A Two-Tiered Approach to OSM Data Collection for Novice Users

Adam Rousell
GIScience Research Group
University of Heidelberg,
Germany
adam.rousell@geog.uni-heidelberg.de

Stefan Hahmann
GIScience Research Group
University of Heidelberg,
Germany
stefan.hahmann@geog.uni-heidelberg.de

Amin Mobasheri
GIScience Research Group
University of Heidelberg,
Germany
amin.mobasheri@geog.uni-heidelberg.de

Abstract

Although OpenStreetMap (OSM) is a widely used crowd generated spatial dataset, it can be difficult for novice users to enter data in a way that conforms to those data already present. It is often the case however that it is these novice users who have a more invested need for relevant data to be present within OSM, such as is the case with users with reduced mobility. In this paper is presented an approach which allows novice users to contribute information which can then be used to enrich the OSM dataset. This is done via a two-tiered approach whereby one user (the Observer) contributes a textual description via an Android app developed as part of a server-client web service. This text is used to create OSM Notes which are then used by experienced OSM users (Editors) to update the information in the OSM. Using such a method means that the collection and entering of information are the responsibility of the people who are more suited for the task – collection for people who know what creates obstacles in the environment, and entering for those people who know how to update the OSM dataset.

Keywords: OpenStreetMap, VGI, Mobile applications, Accessibility, Tagging

1   Introduction

Creating data in Open Street Map (OSM) can be a challenging task for novice users. Previous experience with mapping events has shown that many new editors (people who enter data into OSM) struggle to adapt to the OSM tagging mechanism or are even reluctant to sign up to the service itself [11]. In many instances however, it is important that the person collecting the data is an "expert" in that they know what they are looking at in the real world environment. One example of such an instance is with regards to accessibility whereby an able bodied person may not pay any attention to the presence of steps on a footpath as to them it is only a minor inconvenience to traverse. To a wheelchair user however, such steps may prove to be an impassable obstacle and thus would desire to have it depicted within the OSM. It is likely to be a low occurring scenario where the wheelchair user is an experienced OSM editor and can use existent tools such as Vespucci (Android mobile app) or iD/JOSM (OSM desktop editors) to successfully describe the obstacle in the OSM dataset. Therefore, it is desirable to provide a mobile tool that allows the easy collection of information in the real world which in turn can be transformed into OSM features or attributes. The method proposed in this paper is to make use of an already existing OSM service which allows the submission of relatively unstructured information for use in creating OSM data – the *Notes* service. Merging this with a three part web service/mobile application allows inexperienced users to contribute to the OSM dataset without the need of learning the creation mechanisms.

The remainder of this paper is structured as follows. In the next section a brief review relating to previous studies and the need for data is presented. After that, the basic concepts of the service developed are described followed by a detailing of the system architecture. Finally, an example of the process workflow for data entry and use is presented, and the paper concludes with a summary.

2   Previous Studies & Need for Data

OSM is a crowd generated spatial dataset comprising of numerous features. These features not only have a spatial extent but are also described via key-value pairs known as "tags" associated to them. These tags are free form in that they do not have to follow a predefined ontology. Although this means that the tagging procedure can be seen as easier, it results in semantic interoperability issues [2, 6]. Such semantic issues include vagueness and ambiguity introduced by the personal perceptions of the recorder [1, 9]. Especially in the case of novices who are new to OSM, the concept of selecting an appropriate term to use to describe a particular aspect of a feature can be daunting both because they may not be aware of what tags are already in wide use, and that their own perception of a particular classification may be different to others. Besides the semantics, the geometry may be difficult to handle for novice users, since this requires basic understanding of and the capability to handle point, line and polygon features in GIS like environments as present in current OSM editors.

As discussed in [8], OSM has introduced the Notes system[1] which sits on top of the map and the main database and allows novice users to raise issues in relation to data that is missing or incorrect on the map. Its main goal is that of "encouraging legitimate peripheral participation" [8: 4120], suggesting that the primary use is to allow people who are not particularly involved with OSM to enrich the dataset. Current statistics[2] of the usage of Notes show that 77% of all ~500k opened Notes have been closed indicating that active OSM contributors manage to address the comments enclosed in them.

