

AN OPEN SOURCE APPROACH FOR THE INTRINSIC ASSESSMENT OF THE TEMPORAL ACCURACY, UP-TO-DATENESS AND LINEAGE OF OPENSTREETMAP

M. Minghini*, M. A. Brovelli, F. Frassinelli

Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy -
(marco.minghini, maria.brovelli)@polimi.it, francesco.frassinelli@mail.polimi.it

Commission IV, WG IV/4

KEY WORDS: Data Quality, Lineage, Open Source, OpenStreetMap, Temporal Accuracy, Thematic Accuracy, Up-to-dateness

ABSTRACT:

OpenStreetMap (OSM) is the most popular crowdsourced geographic information project. The main factor that still limits the practical use of OSM is the lack of quality assurance. OSM quality assessment is thus a well-studied topic in literature, with most of the studies evaluating the quality by comparison against reference datasets. In contrast to these extrinsic approaches, OSM intrinsic assessment evaluates the quality by only analysing OSM itself. This study contributes to OSM intrinsic assessment by introducing an open source procedure to evaluate the temporal accuracy, up-to-dateness and lineage of OSM. Two workflows are presented: the first allows accessing the historical evolution of single OSM objects through an interactive web application, while the second aggregates and stores results on a user-defined grid to enable further GIS processing. The methodology is applied on the OSM nodes in the city of Dar es Salaam, Tanzania, by computing the following measures on the cells of an hexagonal grid: total number of nodes, average date of creation and last edit of nodes, average update frequency of nodes, average number of versions of nodes, average and total number of different contributors on nodes. Results highlight the mapping dynamics driven by the Dar Ramani Huria project, focused on increasing flood preparedness and resilience. When moving from the peripheral areas to the city centre, OSM is characterized by a progressively higher density of nodes, created earlier in time and updated by a higher number of contributors, which are all indexes of a general higher data quality.

1. INTRODUCTION

The OpenStreetMap (OSM) project was born in 2004 with the goal of creating and distributing free, crowdsourced geographic data for the whole world (<https://wiki.openstreetmap.org>). After 14 years of development, OSM has certainly become the most popular among crowdsourced geographic information projects (See et al., 2016). While the most easily accessible outputs of the project are cartographic products and maps (especially the main one available at <https://www.openstreetmap.org>), the core output is a geospatial database describing any physical object located on the ground. Featuring a wiki-based approach for map creation and update open to everyone (including non-experts in geography or geoinformatics) and leveraging on mass-market technologies such as GPS receivers and mobile devices, OSM has attracted an ever-growing community of contributors. At the time of writing (April 2018), the project counts more than 4.5 million registered users and 1 million active users, i.e. users who have made at least one edit to the map (<https://wiki.openstreetmap.org/wiki/Stats>).

One of the main factors, which justify the success of the project, is that the OSM database is available under the open access Open Database License (ODbL) (Open Data Commons, 2018), which allows freely using, modifying and building upon the database under non-restrictive conditions such as providing attribution to the OSM contributors. Thanks to the open license of the database and the relatively easy way to access it (see Section 3), OSM is currently used by a multitude of actors and for a huge variety of applications, including map visualization and production, web-based mapping, routing and navigation, update of governmental

maps, disaster management and other humanitarian applications. Mooney and Minghini (2017) provide a detailed overview of this ecosystem of services developed around OSM. The popularity of the project has attracted as well high attention from the academic and research community, with such a large body of literature already available that OSM could almost be considered a field of research in its own right (Jokar Arsanjani et al., 2015a).

Despite this widespread and growing exploitation of OSM, the main factor that prevents an even larger use of the database is the concern about its quality. In fact, while crowdsourced information has substantial advantages compared to authoritative datasets (e.g. richness of local information and fast update cycles), it suffers from a general lack of quality assurance (Goodchild and Li, 2012). This is the reason why quality assessment is among the most studied, yet still debated, topics of research around OSM. A number of studies in specific regions of the world have shown that OSM quality is comparable, if not better, than corresponding governmental datasets (e.g. Haklay, 2010; Brovelli et al., 2016a), although OSM quality may present very heterogeneous patterns across space (Girres and Touya, 2010). OSM quality assessment may concern one or more of the quality parameters available for geospatial datasets, e.g. positional accuracy, thematic accuracy, temporal accuracy, completeness, up-to-dateness and lineage. OSM quality assessment is defined as extrinsic, when it happens through the comparison against an external reference dataset, and intrinsic, when it happens through the sole analysis of OSM itself. The vast majority of literature studies focus on the OSM extrinsic quality assessment, which requires the availability of a reference dataset, provided e.g. by a National Mapping Agency (NMA),

* Corresponding author

and the use of Geographic Information Systems (GIS) algorithms to compare it with the corresponding OSM dataset. Relatively few studies focus instead on OSM intrinsic quality assessment, which requires analysing the temporal evolution of a specific OSM dataset.

This work contributes to the literature on OSM intrinsic quality assessment by analysing the quality parameters of temporal accuracy, up-to-dateness and lineage. This is achieved using two open source workflows, which process the temporal evolution of the OSM database and return two types of results: historical information on single OSM objects through an interactive web application, and aggregated information (e.g. on administrative boundaries or a grid) suitable to extract statistics. The remainder of the paper is structured as follows. Section 2 explores the OSM database in more detail by introducing the OSM data model, the OSM taxonomy and the OSM history. Section 3 revises available literature on OSM quality, with special focus on intrinsic quality assessment. Section 4 introduces the study area of Dar es Salaam, Tanzania, chosen to pay homage to the FOSS4G 2018 conference location where this paper is presented. In Section 5 the two open source software stacks to process the history of the OSM database are presented and applied on Dar es Salaam. Section 6 concludes the paper by discussing the main findings of the work and the directions for future research.

