Citizen-generated spatial data and information: Risks and Opportunities

Peter Mooney*†, Huabo Sun*, Padraig Corcoran† and Lei Yan*
* Institute of Remote Sensing and GIS,
Beijing Key Lab of Spatial Information Integration and its Applications,
Peking University,
Beijing 100871, Peoples Republic of China
email: bj2008.sun@gmail.com
†Geotechnologies Research,
Department of Computer Science,
National University of Ireland Maynooth (NUIM)
Co. Kildare. Ireland
Email: peter.mooney, padraig.corcoran @nuim.ie

***Abstract*—The last few years has seen the use of mobile technology become ubiquituos. Many millions of citizens around the world own smartphones, which they use for both personal and business applications. The majority of smartphones are designed towards Location-based Services (LBS). Consequently these smartphones have on-board GPS devices with the ability of locating the user to an accuracy of a few meters. While at the begining of the growth in popularity of smartphones and LBS technologies most users simply consumed services: where is the nearest coffee shop or subway station? Today, citizens are generating spatial data and information at ever increasing volumes. Examples include: georeferenced photographic collections, location-based social networking such as FourSquare, and volunteered geographic information (VGI). VGI is an exciting movement whereby citizens collect spatial data and information about (their own or another) locality. This content is then shared in collaborative projects such as OpenStreetMap, Geonames, Google Maps Mashups, or WikiMapia. This paper explores the risks involved in using this user-generated spatial data and information with specific emphasis on OpenStreetMap. Most citizens are not specialists in geographic surveying or cartography. Our paper provides results of a large case-study of *high* edit geographical features. We show that user generated spatial data is a very dynamic but has many inconsistencies. This severely limits its use in many security and intelligence applications, for example.**

I. INTRODUCTION

Only two years ago the emergence of micro-blogging, such as Twitter, caused Google to admit that it "was losing out to engines such as Twitter in the race to meet web user demand for real-time information" [1]. Only a short number of years ago the Internet employed the standard provider-consumer model. Content providers produced information and services and these were consumed by end-users. This was also the case for the provision of spatial data and cartographical products and services. Traditionally, as Diaz et al [2] comment, provision of spatial data and information on the Internet followed a *top-down approach*. This scenario mirrored the provider consumer paradigm where only official providers like National Mapping Agencies (NMAs) and other environmental agencies, centrally, managed and deployed data and information resources according to their institutional policies. At this time end users were limited to a pure consumer role [2]. Recent advancements in web technologies have enabled new ways of participation on the web. Collaborative web applications (CWAs) have become a pervasive part of the Internet [3]. Topical forums, blog/article comments, open-source software development, and wikis are all examples of CWAsthose that enable a community of end-users to interact or cooperate towards a common goal. It is not surprising to see how the web has changed the way we communicate, how we do our daily routines, and even our social behaviour. Citizens, experts and non-experts alike, are increasingly participating in the process of generating continuous information and collaborating with others in problem-solving tasks. This highlights the transition of the role of users from just mere data consumers to active participants and providers [2]. VGI has built upon the rapid success of user-generated content on the Internet. Social media carries the advantages of low cost, rapid transmission through a wide community, and user interaction. The crowdsourced approach of OpenStreetMap derives its success from citizens mapping and collecting data and information about their locality. Features being mapped include the location of garbage cans, pedestrian crossings, land cover types, shops, education facilities, to government buildings, roads and river networks. The specific aim of the paper is as follows. Citizens can: collect their own spatial data and information, submit it to a CWA such as OpenStreetMap, and then allow it be used by 3^{rd} parties in web applications, web services, and research. What are the risks associated with this new Web 2.0 evolution of the producer-consumer model?