There are a number of mobile tools in existence which aim at collecting OSM data, but an important factor in encouraging participation is that the tools match the expertise

---

[1] http://wiki. openstreetmap.org/wiki/Notes/
[2] http://resultmaps.neis-one.org/osm-notes

and technical skills of the user who is attempting to contribute data [4]. Therefore, many of these tools which rely on the selection of tags may not be suitable for novice users. In [10] an OSM data collection tool is discussed which was developed for their study with the specific purpose of collecting data relating to routing for disabled users, and as such focused on items such as incline and crossings. One key aspect of their methodology is that the data collected by the tool is not directly inserted into OSM but is instead used in a post-processing stage whereby the data collected is inserted into OSM manually using tools such as JOSM. That process is conceptually similar to the method proposed in this paper, with the difference being that the post-processing stage can be done by anyone with access to the OSM editing tools as opposed to requiring access to the database used by the collection tool.

Data in OSM can be used for a variety of applications ranging from accessibility mapping (i.e. Wheelmap) to routing systems (i.e. OpenRouteService). Within routing systems, one particular topic that requires specific data is that of accessible route planning. For example, people with reduced mobility (such as wheelchair users) face obstacles in the environment that able bodied pedestrians would not have a problem with. These obstacles include steps, surface material and incline, and can all make a route impassable to a wheelchair user. Within the OSM dataset however, such information can be sparse [3, 5, 7].

3   Basic Concepts

The basic concept behind the tool developed is to make it easier for non-OSM users to contribute information that they feel is important to include within OSM but without having to work with the complex tools and structures that actually get the data in there. By using the Notes service provided by OSM those initial contributors simply need to provide a textual (geo-located) description of what is missing or needs

to be changed and then this text becomes visible on the main OSM page where experienced editors can use it to update the dataset. This "two-tier" approach ensures that the individual aspects of data collection and entry are left to the people who are more experienced in them. Thus, this concept to a certain extent decouples in-situ data capturing from the actual OSM tagging processes.

Obviously, there are some drawbacks to the method implemented. Mainly, it relies on two parties for the data collection. This means that there is no guarantee as to the temporal scale from which data is collected to when it is inserted (indeed it may actually not get entered), as well the possibility for uncertainty to be introduced via misinterpretation of the description provided in the Note.

4   System Architecture

The full tagging system that has been developed comprises of three main components: a mobile app, a web-map interface, and a web server. The mobile app focuses on the collection of data, the web-map allows users to view obstacles that have been recorded in an area, and the web server handles all interaction with the database and OSM Notes service via a public API.

4.1   Web Server

The web server is the core component of the tagging service as it manages the storage and retrieval of obstacles, and the communication with other service APIs such as OSM and Mapillary. It runs on an Apache Tomcat platform with data stored in a PostGIS extended PostgreSQL database. Data is returned from the API in the GeoJSON format.

The primary function of the web server is to store obstacles that are submitted via the API and then to create corresponding OSM Notes if suitable. The creation of

Figure 1: The two-tiered architecture
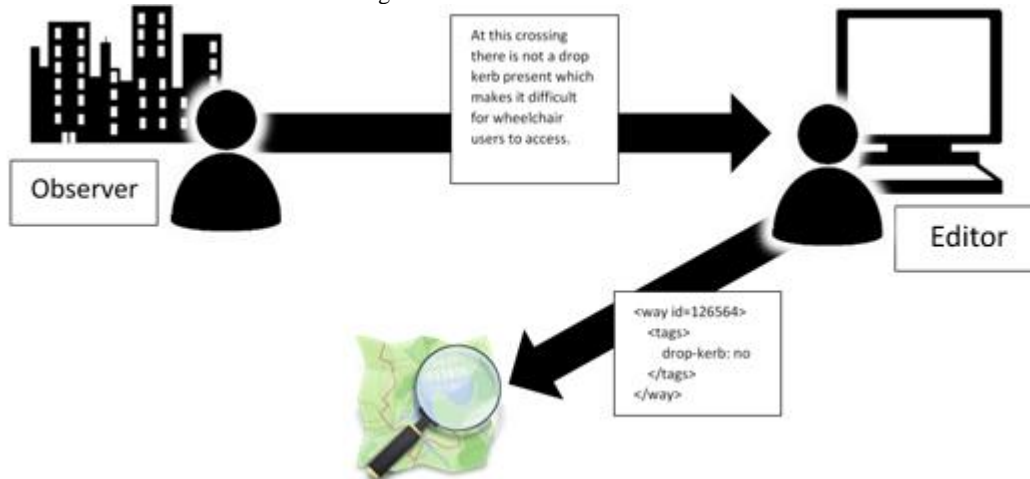
obstacles that are marked as temporary in nature does not generate an OSM Note as those sorts of features are not suitable for the OSM dataset (they can however still be viewed in web map interface discussed later). Within the design process, two requirements identified from discussion with the OSM community were that the Notes should be created from an account (Notes can be created anonymously) so that the source of the information can be traced, and that the Notes should be created from a centralized service as opposed to being generated directly from the devices. The use of a centralized service means that the creation process can be modified or suspended easily at any point.