2. OPENSTREETMAP DATABASE

The objects stored in the OSM geospatial database comply to a vector data model, i.e. they are composed of a geometry and one or more attributes. From the geometrical perspective, objects can be nodes, ways or relations. Nodes are single points representing point-wise features (e.g. trees, benches and trash bins); ways are ordered lists of nodes representing both linear features (e.g. roads, rivers and walls) and polygon features (e.g. buildings, lakes and forests); relations are special structures representing a relationship between two or more nodes, ways and/or other relations (e.g. a bus route, composed of the route itself and all the bus stops) (<https://wiki.openstreetmap.org/wiki/Elements>). The attributes of OSM objects are called tags. Each object must have at least one tag, but there is no limit to the number of tags of an object. Each tag is composed of a key-value pair, where the key describes the object property and the value describes the object value for that key (<https://wiki.openstreetmap.org/wiki/Tags>). As an example, the current OSM tags for the Tour Eiffel in Paris include *man_made=tower* (describing the object type, where *man_made* is the key and *tower* the value) and *name=Tour Eiffel* (describing the object name, where *name* is the key and *Tour Eiffel* the value) (<https://www.openstreetmap.org/way/5013364>). The official list of OSM tags, which specifies the reference tag(s) to be used for any OSM object, is available on the Map Features wiki page (https://wiki.openstreetmap.org/wiki/Map_Features). This OSM taxonomy has been agreed upon over the years and is continuously subject to change, since it reflects a folksonomy approach to tagging based on a negotiation process between contributors (Ballatore and Mooney, 2015). Statistics about the use of tags and their combination with other tags are provided by the popular taginfo service (<https://taginfo.openstreetmap.org>).

OSM objects are created and updated through specific software packages named editors. There is currently a variety of editors, both for desktop and mobile devices, suitable for contributors with different skills and experience. An updated list is available on the wiki page <https://wiki.openstreetmap.org/wiki/Editors>. To add or edit an OSM object, contributors must be registered. Thus, all the edits are stored forever in the OSM database together with

the objects themselves, making it possible to access and study the temporal evolution of each object. The full OSM database, i.e. the set of all existing objects (geometries and tags) is available for download in a single file, weekly updated and named OSM Planet File (<https://planet.osm.org>). Conversely, the same set of all existing objects which also includes their history is named Full History Planet File, is updated weekly and can be downloaded from <http://planet.osm.org/planet/full-history>. In addition, OSM objects with their history can be also accessed through the OSM API (<https://wiki.openstreetmap.org/wiki/API>), which allows to both fetch and save raw data from/to the OSM database.

3. OPENSTREETMAP QUALITY ASSESSMENT

The need to identify methods and techniques to assess the quality of crowdsourced geographic information, including OSM, arose soon after the phenomenon of user-generated geospatial content entered the mainstream (Flanagin and Metzger, 2008). As already mentioned, most of the researchers employed a standard quality evaluation framework based on the comparison of OSM against authoritative datasets, under the assumption that these datasets - produced by trained professionals through rigorous procedures - represent the ground truth. Extrinsic OSM quality assessments mainly extracted quality measures in terms of positional accuracy (see e.g. Haklay, 2010; Girres and Touya, 2010; Koukoletsos et al., 2011; Fan et al., 2014; Brovelli et al., 2017a) and completeness (see e.g. Haklay, 2010; Girres and Touya, 2010; Koukoletsos et al., 2011; Hecht et al., 2013; Fan et al., 2014), followed by thematic and semantic accuracy (see e.g. Girres and Touya, 2010; Fan et al., 2014; Kalantari and La, 2015). The majority of studies assessed the quality of OSM road networks (see e.g. Graser et al., 2014; Brovelli et al., 2016a), while less attention was paid to other thematic layers such as buildings (see e.g. Hecht et al., 2013; Fan et al., 2014; Brovelli et al., 2016b), land use (see e.g. Jokar Arsanjani et al., 2015b; Fonte et al., 2017) and points of interest (see e.g. Girres and Touya, 2010; Jackson et al., 2013).

More recently, researchers have recognized that crowdsourced geographic information, and primarily OSM, is so different than authoritative information (in terms of the creation, management and update process) that it often makes little sense to limit quality assessment to their comparison. For example, in many parts of the world OSM is clearly more accurate, complete and up-to-date than authoritative datasets (Vandecasteele and Devillers, 2015). Also, many times authoritative datasets to be used for comparison are either not available, available under restrictive licenses or too expensive. For all these reasons, authors started to adopt intrinsic approaches to explore the OSM quality. According to Goodchild and Li (2012), three intrinsic methods may be applied to ensure OSM quality: crowdsourcing, i.e. the ability of a crowd to correct mistakes and maximize quality; social measures, based on the contributors' trustworthiness; and geographic measures, focused on checking the consistency of spatially close contributions. A number of quality parameters can be measured through this kind of intrinsic assessments: examples are logical consistency (see e.g. Vandecasteele and Devillers, 2015; Hashemi and Abbaspour, 2015) and temporal accuracy (see e.g. Girres and Touya, 2010; Jokar Arsanjani et al., 2013). Temporal analyses, through which OSM up-to-dateness and lineage can also be studied, are clearly based on the processing of the Full History Planet File or the use of the OSM API. Extracted information includes e.g. the number of users who have edited an object, the number of edits (versions) of the object, the dates of creation and last update of the object.

Antoniou and Skopeliti (2015) identified three families of quality indicators deriving from the intrinsic assessment of OSM, i.e.:

demographic indicators, which correlate quality with population density; socio-economic indicators, which are focused on socio-economic variables such as social deprivation, population age and income; and indicators based on contributors, which analyse the motivation, number, experience and mapping activity of OSM contributors. The authors provide representative references for each of these families. Intrinsic assessment allows to evaluate an additional quality parameter, named usability in the ISO 19113 standard (ISO 2013) and also referred to as fitness for use (Veregin, 1999) or fitness for purpose. This highlights that OSM quality can be also expressed in terms of OSM suitability for a specific end use or application. Barron et al. (2014) defined a framework including more than 25 methods and indicators to study OSM fitness for use based solely on data history. Recently, Antoniou and Skopeliti (2017) adopted an original approach to OSM intrinsic assessment by analysing the temporal evolution of OSM taxonomy (i.e. the mapping guidelines of the project) and editors, and discussing how this evolution can affect data quality.

