

DESIGN AND IMPLEMENTATION OF CROWDSOURCING BASED CHINA'S NATIONAL PUBLIC GEOSPATIAL INFORMATION COLLECTION SYSTEM

Hongping Zhang^{1,2}, Jie Jiang^{2,*}, Wei Huang¹, Liang Yang¹

¹ National Geomatics Center of China, Beijing, China- (zhanghongping, huangwei, yangliang)@ngcc.cn

² School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing, China – (jiangjie_263@126.com)

KEY WORDS: Crowdsourcing, Data collection and update, Public geospatial information

ABSTRACT:

As the basis of public geospatial information service, geospatial information data with wide coverage, strong timeliness, rich content and high positioning accuracy is the key infrastructure of geospatial infrastructure for building smart cities and digital China. How to collect and update massive geospatial information resources fast and efficiently has always been the bottleneck for the rapid development of surveying and mapping geospatial information science and technology and industries, and the construction of China's national spatial data infrastructure. At present, the mainstream of geospatial data collecting and updating in China is that the government and enterprise employ professional engineers to for this work, but the problems of professional and complex data processing process, data management level by level, closed and off-line data update mode lead to a relatively slow update rate and poor data actuality, which seriously restricted the applications of geospatial information service in the fields of government decision-making, planning and construction, resource and environment monitoring, and emergency response. The contributions to the geospatial information data from the public community has also been ignored. This paper addresses the issues of low efficiency on data collection and update occurring in China's national geospatial information service platform, proposing a solution for public geospatial data collection and update based on crowdsourcing. The key technologies of data collection, storage, reviews and publication are studied, and the basic technical process of online data update for the government and the public users is designed. In order to verify the effectiveness and practicality of the mode discussed above, a prototype system is developed by expanding China's national geospatial information service platform, which fulfilled data online collection and reviewing, such as POIs, roads, and residential areas. The system can provide new methods and reference ideas for the collection and update of spatial data for the national geospatial information public service platform. It is a useful supplement to the current spatial data collection and update, and has important significances.

1. INTRODUCTION

The term "crowdsourcing" was first published by Jeff Howe in 2006(Howe, 2006), and it is a type of online collaboration activity in which a company or institution process to a group of non-specific individuals to undertake a task voluntarily. Crowdsourcing develops along with Web 2.0 and User Generated Content (UGC). Since crowdsourcing the features of openness, the participants are from all around the globe, they use the internet and communicate and collaboration. In other words, the public can be the members of a crowdsourcing task. It breaks the traditional production mode, and uses the bottom-up mode of knowledge creation to replace the top-down mode of task assignment (Tulloch, 2014). Good Child presented Volunteered Geographic Information (VGI) based on crowdsourcing to describe geospatial information released by the public, as opposed to the traditional methods of professional surveying engineers. As a special case of crowdsourcing, VGI allows citizens to take the role of sensors, and collect, publish and maintain the data by themselves via the Internet. VGI and citizen science is the key of crowdsourcing (Tulloch, 2014). It is a kind of data with strong timeliness which is different from the traditional surveying and mapping geospatial information data. The development of crowdsourcing mode brings new drivers and challenges to the collection and update of surveying and mapping

geospatial information data. It provides an entirely new way for data collection and update through the open collaboration mode.

Existing applications and studies from different perspectives have proved crowdsourcing geospatial information could be an effective complement to the enrichment and perfection of the current dataset. The data quality and reliability has been considered as the similar precision by comparing them with the professional surveying and mapping data (Ali and Schmid, 2014; Haklay, 2010; Haklay et al, 2010). Almost all commercial map websites opened the function of data online submission, such as Google Maps Maker, Here Map Creator. The surveying and mapping departments of some countries also opened some public geospatial information data to the citizens and updated data with the help of the volunteers. USGS started "The National Map Corps (TNMCorps) " project to make contributions to USGS National Structures Database by updating and verifying structure data in 2010(Tian, 2013). The Ordnance Survey has developed a program that encourages volunteers to provide geospatial information about their local communities.

However, the existing crowdsourcing data project are mainly for the public, the data classification and properties are limited, and missed some requirements of the professionals. This study focuses on the design and implementation of China's national public geospatial information service platform with the help of crowdsourcing mode.

