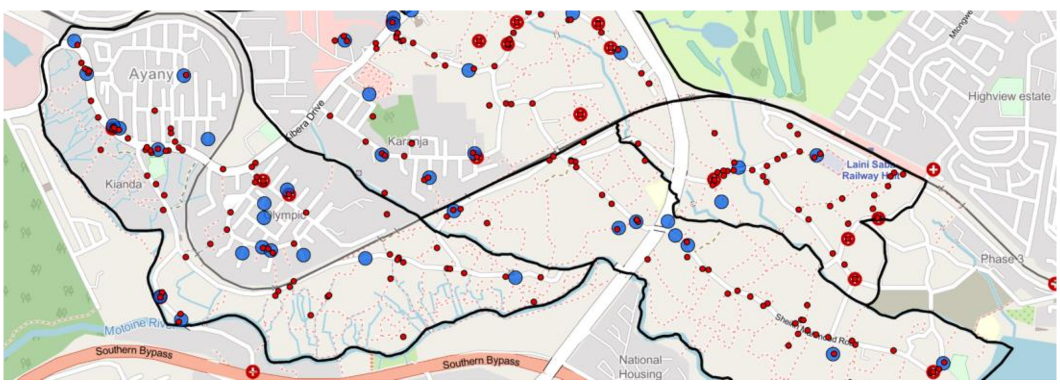
GIMA
Geographical Information Management and Applications

OpenStreetMap data assessment to support health services in the Global South

The cases of Wajir and Nairobi counties, KE

MSc Thesis



Universiteit Utrecht
TU Delft
UNIVERSITY OF TWENTE.
WAGENINGEN UR

Author: Eleftherios Kaltsas (8580138)

Supervisor: dr.ir. R.L.G. (Rob) Lemmens

Nairobi, Kenya | February 2023

Acknowledgments

My relocation from Berlin to Nairobi coincided with the start of this research. This exciting and magnificent journey brought some difficulties to my work that had to do with the time needed for adaptation and the building up of a network of people for support of any kind. Outside of my safety zone, I had to cope with the challenges faced by my two wonderful kids and my precious partner. It is her that I would like to thank first of all for her support and patience and love during my endless nights and mornings of working on this research.

Knowing nobody in Kenya in the beginning, I had to knock on several doors asking for data and information and some dialogue without which research seems sterilized. Boneya Hindada, a GIS officer at the Kenyan Red Cross headquarters, warmly welcomed me and was there whenever I needed him, to share his knowledge and thoughts and I feel lucky and grateful for getting to know him. I would also like to thank Kevin Kamau from the Kenyan Roads Board for his support while searching for data. I would also like to thank my dear friend Christos Koutsouradis from Berlin for sharing his knowledge, ideas, and programming skills.

I would like also to especially thank the head of the Chair of Algorithms and Data Structures of the University of Freiburg Dr. Hannah Bast for sparing her time to answer all my questions while working with SPARQL queries in Qlever, offering at the same time valuable suggestions.

Last but not least I would like to thank my supervisor, dr. ir. R.L.G. (Rob) Lemmens for discreetly guiding me through this process and helping me not to lose my focus but at the same time letting me free to follow my goals and my research instincts.

Abstract

The present research was an attempt to evaluate OpenStreetMap's capacity to support health services in the Global South. A user-oriented workflow was built so that the assessment of data is not generic anymore as oftentimes in the literature. Instead, a fitness-of-use framework was employed for a more precise assessment. This framework was applied in two use cases, the Counties of Wajir and Nairobi in Kenya. The questions that had to be answered concerned first of all the needs of the health workers and the capacity of the OSM vocabulary in the areas of the use cases to satisfy them. Secondly, the research tried to discover if semantic web technologies can enhance that capacity by unifying a fragmented vocabulary and even help build an ontology to promote it. Then the possibility of a proxy for the quality of OSM data through their correlation with humanitarian mapping projects was investigated. Finally, the basic setup of a web application that supports health workers and individuals in the Global South using OSM data was sought after. The use cases chosen were quite different in socioeconomic characteristics. For Nairobi County, it was discovered that the OSM vocabulary has the capacity to reflect the needs of health workers while Wajir is far poorer. The use of triplified OSM data and the SPARQL queries allowed sophisticated searches for objects and attributes that then were easily translated into ontologies. Through those ontologies, it became possible to query for information that was not tagged in OSM. Moreover, it was shown that the existence of completed humanitarian mapping projects in an area has a positive linear correlation with the completeness of health facility objects and some of their attributes. In conclusion, the present research provided a basic setup of a web application that supports health services with the use of OSM data.

Table of Contents

Acknowledgments.. 2

Abstract.. 3

1 Introduction.. 6

1.1 Basic concepts.. 6

1.1.1 Volunteered Geographic Information.. 6

1.1.2 OpenStreetMap.. 7

1.1.3 Fitness-of-Use.. 9

1.2 OpenStreetMap for Humanitarian Use.. 9

1.3 Problem Statement.. 10

1.4 Use Cases of Wajir and Nairobi Counties, KE.. 11

1.5 Research Objectives.. 12

2 Theoretical Background & Related Work.. 15

2.1 OSM Vocabularies.. 15

2.2 Semantic Web Technology for OSM.. 15

2.3 OSM data quality.. 16

2.3.1 Measures of OSM data quality.. 17

2.3.2 Indicators of OSM data quality.. 19

3 Methodology & Tools.. 21

3.1 Users' Needs Analysis.. 21

3.1.1 The Questionnaire.. 21

3.1.2 The interviews.. 23

3.1.3 Kenyan Red Cross report analysis.. 24

3.1.4 Challenges of the User's needs analysis.. 25

3.1.5 Results of user's needs analysis.. 25

3.2 Vocabulary analysis.. 26

3.2.1 Nairobi County.. 26

3.2.2 Wajir.. 28

3.3 Translating users' needs into OSM tags.. 28

3.4 OSM2RDF & Qlever SPARQL Engine.. 29

3.4.1 Building Queries in SPARQL.. 30

3.4.2 Validation of OSM2RDF 30
3.5 Metrics for the OSM data assessment 31
3.5.1 Completeness 31
3.5.2 Accuracy 31
4 Data identification and acquisition 32
4.1 Authoritative data 32
4.1.1 Health facilities 32
4.1.2 WATSAN 32
4.1.3 Road Networks 32
4.2 OSM data 33
4.2.1 Health facilities 33
4.2.2 WATSAN 33
4.2.3 Road Networks 33
4.3 Administrative boundaries 33
5 Analysis & Results 34
5.1 Quality assessment of OSM data 34
5.1.1 Nairobi County Health Facilities 34
5.1.2 Nairobi County WATSAN 49
5.1.3 Nairobi County Road Network 53
5.1.4 Wajir County Health Facilities 58
5.1.5 Wajir County WATSAN 64
5.1.6 Wajir County Road Network 66
5.2 Building an OSM ontology for the Kenyan Health facility classification system 71
5.3 Humanitarian OSM Mapping projects as a proxy for assessing OSM health facility data quality 80
6 Towards an OSM-based web application for health services 85
7 Discussion 89
8 Conclusions 94
Reference List 97
Appendix A 101
Appendix B 117
Appendix C 118
Appendix D 123

1 Introduction

In the context of global climate change, all parts of the world are affected to a larger or smaller extent by unprecedented weather phenomena that can lead amongst others to natural disasters, food and water shortages, and population displacements. In such an environment, where the need for emergency responses is growing bigger, every type of resource is valuable and indispensable. The use of geo-referenced data in humanitarian response to emergencies can be crucial, especially in places where authoritative emergency mechanisms are underdeveloped. Time-saving through immediate and accurate tracking of information can be translated into resource-use optimization. Geo-referenced data account for a big proportion of all data available on the web but retrieving all the relevant information for every case can be tedious. Looking into an ocean of data and then analyzing and visualizing them is mainly a job for geo-professionals.

As a result, humanitarian missions in the Global South are more often than never left with mainstream all-purpose map applications in which the completeness and accuracy of data vary depending on the purposes of each commercial provider or the level of participation in open-source projects. Google, Apple, and Bing (Microsoft) are leading the way for commercial internet global map providers while OpenStreetMap (OSM) and many applications based on it represent the open-source counterpart. As OSM is growing bigger in terms of data completeness, contributing community, and credibility, Bing, the cartographic division of Microsoft is investigating collaboration in various ways.[1] Furthermore, Apple is using OSM data for its map products while contributing through its Apple Data Team (ADT). ADT is responsible for 2.701.062 changesets (a group of edits to the database by a single user over a short period of time) in OSM until the time of writing (21.11.22). It is evident through the examples of Microsoft and Apple that OSM is not just a crowdsourced platform of geographic information but a geodata platform that is credible enough for the two IT giants to utilize and invest in it.

Maps and geo-data on a national level may also be available through their respective geoportals but this is rarely the case in the Global South. This gap left by the slowly evolving mapping agencies is organically being covered by local mapping communities and individuals through OpenStreetMap which is based on crowdsourcing. It has been argued (Jokar Arsanjani et al., 2015a) that OSM can play the role of a Spatial Data Infrastructure (SDI) where the latter is nonexistent or primitive. The basic difference between them is that while for SDIs the standardization of data and metadata is inherent, in crowdsourced projects and particularly in OSM, it is not; at least for metadata. Tagging from the side of the contributors plays the role of metadata but the very nature of OSM and the degrees of freedom offered by default to the contributors in tagging the attributes cannot guarantee homogeneity in their production.

1.1 Basic concepts

1.1.1 Volunteered Geographic Information

The development of Web 2.0 allowed the massive and wide collaboration of experts and amateur cartographers to map the entire world leading to what was called the wikification of GIS (Sui, 2008). The term that prevailed though in describing this phenomenon was Volunteered Geographic Information (VGI) coined by Goodchild (2007). Until then the only spatial data sources to be reckoned with were the National Mapping Agencies. Their credibility is unquestionable due to the strict standardization and their monopolistic nature. Their main disadvantages are the high production costs which lead to long intervals

[1] https://wiki.openstreetmap.org/wiki/Bing_Maps

between updates and sometimes their restrictive licenses (Antoniou & Skopeliti, 2015). Authoritative data is assumed to be produced by professionals, a fact that gives a strong added value to the final product. This could not be the case in VGI projects designed by default for open participation. Although highly experienced contributors are expected to participate in VGI schemes, inexperienced contributors take also part in the data production making the datasets susceptible to inaccuracies and other kinds of mistakes and faulty edits. Like all online communities, VGIs follow the 90-9-1 rule (Anderson et al., 2019) which means that 90 percent of the users are just data consumers, 9 percent are occasional contributors and only 1 percent is producing most of the data. This could be interpreted as the majority of edits being done by experienced users but still, the uncertainty of data quality remains.

1.1.2 OpenStreetMap

The most prominent example of VGI is the OpenStreetMap (OSM) project aiming at creating a free geodatabase of the entire world. The main advantages of OSM are the free, open-licensed data, their constant updating, and the potential of volunteered mapping for non-commercial or difficult-to-reach areas that otherwise wouldn't be mapped at all (Sehra et al., 2014). This very nature of OSM, regarding the openness in editing, is a blessing and a curse at the same time. While the openness of participation allows potentially maximum geographical coverage and data completeness, it may cause at the same time great inconsistency in annotation or a low level of positional and topographical accuracy. According to Quien et al. (2009), the work of non-professionals has a negative impact that cannot be ignored when assessing the overall quality of the data. But when it comes to the costs involved, especially in the Global South, OSM is in many cases the only realistic solution for geoinformation-related tasks.

Data Model

The OpenStreetMap data model is based on two geometry objects i.e., "Node" & "Way" and one object to define the relations between them or abstract features i.e., "Relation" (Figure 1).

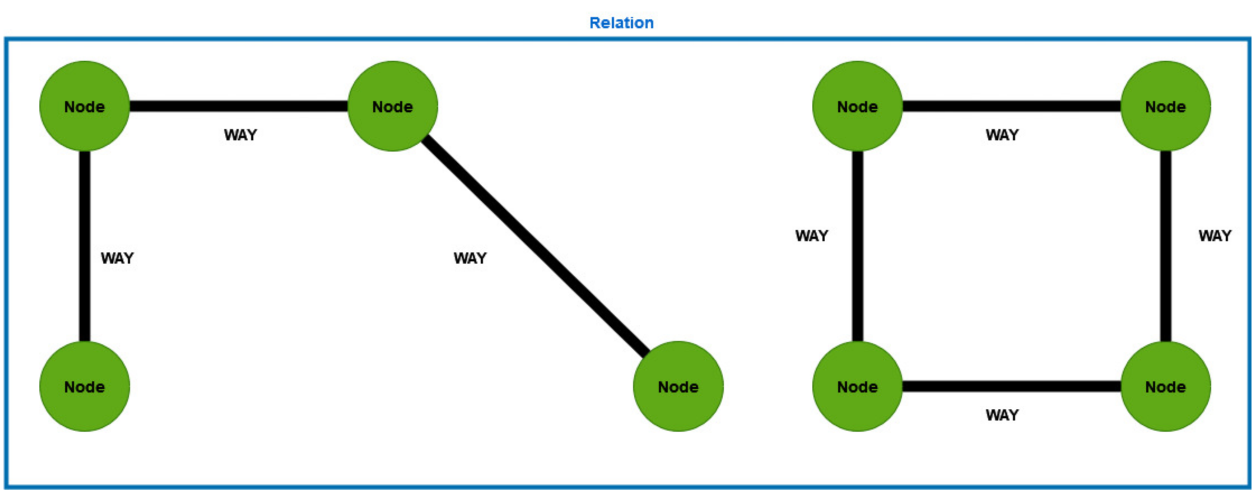


Figure 1. OpenStreetMap data model

A Node represents a pair of coordinates while a Way is a sum of finite nodes with uniform properties in a specific order which is considered an area when the first and last nodes have the same coordinates. Every object of OSM apart from its coordinates is accompanied by a unique ID, time and date of editing, the id of the contributor, and various attributes such as land use, name, etc. called "Tags". OSM's tagging schema is community-driven and it is crucial for the project. Attributes are recorded with the help of key=value pairs (e.g. amenity=universityh; name=ITC - Faculty of Geo-Information Science and Earth Observation) that any user can propose (Haklay & Weber, 2008). OpenStreetMap is open data, licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF). Users can copy, distribute, transmit, and change the data as long as OSM and its contributors are properly credited (Neis and Zielstra, 2014). Until now (November 2022) 9.5 million registered users have contributed more than 8 billion nodes and more than 900 million ways.

Editing Data

To be able to edit (add, modify, delete) data in OSM the user has to be registered first. From then on there are no restrictions. Users can contribute data to OSM in one of the following three ways. Collect and upload GPS data, upload bulk authoritative data in GIS formats like ".shp" from authoritative sources, or trace out lines and polygons from aerial imagery using the OSM editors (Mooney et al., 2010). The editing can be performed by using a variety of editors. Most popular and complete for desktop computers are the iD editor (web-based), JOSM (Windows, macOS, Linux), and Potlach 2 (Windows, macOS)[2]. There are multiple editors available for mobile devices, with Vespucci being the first and only mobile full editor working on Android. According to Behrens et al. (2015), one of the most important barriers in transforming OSM data contribution from a small group pioneer effort to massive mapping participation is the usability of the editing tools. The iD editor has been designed to be used by beginners with no GI background and is the default online editor of OSM since 2013. By trading speed with simplicity and learnability,y the iD editor seems to serve its main goal (Behrens, 2015). On the contrary, the Java OpenStreetMap editor is built for more experienced users. It is a suite that resembles, in terms of interface, a common GIS software, and among other features, it can be used offline and support custom plug-ins (Haklay & Weber, 2008). Potlatch 3 is the newest version of Potlatch 1-2. Both 1 and 2 used to be the standard editors in OSM before iD. Potlatch 3 is now a desktop application aimed at moderately familiar users.[3] Vespucci is an editor for Android that allows doing any kind of editing directly on a mobile device, with full access to data. It makes it possible to edit directly while surveying, without the need to make notes.[4] The statistics provided by OSM for editors' usage show that iD is by far the most popular (77.9%) while most of the edits have been done with JOSM (59.6%) and iD (31.8). [5]

Data Quality

The distance that must be covered for OSM to become a reliable tool for professionals of any domain, regardless of the geographic coordinates of a project's bounding box, is not short. Attempts to assess the OSM data quality started as soon as 2007 (Haklay, 2010a), only three years after the official launch of OSM. They have mainly focused on road networks because the initial focus of OSM was indeed the road networks. The results showed that in many places, OSM data surpass the corresponding authoritative data (Mooney & Corcoran, 2012b). But this usually refers to the urban centers of countries

---

[2] https://wiki.openstreetmap.org/wiki/Editors Accessed 06.05.22
[3] https://wiki.openstreetmap.org/wiki/Potlatch
[4] https://wiki.openstreetmap.org/wiki/Vespucci
[5] https://wiki.openstreetmap.org/wiki/Editor_usage_stats#Tables_and_figures Accessed 15.05.22

with high development indexes. Statistics about OSM world coverage reveal that the Global South is relatively unmapped compared to Europe and North America. While hosting 46% of the world's population, regions with low and medium levels of economic development only have 28% of the buildings and 16% of the roads mapped in OSM (Herfort et al., 2021). These percentages are global, in the statistical sense, so it is expected to have areas with a high level of data completeness and accuracy as a result of organized mapping projects and other areas that are almost unmapped. This heterogenicity makes assessment more complicated in terms of scale. Another form of heterogenicity lies also in the annotation of data. Mappers and contributors may use a variety of vocabulary to describe the same features. This way the existing information cannot be retrieved easily.