Finally, from the purely technical perspective which is extremely relevant to the present work, some notable examples exist of web applications able to process OSM history and return real time, aggregated information useful for intrinsic quality assessment. OSMMatrix (<http://koenigstuhl.geog.uni-heidelberg.de/osmatrix>; Roick et al., 2011) is an open source geospatial web application developed by Heidelberg University and providing a grid-based visualization (organized into hexagonal cells) for a number of attributes related to OSM data and OSM contributors. Recently, Heidelberg University has announced the development of the Ohsome OSM history analytics platform (<https://ohsome.org>), which aims to make the full OSM history database more easily accessible for different data analytics tasks on a global scale. It is based on big data computing technology and is released as open source.

4. STUDY AREA: DAR ES SALAAM, TANZANIA

Dar es Salaam, the former capital of Tanzania, is currently the largest city in East Africa with more than 4 million inhabitants (https://wiki.openstreetmap.org/wiki/Dar_es_Salaam). It is one of the fastest growing cities in the world but is highly comprised of unplanned and informal settlements. Moreover, Dar es Salaam traditionally suffers from flooding, which every year during the rainy season causes many deaths and damages. After the severe flood of May 2015, the Dar Ramani Huria project was started to raise awareness and flood resilience through community-based OSM mapping (<http://ramanihuria.org>). Accurate and detailed maps, previously non existing, were produced for the most flood-prone areas of the city after training university students and local community members. In addition to being available for download and delivered in printed form to the local governing bodies, OSM maps are also exploited in InaSAFE (<http://inasafe.org>), an open source software allowing to run natural disaster scenarios for better planning and response. The Dar Ramani Huria project, still ongoing, is managed by the Tanzanian Commission for Science and Technology, with partners including the City Council of Dar es Salaam, Ardhi University, University of Dar es Salaam, Buni Innovation Hub and the Humanitarian OpenStreetMap Team (HOT), with support from the Red Cross, the Global Facility for Disaster Reduction and Recovery (GFDRR) and the World Bank. Thanks to this initiative, Dar es Salaam is currently one of the most densely OSM-mapped regions in the world.

The boundary of Dar es Salaam, to which the following analyses are limited, is extracted from the corresponding OSM relation (<https://www.openstreetmap.org/relation/7202037>, see Figure 1).

The analyses described in the following are performed on OSM nodes and ways (see Section 5). It must be said that the nodes considered are only those having at least one tag, i.e. the Points Of Interest (POIs); as an example, the corners of a building are not treated as nodes, unless they have associated tags. Also, nodes and ways deleted from OSM are not considered in the analyses. In addition, the edits to nodes and ways considered are only those where a tag was added, changed or deleted; therefore, changes in geometry (e.g. a node moved from a position to another) are not counted. Finally, all the changes made to nodes and ways within a single OSM changeset are counted as one single new version of the object. With these premises, on May 3, 2018 Dar es Salaam includes a total of 129572 nodes and 1156948 ways, on which edits (included object creation) were performed by a total of 1959 contributors. The total number of versions for the nodes and ways are 150716 and 1592221, respectively.

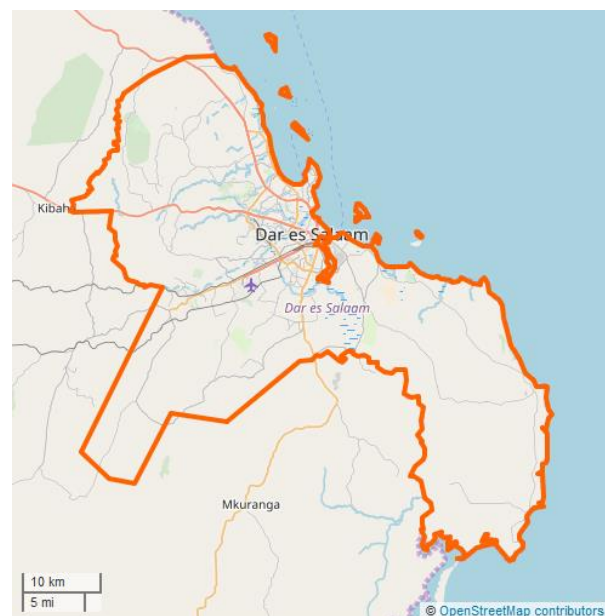


Figure 1. Study area: Dar es Salaam, Tanzania (map data © OpenStreetMap contributors).

5. ARCHITECTURE AND APPLICATION

The analysis of OSM history, which is the basis for OSM intrinsic quality assessment, is performed using two different open source software stacks (see Figure 2). They both allow to process the OSM history to extract measures of temporal accuracy, lineage and up-to-dateness, but they differ in terms of how these results are generated and provided to end users. As a matter of fact, in the first case results are returned in real time and for single OSM nodes and ways through an ad hoc web application named “Is OSM up-to-date?”, while in the second case a more extensive analysis allows to aggregate and store results in a database to be reused for further GIS processing. In the following, the two methods are separately presented and applied to the city of Dar es Salaam.

5.1 Web application

The web application named “Is OSM up-to-date?” is available at <https://is-osm-uptodate.frafra.eu>. It allows to explore the OSM database for any given area and returns real-time information for each OSM node and way available in that area according to a number of attributes: the date of creation (i.e. of first edit) of that node or way; the date of last edit of that node or way; the number

of versions (i.e. the total number of edits) of that node or way; the update frequency of that node or way; and the number of different contributors who have edited that node or way. A search bar allows to search and move the map to any location; users can then zoom and pan the map on their area of interest before fetching OSM data through a special button (see Figure 3). At this point, the available OSM nodes or ways are visualized with different colours according to a bar located in the upper right corner of the interface. From the upper side of the interface, users can switch between the five different attributes analysed and the colour classification of the nodes and ways changes accordingly. On the right side of the interface, two checkboxes allow to switch the map visualization from nodes to ways and a bar allows to reduce the colour of the base map in order to better highlight the visualization of the nodes or ways. Finally, when clicking on a node or way, a popup provides information on the first and last edit, the version, the number of different contributors and the OSM tags available (see Figure 3). A description of the “Is OSM up-to-date?” web application is also available on the wiki page https://wiki.openstreetmap.org/wiki/Is_OSM_up-to-date.