* Corresponding author

2. BACKGROUND

Map world, developed by National Geomatics Center of China (NGCC), is an official one-stop website providing geospatial information services to personal users, enterprises, professional agencies and governments. It has been one of the most visited services among the public welfare services, reaching more than 200 million requests of APIs and services per day. The geospatial information datasets update of the platform are mainly provided by the provincial surveying and mapping department. All the processes are offline and take long time for a whole update in every year. There exist some problems such as low efficiency, poor interaction and collaboration, inability to engage the participation of the professionals at all levels and the public. It is difficult to meet the high timeliness requirement of geospatial information dynamic update to discover changes under the information condition (H, Zhai & Jia H, Peng 2016; Xin, L, Qian 2011).

Besides, “Internet +” epoch has brought new opportunities for the traditional collection mode of surveying and mapping industry. The combination of crowdsourcing mode and professional collection can be applied in the field of spatial features collection and update. An online map data update program of Map World has been applied to realize the data fast update through the participation of the public and professionals from 2018. The study and application of crowdsourcing update will promote the sharing and efficient applications of public geospatial information.

3. DATA COLLECTION AND VERIFICATION PROCESS

Under the crowdsourcing mode, the public and professionals can submit different kinds of geospatial features based on their knowledge background. Unlike other crowdsourcing platform, the data verification in this platform should be done by the professionals from the surveying and mapping departments in the local government. Fig.1 shows the whole data collection and verification process for the system.

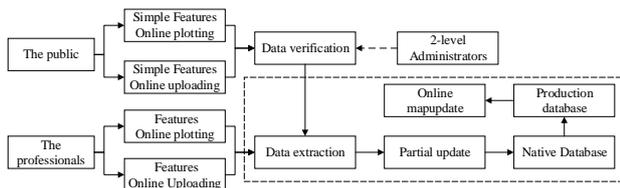


Figure 1. Data collection and verification process

3.1 Data Content

As shown in Table 1, three simple geometric features like the point, line and polygon are defined for data collection.

Features	Data category
Point	Restaurant, Hotel, Shopping, Health, Culture and Education, Sports, Tourist Attractions, Finance and Insurance, Organization, Transport, Life Service, Real estate and Building, Enterprise, Place Name
Line	Road, Railway, Subway, Water, Wall, Tunnel, Boundary
Polygon	Water, Greenbelt, Residence area, Factory, Courtyard, Tourist attraction, Park

Table 1 Features list

For the public, only the general properties are opened for them to edit; but for the ones from the local professional institute, they can edit more information, such as road level and width.

3.2 Data Collection

The users must be logged in before they make contributions. The system offers two methods for data collection: online plotting and uploading. While the map type is satellite and level is 17 or above, users can carry out online plotting. If the data does exist, the users can upload them directly based on the provided template. Whether the data are accepted or not, they are displayed for those who submitted them.

3.3 Data Verification

Considering the quality, accuracy and uniqueness of data acquisition under crowdsourcing mode, background data audit function should be carried out. For the data provided by the public, two-level verification is designed to guarantee the data quality. The local auditor of city level will check the data firstly, and then normalize it, such as address and category. After the preliminary verification, the provincial auditor will check again before submitting them to the platform. For the data provided by the local surveying department, they are considered as qualified and no quality inspection is required. They can be submitted directly to the platform by the provincial auditor.

3.4 Data Extraction

Data extraction refers to exporting reviewed data from the database and transfer the data format as an offline dataset. After received the data from the contributors, the basic information will be stored in the corresponding fields, and the spatial information will be converted to the binary value. A JOSN object is designed to store the extended information that the platform ignored before. The extracted data can be used for native database update, then the map can be updated based on the database. This section is implemented by the professionals from national node.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1 System Architecture Design

The system architecture is given in Figure 2. It is of a three-tiered B/S structure based on the cloud computing environment of Map World, which consists by of a user interface layer, a business logical layer and a data layer.

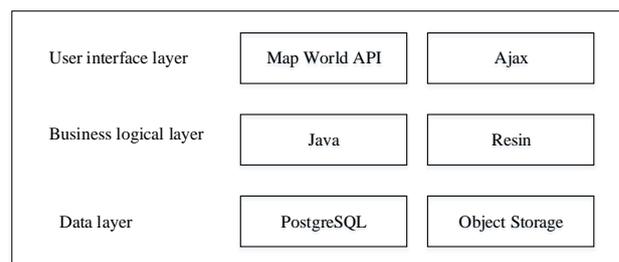


Figure 2. The architecture of the system

4.1.1 User interface layer: Map world JavaScript API and Ajax technology has been used for the browser-side as the interface layer. All the functions and interactions are realized in this layer. It is responsible for providing basic maps for data collection, submitting the collected data to the server and data reviewing, and displaying the data from the server.