1.1.3 Fitness-of-Use

When examining and assessing spatial data, internal and external aspects must be combined. Scholars have developed several methods for evaluating different aspects of internal OSM data quality that fall under the typology of Completeness, Positional Accuracy, Attribute Accuracy, Logical Consistency, Semantic Accuracy, and Temporal Quality (Zhang & Malczewski 2017). Internal quality assessment is reflecting the specifications of data generation (Jokar Arsanjani et al., 2015b) and gives back absolute values based mostly on the comparison between a reference and a test dataset. Beyond the absolute values of those metrics though, the actual quality of OSM data should also be considered. By actual, it is meant the quality which is user-oriented and is thus external. This type of external quality addresses the idea of Fitness-of- Use (FoU) and represents how suitable a dataset is for a certain purpose (Jokar Arsanjani et al., 2015b). If for example, in a given area, 100% of the buildings and roads are mapped along with all their attributes and with high positional and topological accuracy, the dataset could be considered perfect for navigation, but if the freshwater sources are not present on the map, then it might be insufficient for some other purpose. The concept of Fitness-of-Use in OSM data is evident in attempts like the humanitarian orientation of recent OSM-based projects.

1.2 OpenStreetMap for Humanitarian Use

In the last twelve years, several efforts have been made to utilize OSM data for humanitarian causes in the Global South. The response of the local OSM community to the Haiti earthquake of 2010 marked a new era in volunteering mapping and was followed by similar actions (Figure 2). Since then, more than a dozen humanitarian projects have been realized based on the OSM worldwide, not including the numerous projects of the Humanitarian OpenStreetMap Team (HOT) which has even developed its editing tool and training programs. All these projects form a separate domain inside OSM which needs its framework of assessment. Generic assessment metrics are valuable but a specific framework into which only relevant data are evaluated can contribute to timesaving in emergencies as well as in resource management. The special characteristic of humanitarian mapping, which also led OSM to include a separate layer called “Humanitarian” in its interface, is that it is mostly led by local mapping communities with a high level of area knowledge and mapping experience. The mapping projects of these communities are usually subject-specific, and their produced data should have perfect Fitness-of-Use by default. An important part of humanitarian mapping is the health services domain. Improving the access to healthcare for the large populations of the Global South could be assisted by the use of the appropriate geo-datasets. The use of OSM in this direction diminishes the costs for poor countries with meager means but at the same time, an integrated and subject-specific assessment framework is missing. Such an assessment can indicate areas where the health service data are suitable for use and others that have to be gone under mapping projects to meet usability requirements.

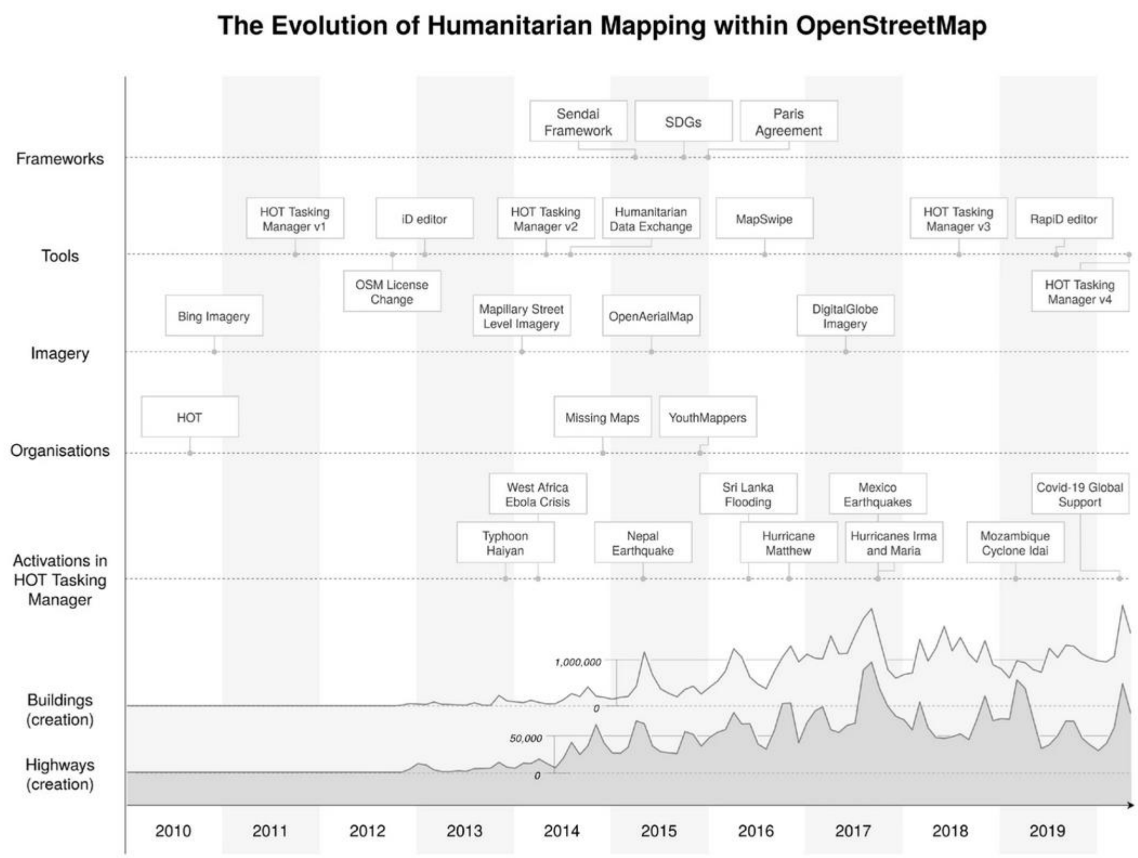


Figure 2. Evolution (Buildings and Highways added) of humanitarian mapping in OSM regarding major disaster activations, the socio-technical development of the community, and global political frameworks. Adapted from Herfort, B., Lautenbach, S., Porto de Albuquerque, J. *et al.* The evolution of humanitarian mapping within the OpenStreetMap community. *Sci Rep* **11**, 3037 (2021), p 2.

1.3 Problem Statement

The very nature of OpenStreetMap, the bottom-up collection and administration of data, gives the maximum degrees of freedom to the contributor to express their perspective of the world regardless of their mapping competence. The different levels of experience, local knowledge, and level of interest in the cause of the community result in different quality or amount of effort which translates into different quality of data produced. If we also take into account the inherent characteristics of each mapper that have to do with their sense of perfection, then we understand that spatial data in OSM can be marked by great heterogenicity. This heterogenicity can be reflected in positional and topological inaccuracy or incompleteness depending on the area addressed. Moreover, it can be strongly reflected in the annotation of the mapped objects. The absence of any top-down standardization or any imposed ontology in the tagging system of OSM makes it impossible to achieve a high level of consistency within the vocabularies used to describe map features and their attributes. For potential end users of OSM like field health workers, this inconsistency can result in an inability to find the information needed. The absence of a way or a tool to query in combination for similar attributes that are annotated following the different personal tastes of every contributor makes OSM less effective than it would be if all information hidden in tags were utilized.

The quality assessment, especially when we are talking about attribute completeness makes sense particularly when we refer to specific domains. So far, the available OSM quality assessment methods are dealing with attribute completeness (and all other parameters of quality) for the total of the data in an AOI. This approach can be highly misleading when in the AOI, domain-specific mapping events or projects have already been implemented. In that case, the overall attribute completeness may be low while the attribute completeness of specific features can be indeed high. Assessing OSM data, not for their overall quality but for their capacity to serve an end user of a specific domain is a concept that has been overlooked by scholars. Domain-specific tools can make it possible for end users to engage in the evaluation and editing process of the data that they are using or make semantic data suitable for use by applications.

1.4 Use Cases of Wajir and Nairobi Counties, KE

Currently, a project of HOT (Humanitarian OpenStreetMap Team) in collaboration with the Kenyan Red Cross (KRC) is using OSM to support resilience of vulnerable communities and investigate the capacity of the healthcare system, in Wajir County of Kenya (Figure 3, Left) while updating and enriching the relevant data to build resilience to disasters. The assessment framework for the Fitness-of-Use is missing there, as stated by the GIS department of the KRC. This gap provides a test field for the present proposed research. Moreover, the various OSM mapping projects implemented in the county of Nairobi (Figure 2, Right), especially the health facility mapping by HOT or the Map Kibera Project (Including the slums of Kibera, Mathare, Mukuru Kwa Njenga, Kawangware, Kangemi, and Korogocho) provide also a suitable testing field that has both strong and weak points. On the one hand, we can assume that the contributors on the field are experienced (or under training and guidance) meaning that more accurate and complete data is expected. On the other hand, heterogenicity is present as a result of the limited spatial extent of each mapping project.

In Figure 4, the map is showing the points of healthcare facilities in OSM for the county of Nairobi. While facilities seem to be widely mapped, areas like Kamulu in the east seem to be left out and this is something to be taken into consideration. By the two use cases of Wajir and Nairobi counties with high contrast in characteristics as seen in Table 1 (urban/rural environment, high/low population density, socioeconomic differences, number of OSM contributors, etc.) this research is expected to cover a wide range of health services environments and capture all the possible needs of health workers on the field.

Table 1. Geographic and economic indicators of Wajir and Nairobi counties

	Area km^2	Population (2019 cencus)	Per Capita GDP US$[6]
Wajir County	55,840.6	781,263	1,587
Nairobi County	703.9	4,397,073	6,344

[6] https://www.knbs.or.ke/download/gross-county-product-2019/

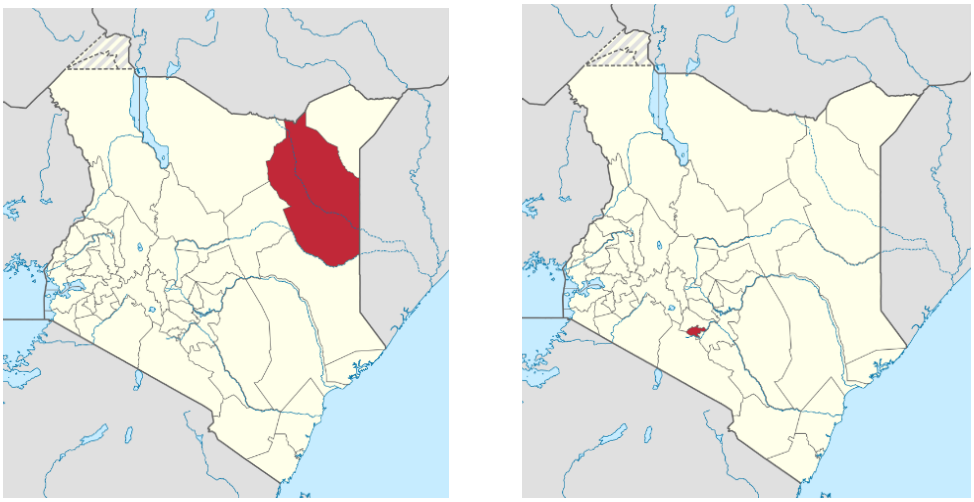


Figure 3. Left: Location of Wajir County in Kenya, Right: Location of Nairobi County in Kenya. Source: Wikipedia

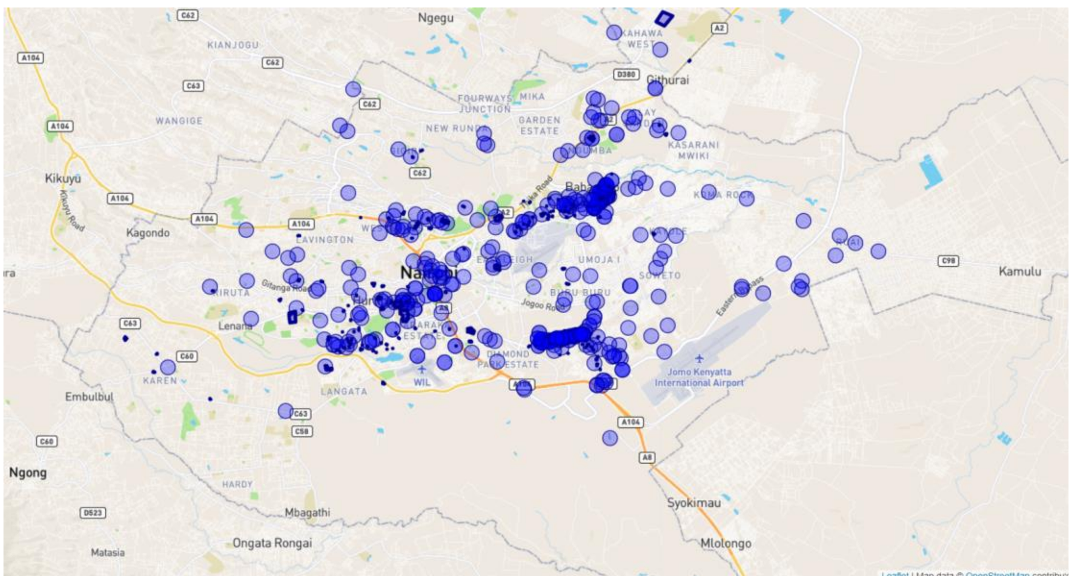


Figure 4. Health facilities of the County of Nairobi in OSM. The map was produced on 20/10/22 with the Qlever SPARQL engine for the Kenya dataset. The setup of Qlever especially for Kenya was kindly provided by Dr. Bast, Albert-Ludwigs-Universität Freiburg.

1.5 Research Objectives

Main Objective

Spatial data related to health services are of crucial importance for humanitarian organizations during emergency response actions. In the Global South context, these data include, mainly but not only, health facilities and road networks. Heath facilities are the main providers of health services, and the road network is the medium to get there or to those in need. The assessment of these datasets in combination with data on public services or infrastructures essential to the end users (e.g., health service field workers) can contribute to a rapid, more inclusive, and overall improved emergency response. The main objective of the proposed research is to:

"Produce a user-oriented assessment framework for the Fitness-of-Use of OSM data related to health services."

Sub-objectives

In this process, integration of intrinsic and extrinsic evaluation is sought after. Before assessing the available data of an Area of Interest (AOI), for their Completeness, Positional Accuracy, Attribute Accuracy, and Logical Consistency, a user-oriented approach is going to be employed. The field needs of humanitarian missions and projects must be defined to assess OSM's Fitness-of-Use for health services. Therefore, the first sub-objective is to**:**

1. **Determine the parts of the OSM datasets that need quality assessment and form the Fitness-of-Use framework for health services**.

Inside these datasets, like in all OSM data, the vocabulary is not consistent for describing the same attributes. Adding more tags and proposing a new vocabulary that better reflects the existing healthcare environment could be an option, but this top-down solution is contrary to OSM values and is doubtful if it would ever be adopted by the users as a whole. Thus, the second sub-objective is to:

2. **Unify the existing heterogeneous vocabulary concerning health services using Semantic Web Technology and make the most information out of the scattered existing data.**

By doing so, the assessment of attribute completeness can be made on a complete basis and give more accurate results. Semantic web technology can also take advantage of Linked Data and retrieve relevant information from external sources.

By assuming that mapping events and projects bring along credibility that comes directly from either the experienced mappers or indirectly from the guidance they provide to new mappers and the reviewing of their changesets, the quality of the produced data is expected to be high. This brings forward the third sub-objective which is to**:**

3. **Figure out and analyze the correlation between the existence of mapping events or projects and the data quality within their area of implementation**.

Data quality assessment will be made for features and attributes according to the user's needs analysis and mainly by comparison with authoritative data. If a strong correlation between mapping projects and Fitness-of-Use quality is found this can be seen as a proxy for OSM data quality assessment in cases and areas where authoritative data are missing or are hard to obtain.

Finally, a web application with editing capabilities will be designed that also provides a general visualization of OSM data quality for specific attributes in an AOI. The fourth sub-objective is to:

4. **Provide the base for the development of a tool that allows field health workers to overview OSM data quality for health services and add or edit features and attributes.**

The aim is to simplify querying and editing with the use of Semantic Web Technology so that the user can contribute without being an expert in OSM or databases. The design of the application will include the RDF dataset of the AOI, SPARQL preset queries, and the python script of the application.

Research Questions

By taking into account the problem statement and the main objective of the research which is to

Produce a user-oriented assessment framework for the Fitness -of-Use of OSM data related to health services.

the following main research question has to be answered.

“To what extent can a user-oriented approach to OSM data assessment be effective in assisting health services in the Global South?”

To reach the sub-objectives into which the main objective breaks down, several research questions and sub-questions have to be answered.

Sub-objective 1

1. What is the capacity of OSM in representing health services data?
   - 1.1 What are the health services data represented in OSM?
   - 1.2 Who are the end users, what are their actual field needs about health services, and how are these translated into OSM data?

Sub-objective 2

2. To what extent can a fragmented health-service OSM vocabulary be unified with the use of RDF?
   - 2.1 How can Semantic Web Technology assist in comparing existing fragmented information in OSM to authoritative data to assess feature and attribute completeness?
   - 2.2 Can the queries be turned into ontologies that promote the existing vocabulary?

Sub-objective 3

3. To what extent, organized mapping projects contribute to OSM data quality improvement?
   - 3.1 What are the metrics of quality assessment for the health-services-related OSM data?
   - 3.2 Is OSM data quality correlated to the completed or ongoing OSM mapping events and projects?

Sub-objective 4

4. How can the assessment of OSM health-services-related data lead to a web application that uses Semantic web Technology to simplify querying and allows editing on the field?

2 Theoretical Background & Related Work

2.1 OSM Vocabularies

Tagging is an organizing technique that uses labels to make sense of a large number of distinct, disparate things following their intended use. OSM tagging system is collaborative and forms the vocabulary of the platform in the absence of an ontology. Looking into the tags makes it possible to analyze what types of distinctions are significant to taggers (Golber & Huberman, 2006). Tagging pairs in the form of `key=value` recommended by the OSM community can be found on the OSM Map Features page and it is the closest possible schema to an OSM ontology. Due to the totally non-restrictive editing policy of the OSM community, it is understood that the proposed tagging system is not always being used. Thus, recommended tags are merged with the intuitive choices of the contributors that are coming from an unbound and uncontrolled vocabulary (Davidovic et al., 2016) leading to heterogenicity due to inconsistency. The global heterogenicity of OSM tagging as proved by Davidovic et al., (2016) has implications for the data quality and the possibility of using OSM tags for applications that need them as input.