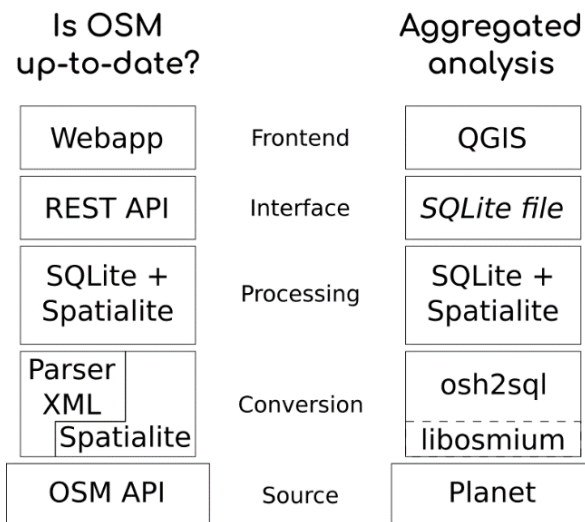


Figure 2. Open source software stacks for OSM intrinsic quality assessment: web application (left) and aggregated analysis (right).

From a more technical perspective, the web application makes use of the OSM API for retrieving the nodes and ways (including their history) for the bounding box specified by the user. Because of the many server requests that have to be executed, the Dugong Python library (<https://bitbucket.org/nikratio/python-dugong>) is used to take advantage of HTTP 1.1 pipelining and to reduce the latency. OSM data are retrieved in the OSM XML format (https://wiki.openstreetmap.org/wiki/OSM_XML) and converted using `spatialite_osm_raw` (an application bundled together with `spatialite-tools`, <https://www.gaia-gis.it/fossil/spatialite-tools>), while an XML parser extracts history-related information. Nodes and ways are then transformed into a GeoJSON file through a single SQL query, by taking advantage of the versatile JSON1 SQLite extension (<https://www.sqlite.org/json1.html>). The result is then served using the `hug` framework (<http://www.hug.rest>), which provides REST APIs and a command line interface, and `uWSGI` (<https://uwsgi-docs.readthedocs.io/en/latest>), which also serves static files such as the user interface. The frontend is based on the `Leaflet` (<http://leafletjs.com>) web mapping library, `jQuery` (<https://jquery.com>), `Bootstrap` (<https://getbootstrap.com>) and `Font Awesome` (<https://fontawesome.com>) for the interface, and `Nominatim` (<https://nominatim.openstreetmap.org>) as the OSM search engine. A `Docker` image (<https://www.docker.com>) is also provided for simple deployment, while the source code of the

web application is licensed under the open source `AGPL v3` (Free Software Foundation, 2007), and available on the repository at <https://github.com/frafra/is-osm-uptodate>.

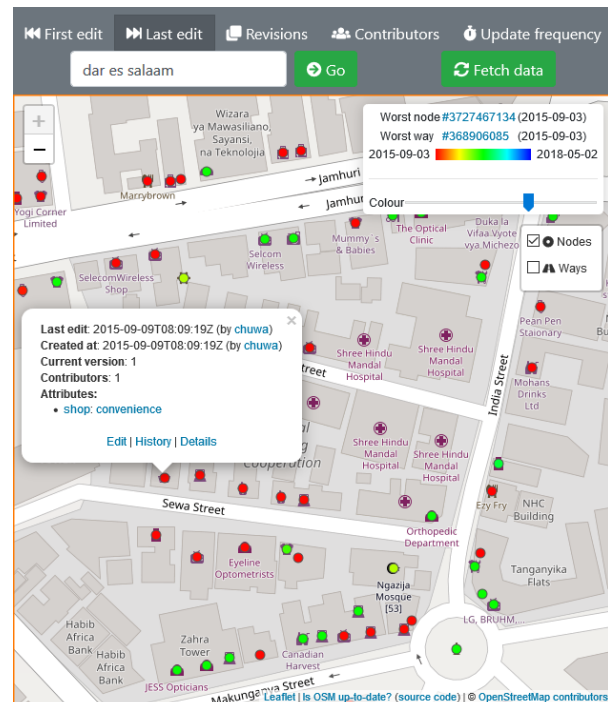


Figure 3. Screenshot of the “Is OSM up-to-date?” application showing a classification of nodes according to the date of last edit for an area in Dar es Salaam, Tanzania (map data © OpenStreetMap contributors).

5.2 Aggregated analysis

Conversely, the second software stack (see Figure 2) makes use of OSM data extracted from the Full History Planet File in the `PBF` format (https://wiki.openstreetmap.org/wiki/PBF_Format), which is more space-efficient than the standard XML format. The file is then converted into a more manageable `SQLite` database (<https://sqlite.org>), enabled with the `Spatialite` extension (<https://www.gaia-gis.it/gaia-sins>), through the use of a Python conversion script developed ad hoc and named `osh2sql` (<https://gist.github.com/frafra/56650708033b6dd5bb3906827d0e58cc>), which in turn is based on the `Osmium` library (<http://osmium.org/libosmium>). At this point, OSM nodes and ways can be associated with the corresponding administrative boundaries or with the cells of any user-defined grid, in order to perform aggregated statistical analyses on each region through common GIS software. In the following, this method is applied to extract intrinsic quality measures for the OSM objects in Dar es Salaam, using `QGIS` (<https://www.qgis.org>). For simplicity the analysis is limited to OSM nodes, while ways are discarded as they can belong to more than one administrative unit or grid cell, e.g. if these units or cells are particularly small or if the ways are particularly long or large (such as a very long road or a very large agricultural or forest area).

The study area of Dar es Salaam (see Figure 1) is mapped onto an hexagonal grid, which is suggested by Hecht et al. (2013) as it offers the advantage of closely approximating the circle while providing a complete coverage of the study area. The side of the hexagon (i.e. the radius of the circle, within which the hexagon is inscribed) is chosen equal to 250 m. This results into a total of 21588 grid cells, where each grid cell has an area of 0.162 km². Figure 4 shows the hexagonal grid classified according to the

number of OSM nodes included in each cell. First, it is clear that a large portion of Dar es Salaam (17052 cells, corresponding to the 79% of the total) does not contain any node, while the density of nodes progressively increases when moving from the rural to the most urbanized areas, particularly the city centre of Dar es Salaam. These are the most flood-prone areas addressed by the mapping activity of the Dar Ramani Huria project.