II. EXPLAINATION OF TECHNOLOGIES INVOLVED

In this paper we investigate one of the most well known examples of a spatial CWA on the Internet - OpenStreetMap (OSM). OSM embodies all of the characteristics of crowdsourcing, VGI, and CWAs and consequently provides a robust case-study for our work. The VGI community is a global crowdsourced (many volunteers working together) community which shares many similarities with the Wikipedia model of information collection. The OpenStreetMap project [4] is a crowdsourced geospatial database with volunteers all over the world. It is probably the most famous VGI project with a mission to create a free, constantly updating and improving, editable map of the world in addition to providing free access to the underlying spatial data (geometry and attributes). Masses of contributors from around the world are volunteering their time and efforts to collaboratively create a detailed base map. Many other volunteers in OpenStreetMap are working on: software development for OpenStreetMap, maintaining the OpenStreetMap Wiki website, organising mapping party events, etc. The growing spatial coverage, and high-quality content, has branched beyond "the converted" and has gained enthusiastic endorsement from the likes of Yahoo, ESRI, MapQuest [5], and Microsoft [6]. A core motivation behind the production of VGI is likely the inaccessibility and cost of accurate sources of geographic information [7], [8]. The capacity of people from around the world to create geographic information has further been assisted by the drop in the price of GPS units and the wide availability of computers and smart mobile devices [7], [8].

A. *Location-based Services*

Location based services (LBS) are a "killer application" [9] in mobile data services thanks to the rapid development in wireless communication and location positioning technologies. This ubiquitous computing paradigm brings great convenience for information access. People with Internet-enabled wireless devices can find out information like: Where is the nearest coffee shop? What time is the next bus at this station? Where is the highest rated seafood restaurant in town? The constraints of mobile environments, the spatial property of location-dependent data, and the mobility of mobile users pose a great challenge for the provision of location-based services to mobile users . The issue of access to geospatial data for LBS is unlikely to disappear in the near future [10]. LBS are often cited as one service likely to continue driving the development of the "Mobile Internet". Irrespective of the range of services encapsulated by the broad "LBS" term some authors [11] comment that "all LBS will continue to require spatial data management capabilities to link position information with other data sources". Other authors [**?**] remark that "ultimately the utility of LBS will be measured by their ability to meet user needs" in the application domain where "*content is king*[12]. The requirements from the LBS technological and user communities for better access to geospatial data can become a significant driver for the geospatial information industry to ensure better access to geospatial data in reality. In this paper we have used the term access in its broadest sense to include access in terms of: cost models for access to geospatial data, accuracy of the data, frequency of updates, and the conditions under which developers can "mash up" or integrate this geospatial data with their own data and information.

III. OVERVIEW OF CURRENT RESEARCH

OSM data is collected by volunteers who are members of the OpenStreetMap community who contribute to the OSM database in the following ways: collecting GPS traces and uploading these traces to the OpenStreetMap data using one of a number of official OSM editors; editing other users' GPS traces; tracing geographic features from aerial imagery (Yahoo! and recently Bing); and bulk upload of spatial data which fits into the OpenStreetMap licensing structure. The percieved lack of cartographical, surveying, and GIS skills of contributors has seen spatial quality in OSM become a major issue. [13] remark that there are no accepted metrics for measuring the quality of OSM or to a wider extent the quality of VGI. Given the dynamic and organic nature of the spatial data contained in the OSM databases the quality of the spatial data can change considerably quickly [14]. The recent study by [15] of OSM and TeleAtlas for Germany shows that "while professional data is not without it's faults the coverage of OSM in rural areas is too small to be seriously considered a sophisticated alternative for *any* applications". However examples of acceptance within the GIS community as a source of spatial data have begun to appear. [16] describe the development of $3D$ models for cities using OSM data combined with Digital Terrain Model (DTM) data. [17] use OSM as sample input data for testing new approaches to ensuring planar and non-planar topologically consistent map simplifications. In [18] OSM data are integrated for the first time into the robot tasks of localization, path planning and autonomous vehicle control. The authors provide encouraging results of this approach with results in outdoor environments demonstrating the effectiveness of OSM data for these tasks.

