VISUAL TOOLS FOR CROWDSOURCING DATA VALIDATION WITHIN THE GLOBELAND30 GEOPORTAL

E. Chuprikova [a,*], H. Wu [b], C. E. Murphy [a], L. Meng [a]

[a] Department of Cartography, Technical University of Munich, Germany - e.chuprikova@tum.de
[b] National Geomatics Center of China, Beijing, China - wuhao@nsdi.gov.cn

Theme Session, ThS18

KEY WORDS: Crowdsourcing, global land cover, visual tools, Globeland30, volunteered geographic information, sense making visualization

ABSTRACT:

This research aims to investigate the role of visualization of the user generated data that can empower the geoportal of GlobeLand30 produced by NGCC (National Geomatics Center of China). The focus is set on the development of a concept of tools that can extend the Geo-tagging functionality and make use of it for different target groups. The anticipated tools should improve the continuous data validation, updating and efficient use of the remotely-sensed data distributed within GlobeLand30.

1. INTRODUCTION

In September 2014 China published global land cover datasets of 30m resolution based on Landsat and similar image data, with multiple classes – GlobeLand30 for two base years 2000 and 2010 (http://www.globeland30.org) (Chen et al., 2015). Although the overall accuracy of GlobeLand30 dataset is higher than 83%, its sustaining improvement is still an issue (Han et al., 2014), especially for small regions. At the same time the amount of user-generated information via the Internet tends to increase, which indicates a great potential for obtaining land cover samples via geo-tagging technique.

Crowdsourcing, the process of obtaining information from contributors amongst the general public (Haklay et al., 2014), is a bridge connecting the users and the data producer. In order to improve this connection two main considerations have to be taken into account: how the data collection process is presented for the users and how the contributed data is transmitted to the producer. In this paper we would like to discuss the issues related to the visualization approach that can improve the usability and empower the GlobeLand30 geoportal with sense-making visual tools for collecting crowdsourcing information related to uncertainties of global land cover classification.

Along with the increasing amount of user-generated information and the improvements in the data acquisition, the overload of information is becoming tangible for data scientists. In many applications the data collection is faster than the ability to use it for decision-making (Keim et al., 2010). Above all, an increasing number of volunteers require effective ways of organizing and providing access to value-added or newly created geospatial data (Kalantari et al., 2014). Consequently, it is necessary to develop representations that enable collaborative reflection, promote mutual visibility of volunteers' efforts and sustain a shared view of the community (Herranz et al., 2014). Moreover, the visualizations should facilitate sense making of heterogeneous information and reflect it in a flexible and interactive way for volunteers to examine both community and global land cover data (Herranz et al., 2014).

Therefore, the research work is focused on the possible improvements for the GlobeLand30 platform that can integrate crowdsourcing in a convenient way and can drive the updating and validation of the data and collect knowledge related to a phenomena, time or event. Above all, this work in progress deals with the development of a concept of tools that can serve to collect, filter, and analyse VGI contribution providing extended functionality for different target groups.

The paper is structured as follows. The subsequent section discusses the related work on public participation tools across application domains. It is followed by a section that analyses design challenges and proposes visualization techniques to improve the GlobeLand30 platform based on the user requirements. Besides, it discusses the data structure and illustrates the application of visualization techniques through a set of design prototypes. Finally, the conclusions are drawn.

2. STATE-OF-THE-ART

Various social networks and Web portals provide an opportunity for the users to involve their observations about everyday life with location-aware content (Cope, 2015). Volunteered geographic information (VGI) is user-generated content of spatially referenced data (Cope, 2015) potentially brings significant results for the spatial analysis of our environment. Among the volunteered geographic communities are OpenStreetMap (http://www.openstreetmap.org/), Wikimapia (wikimapia.org) and Google Map Maker (https://www.google.com/mapmaker). Moreover, the analysis of the VGI data can be used for describing land cover that was captured using a Web-based interface (Comber et al., 2013). One of the examples of a crowdsourcing system for global land cover validation is the Geo-Wiki platform (www.geo-wiki.org) which was developed in 2009 by Fritz et al. (2009). It involves

---

[*] Corresponding author

the scientific community as well as the public in validating global land cover and calibration processes, creation of a hybrid land cover map based on the validated points, and in the development of an education platform (Bastin et al., 2013). The actual application offers to evaluate global land cover disagreement maps and to determine if they are correct based on the imagery and local knowledge (Fritz et al., 2009). In the current phase, the platform is limited to the disagreement maps and it does not provide an opportunity to contribute beyond. Furthermore, Crowdmap is another hosted service for mapping on the Web, focused more on the social mapping experience with a support for multimedia, sharing, and mobile support. Crowdmap allows to share a story on the map along with photos and videos for a specific location or for entire region.