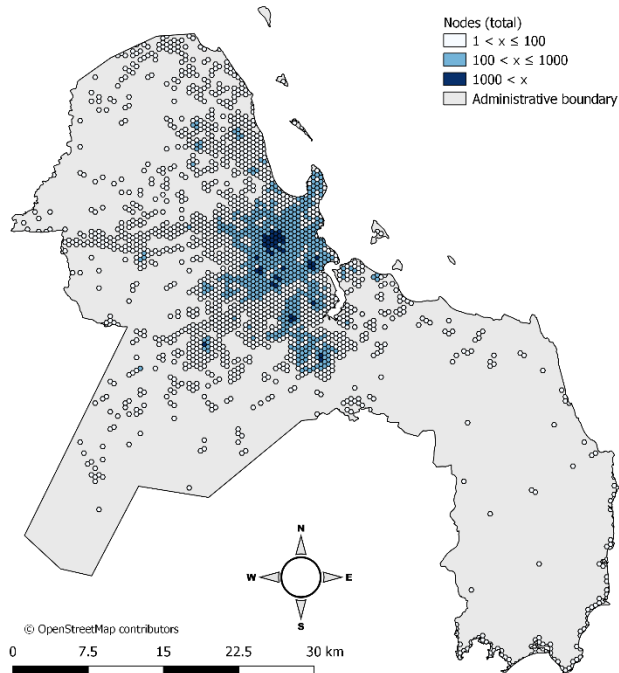


Figure 4. Dar es Salaam grid classified according to the number of OSM nodes (map data © OpenStreetMap contributors).

Figures 5 and 6 show the hexagonal grid classified according to the average date of creation and the average date of last edit of OSM nodes. It is evident that the vast majority of nodes in the city centre of Dar es Salaam were created in 2015, when the Dar Ramani Huria project was initiated. The mapping activity then gradually continued towards the city outskirts in 2016, 2017 and 2018, the latter clearly showing an increased mapping attention on the northernmost area as well as on more or less isolated areas on the west/south-west side of Dar es Salaam (see Figure 5). In particular, the set of cells departing along an almost horizontal line from the city center to the west side of the city correspond to the location of the towers carrying electricity cables (tag *power=tower*). The average date of last edit of nodes confirms the same trend. In fact, only a small portion of the nodes created before the end of 2015 was later updated, while mapping in 2018 focused almost exclusively on the peripheral areas (see Figure 6). To complete this temporal analysis, Figure 7 shows the average update frequency of OSM nodes. Most of the nodes created in 2018 were no longer updated, while the highest frequencies of update occur in specific areas of the city centre, including the one located south-west of the sea inlet.

Figure 8 provides insights on the data lineage by classifying the hexagonal grid according to the average number of versions of OSM nodes. As expected, the most recently created nodes in the city outskirts were in general not updated (average number of version exactly equal to 1), while the transition towards the city centre shows a progressive increase of the average version. For most of the cells this is equal to 2 or 3, while only in few cases it takes values higher than 5. The maximum average number of versions is equal to 19 and occurs for two cells in the city centre.

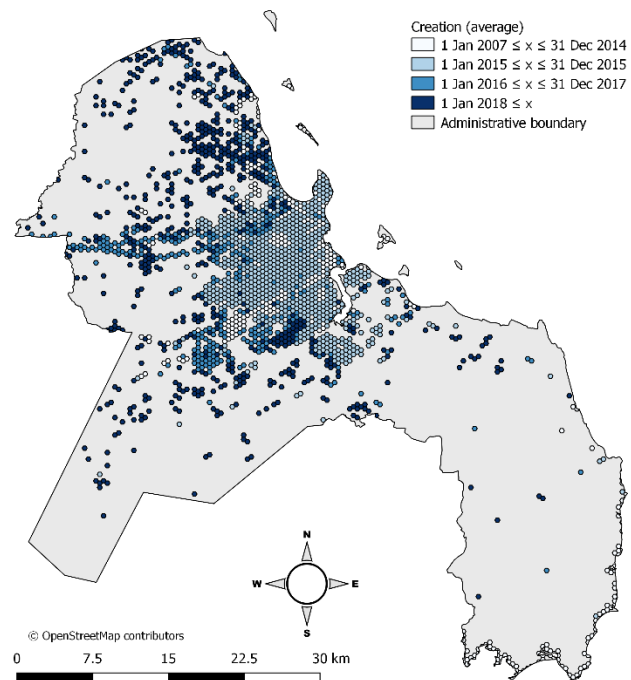


Figure 5. Dar es Salaam grid classified according to the average date of creation of OSM nodes (map data © OpenStreetMap contributors).

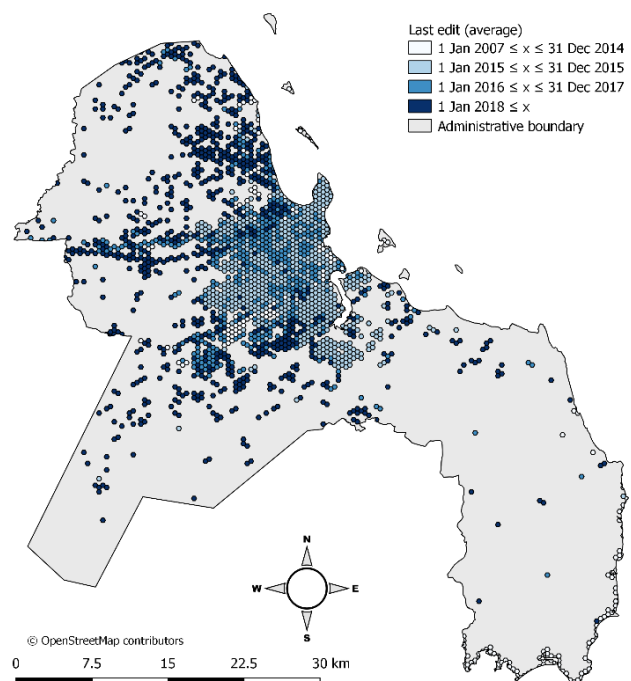


Figure 6. Dar es Salaam grid classified according to the average date of last edit of OSM nodes (map data © OpenStreetMap contributors).