The secondary functionality is to provide access other services such as Mapillary. Within the web-map interface Mapillary tracks and images are shown to the user so that they can be used to assist in the assessment of obstacle risk. To speed up the display of this information and to reduce redundant data, the web server implements an interface to the Mapillary service. This interface collects data from Mapillary based on parameters passed (such as a bounding area) and then processes them before returning data to the client. This involves clipping of tracks to view areas and reducing the number of images returned.

4.2   Mobile App

As a primary data collection point for obstacles in the environment, a mobile app has been developed for Android based devices. It allows users to view obstacles that have been recorded in their vicinity (as well as possible obstacles obtained from OSM) and to record any new obstacles that they come across. When recording an obstacle, they are asked to enter a textual description, mark if the obstacle is temporary (i.e. roadworks that close a footpath) and also provide a photo of the obstacle if they wish. Once recorded, this information is uploaded to the web server when the user chooses to do so via an upload button.

The interface implemented within the Android app makes use of the OSMdroid mapping library, GoogleAPI for positioning, and the ActionbarSherlock library for adding menu and icon based navigation. The app is provided in English, German and Spanish languages which is changed automatically based on the language set for the device.

4.3   Web-map Interface

The web-map interface is the primary location that people can use to view obstacles within a particular area. The main intended use is to allow people to assess whether obstacles they may encounter will cause problems. To assist in this, the user can view the information entered about the obstacle (description, data of entry, imagery etc.) as well as Mapillary images if available. Features from OSM are also shown that may prove to be obstacles (i.e. steps and cobblestone surfaces). In addition to being able to view data entered from the mobile app, it is also possible to record new obstacles and to mark temporary obstacles as removed from the environment. The interface is deployed using OpenLayers and standard web development techniques.

Figure 2: Android app interface (recording an obstacle)
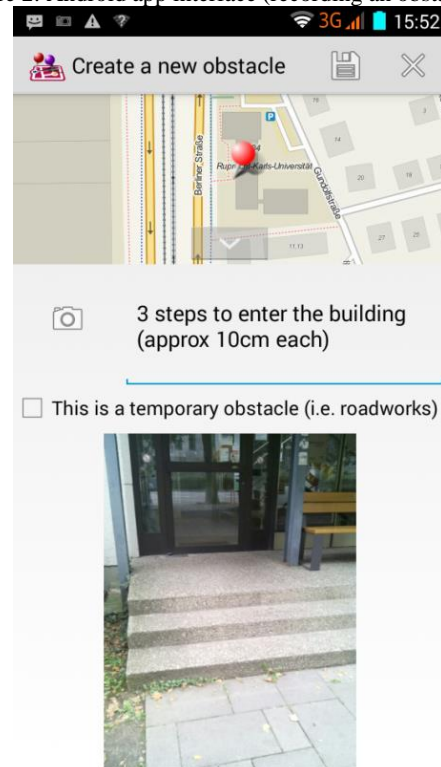
5   Example

For an example as to the workflow of the system, consider a wheelchair user who is visiting a tourist area. Whilst exploring they encounter a footway that has a very rough cobblestone surface that they find difficult to traverse. In the obstacle tagging app they create an obstacle with the description "*The footpath here is a very rough cobblestone surface*" and take a photo of the surface. Later on they also encounter a footpath that has been closed due to building works with the only way to get past being to drop down a kerb onto the road and then go back up the kerb further down to get back onto the footpath. Although this person is able to do that in their manual wheelchair, they are aware that for electric wheelchair users this is more of a problem. So again, they open the app and record the obstacle with a description "*Building works have closed the footpath, and the detour involves traversing a kerb*". They also take a photograph of the kerb and mark the obstacle as being temporary in nature. The user notices however that the location shown on the device does not match their actual location (due to GPS inaccuracy) and so drag the marker identifying the position of the obstacle to the correct location.