A. *Motivation for this Research*

One of the drawbacks of the current literature is that most OSM quality studies are performed as ground-truth comparisons. This involves comparative measurement of a given set of characteristics from OSM against some accepted and trusted ground-truth dataset. While this is, of course, an acceptable method of comparison we feel that OSM in it's current form will always perform poorly in such comparisons. The inhomogeneity of coverage of OSM is a major drawback. With this in mind we feel that there is merit in assessing the quality of OSM in isolation and without comparison to a ground-truth dataset. We now provide a list of potentially useful assessments one can make of OSM which could go some way to answering questions about it's quality and/or suitability for a particular application.

- How does an object structurally change over time? How does the representation of the object change over time? [19], [20] show examples of how object representation in OSM changes radically for some object classes.
- How does the metadata (tags) associated with an object change over time? [21] summarize different techniques employed to study various aspects of tagging: tagging models, tag semantics, generating recommendations using tags, visualizations of tags, applications of tags and

problems associated with tagging usage. [22] argues that as the number of nonspecialist users of GI increases and spatial data are used to answer more questions about the environment the need for users to understand the wider meaning of the data becomes crucial. The use of tagging and metadata is part of this process.

- Is a given region R of OSM topologically consistent over it's lifetime or over some time period t_1 to t_2? The approaches of [23] for checking and maintaining topological consistency of vector datasets undergoing simplification could be applied here.
- Is a given object O valid over some time period t_1 to t_2? Are changes in the object O reflected in other objects U within the same neighbourhood or region? [24] outline a number of measures to quatify map information and distribution of spatial information in a given map.
- Which OSM contributors have edited a given object O or a set of objects? [25] provides a current study of OSM contributor behaviour. They find that the number of OSM objects in an area clearly grows in relation to the number of contributors in the area, but in a non-linear way: most areas only one contributor. There are possible connections with user contributions to Wikipedia. Are user/contributor motivations the same? In one study [26] find that self motivation is a key driver for user contribution for knowledge sharing on Wikipedia

This is not an exhaustive list but presents a flavour of potential research questions. Most of the questions require access to the historical record of objects in OSM. This historical record (or edit paper trail) will allow one to analyse how an object (or group of objects) evolved to their current representation within the OSM database. There are no other examples in the literature, to our knowledge, which analyse the historical record of objects in OSM. In Section IV we provide some examples of an examination of how (1) spatial features are annotated with metadata, and (2) how features are represented as geometric objects within the OSM database.

IV. EXPERIMENTAL DETAILS AND ANALYSIS

In Figure 1 an example of a visualisation of two diferent versions of a polygon are shown. Figure 1 shows the first version (outlined in black) and the current version (66 outlined in red) of a forest in Baden-Württemberg, Germany. The first version was created in February 2008 with 176 nodes. The current version was created in September 2010 with 1049 nodes. There are two distinct and disjoint polygons representing the same object in OSM. Without access to the historical record of edits to this object it would not be possible to visualise how this object has evolved from the first version to it's current form. In this section we discuss: how to access and process the OSM data (section IV-A), analysis of how contributors to OSM annotate spatial features (section IV-B), and then in section IV-C we look at spatial representation as geometric objects.

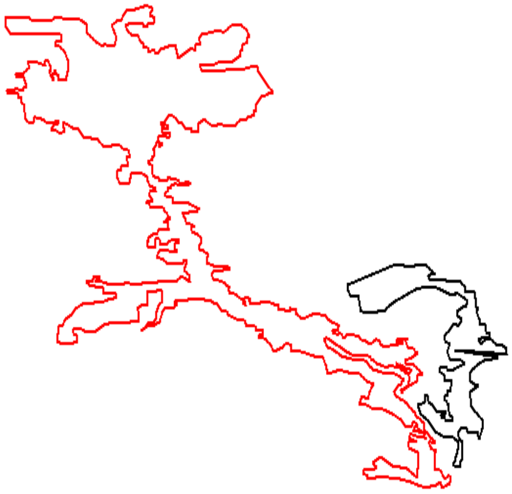

Fig. 1. A visualisation of output from our software. Two GPX files are visualised within QGIS to show the first (black) and current version 66 (red) of an object in OSM

