

# An automated method to assess Data Completeness and Positional Accuracy of OpenStreetMap

Thomas Koukoletsos<sup>1</sup>, Mordechai (Muki) Haklay<sup>2</sup>, Claire Ellul<sup>3</sup>

<sup>1,2,3</sup> University College London,  
Gower Street, London, WC1E 6BT, UK  
+44 20 7679 2745

<sup>1</sup>[thomaskouk@yahoo.gr](mailto:thomaskouk@yahoo.gr), <sup>2</sup>[m.haklay@ucl.ac.uk](mailto:m.haklay@ucl.ac.uk), <sup>3</sup>[c.ellul@ucl.ac.uk](mailto:c.ellul@ucl.ac.uk)

## Abstract

OpenStreetMap (OSM) is an open source mapping application that is based on volunteered effort to create a free and worldwide spatial database. The increasing density, importance and acceptance of OSM increase the importance of understanding data quality, so that potential users can evaluate fitness-for-purpose. When spatial quality analysis is performed through comparison with a reference dataset, a data matching procedure is necessary for the comparison to be meaningful. This matching is usually performed manually at data preparation stage. After this, methods need to be applied to measure quality elements of completeness, positional and attribute accuracy, which should be capable of dealing with OSM's heterogeneity in accuracy, density and attribute information.

So far, research in the UK for OSM (Haklay 2010, Basiouka 2009, Ather 2009), provided valuable information on OSM for selected areas. However, all these studies include manual procedures and methods that hinder repetition of the evaluation in a different and larger area or in the future when OSM data is updated. Furthermore, they measure positional accuracy using a simplified version of the Increasing Buffer Method (Goodchild and Hunter, 1997).

We slightly modify and integrate the Increasing Buffer Method in an automated method that performs data matching and evaluates data completeness and positional accuracy of OSM data, taking into consideration heterogeneity of Volunteered Geographic Information (VGI). We apply the proposed method to the area of greater Liverpool.

## Περίληψη

Το OpenStreetMap (OSM) είναι μία εφαρμογή χαρτογράφησης ανοικτού κώδικα, που στηρίζεται σε εθελοντική προσπάθεια για τη δημιουργία μίας παγκόσμιας χωρικής βάσης δεδομένων. Η αυξανόμενη πυκνότητα, σημασία και αποδοχή του OSM ενισχύουν την ανάγκη για την κατανόηση της χωρικής ποιότητας, ώστε οι ενδεχόμενοι χρήστες να μπορούν να αξιολογήσουν την καταλληλότητά του για το σκοπό τους. Στην περίπτωση αξιολόγησης της χωρικής ποιότητας μετά από σύγκριση με κάποια δεδομένα αναφοράς, απαιτείται μία διαδικασία συσχέτισης των δεδομένων, ώστε η σύγκριση να έχει κάποιο νόημα. Συνήθως, αυτή η συσχέτιση πραγματοποιείται χειροκίνητα στο στάδιο προετοιμασίας των δεδομένων. Στη συνέχεια, απαιτείται η εφαρμογή μεθόδων μέτρησης των χωρικών στοιχείων ποιότητας, όπως η πληρότητα, η χωρική και η περιγραφική ακρίβεια, οι οποίες θα πρέπει να μπορούν να αντιμετωπίσουν την ετερογένεια του OSM στην ακρίβεια, πυκνότητα δεδομένων και σχετική περιγραφική πληροφορία.

Ως τώρα, από σχετική έρευνα για το OSM στο Ηνωμένο Βασίλειο (Haklay 2010, Basiouka 2009, Ather 2009), προέκυψε πολύτιμη πληροφορία για επιλεγμένες περιοχές. Όμως, οι μελέτες αυτές περιλαμβάνουν χειροκίνητες διαδικασίες και μεθόδους που δυσκολεύουν την επανάληψη της ποιοτικής αξιολόγησης σε διαφορετική και μεγαλύτερη περιοχή, ή στο μέλλον όταν το OSM ενημερωθεί περισσότερο. Επιπλέον, η χωρική ακρίβεια υπολογίζεται χρησιμοποιώντας την απλοποιημένη προσέγγιση της Μεθόδου Αυξανόμενης Ζώνης - Increasing Buffer Method (Goodchild and Hunter, 1997).

Τροποποιούμε και ενσωματώνουμε τη Μέθοδο Αυξανόμενης Ζώνης σε μία αυτοματοποιημένη μέθοδο που εκτελεί συσχέτιση δεδομένων, ώστε να αξιολογηθεί η πληρότητα και η χωρική ακρίβεια δεδομένων, λαμβάνοντας υπόψη την ετερογένεια των VGI (Volunteered Geographic Information) δεδομένων. Εφαρμόζουμε την προτεινόμενη μεθοδολογία στην ευρύτερη περιοχή Liverpool.