Above all, a crowdsourcing component is also available within GlobeLand30 as a geotagging tool in which registered users can identify a geometry (point, line or polygon) and assign it to some properties. Besides, the contributors can receive a comment from an expert. Although the geotagging can bring details about misclassified area, it cannot provide extended information and assist in the visual analysis of the contributions.

The Global Land Cover validation is one of the application domains in which visualization mechanisms might be used to empower the public participation. The above summary shows that although visualization is regarded as a powerful tool for empowering citizen participation, most existing platforms mainly focus on displaying citizen information. And there is still very limited support for encouraging the involvement of citizens through visual representations.

3. METHODS

Apart from the creation of global land cover data, it is of paramount importance to communicate this information to the scientific community and to the public. Therefore, the GlobeLand30 platform is a valuable source where users can access the fine resolution GLC (Global Land Cover) data and also contribute to improvements of the data and knowledge collection. To enhance the GlobeLand30, a framework design is proposed that supports crowdsourcing data collection based on intended target audiences and involves visualization of the collected data. The collected results can give a helping hand for the validation process and the overall improvement of the system. Moreover, based on the contributed data, the users may have an opportunity to create their own personalized visualizations for efficient data understanding.

Crowdsourcing is able to assist science for identifying and solving uncertainties that occur and accumulate during data acquisition, data processing and geo-visualization. The concept of uncertainty is broad and can be caused due to several reasons (Rae et al., 2007). This article includes the concepts of accuracy and error as these components can be visualized and validated using the crowdsourcing approach.

3.1 Usability requirements

Observation of land cover at the global scale is of paramount importance for a variety of environmental applications including monitoring of global change (Latham et al., 2014). Therefore, the users are expected to have different requirements for global land cover maps and its assessment (Tsendbazar et al., 2015). The users of the land cover data can be divided into

experts, namely climate modelling community, global forest change analysts, the group of global agriculture monitoring, urban development specialists, producers of improved maps (Tsendbazar et al., 2015) and common citizens interested in land cover data validation. Understanding the interest of these user groups can help to understand the requirements for the validation, the mapping standards, as well as the visualization techniques.

Although the usability requirements are very much dependent on the user group, there are five comprehensive usability attributes defined by Nielsen (1993) which are (a) learnability, (b) efficiency, (c) memorability, (d) errors and (e) satisfaction. The learnability of the software component means that this component is easy to use and the contributors would not have any difficulty to figure out how one tool or function works. Efficiency of the system assumes that all components are working fast and saving time for the contributors using it. The memorability of the application provides the contributors easy-to-use and easy-to-remember design, so that they remember the functions and there is no need to learn it the next time again. In this regard, visual tools can be very helpful, due to visual memory benefits. When it comes to errors in the application components, it is important to provide solutions that can consider potential errors and prevent them from appearing (Nielsen, 1993). Moreover, the satisfactory issue is a crucial factor that assures contributors to stay motivated for further work with the application.

3.2 Data structure

Based on the literature review of the user requirements in the field of land cover mapping, a data structure model is defined (see figure 1). As a result, a workflow is established for the development of the visual tool prototypes.
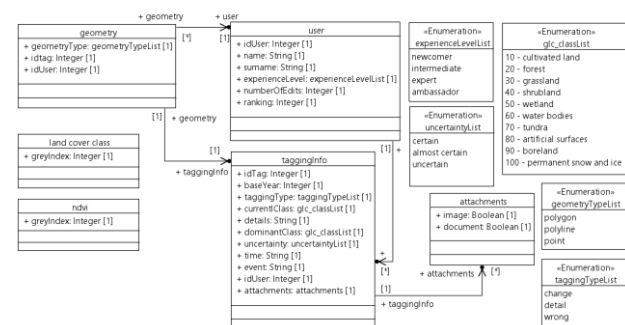