A. Access and Processing OSM data

OSM data is freely available, in OpenStreetMap XML format, from the GeoFabrik website http://download.geofabrik.de/. This data is updated almost hourly so the most up-to-date version of the OpenStreetMap database is always available. We downloaded the OSM XML data for the United Kingdom and Germany on March 4^{th} 2010. We extracted the $15,000$ most heavily edited `ways` (polygons or polylines) from the datasets. To extract these ways the OSM-XML was imported into a PostGRES PostGIS spatial database using the freely available OSM processing tool `osm2pgsql`. The uncompressed OSM-XML files for the UK and Germany are approximately 25 GB in size. Import was performed on an Intel Core i7-2600 processor with 4GB of RAM running Ubuntu Linux 10.10. Total import time to PostGIS was 8.5 hours. Using simple SQL queries we extracted the identification numbers of the $15,000$ ways. Using Python we automatically created a Linux Bash script file with `wget` commands to download the history file for each way directly from the OSM API web service. Each history file is also in OSM-XML format and contains, in temporal order, every consecutive version of the spatial representation of the corresponding way. The total time required to download all $15,000$ files was just under 50 hours. We paused the automated download every 60 minutes for 30 minutes so as not to overload the OSM webservers. Python was used to: query the PostGIS database, build a Linux Bash command line script to download all of the OSM-XML history files, and finally to peform the following processing on the OSM-XML history files. For each way w we compute a number of characteristics for each version wv of w including: wv_n - the number of nodes in wv, wv_T - the set of tags (key,value) pairs annotating wv, wv_u - the user id of the user who created wv, and wv_t the timestamp of the edit of wv. We also store, in PostGIS, the geometry g of the polygon or

TABLE I

OSM-ID 26164873 FROM GERMANY. THE TABLE SHOWS THE CHANGES IN TAGS DESCRIBING THE WOOD OR FOREST REPRESENTED BY THE POLYGON

Version	**Date**	**UserID**	**Tag Action**
1	09 − 08 − 08	24748	landuse=forest (added)
2	18 − 08 − 08	8732	natural=wood (added)
32	16 − 01 − 09	24748	natural=wood (deleted)
53	19 − 05 − 09	100946	area=yes (added)
71	01 − 08 − 09	53563	area=yes (deleted)

polyline representing wv of w.

B. Annotation of spatial features

Tagging has recently become popular as a means for annotating and organizing web content, particularly in the context of community generated media. Tags are collections of keywords that are attached to web content to help describe the entry. Tags are also attached to spatial content in VGI. Park et al [27] remark that in the vast majority of CWA (including OSM to a certain extent) there is no globally agreed list of tags user can choose from. Consequently different users use different tags to describe the same web resources (or spatial features) and even a single users tagging practice may vary over time. OSM does maintain a community agreed list of tags that users can choose from. These are listed on the Map Features page of OSM [28]. However, contributors are free to add their own additional tags to annotate spatial features. In Table I an example is shown where the tags describing the wood or forest represented by a polygon change. There are four unique contributors to this polygon. In Table II another example is shown where the tags change frequently on an object. There is disagreement amongst the two users involved in the editing of the feature as to if the feature is a “natural” feature as a “wood” or is a landuse representation as a forest. Figure 2 shows a plot of the number of unique contributors to the $15,000$ features in our case-study area against the number of tags associated with each feature. The number of tags is taken from the final (or current) version of the feature. The plot shows a very interesting pattern (correlation -0.12). As the number of contributors increase there is a decrease in the number of tags assigned to each feature. Figure 3 shows a 2-D heat map histogram illustrating the distribution of the number of versions (x axis) against the number of unique contributors (y). As the number of contributors increases this is not correlated with a increase in the number of versions created for each object. The number of unique contributors is tightly clustered around the mean ($\mu = 5.89$) while the number of versions created by these contributors has a mean ($\mu = 5.89$). This result dispels the anecdotal belief that OSM contains “many different contributors created lots of different versions of the same feature”.