When they return to their hotel they upload the obstacles to the server using a WiFi connection. When received by the server, both obstacles are stored in the database with their images. The server uses the OSM Notes API to create a Note on OSM for the cobblestone surface obstacle at the location that the obstacle was recorded at. The text in the note

comprises of the textual description provided and a link to viewing the obstacle in the web-map interface so that the image (if present) can be viewed. As the building work obstacle was marked as temporary, no Note is created for that but it is still stored in the obstacle database for viewing in the web-map interface.

The next day, a wheelchair user also visiting the same tourist location goes to the web-map interface to see if there are any obstacles in the area. They see that a footpath has been marked as having a cobblestone surface, but as they do not mind a short distance of this surface they look at Mapillary images on the web-map that are also present for the footpath. After viewing the images they identify that the surface continues for some distance and thus decide to avoid that footway.

encountered in the environment via a custom made Android mobile app. Data is entered into the OSM by Editors who make use of this textual description which appears on the map as a Note. By using such an approach, errors introduced can hopefully be reduced as it would be more experienced OSM Editors who edit the OSM dataset as opposed to the novice Observers. The service decouples first hand data acquisition from the translation into tags that are meaningful for OSM. In addition, by raising the Notes in the OSM, the Editors in areas may become more aware of the importance of mapping particular aspects which affect the accessibility of places.

It should be noted that the use of this approach is not without its shortcomings in the form of a delay in the entering of data being entered after it is collected and the possibility of introducing uncertainty due to vague descriptions provided by

Figure 3: Web-map interface
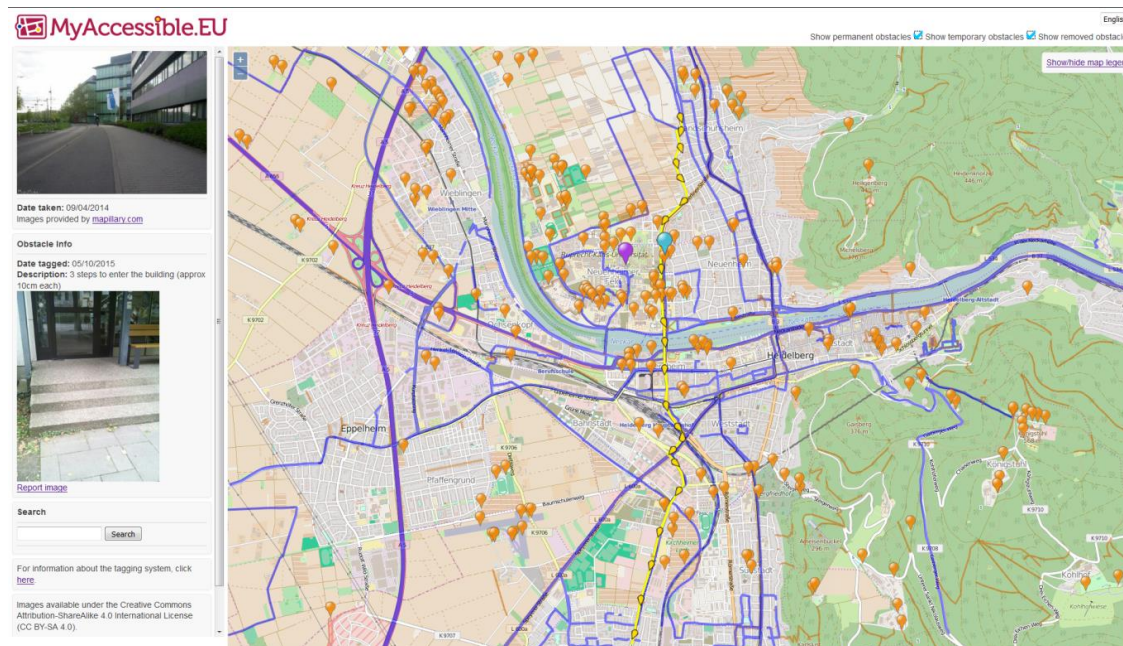
Around the same time, an OSM Editor who has subscribed to an RSS feed of Notes that are created in their local area notices that a new Note has been submitted about the surface of a footpath. Based on the description and imagery (accessed from the web-interface through a link in the note), they open their preferred OSM editor tool and add the *surface=cobblestone* and *smoothness=bad* to the feature. After doing so, they mark the Note as being complete and the data is present within the OSM dataset for anyone to use and view (i.e. routing services).