4.1.2 Business logical layer: The system uses JAVA language as programming tool for developing the sever as the business logical layer to with front-end users and the database. It is responsible for format processing after receiving the submitted data, and encoding data with the administration code.

4.1.3 Data layer: The crowdsourcing spatial and attribute data is stored in the PostgreSQL database with the plugin of PostGIS, and the pictures related to the data are stored in the object store database. The database access interface is provided in this layer.

4.2 System Implementation Technique Routine

The technique routine of the system can be seen in Figure 3. As a sub-system for Map World platform, the users of this system should be connected to the user center of Map World, and the latter provides a single sign-on (SSO) mechanism. The online plotting and uploading modules are developed on the basis of Map World JavaScript API 4.0. PostgreSQL is integrated in the system, since it is open source and enables the application to store the relational data tables. Moreover, the spatial entity type is supported in the database, such as point, line, polygon, circle. With the help of PostGIS, some spatial manipulation functions can be applied. This is useful for data visualization and extraction. Besides the spatial and attribute data stored in PostgreSQL, the Object storage database in the cloud is used for the pictures. The reviewed crowdsourcing data will be extracted every as the update data source and integrated with the native database stored in the FileGDB Database. Then the map, POIs and some functions related to the data will be updated.

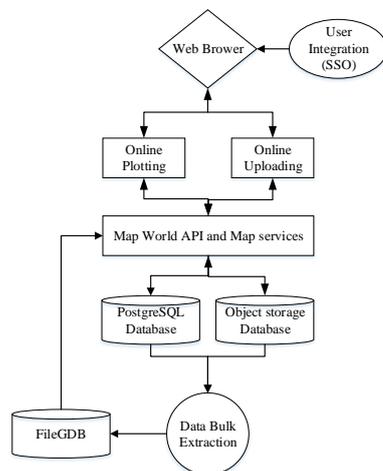


Figure 3. The technique routine of the system

4.3 Functional Design and Implementation

According to the system architecture design, the construction of crowdsourcing geospatial information collection system is divided into two parts: Web Online crowdsourcing data acquisition system and Web Online Data reviewing and management system. The system is divided into six modules

based on the functions, as shown in Figure 4. Each module has detailed functional design.

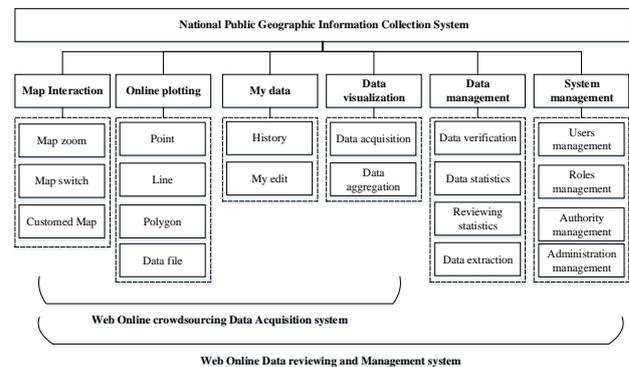


Figure 4. The functions and structure of the system

4.3.1 Map interaction: Map interaction is the basis for data collection, it includes the functions of map zoom, map switch, and adding a customized map layer. In general, the vector map and satellite map of Map World are provided as the default layers. If the user has a better online image URL with the same spatial coordinates and the interface supporting OGC WMTS, the new layer can be set as the working background map.

4.3.2 Online plotting: As shown in Figure 5, this module mainly provides tools such as spatial data editing and batch uploading EXCEL/CSV file confirmed to the template requirements. If there is only address and the coordinates do not exist in the uploaded file, the system will transfer the address to coordinates automatically via the geocoding API of Map World. Spatial data editing supports simple operations of point, line, and polygon. For example, drawing a point in the map window, the point details page will be opened, and some mandatory fields need to be completed by the public, such as name, category. As to the optional fields, such as secondary category, data tags, these should be filled by the professionals if there exists the information. The point, nodes of line and polygons can be modified at any time. After finishing editing, the data can be submitted to the server. Before writing the data to the database, the system will pre-check the data in the aspect of whether some information is missed, the length or format is wrong.