C. Spatial Representation

In many features the spatial representation of the geometry of the feature is not consistent over the entire perimeter

TABLE II

OSM-ID 24015216 FROM GERMANY. THE TABLE SHOWS THE CHANGES IN TAGS DESCRIBING THE WOOD OR FOREST REPRESENTED BY THE POLYGON

Version	**Date**	**UserID**	**Tag Action**
1	28 − 04 − 08	27675	landuse=forest (added)
1	28 − 04 − 08	27675	natural=wood (added)
5	28 − 04 − 08	27675	landuse=forest (deleted)
12	29 − 04 − 08	27675	landuse=forest (added)
12	29 − 04 − 08	27675	natural=wood (deleted)
28	12 − 07 − 08	27675	highway=primary (added)
28	12 − 07 − 08	27675	natural=wood (added)
31	12 − 07 − 08	27675	highway=primary (deleted)
35	13 − 07 − 08	24748	natural=wood (deleted)
36	14 − 08 − 08	24748	natural=wood (added)
36	30 − 08 − 08	24748	natural=wood (deleted)

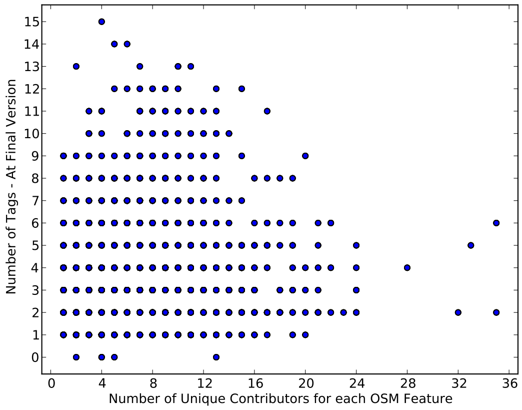

Fig. 2. Scatter plot of the number of contributors (x) against the number of tags for each feature (y)

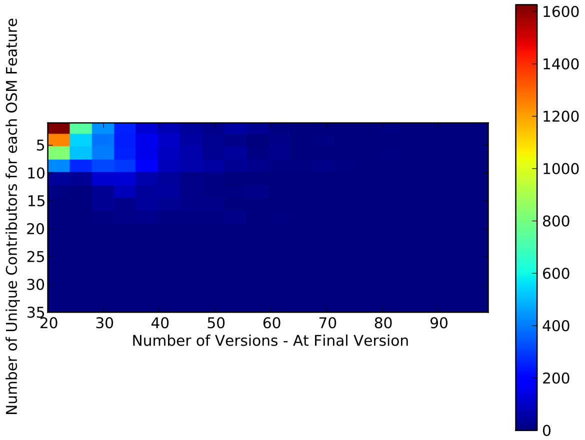

Fig. 3. A 2-D heat map histogram showing the distribution of the number of versions (x axis) against the number of unique contributors (y)

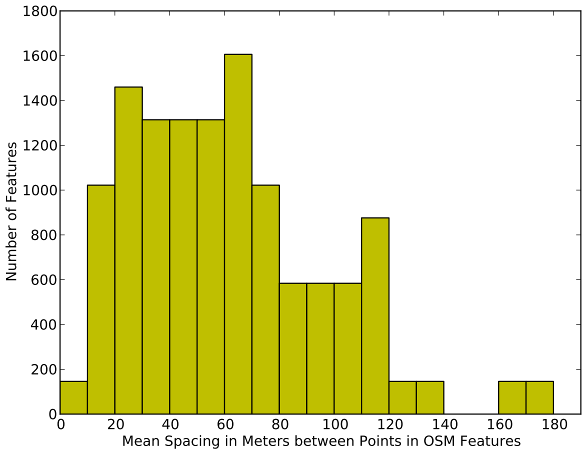

Fig. 4. Distribution of the mean spacing (meters) between sample points in the features in our case study

