# OSM-CAT: A Java tool for OpenStreetMap Contributor Analysis

Peter Smith and Peter Mooney

Department of Computer Science, National University of Ireland Maynooth,
Co. Kildare. Ireland
Tel. +353-1-708 3847 Fax +353-1-708 3848
Email: peadar81@gmail.com
, peter.mooney@nuim.ie
http://www.cs.nuim.ie/~pmooney/OSM-CAT

**Summary:** In this paper we describe a new Java-based software tool called OSM-CAT (OpenStreetMap Contributor Analysis Tool). OSM-CAT allows GIS researchers to analyse OSM-XML data (for any given area). OSM-CAT produces a large set of output including: breakdown and analysis of who contributed to the OSM data in the chosen area, what types of contributions were made, when these contributions were made, and the accuracy of tagging based on the rules outlined in the OSM Map Features Ontology. It removes the requirement for the user to have any knowledge of handling of OSM-XML data.

**KEYWORDS:** OpenStreetMap, Quality, Web GIS, VGI

## 1. Introduction

OpenStreetMap (OSM) is the most famous example of Volunteered Geographic Information (VGI) (Goodchild; 2007) on the Internet today. It hardly requires any introduction to the GIS community. (Sui and Goodchild; 2011). Papers such as Ciepluch et al. (2009), Haklay (2010), Haklay et al. (2010), Girres and Touya (2010), Over et al. (2010), and Mooney and Corcoran (2012) have firmly placed OSM in the research agenda of GIS. However there still appears to be some reluctance amongst GIS researchers to use OSM for "serious geomatics" (Over et al.; 2010) or geographical analysis such as landcover change (Ludwig et al.; 2011). As reported in previous papers by our research group (Mooney et al.; 2010; Mooney and Corcoran; 2011b, 2012) one of the reasons for this reluctance is the availability of raw OSM spatial data in OSM-XML format which many researchers find difficult to work with (Mooney and Corcoran; 2011a) and XML in general (Dunfey et al.; 2006). Other reasons include concerns about the quality of the OSM data, OSM contributor skills and backgrounds, and inhomogeneity of coverage. In this paper we describe the functionality of a new Java-based software tool called OSM-CAT (OpenStreetMap Contributor Analysis Tool). OSM-CAT allows GIS researchers to analyse OSM-XML data (for any given area). OSM-CAT produces a large set of output including: breakdown and analysis of who contributed to the OSM data in the chosen area, what types of contributions were made, when these contributions were made, and the accuracy of tagging. OSM-CAT has been developed as part of an MSc in Computer Science and Engineering with the goal of providing users with a tool to analsye characteristics of contributions to OSM for small areas. OSM-CAT is not intended as a replacement for tools such as OSMOSIS or osm2pgsql or TagInfo and others which are already used widely in the OSM community. Rather this adds to the list of available tools by providing specific application functionality. The software is available to download at http://www.cs.nuim.ie/~pmooney/OSM-CAT

## 2. Overview of the OSM-CAT tool

There are a number of similar tools available which provide information on contributions to OpenStreetMap. AltogetherLost (2011) provides a graphical interface to statistics about the global

OpenStreetMap database with some information on contributors such as number of members who are the last modifier of at least one OSM object (Node, Way or Relation) in the past day. Neis (2011c) allows the comparison between any two OSM contributors on competitions such as "Which OSM contributor edited most nodes?". How did you contribute to OSM by Neis (2011a) is similiar as it displays the summary of edits performed by a single contributor. The OSM Heat Map (Neis; 2011b) overlays a heat map on a web-based map to indicate where a given contributor carried out the most edits. Changesets are an important concept in contributions to OSM. A changeset is a group of edits made within a certain time by one user. A changeset has a maximum capacity (currently 50, 000 edits) and maximum lifetime (currently 24 hours). Edits can only be added to a changeset as long as it is still open. A changeset can either be closed explicitly, or it closes itself if no edits are added to it for a period of inactivity (currently one hour). The OpenStreetMap Watch List (OWL) by Amos (2011) shows changes to areas by comparing changesets for that area. While all of these tools are very useful we felt that there was a gap in the available tools in accessing more detailed information about contributions and the data for given OSM area.

### 2.1. Functionality of OSM-CAT

The functionality provided by OSM-CAT includes:

- Summary Information: OSM Primitive Data Overview including total number of edits, nodes, ways, relations, and changesets
- Changeset Information: An analysis of the number of changesets produced by each contributor. This information is provided in tabular and graphical formats
- Activity in Changesets: This provides a summary of the timespan between changesets (longest and shortest) for each contributor
- Overall Contributions: Tabular and graphical summary of the total number of contributions per month since oldest edit in chosen area
- Individual Contribution Rates: Number of edits per month from the top N contributors (tabular and graphical)
- Map Features and Tag Validity: OSM-CAT provides a summary of the number of valid and invalid assignments to the tag keys for all qualifying objects.
- Polygon Validity: Using the Java Topology Suite library OSM-CAT outlines any polygons (ways) which are invalid such as those with self intersections.