Figure 5. The interaction of online plotting

4.3.3 My data: The submitted data or those being edited can be managed by my history and my edit. In my history, the data can be filtered by name, data type, reviewed progress. If the data is not accepted, the rejection reason is also given here. The data being edited are temporarily stored in my edit, so that they can be submitted in bulk. If the submission task fails, the reason for the failure is given for each data in this page. Then the user can update the data according to the tips. Even the accepted data, the user can also modify them. But if this kind of data is modified and re-submitted, the verification status is changed as "submitted".

4.3.4 Data visualization: The Amount of data collected and uploaded by the public and professionals is huge, and it is lack of hierarchical organization and scale information. Therefore, a method of data aggregation is applied integrated administration code, map level and data density. When the point is submitted, the administration code is added. For the line and polygon, the central point is used to represent the feature. It is well-known that the map is divided into 18 levels according to the details of the features. The submitted data is also divided into 4 levels for visualization according to the administration code of different levels, respectively is: level 1 to level 6, level 7 to level 10, level 11 to 16, and level 17 to 18. In level 17 to 18, all the data in the field of view are displayed, while the aggregation number is shown in other levels. The visualization effect of different levels can be seen in Figure 6.



Figure 6. Visualization effect of different levels

4.3.5 Data management: Data management includes data verification, data statistics, reviewing statistics and data extraction. These functions are mainly for the auditors. Two-level verification mechanism is designed, so the auditor of different administration levels can review the corresponding data in the same administrative division. Each auditor performs different operations according to different authorities. Only the data accepted by the national auditors can be used for map update. The process of data submission and reviewing can be calculated at any time so as to know the state of the crowdsourcing data. As to the data extraction, it is used to transfer the accepted data to the offline environment for native database update.

4.3.6 System management: System managements module is composed of user management, roles management, authority management, and Administrative division management. The users in the user center of Map World are divided in to the ordinary ones and the professionals from the local departments. All the users are from the user center, and the basic information of users is stored in this system. Five types of roles are designed in this system, including the public, the professionals, auditors, administrators, and a super administrator. The auditors and administrators are granted the authority of reviewing data and setting auditor according to the administrative divisions. All the functions are considered as authorities, such as data extraction, data verification. Since the roles and authorities are closely related to the administrative divisions, a dynamic and expansive management function is used for adjusting the data conveniently.

5. CONCLUSION

Since the crowdsourcing information has the advantages of real-time, rich semantic information and openness, a growing number of applications make use this kind of data as update source, such as industry data collection, and emergency relief. But in the field of professional surveying in China, the crowdsourcing geospatial information data has not been used extensively. This article studied the collection and verification process of crowdsourcing geospatial information integrated the requirements of the public and the professionals. Based on the cloud computing environment and Map API of Map World, a prototype system is designed and developed, realizing the functions of map interaction, data online collection, visualization, reviewing, and management. The publication of this system has an industry demonstration value to create feasible technical means and mode that can be popularized and applied in order to promote the comprehensive upgrading of geospatial information collection from the professionalization to popularization. Future work will focus on the further promotion and application, and trying to realize updating the map data online.

ACKNOWLEDGEMENTS (OPTIONAL)

Financial supports from the National Key R&D Program of China (No.2017YFB0503700) are gratefully acknowledged.

REFERENCES

- Ali, A.L., Schmid, F., 2014. Data Quality Assurance for Volunteered Geographic Information, in: Duckham, M., Pebesma, E., Stewart, K., Frank, A.U. (Eds.), *Geographic Information Science*. Springer International Publishing, Cham, pp. 126–141. https://doi.org/10.1007/978-3-319-11593-1_9
- Haklay, M., 2010. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design* 37, 682–703. <https://doi.org/10.1068/b35097>
- Haklay, M. (Muki), Basiouka, S., Antoniou, V., Ather, A., 2010. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal* 47, 315–322. <https://doi.org/10.1179/000870410X12911304958827>
- Howe, J., 2006. The rise of crowdsourcing. *Wired magazine* 14, 1–4.
- Tulloch, D., 2014. Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice. *International Journal of Geographical Information Science* 28, 847–849. <https://doi.org/10.1080/13658816.2013.873440>
- Tian, W., 2013. *Geospatial Information Change Detection and Update Based on Volunteered Geospatial Information*. PhD Thesis, Wuhan: Wuhan University
- Zhai, H., & Peng, J. H., 2016. Study of Spontaneous Geographic Information Collection Platform. *HUNAN AGRICULTURAL SCIENCES* 2016(4): 94–97.
- Qian, X. L., 2011. *Research on the Representation and Management of Geospatial Data from Volunteered Geographic Information*. PhD Thesis, Wuhan: Wuhan University.