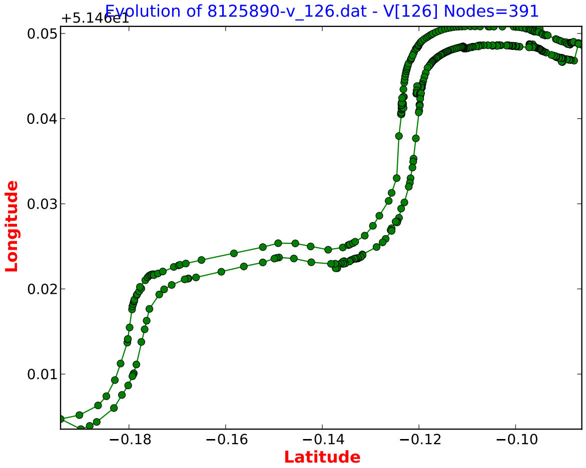

Fig. 5. The River Thames in London, England modelled as a polygon

of the corresponding polygon or polygon. An example of this is illustrated in Figure 5 representing a portion of the riverbank of the River Thames in London, England. The polygon representing the riverbank has been edited 126 times. There is a greater concentration of nodes at the north-east section of the river (close to London Bridge). On the bend sections from the south-west section the sampling points are more spaced out and scattered. In Figure 6 a river, close to Maidenhead, England, is shown. In total 839 nodes are used to represent the river feature. The representation is much more consistent with sampling points distributed evently along the feature. Differences in the spatial representation of polygons can have issues when the spatial data is generalised, viewed at different scales, etc [14]. The historgram shown in Figure 4 shows the distribution ($\mu = 112.34$) of mean spacing (in meters) between sample points in all of the features. The histogram shows that there is wide variance amongst all of the features in our case study. Features with mean spacing of 30 meters or less could be considered as "well represented" and we speculate that these features are the result of ground-sampling campaigns by OSM contributors. Features with very large mean spacing between sample point can be the result of: poor sampling, tracing of features from low-resolution aerial imagery, or important of spatial data from other datasets which may have already undergone simplification or generalisation. An interesting topic for future work is supported by the graph illustrated in Figure 7. In this simple plot each unique user who contributes to the River Thames polygon has their user-id mapped to their position in the sequence of edits. The plot shows almost 21 unique contributors in the first 30 versions. Then the editing is dominated by a single contributor until version 95. For the remaining 30 versions new contributors arrive combined with updated contributions from previous contributors earlier in the evolution of the feature.

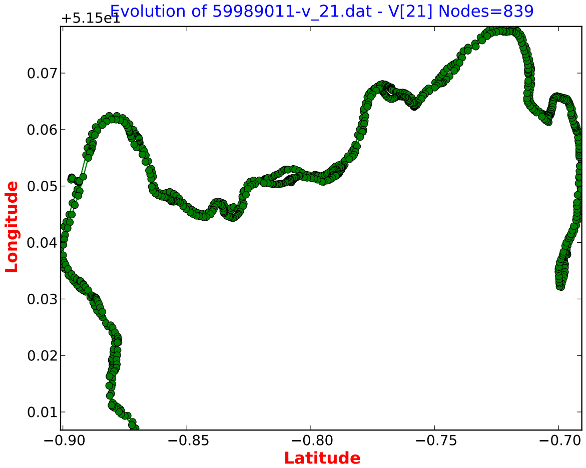

Fig. 6. A river at Maidenhead, England at it's current version 21 created on 01/02/2011

V. CONCLUSION

The usage of citizen generated spatial data information is very much use-case dependent. Subsequently, researchers, GIS experts, data managers, etc should be careful about generalising. For OSM, for example, there are many variables and issues to consider before one can definitely say "OSM is a very good choice for X" or a "bad choice for Y". The precise spatial data and information requirements of the task/problem in hand must be carefully considered. We can state with certainty that the variability in coverage of OSM prevents its usage as a homogeneously consistent data source over large geographical areas. We have shown a number of examples, taken from a large subset of OSM, that there is great variability in the: spatial respresentation, annotation, and number of contributors that each feature has.