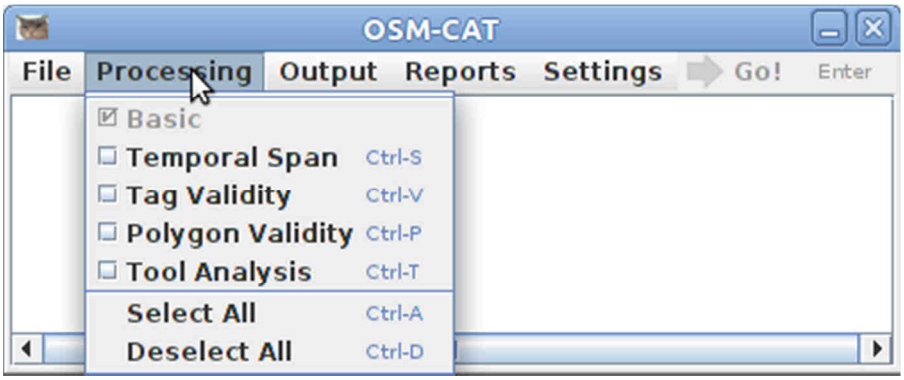


*Figure 1: The graphic user interface to OSM with the functionality menu extended to show all options available.*

**3. Case-study example usage.**

In this case-study we used OSM-CAT to perform analysis on OSM-XML data from 60 European cities (3KM square grids). The OSM-XML data was automatically downloaded from the OpenStreetMap servers on November 10th 2011. Cities were chosen to provide a very good distribution of contribution types and spatial coverage to OSM. The purpose of the case-study is to show the types of statistics that can be generated by OSM-CAT, which may otherwise be difficult for those not familiar with OSM-XML to generate. There is no scope in this paper to discuss the implications of the results but we comment on some of the statistics below.

- The top 10 locations sorted in order of total number of nodes are: 1. Paris 244, 872, 2. Lyon 194, 913, 3. Prague 152, 282, 4. Ruhr-NorthRhine-Westphalia 138, 836, 5. Karlsruhe 115, 992, 6. Bratislava 114, 774, 7. Wien 106, 640, 8. Stuggart 106, 619, 9. Helsinki 105, 643, and 10. Salamanca 103, 870.

- The top 10 locations sorted in order of the total number of contributors are as follows. Their ranking overall in terms of number of nodes is provided in brackets. London 553 (16th), Wien 511 (7th),Munich 510 (13th), Berlin 506 (23rd), Paris 322 (1st), Meunster 318 (36th), Stuggart 293 (8th),Karlsruhe 284 (5th), Moscow 263 (19th), and Zurich 257 (20th).

- The top 10 locations sorted in order of the total number of Points-of-Interest (POI) nodes (nodes with a “name” attribute): London 5, 072 ,Wien 5, 051 ,Munich 3, 831 ,Paris 2, 998 ,Berlin 2, 699 ,Helsinki 2, 132 ,Zurich 2, 090 ,Prague 1, 823 ,Stuggart 1, 767 ,Copenhagen 1, 752

- The top 10 locations sorted in order of the users with the largest number of changesets for that location are provided as follows. In each case the users changesets are expressed as a percentage of the total number of changesets for that location. Birmingham 1817 (69%), Helsinki 1355 (35%), Copenhagen 1355 (34%), Milano 1152 (52%), Zagreb 918 (57%), Paris 833 (26%), Oslo 810 (52%), Dublin 690 (25%), Innsbruck 651 (35%), and Brussels 631 (36%). This analysis can help identify dominant contributors for a given area.

- OSM-CAT checks tags (key value pairs) against the community agreed ontology of tagging on OSM Map Features Wiki page. There were a number of recurring problems. These are summarized as follows: “railway=platform” has been replaced by “public-transport=platform” and this problem occurs 213times in 43 cities. For mapping fire-hydrants “amenity=fire hydrant” occurs 416 times in 39 cities. This has changed to “emergency=fire hydrant”. The tagging for motorcycle parking “amenity=motorcycle parking” occurs in 49 cities on 318 occasions. This tagging is not officially sanctioned in the OSM Map Features ontology. Other frequently occurring problems regarding assignment of objects to the
  wrong class include: “building=hospital”, “building=church”,“access=restricted”.