Finally, Figures 9 and 10 show the hexagonal grid classified according to the average number and the total number of different contributors on OSM nodes, respectively. For 2390 cells (53% of the total number of cells having at least one node) the average number of different contributors is equal to 1, i.e. all the nodes in those cells were created and never updated, or they were all created and updated by the same contributor. Once again, these cells are mostly located in the outskirts of Dar es Salaam, while

a higher average number of different contributors occurs when moving towards the centre. In this case, the maximum value is equal to 17.3. The total number of different contributors, which is a measure of the “popularity” of the areas to be mapped, clearly reflects the progressive shift from the periphery of the city (where most of the cells have a value equal to 1) to the centre, where the values increase up to 64 (see Figure 10).

All the figures included in this section are generated from QGIS. The SQLite database table containing the aggregated data for Dar es Salaam (total number of nodes, average date of creation and last edit of OSM nodes, average update frequency of OSM nodes, average number of versions of OSM nodes, average and total number of different contributors on OSM nodes) is available for download at <https://frafra.eu/archive/osm/dar-es-salaam.zip>.

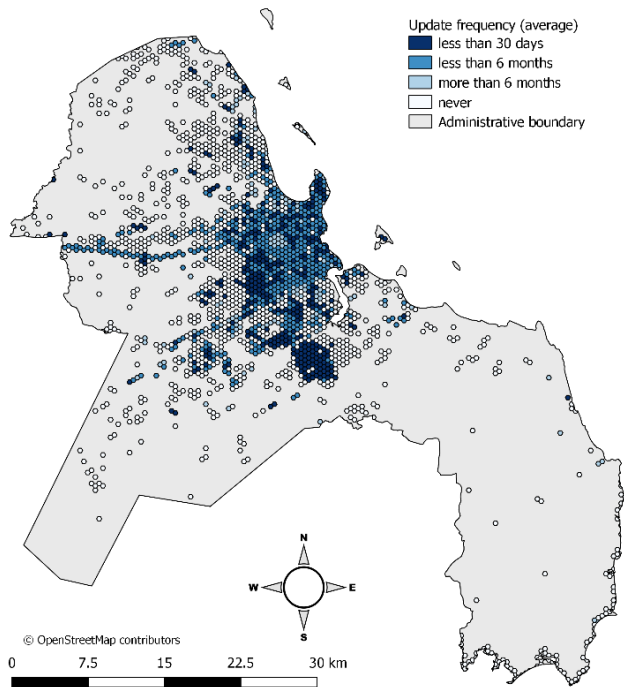


Figure 7. Dar es Salaam grid classified according to the average update frequency of OSM nodes (map data © OpenStreetMap contributors).

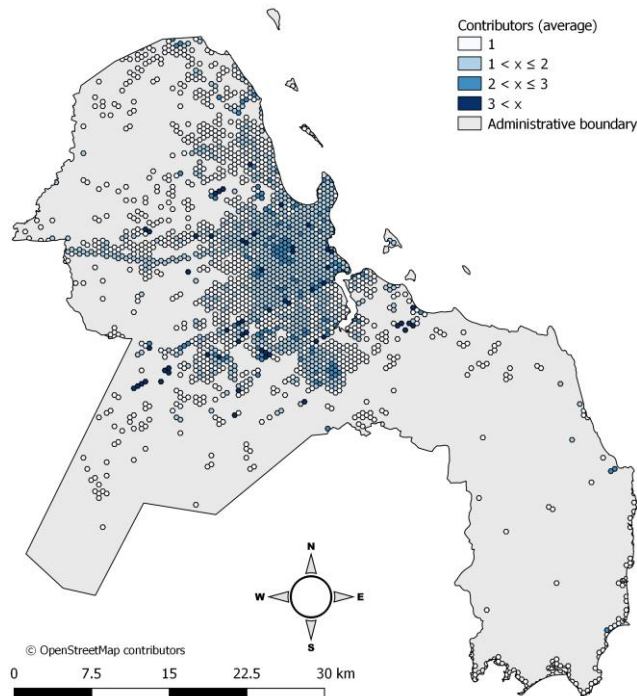


Figure 9. Dar es Salaam grid classified according to the average number of different contributors on OSM nodes (map data © OpenStreetMap contributors).

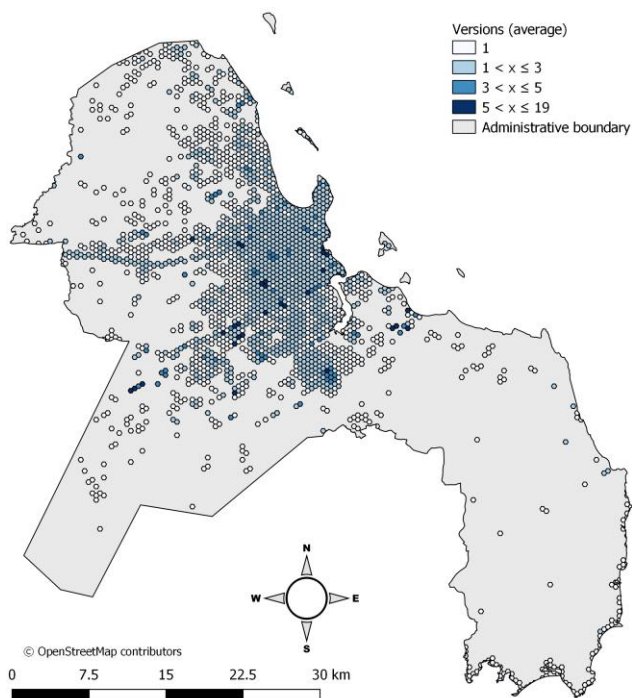


Figure 8. Dar es Salaam grid classified according to the average number of versions of OSM nodes (map data © OpenStreetMap contributors).