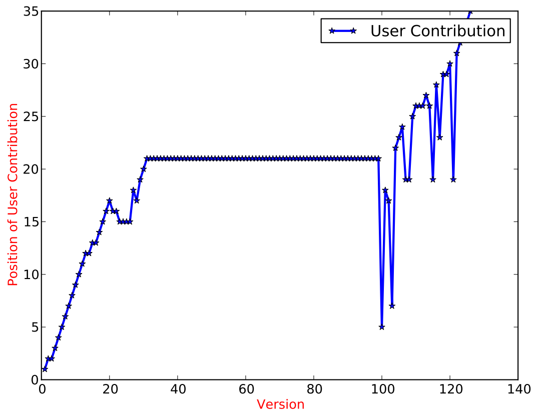

Fig. 7. The contributors from contributors to the creation of the polygon for the River Thames riverbank in Figure 5

ACKNOWLEDGMENT

The work presented in this paper is supported by The Scholarship for International Young Scientists of NSFC (National Natural Science Foundation of China) (ID: 41050110441). Peter Mooney is also funded by the Irish Environmental Protection Agency STRIVE programme (grant 2008-FS-DM-14-S4). Padraig Corcoran is part of STRAT-AG which is a Strategic Research Cluster grant (07/SRC/I1168) funded by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge all of these supports.

REFERENCES

[1] R. Wray, "Google falling behind twitter," The Guardian Online: Tuesday 19 May 2009. http://www.guardian.co.uk/business/2009/may/19/google-twitter-partnership, 2009.

[2] L. Daz, C. Granell, M. Gould, and J. Huerta, "Managing user-generated information in geospatial cyberinfrastructures," *Future Generation Computer Systems*, vol. 27, no. 3, pp. 304 – 314, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/B6V06-511TN6Y-1/2/71ddab2f32a6a8d32ff78cdfc9e8b69e

[3] A. G. West, J. Chang, K. K. Venkatasubramanian, and I. Lee, "Trust in collaborative web applications," *Future Generation Computer Systems*, vol. In Press, Corrected Proof, pp. –, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/B6V06-526DWDM-2/2/ff5fb978d8a1871a0d3bb1f3d84b7ebb

[4] OpenStreetMap, "The OpenStreetMap project," Online: http://www.openstreetmap.org, February 2011.

[5] MapQuest, "Aol corporate communications - mapquest 'opens up' in europe with open-source mapping with uk launch," AOL Online Blog report http://corp.aol.com/2010/07/09/mapquest-opens-up-in-europe-with-open-source-mapping-with-uk-l/, July 2010.

[6] Bing Maps, "Bing engages open maps community," Microsoft Bing Maps Online Blog report http://www.bing.com/community/site_blogs/b/maps/archive/2010/11/23/bing-engages-open-maps-community.aspx, November 2010.

[7] M. M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

[8] M. Zook, M. Graham, T. Shelton, and S. Gorman, "Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake," *World Medical & Health Policy*, vol. 2, no. 2, p. Article 2, 2010.

[9] J. Xu, B. Zheng, and W.-C. Lee, "Data access techniques for location-based services," *Mobile Data Management, IEEE International Conference on*, vol. 0, p. 64, 2006.

[10] P. Mooney and P. Corcoran, "Geospatial data issues in the provision of location-based services," in *Proceedings of the 7th International Symposium on LBS & TeleCartography*, G. Gartner, Ed., South China Normal University, Guangzhou, China, 2010.