The reasoning for the deviance of OSM tags from the suggested vocabulary can be found in the work of several scholars. According to Špiranec & Ivanjko, (2013), this is a matter of the expertise level of the tagger. Golber and Huberman (2006) claim that this is because some taggers pursue their personal goals, not the community's benefit. Nevertheless, it is not only the users to blame. Some geographic elements may be more ambiguous in their classification than others since they are harder to grasp, which may cause the OSM community to lack unanimity. The two major categories of annotation inconsistency are described in the works of Ballatore et al. 2013; Mooney and Corcoran 2012b; Vandecasteele & Devillers 2015. Different names (e.g., "dispensary" and "pharmacy") can be used to represent similar geographic phenomena (synonymy) while the same name (e.g. hospital) can be used to describe different (e.g. hospital, clinic, health center) geographic phenomena (polysemy). Moreover, as Smith & Mark (2001) point out the locality of some concepts can also be a source of tagging inconsistency. Additionally, the evolution of the proposed vocabulary on behalf of the OSM Foundation means that new tags are used for concepts that have already been tagged otherwise (Vandecasteele & Devillers 2015). This creates inconsistency in tagging the same concept because of the time needed to replace old tags and because of the lagging in time until the community adopts the change.

2.2 Semantic Web Technology for OSM

Several methods (Table 2) proposed based on Semantic Web Technology attempt to solve tagging inconsistency problems by constructing schemas that work as ontologies (Vandecasteele & Devillers 2015). Beyond solving or minimizing these problems Semantic Web Technology and Linked Data can improve also the user's experience in a context-centric approach. This implies that users may quickly acquire additional information, such as brief summaries of explanatory text, relevant geographic or demographic data, or details about connected individuals. Additionally, they can find appropriate internet materials like photographs or Web sites (Simon et al., 2011).

The first attempt in the literature to create a path for converting OMS to RDF is found in Stadler et al. (2012). Before that, many tools were built to convert non-spatial RDB data to RDF. This project was named LinkedGeoData (LGD) and its main objective was to build an analogous DBpedia for OSM. The author claimed live synchronization of data to be an advantage rejected by Anelli et al. (2016) who describe the main disadvantage of this method as the time lag between the triplified dump of data and the real-time OSM data. As a solution to this, Anelli et al. (2016) proposed Linked Open Street Map (LOSM) which acts as a translator from SPARQL to Overpass API. This tool is supposed to retrieve real-time OSM data while querying triples. Another important disadvantage of LGD is that there was no support for Relations.

In 2012 the GeoKnowledge project, funded by the EU emerged with the ambition to integrate all geodata on the web by applying the RDF model and the GeoSPARQL standard. This is basically the core idea of Athanasiou et al. (2014) for representing and querying geospatial data. Their tool called GeoKnowledge Generator[7] tiplifies geospatial data using the GeoSPARQL, WGS84, or Virtuoso RDF vocabulary. The Generator builds a triple store this way which can be enriched by other non-spatial data in RDF and data from RDBs. This approach does not provide the triples ready, but it aims to be user-centric and let the user triplify the data they need. In this course, the user is being assisted by the appropriate embedded software for querying, linking, enriching, and visualizing data.

A decade after the emergence of LGD, a tool called osm2rdf that converts OpenStreetMap (OSM) data to RDF triples was developed at Freiburg University. With the help of a SPARQL endpoint, the user can form SPARQL queries for that data keeping all their attributes including geometry (Bast et al., 2021). The authors claim that no data is missing after the conversion to RDF with the database being updated every week for the whole planet. They refer to some lack of speed during the precomputation of the spatial triples, but this doesn't seem to affect the user.

Chen et al. (2017) proposed an interesting method that could potentially overcome the problem of linking the transformed OSM to RDF data only to other RDF sources. They take the case of Wikidata JSON data and claim to transform it to JSON-LD which is a concrete RDF syntax.[8] In their paper, Chen et al. do not demonstrate querying examples or use cases in detail, nor has our research come to any results referring to this method. So, it remains a question of to what extent this attempt had the expected results.

2.3 OSM data quality

According to Antoniou & Skopeliti, (2015) and Senaratne et al., (2016), quality assessment of OSM data can be performed through measurements as well as indicators. According to the ISO principles and recommendations, quality measures are those components that may be used to compare contributed geographical data to authoritative data in order to determine whether there is a disagreement between the data and the ground truth (Senaratne et al.,2016). Authoritative data though, cannot always be available. To address this problem, researchers have looked at more intrinsic approaches to evaluate VGI quality by looking into various proxies.

[7] http://generator.geoknow.eu
[8] https://www.w3.org/TR/json-ld11/

Table 2. Synthesis Matrix containing literature on triplification of OSM data

	Data formats	Query Language	pros	cons	Links to
Stadler et al. 2012 **LGD**	OSM to RDF	SPARQL	Being the first at that time to triplify OSM Live synchronization (?)	Relations are not supported. Linking only to RDF. Update of triples every 6 months Many OSM data filtered out	Spatial/Non-spatial data
Athanasiou et al. 2014 **GeoKnowledge**	Shapefiles, spatial tables, and non-spatial data to RDF	geoSPARQL	User-centric	Complicated using many stand-alone tools tools	Spatial/Non-spatial data
Anelli et al. 2016 **LOSM**	RDF	Translates SPARQL to Overpass API	Real-time Querying	A tool working on top of another tool with unknown interoperability	N/A
Bast et al. 2021 **osm2rdf**	OSM to RDF	SPARQL	Complete OSM data to triples Keeping all attributes Weekly update	Linking only to RDF	Spatial/Non-spatial data
Chen et al. 2017 **CrowdGeoKG**	OSM to RDF & JSON-LD	SPARQL	Converts OSM to RDF and JSON from Wikidata to JSON-LD. Both then can be queried with SPARQL.	No available examples and tests. No evidence of the usability of the described method.	Spatial/Non-spatial data

2.3.1 Measures of OSM data quality

Geographic data, according to the 2013 revision of the International Standardization Organization (ISO) are being increasingly shared and utilized. The myriads of web or mobile applications that need geodata can be quite different in context from their initial production aim. Since the value of data is directly correlated with its quality, knowledge about the quality of the geographic data that is readily accessible is essential to the process of choosing a data collection. A consumer of geographic data could have a variety of data sets to pick from. In order to evaluate which data collection best satisfies their expectations, it is required to compare the quality of the data sets (ISO, 2013). ISO defines in this context 6 elements of data quality: Completeness; Logical Consistency; Positional Accuracy; Thematic Accuracy; Temporal Accuracy; Usability. These are shown in the diagram of Figure 5.

Of course, not all the metrics of these measures described by ISO are relevant to OSM because it is not a standard geodatabase with structure and metadata (Hashemi & Abbaspour, 2015). The measures of OSM data quality when following the ISO 19517 typology are adapted by many scholars as follows (Baron et al., 2014; Kaur et al., 2017; Borkowska & Pokonieczny, 2022; Fonte et al., 2017; Girres & Touya, 2010; Zhang & Malczewski, 2017;):

1. Completeness: The absence of data or the excess of data through duplications are considered errors.
2. Logical consistency: Topological correctness and relations between objects are assessed.
3. Positional accuracy: The degree to which the result of position measurement is near to an agreed true value.
4. Thematic accuracy: Attribute completeness and Semantic Accuracy
5. Temporal Quality: The Validity of the time of editing (Currentness).

Usability cannot be implemented as a separate measure for OSM but as an overarching one for the other five. This is because OSM datasets are not created for a specific reason in their totality so usability cannot be assessed out of context (Devillers et al., 2007; Fonte et al., 2017). This means that by defining the use or context then other measures can be used more precisely to quantify the usability of specific OSM data (Figure 6).

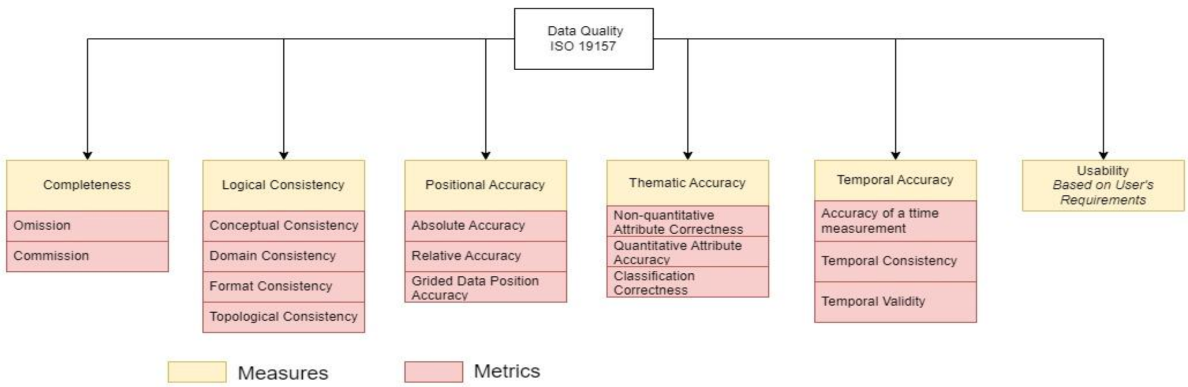


Figure 5. Quality measures according to ISO 19157:2013(E) for Geographic Data

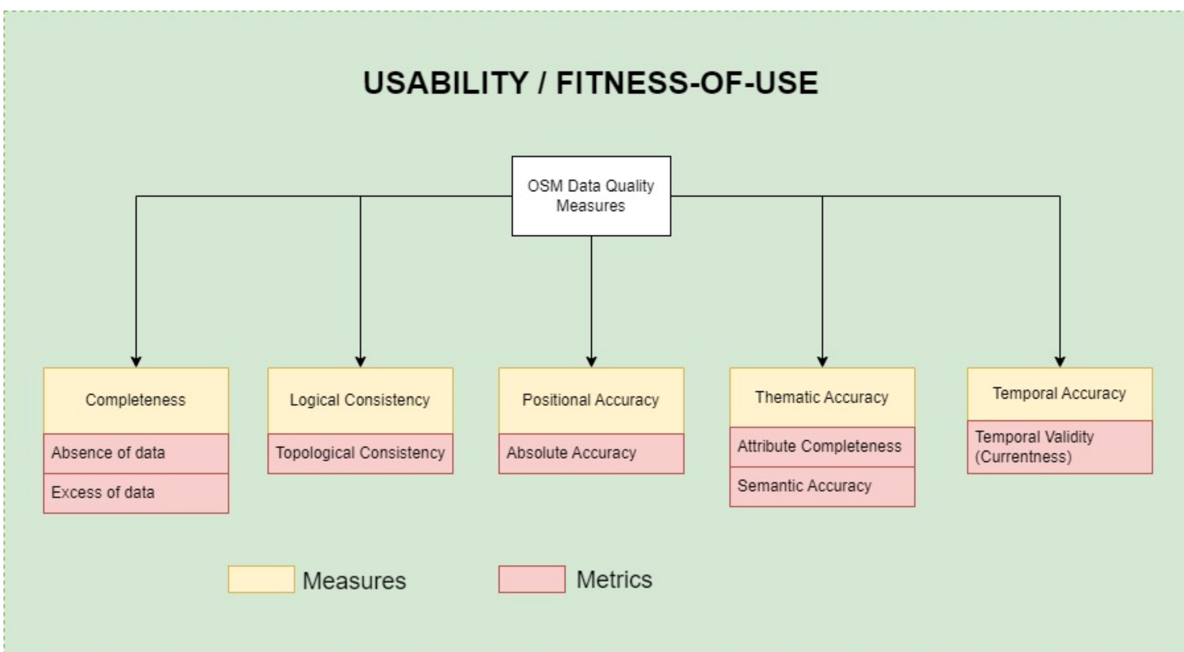


Figure 6. OSM Data Quality Measures & Usability

The most popular method for **Data Completeness** is unit-based (Zhang & Malczewski, 2017; Barron et al., 2014; Girres & Touya, 2010; Haklay, 2010; Jokar Arsanjani, et al., 2013; Jokar Arsanjani et al., 2015b; Jackson et al., 2013). In this method, the number of objects is counted or the total length or area is calculated. In all these cases a reference dataset for comparison is needed.

For **Positional Accuracy**, Euclidean distance is commonly used for points (Girres & Touya, 2010; Jackson et al., 2013) while buffers are used for ways (Jokar Arsanjani et al., 2013; Zhang & Malczewski, 2017; Ludwig et al., 2011). There are also other methods like bidimensional regression and G-statistics that may be considered to investigate the usability of the data for mapping applications (Helbich et al., 2012).

Logical consistency deals with the topological and logical relationships between objects. For nodes where geometrical distortion is not a problem, assessment is usually unnecessary (Jokar Arsanjani et al., 2013). There are cases though that the position of a node must be assessed in relation to an administrative boundary when those two are in proximity. The relation of objects (Zhang & Malczewski, 2017) but also the topological correctness of polygons and polylines that are more susceptible to this kind of error, are mostly done through visual comparison between a test and a reference dataset (Jokar Arsanjani et al., 2013).

Thematic Accuracy consists of Attribute Completeness and Semantic Accuracy. Attribute Completeness is assessing if the objects are tagged with their main or secondary attributes (Girres & Touya, 2010; Ludwig et al., 2011) while Semantic Accuracy assesses if the semantics of the objects reflect reality (Girres & Touya, 2010). These two assessments correspond to the Quantitative and the Non-quantitative Attribute Correctness of ISO 19157.

Temporal Accuracy for OSM data is basically referring to temporal validity or currentness. Girres & Touya (2010) suggest dividing the dataset into thematic areas and then counting the objects between one time in the past and the present day. This method can have a reliable result only when combined with the assessment of completeness.

2.3.2 Indicators of OSM data quality

According to Devillers et al. (2014), an indicator is helping the researcher to have an overview of something by looking only at a small part of it helping at the same time in decision-making. Purpose, usage, and lineage are the indicators for geographic data quality proposed by ISO standards. It is hard to define purpose in VGIs, especially in OSM which started as a road map of the world but now is in the direction of containing every object of the real world. Usage deals with the applications that use the data. If for example OSM is used by many (and this has to be defined) navigation applications with a great number of loyal end-users, this probably means that the road network and the representation of the POIs of the dataset are of high quality. Lineage is also an indicator of quality for OSM that measures the activity on a dataset by tracking editing history (Keßler et al., 2011; Girres & Touya2010; Sanarante et al., 2016). The trustworthiness of users can then be estimated based on the period that the changesets are staying the same or being revised. Credibility is another quality indicator proposed for VGIs, including trustworthiness and expertise (Flanagin & Metzger, 2008).

Another indicator that works by deviating the reference datasets that sometimes are unavailable, based on Linus' Law was proposed by Haklay et al. (2010b). Linus' Law was originally formulated to assess open-source software and stated, ‘Given enough eyeballs, all bugs are shallow’ (Raymond, 2001, p. 19). For VGIs, according to Haklay et al., (2010b) that means, the more mappers the better the quality of the respective dataset even if not in a linear relation. This assumption was challenged by Zhang & Malczewski, (2017) and by McGough et al., (2022) with the latter suggesting an Agent-Based Modeling method to include factors such as experience and socioeconomic or demographic backgrounds for assessing data quality. So far there has been no attempt to come up with a proxy for OSM quality through the presence of organized mapping projects as an indicator of credibility.

3 Methodology & Tools

The present research involves several stages in the way to reaching its objectives. The user's needs analysis is the first part which builds the context for all the other parts and consists of a Questionnaire, several interviews, and a report analysis. Based on that the test datasets are going to be produced with the source being Kenya's OSM dataset in RDF format. Using Qlever SPARQL engine and RDFlib for Python the aim is to construct the most complete test dataset possible containing all hidden information. These datasets are then going to be compared to authoritative data to assess health services OSM data. The queries used for the assessment can also be a guide to building an ontology for promoting existing vocabulary. Then by using OSMCha and querying for humanitarian hashtags the goal is to find areas of the use cases where humanitarian mapping projects have been implemented and see if they are correlated with high-quality health services data. Finally, the setup of a web application will be proposed using the outcomes of the previous stages. The workflow described is visualized in Figure 7.

3.1 Users' Needs Analysis

In order to acquire the end user's perspective on what are the essential OSM data for supporting health services in the Global South, the first step would be to define the end user. There are mainly three categories of users. Health field workers of governmental or non-governmental organizations, the GIS department (if existent) of these organizations, and those involved in projects dedicated to enriching OSM data for humanitarian use with the most prominent being the Humanitarian OpenStreetMap Team (HOT), MissingMaps, and YouthMappers. In the case of Kenya, the GIS department of the Red Cross along with HOT and YouthMappers are working in close collaboration to form a large community of OSM contributors with the common aim of resilience for vulnerable communities across the country. At the date of writing a HOT project is in process for mapping Wajir County with the collaboration of the Kenyan Red Cross and YouthMappers. Mapping on the field was followed by a mapathon that took place in Nairobi's Red Cross Headquarters on 19.11.22 using HOT's Tasking Manager for editing. This large community of OSM mappers is the focus group for the first part of the user's need analysis. By assumption, the individuals in this group have the basic knowledge to answer a questionnaire related to OSM data. Furthermore, the official report of the Red Cross concerning the field work while mapping Wajir will be analyzed for conclusions about users' needs. The second part of the user's needs analysis is qualitative and involves interviews with health field workers of the Kenyan Red Cross. The aim is to understand their needs for geo-information while on the field which they either search for themselves or are provided to them by their GIS department.

3.1.1 The Questionnaire

The group of people that received the questionnaire was heterogenous, consisting of GIS experts, researchers in the field of health services, health workers, and HOT mappers. This means a different level of understanding of the spatial dimension of health services data. For this reason, the level of the questions in terms of content and terminology was kept to that of general knowledge. The questionnaire is divided into two parts. The first asks questions about spatial data and the second specifically about OSM.