Figure 1. UML (Unified Modelling Language) schema of the data structure for visual tools implementation

3.3 Visualization techniques

The important role of geovisualization tools as a medium to enhance the interface communication was proved by several researchers (Kraak, 2009, Keim et al., 2010, Kim, 2009). With the improvement of the devices for data collection and its storage, it has become faster to collect and store data than to use it for decision-making (Keim et al., 2010). For this reason, the data should be utilized in an efficient way to take the advantage of it and allow to reveal hidden opportunities of available resources. Apparently, the visualization techniques can assist the target groups in comprehending complex land cover information. While the visualization of land cover classification can provide a bird's eye view of the overall area, it does not

provide quantitative estimates of different land cover types on a particular area of interest. Besides, it cannot be analysed with additional information layers. To address the challenges of flexible, interactive ways for volunteers to explore global land cover data, this paper presents the visualization techniques that can be implemented within the GlobeLand30 platform and can drive the integration of both community participation and individual sense making.

During the development of the prototypes it is important to consider the following issues: collaboration – how visual tools can enable different parties to work together and communication - how visual tools can facilitate effective transfer of spatially and temporally-related information and knowledge (Keim et al., 2010). The visual techniques and tools for the interactive land cover data analysis and crowdsourcing data validation are tested within a Web-based GIS application (see the figure 1) established using free and open-source software and libraries (including PostGIS, OpenLayers, GeoExt.js, Ext.js and Geoserver).
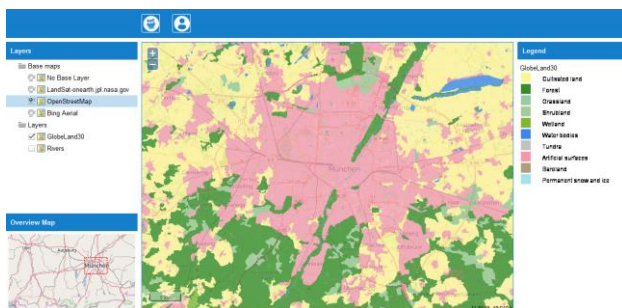

Figure 1. WebGIS test environment

The interface allows the contributors to select an area of validation. Based on the selection, the pop-up window presents the distribution of the classes on the selected area and shows additional information (see the figure 2) that can assist in the validation process (Normalized Difference Vegetation Index (NDVI) for example). Following Bastin et al. (2013), selected dominant class can be assigned with an uncertainty value, which indicates that additional information might be needed for a more precise verification.
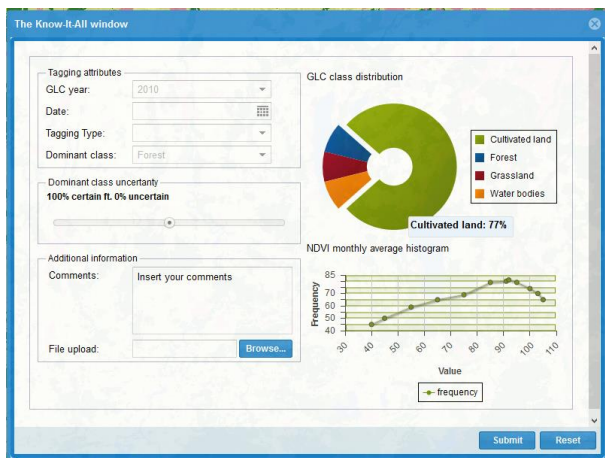

Figure 2. User contribution form

Moreover, the community is a crucial factor of any crowdsourcing supporting platform. Therefore in order to

support the community issues, it is of paramount importance for volunteers to access and collectively reflect within the community scope (Herranz et al., 2014). As it can be seen from the design prototype (see the figure 3) the contributors can observe the visual analysis about the number of contributions throughout the year and their ranking according to the contributions of other users. Therefore, the history of contributions may motivate the users to take part more actively in the validation process.
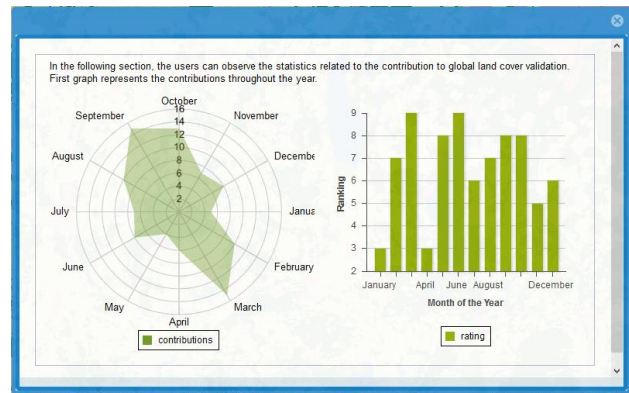