6 Summary

This paper has presented a novel service which makes use of the OSM Notes system to allow novice users to contribute important information to OSM without the need to directly edit dataset themselves. In the approach described, the Observers provide a textual description of obstacles

Observers. The process of allowing two-way communication between the Observer and the Editor is currently being investigated as a means of reducing the latter shortcoming.

7 Future Work

The next stages of the service development is to perform an analysis on data collection and entering as a means of assessing the effectiveness of the technique. This will investigate the number of obstacles that can be entered successfully into the OSM via the Notes created, and the temporal span from recording to insertion.

Also, the provision of two-way communication between the Observer and the Editor via the Notes comment facility will be investigated. It is perceived that allowing for two-way communication will remove some of the data quality issues introduced by misinterpretation of the information provided as

the Editor will be able to ask for clarification from the Observer.

References

[1] A.L. Ali, F. Schmidt, Z. Falomir, C. Freska. Towards Rule-Guided Classification for Volunteered Geographic Information. In *IPSRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W5*. ISPRS Geospatial Week, La Grande Motte, 2015

[2] A. Baglatzi, M. Kokla, M. Kavouras. Semantifying OpenStreetMap. In *Proceedings of the 5th International Terra Cognita Workshop*, 2012

[3] M. Bakillah, A. Mobasheri, A. Rousell, S.Hahmann, J. Jokar-Arsanjani, S.H.L. Liang. Toward a Collective Tagging Android Application for Gathering Accessibility-Related Geospatial Data in European Cities. In *The 18th AGILE International Conference on Geographic Information Science*. Lisbon, Portugal, 2015

[4] A. Kashian, A. Rajabifard, K.-F. Richter. Roadplex: a Mobile VGI Game to Collect and Validate Data for POIs. In *IPSRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-2, 2014*, ISPRS Technical Commission II Symposium, Toronto, 2014

[5] A. Mobasheri, M. Bakillah, A. Rousell, S. Hahmann, A. Zipf. On the Completeness of Sidewalk Information in OpenStreetMap; a Case Study of Germany. In *The 18th AGILE International Conference on Geographic Information Science*. Lisbon, Portugal, 2015

[6] P. Mooney, P. Corcoran. The Annotation Process in OpenStreetMap. *Transactions in GIS*, 16(4), pp. 561-579, 2012

[7] P. Neis, D. Zielstra. Generation of a tailored routing network for disabled people based on collaboratively collected geodata. *Applied Geography*, 47, pp 70-77, 2014

[8] L. Palen, S.T. Jennings Anderson, M. Barrenechea. Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. Seoul, Republic of Korea, 2015

[9] S.H. Pazoky, F. Karimipour, F. Hakimpour. You Describe It, I Will Name It: an Approach to Alleviate the Effect of Users' Semantics in Assigning Tags to Features in VGI. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3/W3*. ISPRS Geospatial Week, La Grande Motte, 2015

[10] F. Prandi, M. Soave, F. Devigili, R. De Amicis, A. Astyakopoulos. Collaboratively Collected Geodata to Support Routing Service for Disabled People. In *Proceedings of the 11th international Symposium on Location-Based Services*. Vienna, Austria, 2014

[11] A. Vandecasteele, R. Devillers. Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. In *OpenStreetMap in GIScience* (pp. 59-80). Springer International Publishing, 2015