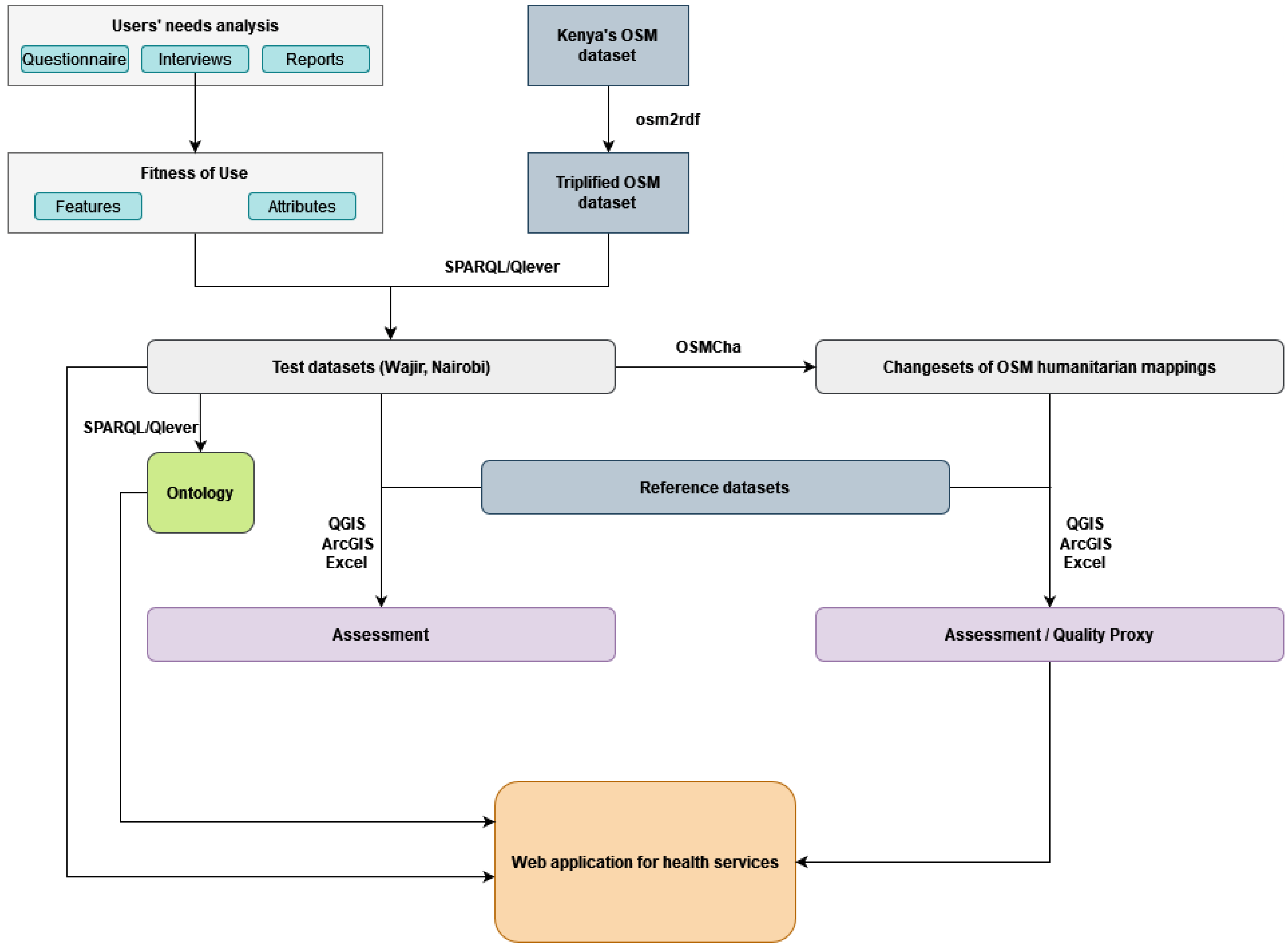


Figure 7. The workflow of the research shows all stages and their interconnections.

The first two questions are introductory asking about the Organization and position of the responder. The first part follows. Here are the questions and their descriptions indicating their goal.

Q3: Do you use spatial data for operational reasons?

The answer to this question varies between never, rarely, often, and always. The answer "never" will rule out the questionnaire from the evaluation of the results. The other three answers will give respectively a multiplier of 0.33, 0.66, and 1.00 to the whole answer sheet as a weight to the multiple-choice questions.

Q4: Do you have a GIS department/ specialist in your organization?

Answering yes to this question will give a multiplier of 1.1 to the multiple-choice questions.

Q5: What sources of spatial data do you use?

Q6: What type of spatial data do you use?

Questions 3 and 4 are put to give information on possible reference datasets or datasets that can be used in combination with OSM.

Q7: What of the following would you describe as health-service-related spatial data?

The possible answers are Health facilities, Road Networks, Water Resources or Other to be specified by the responder. This is the first core question for the user's needs analysis. The answers with the weights applied will give reveal the features that the user needs from OSM. The preselection of answers to this question can make the questionnaire look biased and it somehow is. This happened because of the heterogeneous background of the focus group. The two first answers, health facilities, and road networks are considered by assumption absolutely necessary. With water resources as the third option, the responder is expected to think of other important but not obvious features related to health services.

Q8: How important is the following information about health facilities for your operational needs?

Q9: What other information about health facilities would you consider important?

The answers to question 8 vary between Not important, Little important, Very important and Absolutely important for the geometry and some attributes of health facilities (Geographic Location, Classification (ownership, level, etc.), Type of services/departments, Number of beds /cots, Equipment available, Phone number). Question number 7 tries to capture other important information about health facilities that can be translated into attributes.

3.1.2 The interviews

In the need to identify the users of spatial data and potentially OSM data related to health services, which should also be defined, several interviews took place at the Kenyan Red Cross Headquarters and the regional branch both in Nairobi. The positions of the people interviewed span from project officers, paramedics, and emergency medical services operators to Red Cross Action team members and Monitoring & Evaluation focal persons. The advantage of interviewing health services professionals from the Red Cross is that the organization has valuable collective experience in addressing all types of health-related issues and situations, runs numerous projects for assisting vulnerable populations, and has a structure dependent on data including spatial data. During the interviews, it was obvious that the Red Cross has its own system of dealing with the spatial dimension of all kinds of situations. The interviewees were most of the time reluctant to change that even if they could identify inherent problems and weaknesses. This is of little importance for the present research because the main aim is to build the framework for assisting all relevant organizations in using OSM data for health services and not only the Red Cross which has comparatively plenty of resources. In other words, the interviews were not to answer what is missing or what should be changed inside the Red Cross and the handling and use of spatial data but more to identify the needs of such data and potential users. The interviews started with one Project Officer and the choice of the next interviewee was made by the previous one depending on the answers they had difficulty responding to. This can be seen as an organic user identification process. The interviews were based only on open questions and in all cases turned into discussions. The questions on which the interviews were based are as follows.

1. Can you describe your duties during fieldwork (or office for coordinators) and missions?
2. Do you use any data and for what reason? Where do you get them from?
3. What type of information do you need while on a mission?
4. Do you use any navigation map?
5. I will show and explain some data (tags from the vocabulary analysis) that exist in OpenStreetMap can you recognize any relevance to your work?

3.1.3 Kenyan Red Cross report analysis

At the time of writing, a mapping project of the Kenyan Red Cross in collaboration with the Humanitarian OpenStreetMap Team was taking place in Wajir County in the Northeast of the country. This project consisted mainly of fieldwork mapping and mapping events through the Tasking Manager of HOT. Through analyzing the report, information about the use of OSM health services data by the KRC can be extracted.

Background of the mission

Kenya Red Cross Society, through the International Center for Humanitarian Affairs (ICHA), is implementing a Community Participatory Mapping in Wajir County. This 1-year project is funded by the Humanitarian OpenStreetMap (HOTOSM) ESA hub focusing on using geo-data to build disaster resilience and reduce climate-related hazards' impacts.

Objectives of the mission

According to the report the project has three main objectives:

1. Increased community understanding of their risk, vulnerabilities, and capacities.
2. Enrich OpenStreetMap (OSM) data by increasing contributions from new community members. Plan on adding 5000 features to OSM, including building footprints, roads, and social amenities.
3. Foster the use of data-driven decision-making through data for proactive response to disasters.

The second objective which is about enriching OSM data gives an overview of the needs of humanitarian mapping in general but not for health services specifically. From the list of public amenities to be mapped and edited during the mission those related to health services were:

- Health facilities
- Pharmacies
- Water points/ facilities

3.1.4 Challenges of the User's needs analysis

The questionnaires were sent to people without prior personal contact. Out of the 45 sent 31 of them, almost 70%, were never received back. From the rest, there were no significant variations in the answers from which can be assumed that they were biased. Every data type indicated by the questionnaire was considered important with only a small variance in impotence. Thus, the results were ruled out as a whole but were used as a reference during the interviews.

3.1.5 Results of user's needs analysis

The analysis was focused on the interviews with the prior knowledge that the Kenyan Red Cross in their mapping project in Wajir County through the HOT platform is interested in health facilities and water points.

Project officer of NCDs

What was obvious from the interview is that spatial data for people that are planning actions and projects are valuable. The basic need was to know where the health facilities that offer screening and treatment for Non-Communicable Diseases (NCDs) are. On the spot of the interview, there was a detailed tag search containing keys or values of NCDs (diabetes, hypertension, cancer, mental health) without any results. Then a proxy for identifying the available services of a facility through its level according to the Kenyan official classification. The answer was that levels 4, 5 & 6 are supposed to have all the services concerning diabetes and hypertension while 5 & 6 should be able to deal with mental health issues. According to the same source, populations in informal settlements and refugee camps around the country need to have a tool to guide them on where to reach for help for NCDs. Data on where pubs that offer alcohol and processed food are located or areas without public spaces for exercise (main causes of diabetes and hypertension) are also considered important by the NCDs project officer.

Emergency Medical Services Operator (All Kenya).

According to the interviewee, the Red Cross has an integrated system of locating the ambulances available regarding the proximity to an event (Zone Control Service) and then sending pins of multiple destinations using google maps to the driver's mobile phone. The primary need here concerning geographic data is the road network and the location of facilities. The interviewee agreed that google maps are not always reliable when talking about rural areas. The administrative level of the health facilities according to her is valuable information theoretically but usually does not reflect the ground truth due to the limited resources available. The number of beds for a facility can be useful only in the case of real-time information. The first decision that has to be made when transferring a patient to a health facility is the type of operator distinguishing between the public and private sectors.

Red Cross Action Team (Nairobi County)

The interviewee stressed ones more the first decision that has to be made when casualties are involved; public or private sector facility for the patients. The additional information here was that the private sector is divided into Profit and Non-Profit organizations. This is crucial for the acceptance or not of patients without insurance. In any case, patients are accepted in all facilities for stabilization, regardless of their insurance status. According to him, navigation in the city of Nairobi is not necessary because the drivers are experienced enough with him being a former bus driver. The information they seek in google maps is usually the traffic conditions to choose the fastest route. The need to know the services of each facility is crucial and the level of administration is a proxy to this information but as Action Team they work with 10-12 health facilities that they know very well and have an emergency department. We searched for these in OpenStreetMap to realize that a very important for the Action Team hospital (South B Hospital) was not mapped at all and we added it to the map at the end of the interview. This initiated great interest and the possibility of an OpenStreetMap contribution seminar among health workers was discussed.

Monitoring & Evaluation Focal Person

As a person who is doing fieldwork to check the ground truth and compare it with the objectives of implemented projects, the interviewee has to be in contact with health facilities and individuals. The contact information and location of every facility are considered to be important. Through the interview, it was clear that individuals can be seen as end users of OSM health facility data. According to the interviewee, information about the services offered to the patent by a facility and its proximity to them can help the health system to work more efficiently.

Project Implementation Assistant

In most field projects the Red Cross has its own water and sanitation (watsan) team. The deployment of the team is most of the time the first thing to happen on the field. This information coming from the interviewee shows the importance of watsan services and subsequently the need to know their location before deploying a mission when there are limited resources. Watsan services are also the first line of action in preventive medical policies so knowing the capacities in every place or region is necessary for planning. When on mission health workers need also information on health facilities' location and capacities (departments, medical staff, equipment, etc.).

3.2 Vocabulary analysis

3.2.1 Nairobi County

With the help of the interactive context-sensitive autocompletion feature of the Qlever SPARQL engine, all health-services-related vocabulary, containing keys and values, can be comprehensively studied.

Health Facilities

In the case of Nairobi County, 63 keys were found related to health facilities with 280 corresponding values. Of them, 13 keys were found in the 108 different "amenity" keys. In Appendix A all keys with their respective values can be viewed.

By studying this vocabulary there are several takeaways to be noted.

- Keys that use underscore instead of space are coming usually from organized mapping projects in the informal settlements of Nairobi (Figures 8 & 9).
- The English colonial past of Kenya has an impact on the spelling of specific words. Most of the time but not always, and this is the major problem, it is “centre” instead of “center” which is crucial when this refers to medical or health center. This issue is known in the literature as the locality of the vocabulary.
- The organized mapping projects in Nairobi county prefer precise tagging with keys such as `health_facility:type` which gives back 24 different health facilities while individual contributors tend to prefer healthcare or amenity that correspond to 13 different values each.
- There are several spelling mistakes like in the key `psychosocial_support_counselling` but this is typical in OSM as well as the capital initial letters in words that are not suggested by OSM.
- Except for the tags used only for health facilities, some keys are generic and are used for various features but can also give valuable information.

  operator: description | operator:type |operational_status | operational_status:availability | emergency | operational_status:electricity | opening_hours | name | email

Figure 8. Tagging examples of health facilities followed during humanitarian mapping projects in the informal settlements of Nairobi

WATSAN (water and sanitation)

By examining the vocabulary of Nairobi for WATSAN facilities it seems that there is a specific taxonomy followed for the tagging which is the result of humanitarian mapping projects. Colons and underscores are used to form a sophisticated and organized OSM vocabulary for the specific domain. Using the key watsan in the tags, which is the abbreviation for “water and sanitation” the goal is to distinguish between water features for personal use and features that represent agricultural use or natural resources in general. There are 79 different watsan keys for Nairobi most of them with binary values (yes, no). Detailed descriptions can be found in Appendix A.

Here are two typical examples of water and sanitation (watsan) points tagging through informal settlements mapping projects.

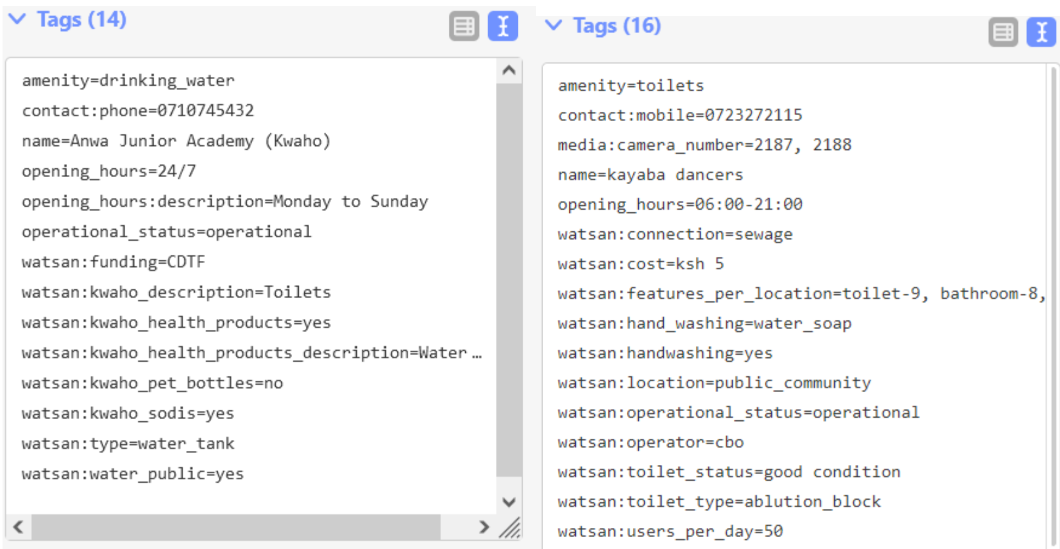


Figure 9. Tagging examples of WATSAN POIs followed during humanitarian mapping projects in the informal settlements of Nairobi.

3.2.2 Wajir

Health Facilities

The analysis for Wajir County (October 2022) showed that the vocabulary is relatively poor compared to Nairobi's. Only three keys for health facilities 2 of which are dedicated and one generic (amenity) with 9 values and they are those that are typically used outside organized humanitarian mapping projects. For a detailed view see Appendix A.

Water-related features

Similarly, for water and sanitation the vocabulary in Wajir County is much poorer compared to the one in Nairobi. Only 6 keys with 26 corresponding values give general information about water and no WATSAN tagging (see Appendix A).

3.3 Translating users' needs into OSM tags

Having the results of the users' needs analysis and the OSM vocabulary analysis made it possible to proceed to the translation of the former to the latter. The results of this process can be seen in Table 3.

Table 3. Translating the needs of the users to OSM tags according to the local OSM health facility and WATSAN vocabulary

Need	**OSM data**
Navigation	**highway**=*
Traffic	N/A
Emergency	**emergency**=yes
Health facility	**amenity**=("pharmacy", "hospital", "clinic", "dentist" "laboratory", "doctors", "medical transport service" "traditional health centre", "nursing_home") **healthcare**=("pharmacy", "hospital", "clinic", "dentist", "laboratory", "doctor", "nursing_home", "alternative", "yes", "hospice") **health_facility:type**=(chemist_dispensing, chemist_over_the_counter, pharmacy, dispensing, nursing_home_with_maternity, Dispensing drugs, chemist_over_the_counter;chemist_dispensing, health_center, health_center;hospital) **dispensing**=yes
Level of Administration (1-6)	**health_facility:type**=(chemist_dispensing, chemist_over_the_counter, pharmacy, dispensing, nursing_home_with_maternity, Dispensing drugs, chemist_over_the_counter;chemist_dispensing, health_center, health_center;hospital) **dispensing**=yes part of the **name** is: Community Health Unit / Dispensary / Health Centre / Sub-District / Sub-County Hospital & District / County / Provincial / Referal Hospital **medical_staff:*=***
Departments / Medical Services / Treatments	**medical_service**=*
Contact	**contact:phone**=* **contact**=*
Operator type	operator:type=* operator=*
Equipment	N/A
Water and Sanitation	**watsan:*=***

3.4 OSM2RDF & Qlever SPARQL Engine

Several ways have so far been suggested to query OSM data. The simplest if not simplistic one is the pointing arrow with the question mark in openstreetmap.org. It may not be really a query tool, but it is useful to find what data exist in an area of interest. Overpass turbo is a well-established and effective option, but its querying language is difficult when it comes to advanced and complicated requests. OSMcha is a sufficient tool for querying data with various filters provided in the GUI without the need of scripting the requests. The disadvantage of OSMcha is that the results are not visualized to give an overview (only one by one) nor are they downloadable to work with them. Downloading OSM data and inserting them into a PostgreSQL database with PostGIS extension can be a solution but it is resource intensive.