Figure 3. User statistical information

As far as geo-visualization is concerned there are some challenges within WebGIS that should be overcome with the further development of additional visualizations for the Global Land Cover platform. As for example, some visualization techniques can handle only a limited number of geometrical entries, therefore scientific visualization techniques could be involved.

4. CONCLUSION

Global land cover and its validation has been a topic of significant interest within the environmental community, authorities and scientific institutions. Recently published GlobeLand30 has emerged as an effective platform for the sharing and distribution of high resolution remotely sensed data. Based on the platform GlobeLand30, an approach to support crowdsourcing validation and collection of new knowledge via visual tools could be established.

The research is a work in progress that aims at improving the usability of the GlobeLand30 platform by integrating visualization techniques in a convenient way and facilitating the updating and validation of the data as well as collecting knowledge related to a land cover phenomena, time or event. The designed prototypes show how the additional visual tools can enhance the sense making for the contributors and the community.

The future development will be focused on the improvement and extension of the applied visualization techniques. Furthermore, the user-friendliness and fitness of the user graphic interface is to be evaluated using an eye-tracking system.

REFERENCES

Bastin, L., Skøien, J., and Schulz, M., 2013. UncertWeb. The Uncertainty Enabled Model Web. ICT for Environmental Services and Climate Change Adaptation Deliverable.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J., 2015. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 103, pp. 7-27.

Comber, A., See, L., Fritz, S., Van der Velde, M., Perger, C., and Foody, G., 2013. Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 23, pp. 37-48.

Cope, M., 2015. Commentary: Geographies of digital lives: Trajectories in the production of knowledge with user-generated content. *Landscape and Urban Planning*, Vol. 142, pp. 212-214.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., and Obersteiner, M., 2009. Geo-wiki.org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, Vol. 1(3), pp. 345-354.

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., and Mooney, P., 2014. Crowdsourced geographic information use in government, Report to GFDRR (World Bank). London

Han, G., Chen, J., He, C., Li, S., Wu, H., Liao, A., and Peng, S., 2014. A web-based system for supporting global land cover data production. *ISPRS Journal of Photogrammetry and Remote Sensing*, Val. 103, pp. 66-80.

Herranz, S., Romero-Gómez, R., and Díaz, P., 2014. Visualization Techniques to Empower Communities of Volunteers in Emergency Management, *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, pp. 335-344.

Kalantari, M., Rajabifard, A., Olfat, H., and Williamson, I., 2014. Geospatial metadata 2.0 - An approach for volunteered geographic information. *Computers, Environment and Urban Systems*, Vol. 48, pp. 35-48.

Keim, D. A., J. Kohlhammer, G. Ellis, and F. Mannsmann. Eds. 2010. Mastering the information age – solving problems with visual analytics. Goslar: Eurographics Association.

Kim, C., 2009. Spatial data mining, geovisualization. In Thrift, R. K., editor, International Encyclopedia of Human Geography, pp. 332-336. Elsevier, Oxford.

Kraak, M.-J., 2009. Geovisualization. In Thrift, R. K., editor, International Encyclopedia of Human Geography, pp. 468-480. Elsevier, Oxford.

Latham, J., Cumani, R., Rosati, I., and Bloise, M., 2014. Global Land Cover SHARE. Deliverable.

Nielsen, J., 1993. Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Rae, C., Rothley, K., and Dragicevic, S., 2007. Implications of error and uncertainty for an environmental planning scenario: A sensitivity analysis of GIS-based variables in a reserve design exercise. *Landscape and Urban Planning*, Vol. 79(3-4), pp. 210-217.

Tsendbazar, N. E., de Bruin, S., and Herold, M., 2015. Assessing global land cover reference datasets for different user communities. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 103, pp. 93-114.

Revised April 2016