**3. Summary, conclusions, and future work**

The role of contributors to VGI and OpenStreetMap is crucial. Coleman et al. (2009) argue that the participants in the production process of VGI are both users and producers or “prosumers”. Assessing the credibility of contributors is important to evaluating the overall reliability of their contribution. OSM-CAT provides a user-friendly means of performing analysis of contributions to OSM for a given area and identifying who the key contributors are, when they have contributed, and how accurate their tagging is (against the OSM ontology). In the case-study example we showed some results of using OSM-CAT to analyse contributions to OpenStreetMap in 60 European cities. OSM-CAT processed, and computed all relevant statistics, for all files (up to 100MB) in less than 3 minutes. OSM-CAT is a fully Java-based tool. Currently OSM-CAT is desktop-based. We will look to extend the functionality to a web-based version of OSM-CAT. This will allow us to build a Web Map interface for geographical area selection. Users will also be offered functionality to upload old

versions of OSM-XML data where the changes in contributions over time can be compared with the current snapshot of the OSM data for that location.

## 7. References


AltogetherLost (2011). OpenStreetMap members, nodes, and ways statistics, Webservice http://osmstats.altogetherlost.com/.

Amos, M. (2011). OWL (openstreetmap watch list), Webservice http://matt.dev.openstreetmap.org/owl_viewer/.

Ciepluch, B., Mooney, P., Jacob, R. and Winstanley, A. C. (2009). Using openstreetmap to deliver location-based environmental information in ireland, SIGSPATIAL Special 1(3): 17–22.

Coleman, D. J., Georgiadou, Y. and Labonte, J. (2009). Volunteered geographic information : The nature and motivation of produsers, International Journal of Spatial Data Infrastructures Research 4: 332–358.

De Leeuw, J., Nagda, S., Georgiadou, Y. and DeBlois, M. (2011). An assessment of the accuracy of volunteered road map production in western kenya, Remote Sensing 3: 247–256.

Dunfey, R. I., Gittings, B. M. and Batcheller, J. K. (2006). Towards an open architecture for vector gis, Computers and Geosciences 32(10): 1720 – 1732.

Girres, J.-F. and Touya, G. (2010). Quality assessment of the french openstreetmap dataset, Transactions in GIS 14(4): 435-459.

Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography, GeoJournal 69:

Haklay, M. (2010). How good is volunteered geographical information? a comparative study of open-streetmap and ordnance survey datasets, Environment and Planning B: Planning and Design 37(4): 682 –703.

Haklay, M., Basiouka, S., Antoniou, V. and 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(4): 315 – 322.

Ludwig, I., Voss, A. and Krause-Traudes, M. (2011). A comparison of the street networks of navteq and osm in germany, in S. Geertman, W. Reinhardt and F. Toppen (eds), Advancing Geoinformation Science for a Changing World, Vol. 1 of Lecture Notes in Geoinformation and Cartography, Springer

Mooney, P. and Corcoran, P. (2011a). Accessing the history of objects in openstreetmap, in S. Geertman,W. Reinhardt and F. Toppen (eds), Proceedings AGILE 2011: The 14th AGILE International Conference on Geographic Information Science, Springer Verlag, Utrecht, The Netherlands, p. P155.

Mooney, P. and Corcoran, P. (2011b). Annotating spatial features in openstreetmap, Proceedings of the 19th annual gis research uk (gisruk) 2011, Portsmouth, England.

Mooney, P. and Corcoran, P. (2012). The annotation process in openstreetmap, Transactions in GIS To Appear.

Mooney, P., Corcoran, P. and Winstanley, A. C. (2010). Towards quality metrics for openstreetmap, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS ’10, ACM, New York, NY, USA, pp. 514–517.

Neis, P. (2011a). OSM-hdyc: how did you contribute to openstreetmap?, Webservice http://hdyc.neis-one.org/.

Neis, P. (2011b). OSM-yohm: your osm heat map!, Webservice http://yosmhm.neis-one.org/.

Neis, P. (2011c). OSMFight compare the contributors of any two osm users, Webservice http://osmfight.neis-one.org/.

OpenStreetMap (2011). The osm Map Features ontology page, Online at http://wiki.openstreetmap.org/wiki/Map_Features.

Over, M., Schilling, A., Neubauer, S. and Zipf, A. (2010). Generating web-based 3d city models from openstreetmap: The current situation in germany, Computers Environment and Urban Systems 34(6): 496– 507.

Sui, D. and Goodchild, M. (2011). The convergence of gis and social media: challenges for giscience, International Journal of Geographical Information Science iFirst(0): 1–12.

TagInfo (2011). The openstreetmap ’taginfo’ web service, Online - dynamically updating - http://taginfo.openstreetmap.org/.

**8. Biography**

*Peter Smith is an MSc student in Software Engineering at the Department of Computer Science NUI Maynooth. He is currently working as a software developer with Ericsson Research and Development, Athlone, Ireland. Dr. Peter Mooney is a research fellow at the Department of Computer Science NUI Maynooth and he is funded by the Irish Environmental Protection Agency STRIVE programme (grant 2008-FS-DM-14-S4).*