Querying in SPARQL takes advantage of Semantic Web Technology but presupposes that OSM data are converted to RDF datasets. Of the tools reviewed doing this conversion, the osm2rdf tool is the most up-to-date and fully functional covering OSM data globally. Developed at Freiburg University osm2rdf converts OpenStreetMap (OSM) data to RDF triples (ttl files). All data transformed in RDF are available for downloading at the global, continental, or national level. The datasets are downloaded as a compressed package after which they can be extracted and loaded into a semantic graph database like GraphDB or a local server. The triplified data represent all objects and relations known, therefore, we consider the RDF datasets to be complete. The creators of osm2rdf claim that no data is missing after the conversion to RDF with the database being updated every week for the whole planet. The produced data keep all their attributes including geometry (Bast et al., 2021).

An additional advantage of this method is that the queries in SPARQL, the query language used for RDF data, can ask for complementary information from external sources (e.g., Wikidata) through federation. With the help of a SPARQL endpoint called Qlever also designed and developed by Bast et al. (2021), the user can query and retrieve all OSM data. Qlever is a high-quality SPARQL engine that can preload OSM triplified data or Wikidata data amongst others. It also provides an auto-completion function and the option to see the results of the query on an interactive map using Leaflet. There are also querying examples available in Qlever that can assist the beginner in learning SPARQL. The main disadvantage of Qlever is that does not support yet several expressions used for advanced SPARQL querying. For this reason, the most advanced queries will be constructed in RDFlib, the RDF library for python against the turtle file stored in a local hard drive.

3.4.1 Building Queries in SPARQL

SPARQL is a query language for RDF datasets that are built on the subject-predicate-object model. The queries follow exactly the same model, and they can take the form of SELECT, CONSTRUCT, ASK, and DESCRIBE. In the present research, only the first two were utilized. The SELECT form was used to get the values of the OSM objects. The CONSTRUCT form was used to extract sub-graphs from the original OSM ttl file of Kenya.

3.4.2 Validation of OSM2RDF

To validate the results of the SPARQL queries regular comparisons were made by querying for the same objects in QGIS using the QuickOSM plugin. The results were all of the times identical except for a case in Wajir County where an ongoing mapping project was feeding OSM with new data and those could be seen in real-time in QuickOSM because it works with Overpass Turbo. The triplified OSM dataset though, provided by Freiburg University is not real-time data, so in that case it took a few months until the results were identical. Overall, the completeness of the results coming from SPARQL queries against the OSM ttl file of Kenya was perfect with the only drawback being the update latency.

3.5 Metrics for the OSM data assessment

After the users' needs analysis, the results were used to narrow down the data that had to be assessed for the objective of the research. The relevant literature review provided a wide range of metrics out of which the most adequate to the data was chosen.

3.5.1 Completeness

Completeness, when talking about spatial data can have many interpretations. The two basic categories are the completeness of objects and the completeness of the attributes of those objects. In all cases, there is a comparison involved with some reference dataset and the assessment is unit-based as it counts objects or lengths.

Completeness of POI or OSM objects

According to the relevant literature,[9] the completeness of OSM data can be calculated by comparing the number of objects and compare to the respective objects in the reference dataset. In the present research, authoritative data were used when available and the percentage of completeness was calculated by dividing the number of overlapping objects by the number of objects in the reference dataset. This is the part of the assessment called omission and is a widely used metric for assessing OSM quality of data. There is also another relevant metric called commission which counts the number of objects that exist only in the OSM dataset. High commission combined with high completeness means high quality of OSM data even higher than the reference dataset.

Completeness of attributes

For assessing the attribute completeness of the object of the OSM datasets the "tag presence" method was used from the relevant literature.[10] This means the percentage of objects that had at least one tag for the attributes indicated as important by the users' needs analysis. Here there was no comparison with authoritative data, but the same percentages can be calculated for the reference dataset separately.

Completeness of road networks

This is another case of object completeness by counting the length of road segments in OSM and then comparing them to the authoritative data following the same literature as in the completeness of the OSM objects.

3.5.2 Accuracy

Accuracy of attributes

Except for the completeness of attributes, their accuracy in reflecting reality is of equal if not higher importance. The ground truth of the attributes was considered to be the authoritative data meaning 100% accuracy by convention. Then the values of the keys that were included in the users' needs analysis were compared to the reference data. This process was manual because an automated text-matching process skips semantic matching. For example, a faith-based organization for the scope of the present research is considered to be the same as a religious institution as types of ownership of a health facility.

[9] Barron, Neis, & Zipf (2014), Girres & Touya (2010), Haklay (2010), Jackson et al., (2013), Jokar Arsanjani, Barron, Bakillah, & Helbich (2013), Jokar Arsanjani, Mooney, Zipf, & Schauss (2015)

[10] Girres & Touya (2010), Ludwig et l. (2011)

4 Data identification and acquisition

4.1 Authoritative data

According to the User Identification and the following users' needs analysis, it is possible to narrow down the data needed for the quality assessment of OSM. For the comparison of the identified OSM data to their authoritative counterpart, contacts with the Kenyan authorities had been necessary. Health facility data for the counties of Wajir and Nairobi were requested from the Ministry of Health. Road networks of the counties were requested by the Kenya Roads Board and Water and Sanitation data from the County Governments.

4.1.1 Health facilities

According to the users' needs analysis, what is needed for the assessment are the services provided for each facility along with the type and/or the level of administration. The operator type, being private or public is also desirable. The official process of acquiring those data from the ministry of health is sending an official request by post and then waiting for an answer. This process was followed but unfortunately without results. The alternative was to scrape data from the Kenya Master Health Facility List (KMHFL), an application with all health facilities and community units in Kenya that is owned by the Kenyan Ministry of Health. The data of this portal were last updated in 2020. From the website's API a request was made asking for the data needed then these were exported to JSON files and transformed to CSVs with coordinates. The scraping was done using Python (Appendix B) The CSV's were used to be easily edited and then transformed into SHP files using "MyGeodata Cloud" to then be loaded to QGIS and ArcGIS pro.

4.1.2 WATSAN

There were no authoritative WATSAN data available for entire Kenya. Nevertheless, available WATSAN data for the informal settlements of Nairobi are available through the UN-Habitat portal. This data set will be used as a reference for the assessment. These data include amongst others, water points, communal sanitation points, and handwashing points which are also identified as users' needs. For Wajir County, there are no WATSAN data, and this is not a surprise because the population is mostly rural. Data for water points were requested but with no results until the time of writing. What was only available was a map of Wajir water points through a Policy Brief of the County of Wajir (Bedelian, 2019). The map was georeferenced and digitized on QGIS to get the coordinates of the water points in four different layers attributed as boreholes, pans, mega pans, and shallow wells.

4.1.3 Road Networks

For the road network data, the same process of an official application was followed for Nairobi and Wajir Counties. A "Terms and Conditions" document was signed and then the data were available for download in SHP format through a link sent by the Kenya Roads Board.

4.2 OSM data

4.2.1 Health facilities

Using Qlever and SPARQL queries (Appendix C), it is possible to extract CSV files containing all facilities within the OSM relations of choice along with all the tags that are relevant to the users' needs analysis. For the cases of Nairobi and Wajir counties (OSM relations), we queried for health facilities with coordinates and names so that they could be comparable to authoritative data for indicating the completeness of OSM data. Another query called for health facilities containing tags indicating the services offered. Contact details and operator-type tags were also queried separately. After downloading all respective CSV files, the data were closely examined for duplicates that were removed. OSM data for health facilities in Kenya also include "Traditional Healers" and "Herbalists". These categories were also removed from the datasets because there are not comparable to any category of the KMHFL. After cleaning up the data the CSV files were transformed into SHP files using again "MyGeodata Cloud".

4.2.2 WATSAN

Similar to the Health Facilities we queried for tags related to WATSAN and reflect the users' needs (Appendix C). For Nairobi County, there are also WATSAN dedicated tags while for Wajir we could only look up for "water" or "drinking water" or indirectly by including schools which are usually points with access to drinking water and basic sanitation. The results of the queries were again extracted in CSV files with coordinates, using Qlever, and were then transformed into SHP files.

4.2.3 Road Networks

Road networks for Wajir and Nairobi Counties were downloaded for each as layers in QGIS using QuickOSM extension. From all the tags with the key highway, only those values which represent roadways were chosen. These values are motorway, motorway_link, trunk, trunk_link, primary, primary_link, secondary, secondary_link, tertiary, unclassified, residential, and service.

4.3 Administrative boundaries

For the analysis of the data, auxiliary data are needed. Administrative boundaries of Kenya in the county, the constituency (sub-county), and the ward levels were downloaded from the United Nations Office for the Coordination of Humanitarian Affairs (**OCHA**) through their Humanitarian Data Exchange portal.[11]

[11] https://data.humdata.org/

5 Analysis & Results

After setting up the theoretical framework and acquiring all available data the next step was the assessment of the OSM data that coincide with the needs of the identified users. Then the SPARQL queries used for the assessment were modified to build an ontology for the classification of the health facilities in Nairobi County based on the local OSM vocabulary. Finally, a proxy of health facility OSM data quality was proposed

5.1 Quality assessment of OSM data

The assessment was first divided geographically between the Counties of Nairobi and Wajir. For each county the assessment included object completeness of the health facilities, attribute completeness and accuracy of the facilities, completeness of the WATSAN (water and sanitation) or just water points, and completeness of the road network.

5.1.1 Nairobi County Health Facilities

Object Completeness

In this part, a comparison is made between the number of OSM objects representing health facilities in Nairobi and the number of health facilities coming from authoritative data. This is done as an attempt to assess the completeness of the OSM data. After studying in detail the OSM vocabulary of Nairobi, it was possible to build a query (part of it in Figure 9. See Appendix C for the entire query) in SPARQL, calling for all types of health facilities from the triplified OSM data. That was made using the "union" condition of several OSM keys and values. The "distinct" modifier was used to eliminate the duplicates that may existed in the results and the "regex" filter to exclude "polylines" from the results and keep only the "multipolygon" version of each object. Then the results were downloaded in CSV format containing the names of the health facilities and their coordinates. The last cleanup of the data was performed in this format by removing duplicate names with almost identical coordinates but different OSM id as well as all alternative medicine objects like "herbalist", and "traditional healer" that are not comparable with the authoritative data from the Kenyan Ministry of Health (KMH). Duplicates of generic names like "Chemist" or "Clinic" were not removed, nor were health facility branches of the same operator sharing the same name. This process resulted in 961 OSM objects (points and polygons) in Nairobi that represent health facilities with distinct coordinates and names.

For Nairobi County, the Authoritative data drawn from the Kenya Master Health Facility List[12] (Kenyan Ministry of Health) show that in Nairobi County exist 1.288 health facilities of every type with 1122 of them having coordinates. The remaining 166 health facilities cannot be visualized in the authoritative data map layer, but they are taken into account when calculating completeness for OSM data. According to part of the literature[13] that implies 74,6% completeness.

[12] https://kmhfl.health.go.ke/#/home

[13] Barron, Neis, & Zipf (2014)Girres & Touya (2010), Haklay (2010), Jackson et al., (2013), Jokar Arsanjani, Barron, Bakillah, & Helbich (2013), Jokar Arsanjani, Mooney, Zipf, & Schauss (2015)

When visualizing those numbers on the map (Figure 10) it is easy to observe that this percentage is not reflecting the ground truth. There are several takeaways from its visual examination. Obviously, there is no high overlap of data, which means that the completeness as measured only by the number of objects is highly doubtful. The OSM data seem to form clusters in the areas of the informal settlements of Nairobi outnumbering the authoritative data. In all other areas, it seems that this phenomenon is reversed. Regarding their spatial distribution, this inconsistency in OSM data can be studied by zooming in on lower administration levels. Nairobi County consists of 17 sub-counties or constituencies which are then divided into 85 wards.