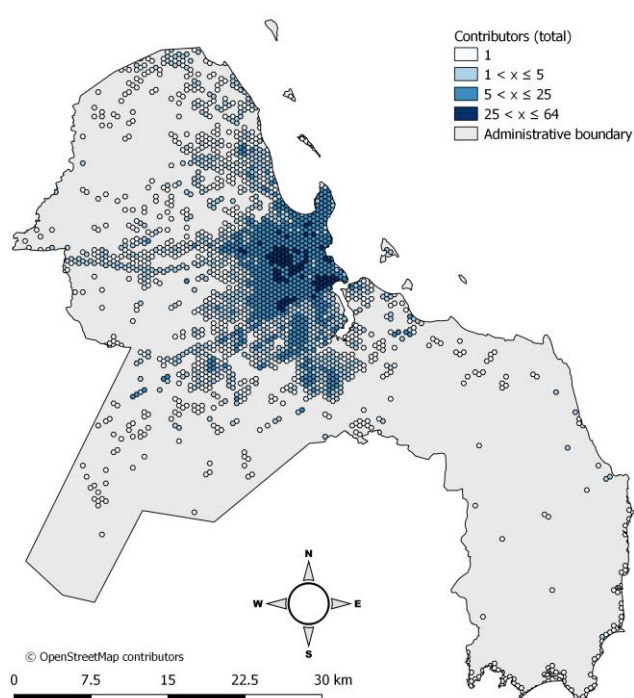


Figure 10. Dar es Salaam grid classified according to the total number of different contributors on OSM nodes (map data © OpenStreetMap contributors).

6. CONCLUSIONS

Crowdsourced geographic information has become a key player in the mainstream domain of geomatics and geoinformatics, and OSM is by far the most representative example in this regard. In spite of this, while OSM has gained a great deal of attention from the academic and research community, a systematic adoption and exploitation is still far from becoming a reality. Antoniou and Skopeliti (2015) suggest that the main reason for this is the lack of a stable and standardized way to evaluate OSM quality. In fact, traditional extrinsic quality assessments based on the comparison against reference datasets may not be sufficient to evaluate and understand such a complex data source like OSM. Therefore, this work has contributed to the more recent trend of exploring the OSM database intrinsically, i.e. by only looking at the database itself, to draw conclusions on OSM quality, or fitness for use.

The intrinsic quality assessment has focused on the city of Dar es Salaam, the venue of the FOSS4G 2018 international conference, and has studied the quality elements of temporal accuracy, up-to-dateness and lineage through both the use of the OSM API and the processing of the Full History Planet File. The study of OSM temporal evolution has shed light on the dynamics of the mapping process happened in Dar es Salaam, which was strongly driven by the Dar Ramani Huria project started in 2015 to increase flood preparedness and resilience. As a matter of fact, the training and educational activities supported by the project have formed an active community of OSM contributors which was almost absent before. The findings of the work, which are limited to OSM nodes but in all likelihood apply also to OSM ways and relations, show that the OSM development in Dar es Salaam is still at a young stage. The density of data (which is a proxy for its variety and detail) progressively increases from the periphery to the city centre. While in the periphery OSM data has been created by few contributors, recently, and shows few updates over time, the city centre (when mapping started earlier) is characterized by higher numbers of active contributors and higher frequencies of update, although in some cases large areas were mapped by only one or few different contributors. This is clearly an effect of the Dar Ramani Huria project, which focused its primary attention on mapping infrastructure networks (drainage systems, electricity networks, etc.) and commercial POIs in the most flood-prone areas within and close to the city centre. Thus, OSM fitness for use for the specific requirements of the Dar Ramani Huria project can be considered high, since OSM temporal accuracy, lineage and up-to-dateness are high as well. Conversely, a decreasing mapping attention is evident when moving towards the most rural areas, where only basic OSM objects such as building footprints, roads and the hydrographic network are usually present.

It must be recognized that, despite intrinsic quality approaches provide a new and innovative way to understand and describe OSM, they are similar to extrinsic approaches in that they offer a way to evaluate the quality of already-created data, where this data – being crowdsourced – substantially differs from traditional sources. By reversing this approach, another way to implicitly assess the quality of these crowdsourced datasets is before data is created, i.e. through the introduction of specific, project-based data collection protocols (Mooney et al., 2016; Minghini et al., 2017). Since the purpose of the Dar Ramani Huria project was exactly to train and form OSM contributors to achieve specific mapping results, it is expected that the data collection protocols developed for the specific needs of the project were themselves a guarantee of the quality of the data to be collected.

The work described in this paper might be extended in a number of directions. First, the intrinsic approach might be differentiated

according to specific categories of OSM nodes, e.g. addresses, commercial activities, natural elements and street furniture, which clearly have different characteristics in terms of temporal evolution. For example, an address is usually mapped once (with all the required tags) and is not later updated, while a commercial activity may be edited multiple times after creation in order to reflect changes or add complementary tags (e.g., for a restaurant, the cuisine, the opening hours, the presence of a delivery service, and the contact information). OSM intrinsic quality assessment might also benefit from an evaluation of the single contributors. In fact, it is clear that based on elements such as the number of OSM edits performed, the type of OSM objects mapped, etc., it is possible to estimate the experience of each contributor and thus the reliability of the contributed data. This kind of approach has been already proposed in literature (see e.g. Barron et al., 2014). Finally, intrinsic quality assessment might be also enriched by quantitatively evaluating demographic and territorial conditions, e.g. population density, land cover/land use and the altitude of the terrain, and their correlation with data quality parameters.

Lastly, the processing workflows presented in this study confirm how powerful, flexible and fully customizable are open source software solutions (including FOSS4G) to study such a complex and diversified database like OSM, which, as documented in the case study of Dar es Salaam, can shed light on a number of other environmental, political, social and societal dynamics (Brovelli et al., 2017b).

7. ACKNOWLEDGEMENTS

The authors would like to thank all the contributors of the OSM project, and particularly those active in Dar es Salaam.

