system, but the data update situation is not optimistic, the update mechanism lacks of systematic planning, which is difficult to ensure that the data not out of time. In the basic geographic data update mode, a lot of local professional surveying and mapping departments according to certain update period update complete coverage of the database, which is very easy to cause the inconsistency between various database versions. The linkages between fundamental geographic database and application database are not established, which cause the problem that lack of the updating linkage between databases. For example, municipal departments had built a new road, when surveying and mapping departments go to field surveying and mapping, the change information could be found. There is no updating channel for departments of surveying and mapping to the application departments, which increase the cost of time and energy. The traditional version update mode directly influences the efficiency updating of electronic map, which cause the update frequency of electronic map on the market at present also can only do the important city update 2 times a year, provincial capital city and general cities updated 1 time a year, as many as billions of POI data accuracy is difficult to guarantee.

Based on the above situation, this paper tried to design a mechanism of surveying and mapping in accordance with

China's situation, which can be described as followed:

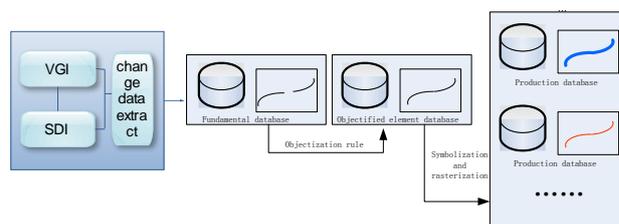


Figure 4. The updating flow

5. CONCLUSION

This paper gives a mechanism from VGI to SDI, explores the possibility of VGI as complement to detect changes and provide more details information for SDI, in turn, from SDI to VGI. The mechanism or cycle includes collecting VGI data, preprocessing data, matching the homonymous elements between VGI and SDI vector data, detecting changes, verifying the credibility of the changes, updating SDI database, updating the products to end-users. The paper discusses the method of data collection and preprocessing, the matching algorithms of homonymous elements, the detection of change and the definition of change types, the confirmation and publication for SDI and the ring for updating. Then, taking one professional department in china as example, connecting with data updating workflow and requirements, the paper discusses the proposed mechanism for the possible application prospect and problems.

6. REFERENCES:

- Coleman, D. J. and P. Eng, 2010. Volunteered Geographic Information in Spatial Data Infrastructure: An Early Look At Opportunities And Constraints, *GSDI*.
- Coleman, D. J. and Y. Georgiadou, et al, 2009. Volunteered Geographic Information: the nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research* 4 (2009), pp. 332-358.
- Coleman, D. and B. Sabone, et al, 2010. Volunteering geographic information to authoritative databases: Linking contributor motivations to program characteristics. *Geomatica* 64, pp.27-40.
- Goodchild, M. F, 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69 (4), pp.211-221.
- Mooney, P. and P. Corcoran, 2011. Can Volunteered Geographic Information Be a Participant in eEnvironment and SDI? *Environmental Software Systems. Frameworks of eEnvironment* 359, pp.115-122.
- Saalfeld, A, 1988. Conflation Automated map compilation. *International Journal of Geographical Information System* 2 (3), pp.217-228.
- Zook, M. and M. Graham, et al, 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy* 2 (2), pp. 7.

7. ACKNOWLEDGEMENTS

This paper is supported by the following projects, including Fundamental Research Funds for the Central Universities(No.214275597) and the Key Technology and Application of Location Based Sensor Network and Holo Information Map (2012BAH35B03).