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct ?osm_id ?hasgeometry ?name WHERE {
{SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare ?healthcare .
      ?osm_id osmkey:name ?name .
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
 ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare ?healthcare .
  ?osm_id osmkey:name ?name
    filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:dispensing ?dispensing .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
```

Figure 9. Part of the SPARQL query was used to capture all the health facilities from the triplified OSM dataset of Nairobi County.

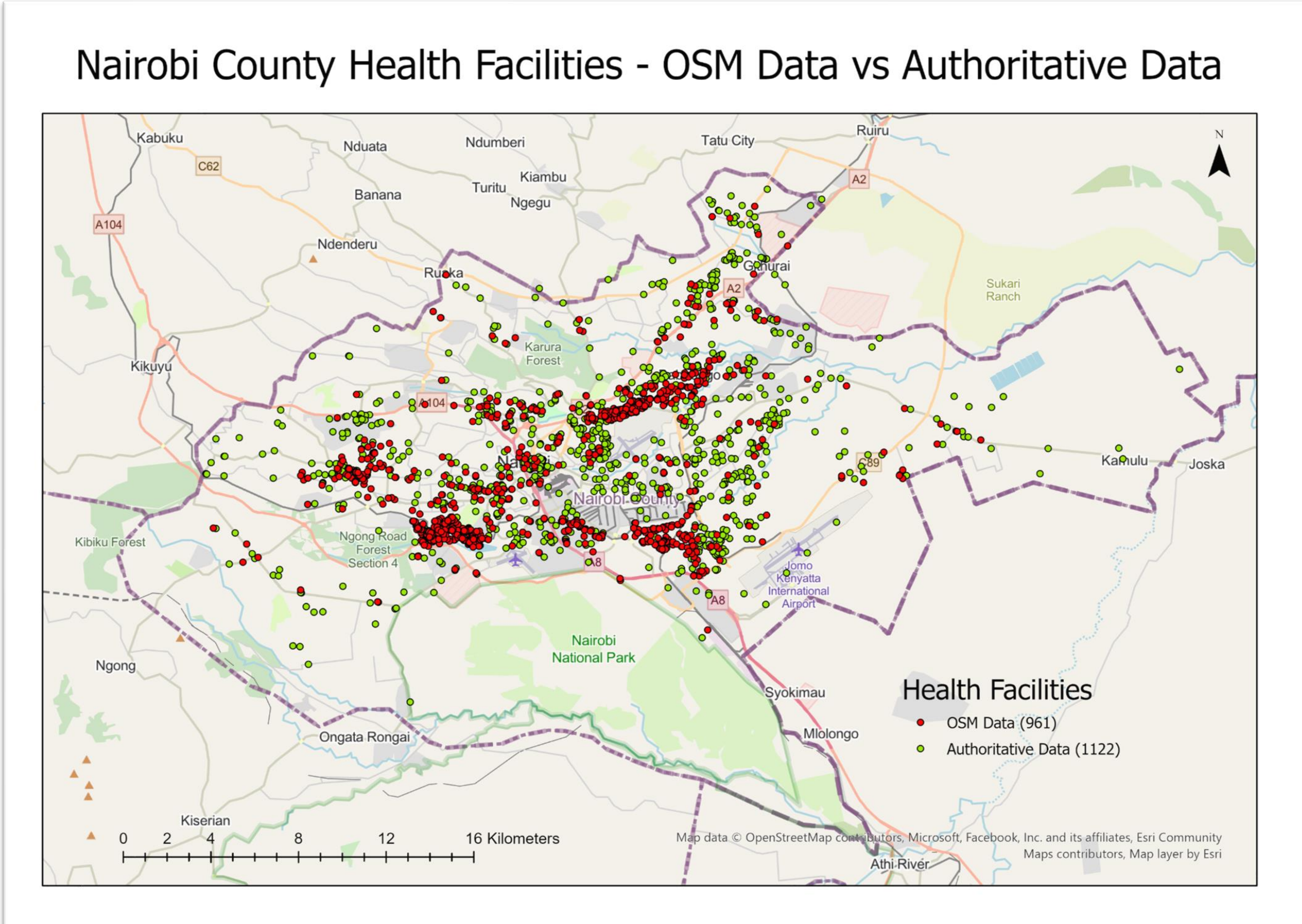


Figure 10. Map of Nairobi County showing the health facilities according to OSM (red) and the Kenyan Ministry of Health (green).

This clustering in the informal settlements can be observed also in Table 2 where we can see the number of OSM health facilities objects in each constituency level of administration along with the contained informal settlements. In the constituencies with informal settlements, there is a significantly higher percentage of OSM object completeness and a high number of health facility objects that outnumber the authoritative data. To get those results (Table 2 & Figure 3) we compared health facilities between OSM and the authoritative dataset in every constituency separately based on matching or similar names.

Table 4. Completeness of Health Facilities in OSM for Nairobi County at the constituency level, compared to KMH data. In bold are the constituencies having informal settlements inside their administrative boundaries.

Constituency (Informal Settlement)	Number of health facilities in authoritative data	Overlapping - Number of health facilities in both datasets	% Completeness – Percentage of authoritative data objects that exist in OSM	OSM
Mathare (Mathare)	**15**	**13**	**86.67**	**139**
Embakasi North (Korogocho)	**24**	**8**	**30**	**20**
Ruaraka	55	7	12.72	68
Kamukunji	72	6	8.33	7
Makadara (Mukuru)	**55**	**14**	**25.45**	**47**
Embakasi West	52	5	9.61	7
Embakasi Central	56	4	7.14	4
Embakasi East	104	16	15.38	19
Kasarani	75	20	26.67	22
Roysambu	111	11	9.91	17
Westlands (Various villages)	**133**	**27**	**20.3**	**42**
Dagoreti South	54	2	3.7	13
Dagoreti North (Kawangware)	**104**	**28**	**26.92**	**128**
Langata (Various villages)	**91**	**17**	**18.68**	**24**
Kibra (Kibera)	**102**	**54**	**52.94**	**253**
Starehe (Mukuru)	**129**	**28**	**21.70**	**70**
Embakasi South (Mukuru)	**54**	**15**	**27.77**	**81**
Total	**1286**	**275**	**21.38**	**961**

It is interesting to observe the numbers in Table 5 for Mathare. The only constituency that is entirely an informal settlement is Mathare. Looking at the completeness of health facility OSM data there, we see that out of the 15 health facilities present in the authoritative data 13 are also present in OSM of which 8 match in name while the remaining 5 had to be found through the description of the location in the authoritative data or rational assumptions such as analysis of abbreviations (e.g SHOFCO is standing for the Shining Hope for Communities NGO). Nevertheless, the completeness of OSM data is significantly high compared to other constituencies, reaching 86,67%. At the same time, 126 health facilities objects in OSM do not appear in the authoritative data. Those numbers call for looking into the other informal settlements of Nairobi County but not as a part of a constituency. To do so it is necessary to zoom in more and reach the ward level of administrative boundaries. This is due to the very fact that those settlements are informal, so they spread across any level of administrative units. In Table 3 we can see the 5 major informal settlements of Nairobi, their respective OSM and authoritative data their overlapping, and OSM completeness (omission) and commission.

Table 5. Completeness of Health Facilities in OSM for Nairobi's largest informal settlements compared to KMH data.

Informal Settlement	Number of health facilities in authoritative data	Overlapping - Number of health facilities in both datasets	% Completeness – Percentage of authoritative data objects that exist in OSM	OSM	Commission[14]
Kibera	42	31	73.81	218	187
Mathare	15	13	86.67	139	126
Korogocho	2	2	100	17	15
Mukuru	22	14	63.63	108	94
Kawangware	18	4	22.22	67	63
Total	**99**	**64**	**64.64**	**549**	**485**

For Kibera, completeness is 73.81%, while for the whole constituency of Kibra (Kibera is part of Kibra) is 52.94% while 187 health facilities in OSM do not appear in the authoritative data. For Korogocho the 100% completeness is not of much importance as there are only two objects in the authoritative data, but the extra 15 facilities are showing a possible higher quality of the OSM dataset. In Figure 11 the chart shows the outnumbering of authoritative data by OSM health facilities data. The overall OSM health facilities completeness in the informal settlements accounts for 64.64% which is three times higher than the average completeness of the whole

[14] Commission as explained in paragraph 3.3.1 (Figure 5) counts the OSM objects that do not appear in the authoritative data and is part of the ISO 19157 for Geographic Data quality assessment.

county (21.38%) and five times more than the areas without informal settlements (12.26%). Figures 13-17 visualize the completeness assessment on the maps of the informal settlements.

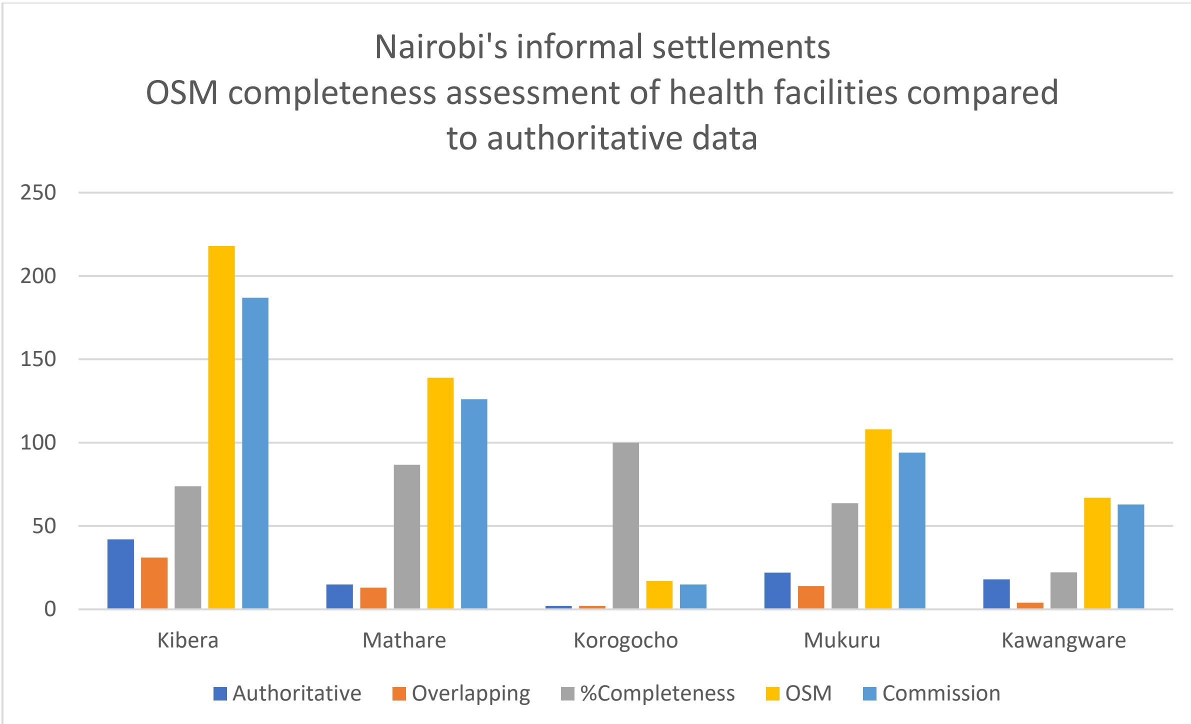


Figure 11. Chart showing four metrics of the completeness of OSM data compared to the authoritative dataset for Nairobi's Health Facilities. The number of objects for both datasets, their overlap, the completeness of OSM based on the authoritative data, and the excess of data (commission) in OSM.