[11] S. Jacoby, J. Smith, L. Ting, and W. I.P., "Developing a common spatial data infrastructure between state and local governments - a case study from victoria, australia," *International Journal of Geographical Information Science*, vol. 16, pp. 305–322, 2002.

[12] C. S. Jensen, "Database aspects of location-based services," in *Location-Based Services*, J. H. Schiller and A. Voisard, Eds. Morgan Kaufmann, 2004, pp. 115–148.

[13] P. Mooney, P. Corcoran, and A. C. Winstanley, "Towards quality metrics for openstreetmap," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 514–517.

[14] ——, "A study of data representation of natural features in openstreetmap," in *Proceedings of the 6th GIScience International Conference on Geographic Information Science*, ser. GIScience 2010. University of Zurich, 2010, p. p150.

[15] D. Zielstra and A. Zipf, "A comparative study of proprietary geodata and volunteered geographic information for germany," in *Proceedings AGILE 2010: The 13th AGILE International Conference on Geographic Information Science*, M. Painho, M. Y. Santos, and H. Pundt, Eds. Guimarães, Portugal: Springer Verlag, 2010.

[16] M. Over, A. Schilling, S. Neubauer, and A. Zipf, "Generating web-based 3d city models from openstreetmap: The current situation in germany," *Computers, Environment and Urban Systems*, vol. 34, no. 6, pp. 496 – 507, 2010, geoVisualization and the Digital City - Special issue of the International Cartographic Association Commission on GeoVisualization.

[17] P. Corcoran, P. Mooney, and A. C. Winstanley, "Planar and non-planar topologically consistent vector map simplification," *International Journal of Geographic Information Science.*, vol. x, p. to appear, January 2011.

[18] M. Hentschel and B. Wagner, "Autonomous robot navigation based on openstreetmap geodata," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, 2010, pp. 1645 –1650.

[19] P. Mooney, P. Corcoran, and A. Winstanley, "A study of data representation of natural features in openstreetmap," in *Proceedings of GIScience 2010: the* 6^{th} *International Conference on Geographic Information Science*, R. Purves and R. Weibel, Eds. Published online at http://www.giscience2010.org, 2010, p. p150.

[20] P. Mooney, P. Corcoran, and A. C. Winstanley, "Towards quality metrics for openstreetmap," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 514–517.

[21] M. Gupta, R. Li, Z. Yin, and J. Han, "Survey on social tagging techniques," *SIGKDD Explor. Newsl.*, vol. 12, pp. 58–72, November 2010.

[22] A. Comber, P. Fisher, and R. Wadsworth, "What is land cover?" *Environment and Planning B: Planning and Design*, vol. 32, no. 1, pp. 199–209, 2005.

[23] P. Corcoran, P. Mooney, and A. C. Winstanley, "Planar and non-planar topologically consistent vector map simplification," *International Journal of Geographical Information Science*, vol. x, no. x, p. To Appear, 2011.

[24] L. Harrie and H. Stigmar, "An evaluation of measures for quantifying map information," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 3, pp. 266 – 274, 2010.

[25] M. Haklay, S. Basiouka, V. Antoniou, and A. Ather, "How many volunteers does it take to map an area well? the validity of linus' law to volunteered geographic information," *The Cartographic Journal*, vol. 47, no. 4, pp. 315 – 322, 2010.

[26] H.-L. Yang and C.-Y. Lai, "Motivations of wikipedia content contributors," *Comput. Hum. Behav.*, vol. 26, pp. 1377–1383, November 2010.

[27] J. Park, B.-C. Choi, and K. Kim, "A vector space approach to tag cloud similarity ranking," *Information Processing Letters*, vol. 110, no. 12-13, pp. 489 – 496, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6V0F-4YNT4BR-1/2/4e516ad758cdf09a06c2d329577eadfd

[28] OSM, "The map features page on openstreetmap.org - the community agreed ontology of tags," Online Wiki Page: http://wiki.openstreetmap.org/wiki/Map_Features (checked February 2011), February 2011.