8. REFERENCES

- Antoniou, V., and Skopeliti, A., 2015. Measures and indicators of VGI quality: An overview. In: *The ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, La Grande Motte, France, Vol. II-3/W5, pp. 345-351.
- Antoniou, V., and Skopeliti, A., 2017. The Impact of the Contribution Micro-environment on Data Quality: The Case of OSM. In: *Mapping and the Citizen Sensor*. Ubiquity Press, London, pp. 165-196.
- Ballatore, A., and Mooney, P., 2015. Conceptualising the geographic world: the dimensions of negotiation in crowdsourced cartography. *International Journal of Geographical Information Science*, 29(12), pp. 2310-2327.
- Barron, C., Neis, P., and Zipf, A., 2014. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), pp. 877-895.
- Brovelli, M.A., Minghini, M., and Molinari, M.E., 2016a. An automated GRASS-based procedure to assess the geometrical accuracy of the OpenStreetMap Paris road network. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Prague, Czech Republic, Vol. XLI-B7, pp. 919-925.
- Brovelli, M.A., Minghini, M., Molinari, M.E., and Zamboni, G., 2016b. Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: the method and a case study. In: *The International Archives of*

- Photogrammetry, Remote Sensing and Spatial Information Sciences*, Prague, Czech Republic, Vol. XLI-B2, pp. 615-620.
- Brovelli, M.A., Minghini, M., Molinari, M., and Mooney, P., 2017a. Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Transactions in GIS*, 21(2), pp. 191-206.
- Brovelli, M.A., Minghini, M., Moreno-Sanchez, R., and Oliveira, R., 2017b. Free and open source software for geospatial applications (FOSS4G) to support Future Earth. *International Journal of Digital Earth*, 10(4), pp. 386-404.
- Fan, H., Zipf, A., Fu, Q., and Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4), pp. 700-719.
- Flanagin, A., and Metzger, M., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72(3), pp.137-148.
- Fonte, C.C., Minghini, M., Patriarca, J., Antoniou, V., See, L., and Skopeliti, A., 2017. Generating up-to-date and detailed land use and land cover maps using OpenStreetMap and GlobeLand30. *ISPRS International Journal of Geo-Information*, 6(4), 125.
- Free Software Foundation, 2007. GNU Affero General Public License <https://www.gnu.org/licenses/agpl-3.0.en.html> (30 April 2018).
- Girres, J.F., and Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), pp. 435-459.
- Goodchild, M.F., and Li, L., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, pp. 110-120.
- Graser, A., Straub, M., and Dragaschnig, M., 2014. Towards an open source analysis toolbox for street network comparison: Indicators, tools and results of a comparison of OSM and the official Austrian reference graph. *Transactions in GIS*, 18(4), pp. 510-526.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37(4), pp. 682-703.
- Hashemi, P., and Abbaspour, R.A., 2015. Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept. In: *OpenStreetMap in GIScience*. Springer, Cham, pp. 19-36.
- Hecht, R., Kunze, C., and Hahmann, S., 2013. Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS International Journal of Geo-Information*, 2(4), pp. 1066-1091.
- International Organization for Standardization, 2013. *ISO 19157:2013 Geographic information - Data quality*, Geneva: ISO.
- Jackson, S.P., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., and Stefanidis, A., 2013. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2(2), pp. 507-530.
- Jokar Arsanjani, J., Barron, C., Bakillah, M., and Helbich, M., 2013. Assessing the Quality of OpenStreetMap Contributors together with their Contributions. In: *16th AGILE international conference of Geographic Information Science*, Leuven, Belgium, pp. 14-17.
- Jokar Arsanjani, J., Zipf, A., Mooney, P., and Helbich, M., 2015a. An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. In: *OpenStreetMap in GIScience*. Springer, Cham, pp. 1-15.
- Jokar Arsanjani, J., Mooney, P., Zipf, A., and Schauss, A., 2015b. Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In: *OpenStreetMap in GIScience*. Springer, Cham, pp. 37-58.
- Kalantari, M., and La, V., 2015. Assessing OpenStreetMap as an Open Property Map. In: *OpenStreetMap in GIScience*. Springer, Cham, pp. 255-272.
- Koukoletsos, T., Haklay, M., and Ellul, C., 2011. An automated method to assess Data Completeness and Positional Accuracy of OpenStreetMap. *GeoComputation*, 3, pp. 236-241.
- Minghini, M., Antoniou, V., Fonte, C.C., Estima, J., Olteanu-Raimond, A.-M., See, L., Laakso, M., Skopeliti, A., Mooney, P., Jokar Arsanjani, J., and Lupia, F., 2017. The relevance of protocols for VGI collection. In: *Mapping and the Citizen Sensor*. Ubiquity Press, London, pp. 223-247.
- Mooney, P., Minghini, M., Laakso, M., Antoniou, V., Olteanu-Raimond, A.M., and Skopeliti, A., 2016. Towards a Protocol for the Collection of VGI Vector Data. *ISPRS International Journal of Geo-Information*, 5(11), 217.
- Mooney, P., and Minghini, M., 2017. A review of OpenStreetMap data. In: *Mapping and the Citizen Sensor*. Ubiquity Press, London, pp. 37-59.
- Open Data Commons, 2018. Open Database License (ODbL) v1.0. <https://opendatacommons.org/licenses/odbl/1.0> (30 Apr. 2018).
- Roick, O., Hagenauer, J., and Zipf, A., 2011. OSMMatrix – grid-based analysis and visualization of OpenStreetMap. In: *Proceedings of the first European State of the Map Conference*, Vienna, Austria, pp. 44-54.
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.Y., Milčinski, G., Nikšič, M., Painho, M., Pödör, A., Olteanu-Raimond, A.-M., and Rutzinger, M., 2016. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, 5(5), 55.
- Vandecasteele, A., and Devillers, R., 2015. Improving volunteered geographic information quality using a tag recommender system: The case of OpenStreetMap. In: *OpenStreetMap in GIScience*. Springer, Cham, pp. 59-80.
- Veregin, H., 1999. Data quality parameters. In: *Geographical Information Systems: Principles and Technical Issues*. John Wiley and Sons, New York, pp. 177-89.