Nairobi County Constituencies - OSM Health Facility Completeness

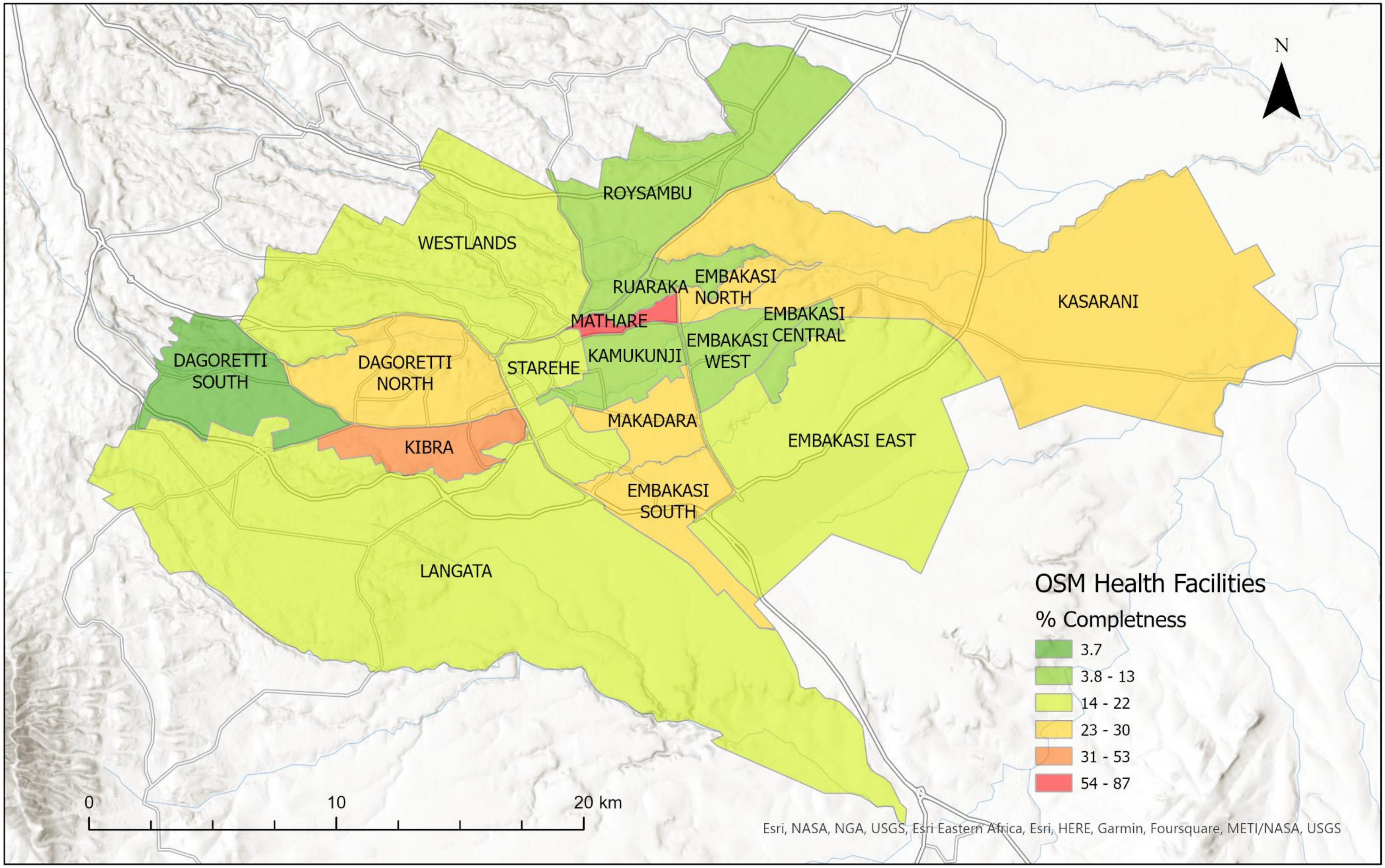


Figure 12. Nairobi County - Level of OSM Health Facilities Completeness in Constituency level. Comparison made with Health Facilities data from the Kenyan Ministry of Health .

Kibera Informal Settlement Health Facilities - OSM vs Authoritative Data

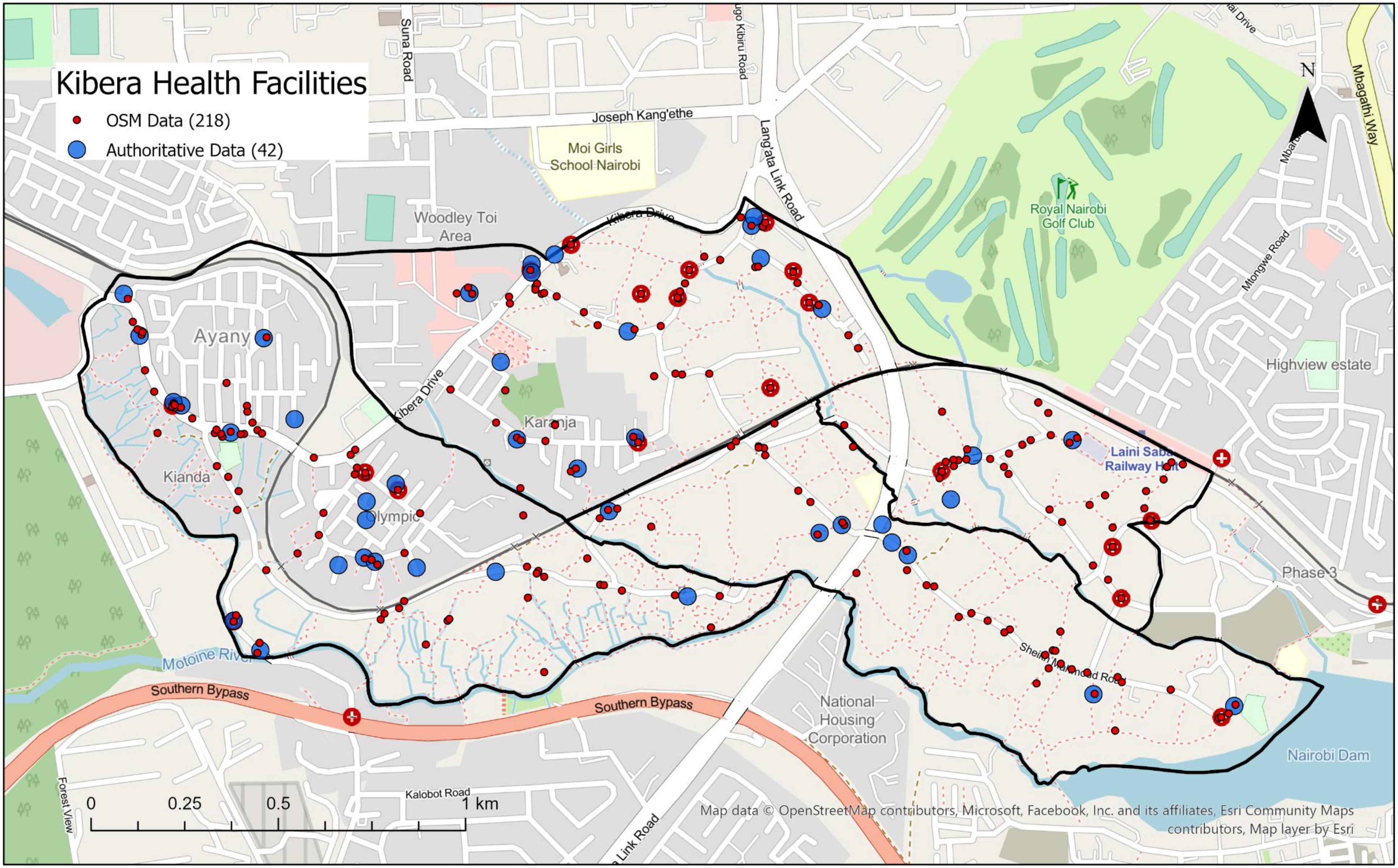


Figure 13. Map of Kibera informal settlement in Nairobi County showing the location and the number of health facilities according to OSM and authoritative datasets. The signs with the cross are part of the OSM basemap and are not part of the legend.

Mathare Informal Settlement Health Facilities - OSM vs Authoritative Data

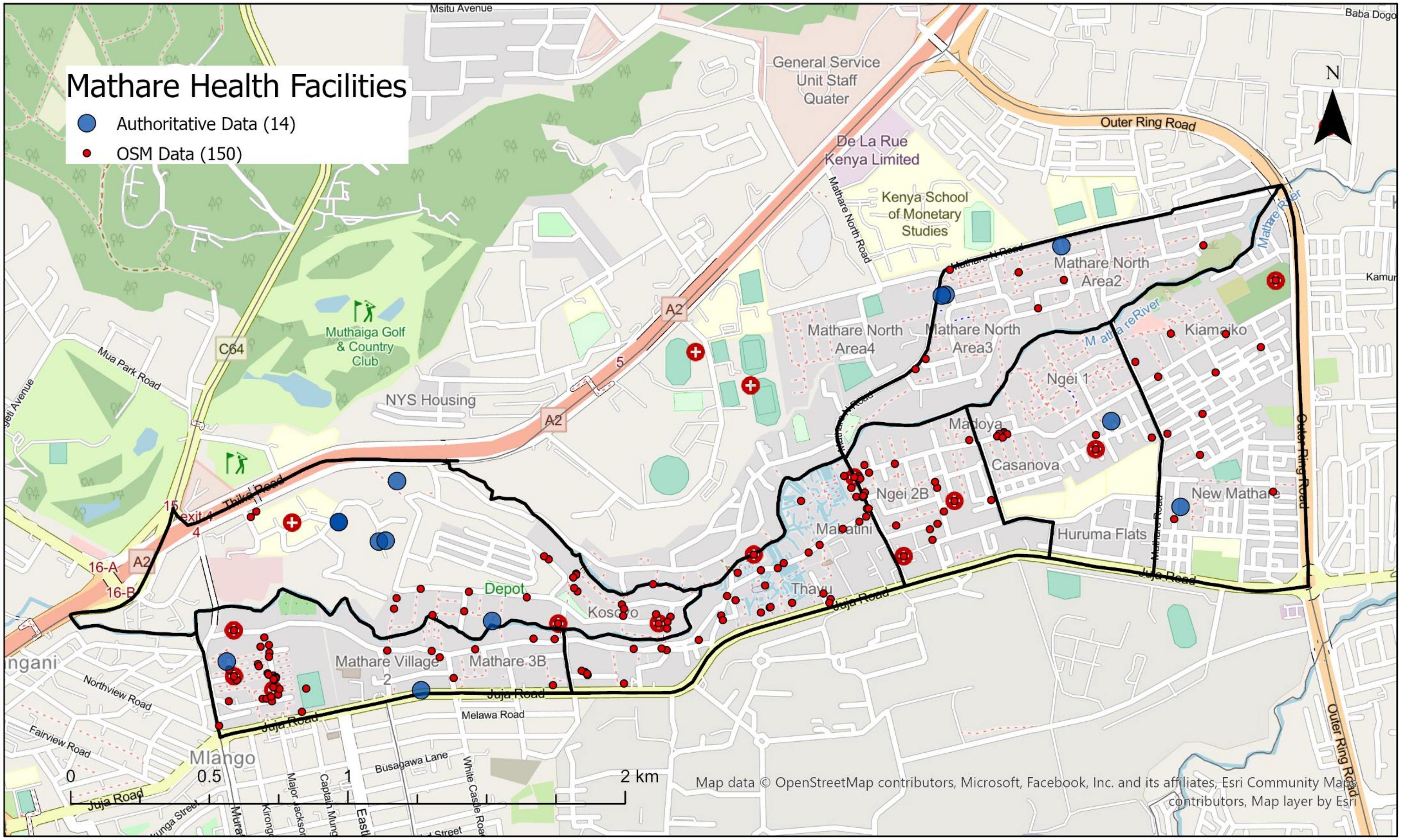


Figure 14. Map of Mathare informal settlement in Nairobi County showing the location and the number of health facilities according to OSM and authoritative datasets. The signs with the cross are part of the OSM basemap and are not part of the legend.

Korogocho Informal Settlement Health Facilities OSM vs Authoritative Data

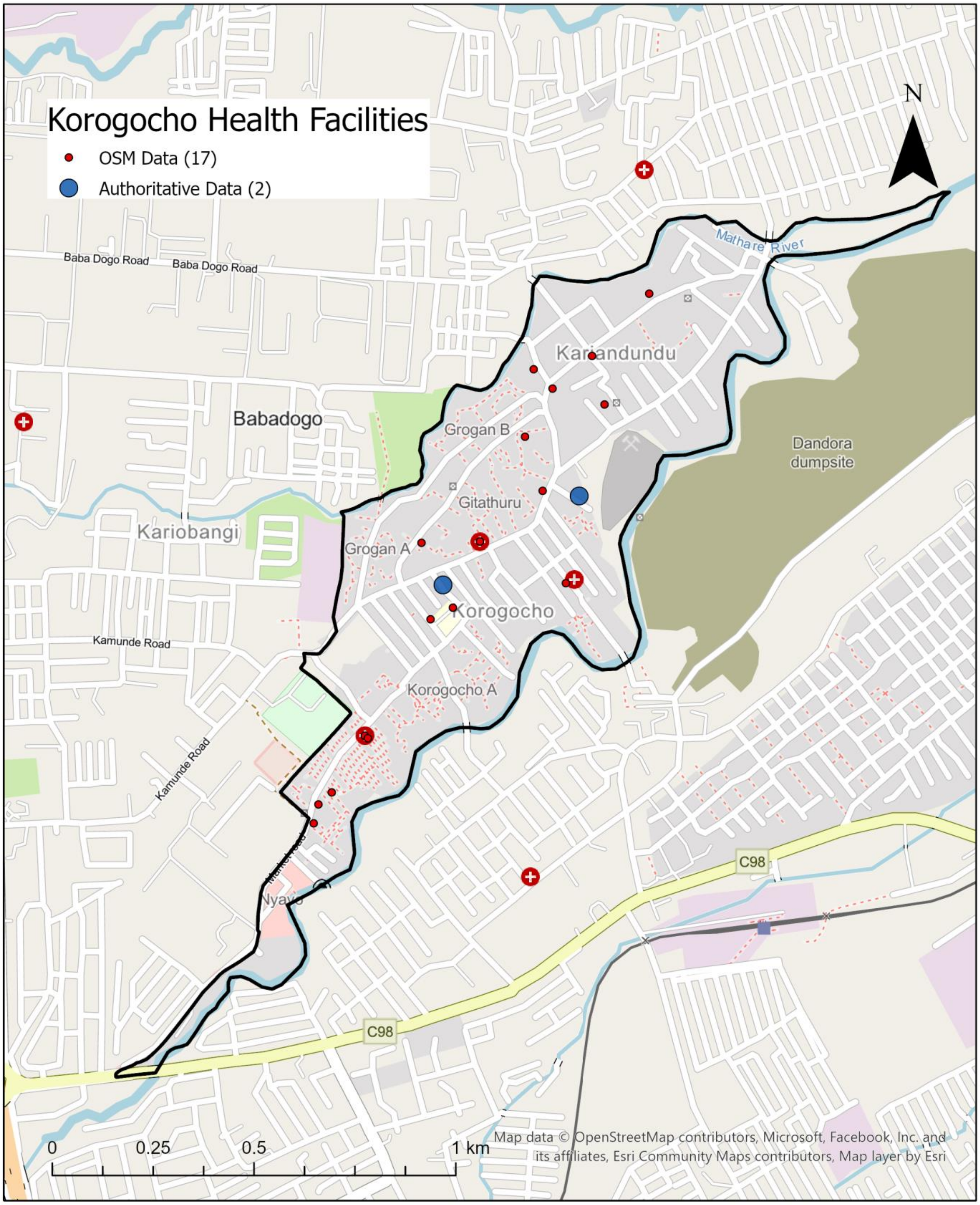


Figure 15. Map of Korogocho informal settlement in Nairobi County showing the location and the number of health facilities according to OSM and authoritative datasets. The signs with the cross are part of the OSM basemap and are not part of the legend.

Kawangware Informal Settlement Health Facilities
OSM Data vs Authoritative Data

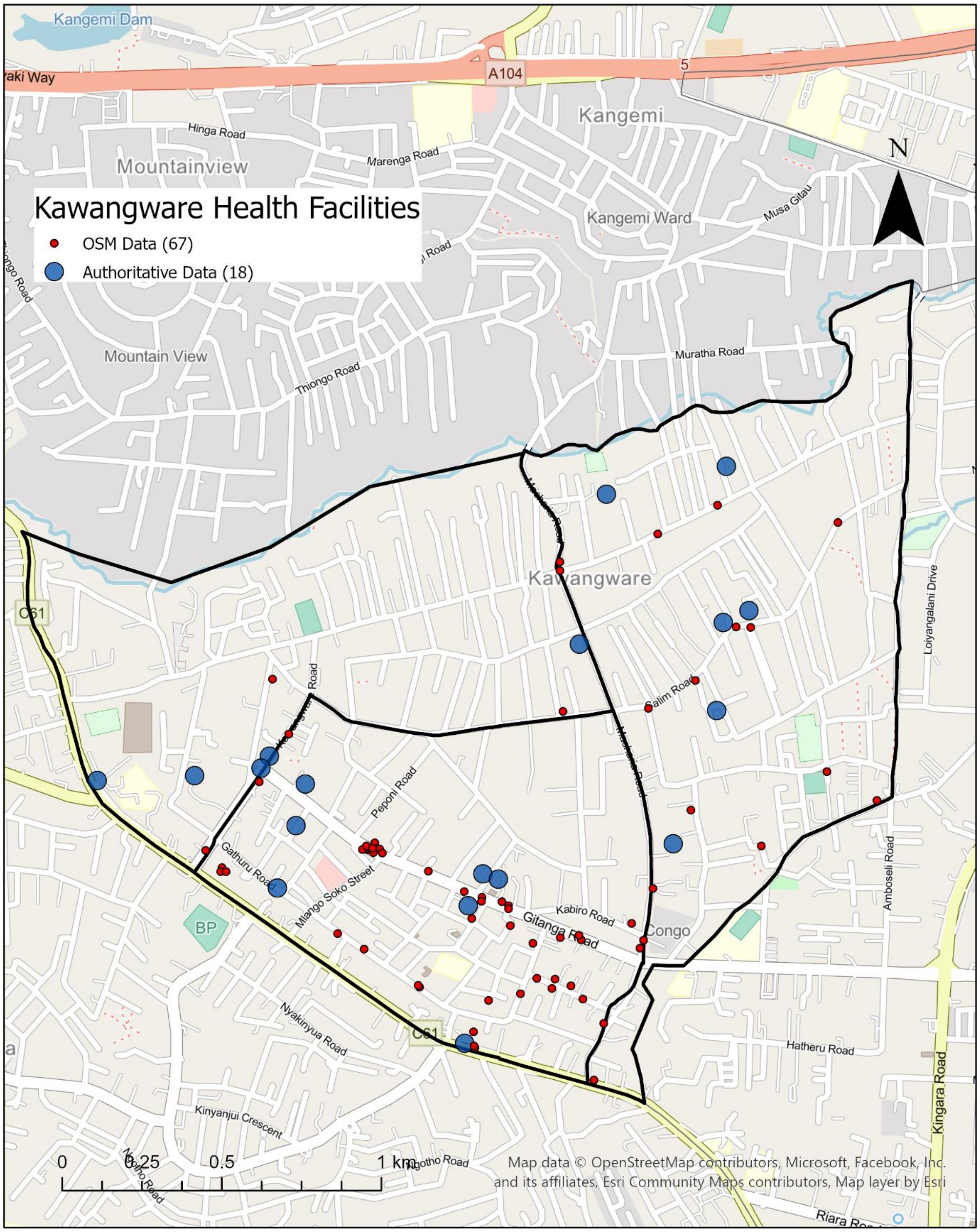


Figure 16. Map of Kawangware informal settlement in Nairobi County showing the location and the number of health facilities according to OSM and authoritative datasets.

Mukuru Informal Settlement (part) Health Facilities
OSM Data vs Authoritative Data

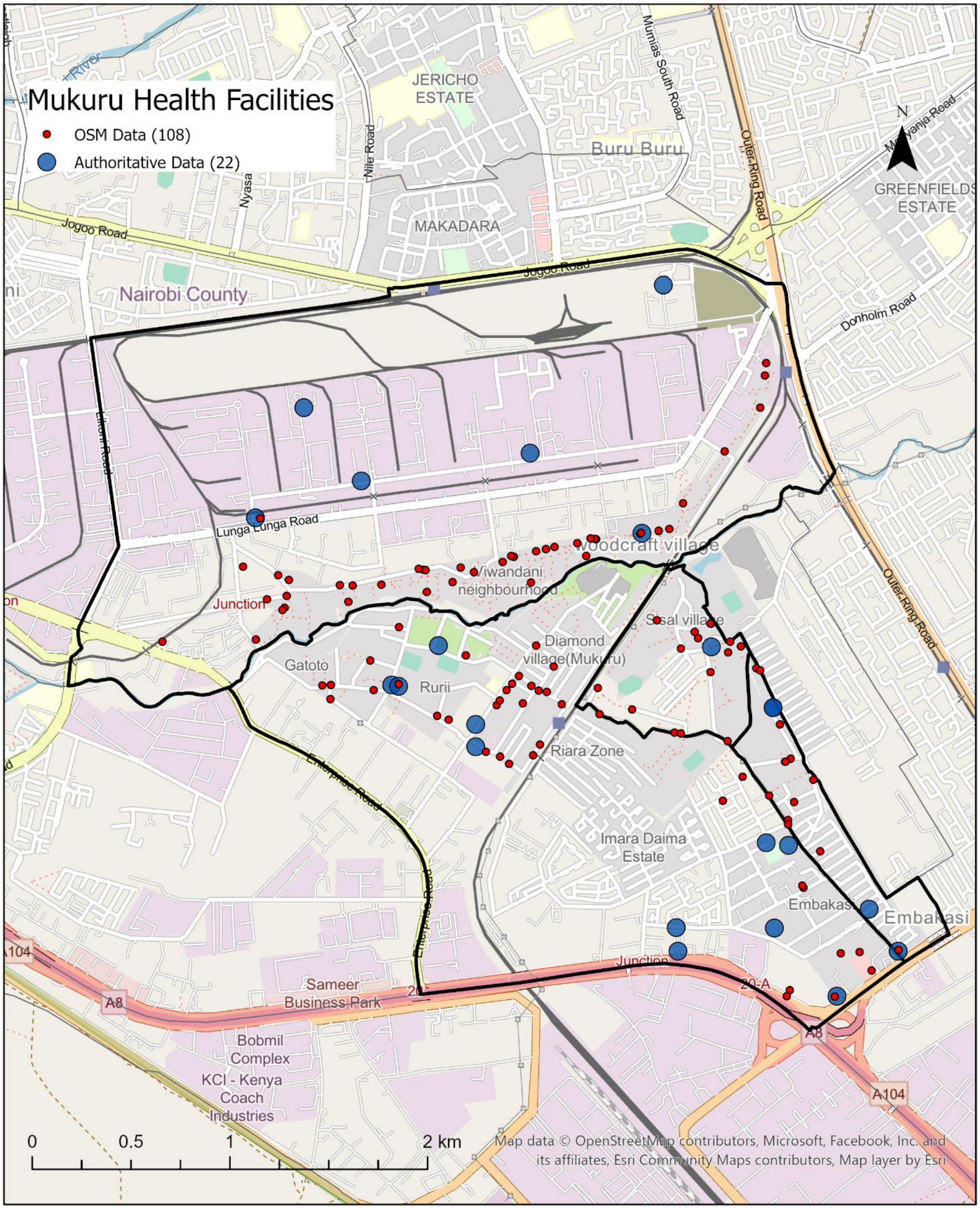


Figure 17. Map of Mukuru informal settlement in Nairobi County showing the location and the number of health facilities according to OSM and authoritative datasets.

Attribute completeness

The important attributes of the health facilities as identified by the users' needs analysis are the level according to the national classification system, the type of operator, the medical services offered, the contact details, and the equipment available. Table 4 and Figure 18 show the results of this type of quality assessment.

Classification Level

There are no tags in Nairobi's OSM vocabulary implying the level of the health facility according to the national classification system.

In the authoritative data, 783 out of 1289 (**60.75%**) facilities are classified according to the Kenya Essential Package for Health (KEPH) designed by the Kenyan Ministry of Health with levels 2-6. Level 1 is the body of volunteers countrywide and they are not included in the authoritative dataset.

Operator Type / Ownership

In OSM 480 out of 961 facilities in Nairobi County indicate the type of operator which is **49.94**%. In Mathare informal settlement, the percentage is **51.08%**

In the authoritative data of the Ministry of Health, this percentage is **100%**.

Contact

In OSM out of the 961 health facilities in Nairobi County, 75 have a tag with the key "`contact`" which is **7.8%.** In the Mathare informal settlement, there is only one facility having contact details.

For the Authoritative data of the Ministry of Health, this percentage is **100%**.

Medical services

There are 35 valid tags (out of 37) in OSM using the key "`medical_service:*`" indicating the services provided in each facility by 35 unique values. In total, those tags are applied in 319 out of the 961 health facilities of OSM Nairobi County which is **33.19%.** If we zoom in on the informal settlements this percentage is rising. In Mathare 122 out of 139 health facilities in OSM include at least one tag with "`medical_service:*`" as a key which is **87.76%**. For the authoritative data of the Ministry of Health, 574 out of 1289 facilities are accompanied by information concerning the offered services which is **55.46%**.

Available equipment

There are no tags in Nairobi's OSM vocabulary describing the available equipment in health facilities. Similarly, there are no such tags in the data of the Ministry of Health.

Table 4. Health facilities of Nairobi County - Attribute completeness for the authoritative and OSM datasets measured in the percentage of objects that contain the attributes as described by the users' needs analysis.

	Authoritative %	OSM %
Classification	60.75	N/A
Operator	100	49.94
Contact	100	7.8
Services	55.46	33.19
Equipment	N/A	N/A

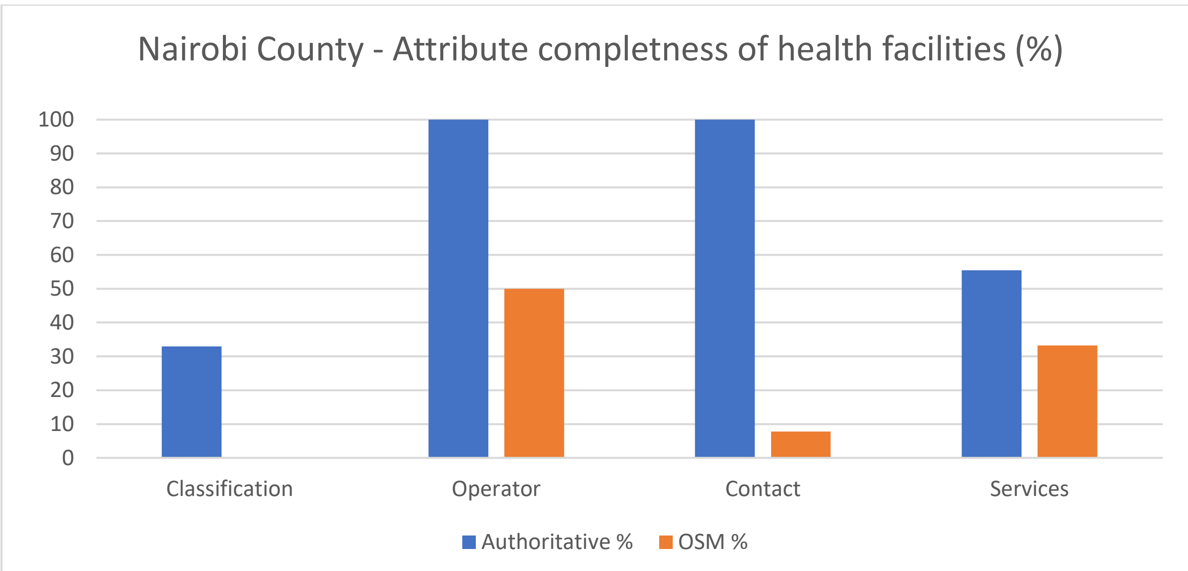


Figure 18. Chart showing the comparison of the attribute completeness of the health facilities in Nairobi's County OSM and authoritative datasets.

Attribute Accuracy

This metric measures to what extent the attributes of OSM reflect in meaning those of the Authoritative dataset which are considered as the ground truth. By default, the Attribute accuracy of the authoritative data is 100%. Table 5 shows the results of the assessment.

Classification of Health Facilities

In the authoritative data the information about the class, according to the national classification system, is given by the tags "Level 2-6". In OSM Kenya the vocabulary has no tags indicating the Classification of the Health Facilities. Thus, OSM's data Attribute Accuracy for the health facilities classification is not available.

Medical Services

In the authoritative data, there are 129 distinct tags describing the medical services of the Health Facilities while in OSM there are 35 valid tags. From the 129 distinct authoritative tags, 21 are used in OSM showing a **16.27%** of relative attribute accuracy.

Operator Type / Ownership

In the authoritative data, 14 different tags are describing the ownership status. This is described in OSM Kenya by the key "`operator_type`". The matching to the authoritative data keys are 3 meaning 21.43%. However, all of the tags in the authoritative can be grouped into four categories i.e. public, private, NGand O, religious. The same can be done for the OSM "`operator_type`" values. So according to the user needs analysis which indicates as important information the distinction between private, public,c, NGOs, or faith-based facilities the attribute accuracy can be considered **100%**.

Contact

Both for the authoritative and the OSM data "contact" is used as a tag. The attribute accuracy of OSM here is 100%.

Table 5. Health facilities of Nairobi County - Attribute Accuracy for the authoritative and OSM datasets measured in the percentage of matching OSM values with the respective Authoritative tags.

	Authoritative %	OSM %
Classification	100	N/A
Operator	100	100
Contact	100	100
Services	100	16.27

Positional accuracy

Positional accuracy was not identified as a need of the users because according to them even at a distance of up to 20 meters, accuracy does not play a significant role. However, some interesting numbers are coming out from the analysis of the authoritative data. 12.88% of the health facilities do not have coordinates while this percentage for OSM is 0%. Furthermore, form the 1122 KMHFL objects that have coordinates 203 are completely misplaced. This means that the coordinates did not match the descriptive location given in the facility's description coming from the authoritative dataset. In total the authoritative data have a percentage of 70.8% inherent positional accuracy. These descriptions of locations in the authoritative data have proved really helpful for checking suspicious pairs of coordinates and then crosschecking with OSM and/or Google Maps. For those objects, the address in their description had always been matching with the OSM coordinates for the same object within 10m of distance. So, from the 275 matching health facilities (with or without coordinates) of the two datasets the positional relative accuracy of OSM has been 100%.

5.1.2 Nairobi County WATSAN

The present research, when referring to WATSAN is about drinking water points, communal sanitation points, and handwashing facilities as defined in the users' needs analysis. It must be noted that it was not possible until the moment of writing to find and acquire authoritative data for Nairobi County. For this reason, data from UN-Habitat were used as a reference dataset. UN-Habitat has produced WATSAN data for Nairobi's largest informal settlements (Mathare, Kibera, Kawangware) which are also the main clusters of WATSAN data for OSM (Figures 22 & 23). This is not a surprise as the informal settlements are super dense urban areas where access to clean water, proper sanitation, and hygiene are major concerns for public health. Specifically, for OSM, the Informal Settlements in Nairobi are the sole humanitarian mapping projects grounds.

Completeness

The fact that water and sanitation points usually do not have names as the health facilities do, makes it impossible to measure the overlapping of OSM data and the reference dataset except in the case of Kawangware where there is 100% overlapping (see appendix D) may imply data borrowing between the two datasets. The coordinates of the objects in Kibera and Mathare cannot provide any assistance in calculating the overlapping because of the density of the points. The only metric available is comparing the net number of points in each of the three categories of interest (water points, communal sanitation points, handwashing points). In Figures 19, 20 & 21 we can see the results of the comparisons (see Appendix D for a detailed map overview of those numbers).

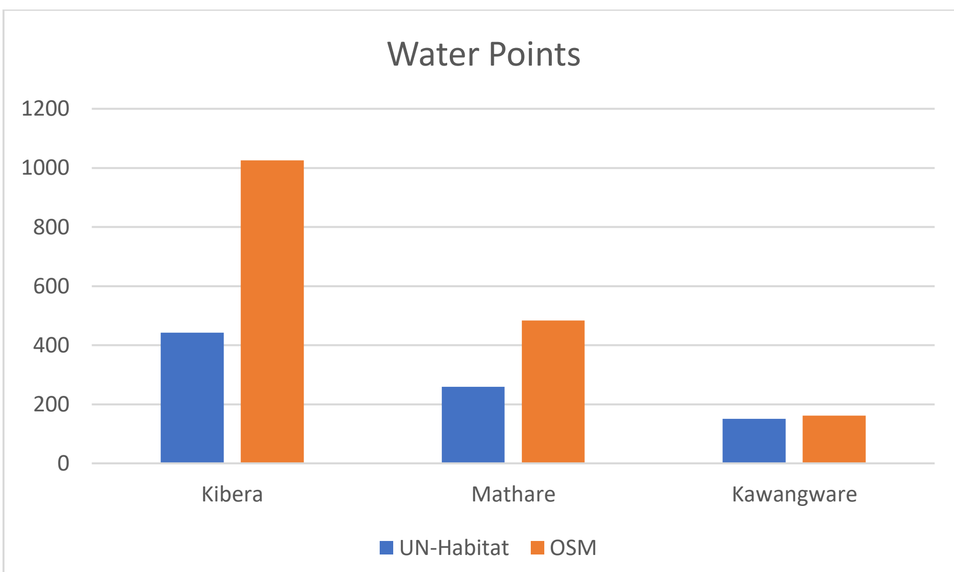


Figure 19. Comparison of the number of water points in three informal settlements of Nairobi between UN-Habitat and OSM datasets.

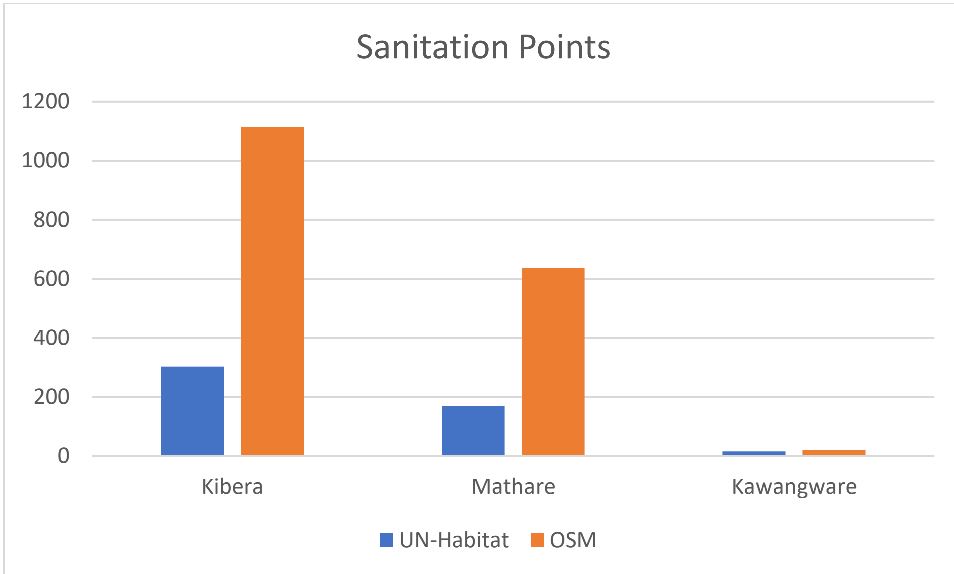


Figure 20. Comparison of the number of communal sanitation points in three informal settlements of Nairobi between UN-Habitat and OSM datasets.

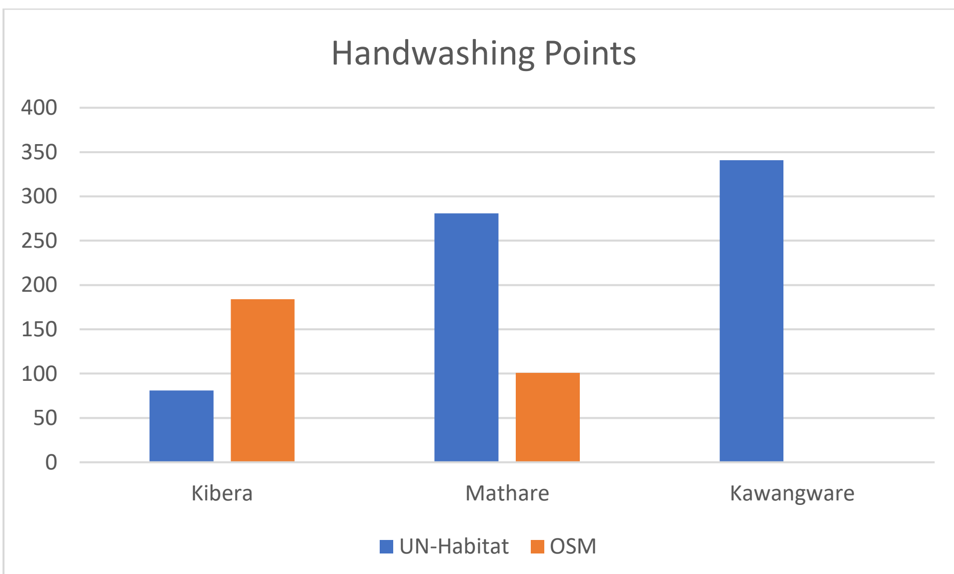
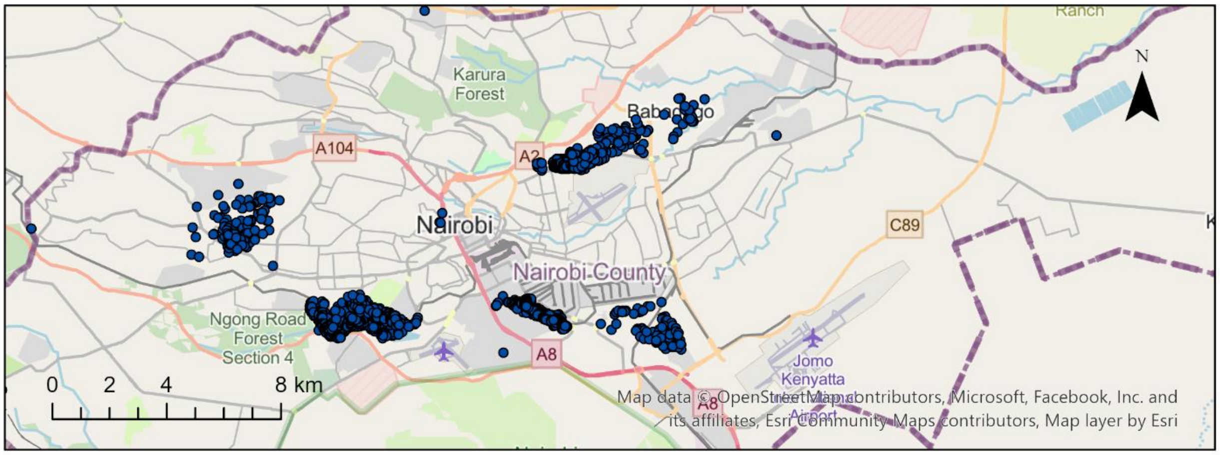
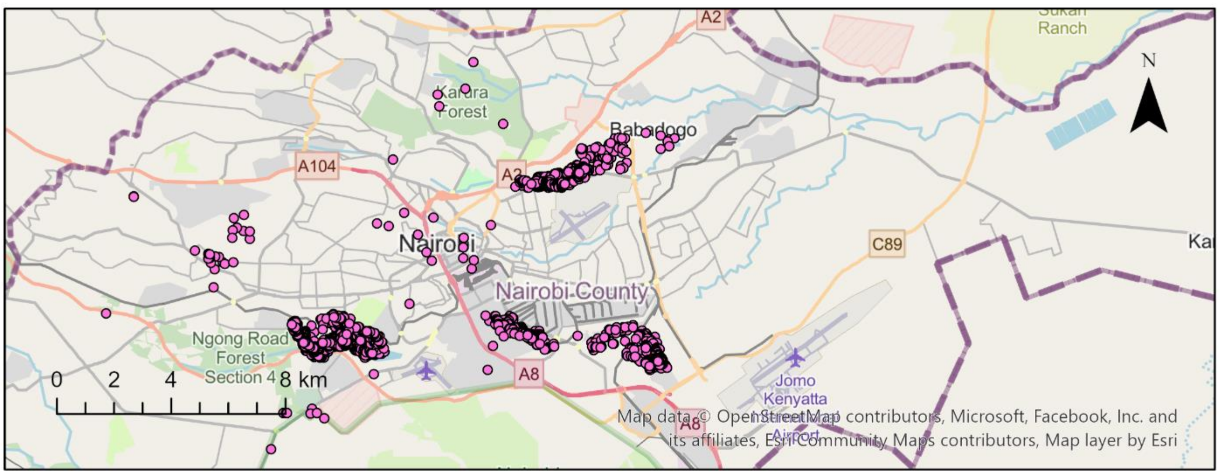


Figure 21. Comparison of the number of handwashing points in three informal settlements of Nairobi between UN-Habitat and OSM datasets.

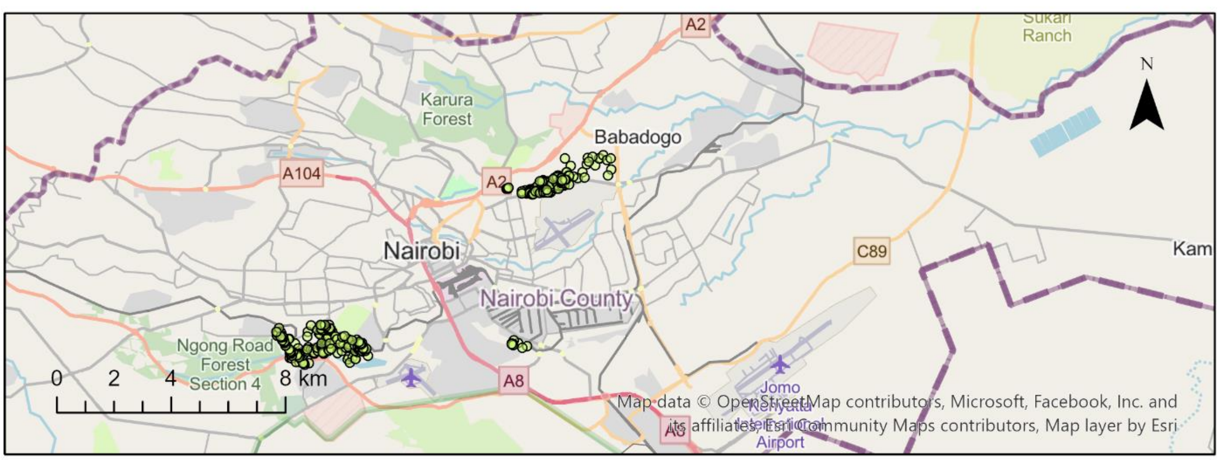
Nairobi County
OSM WATSAN Data



Water Points



Communal Sanitation Points



Handwashing Points

Figure 22. Maps of Nairobi County showing the clustering of OSM data concerning, water points, communal sanitation, and handwashing points.

Nairobi County - UN-Habitat WATSAN Points

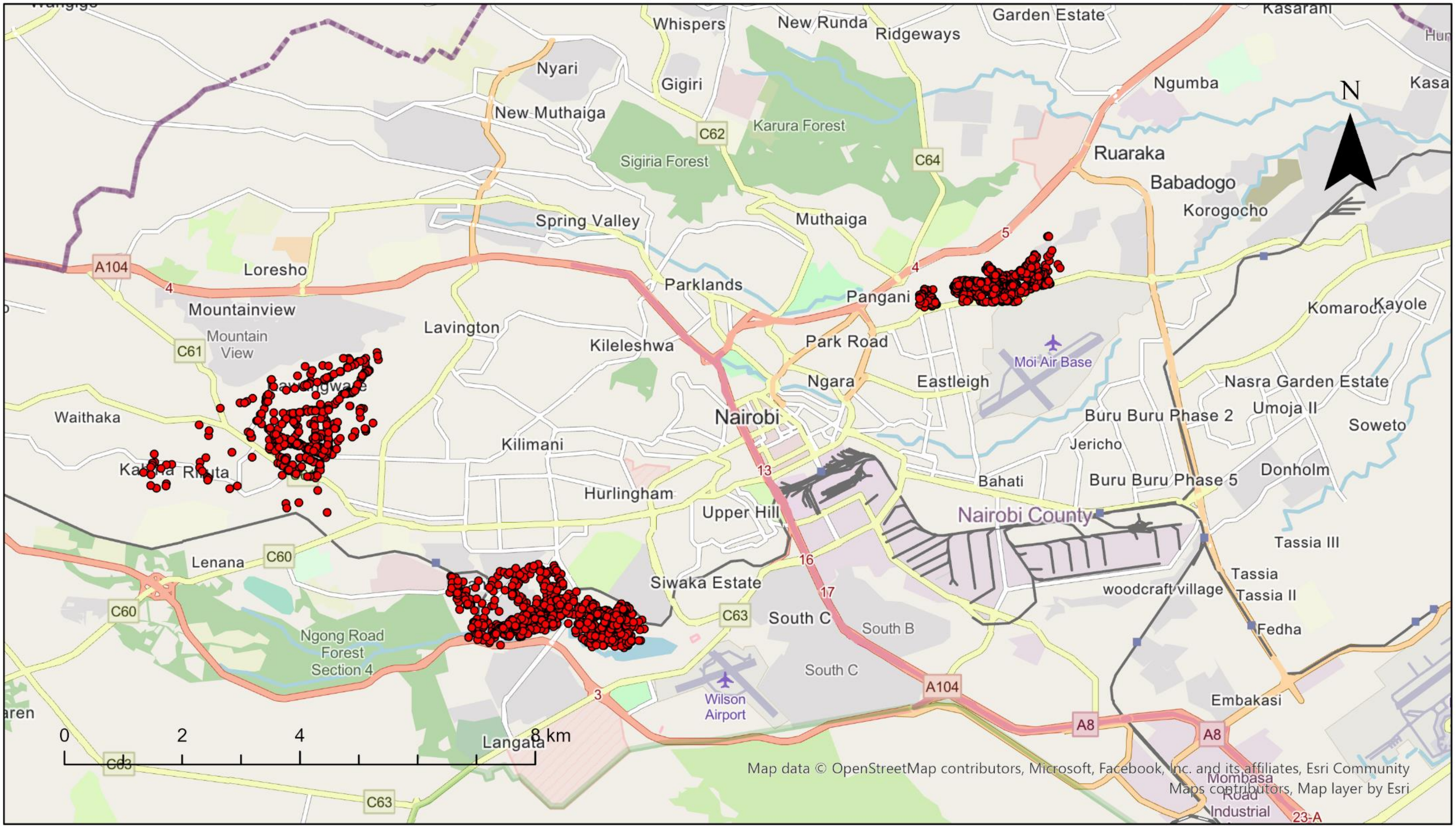


Figure 23. Map showing the areas where the WATSAN data collection of the UN Habitat took place i.e., Kawangware, Kibera, Mathare informal settlements

5.1.3 Nairobi County Road Network

The authoritative data for the road network of Nairobi County, coming from the Kenya Roads Board, give a total length of 4713.91 Km. In OSM the whole road network of Nairobi County accounts for 7066.28 km. The overlapping of the two networks amounts to 4392.29km which means 93.2% completeness of the OSM road network for Nairobi County. This calculation was made by using the "select by location" geoprocessing tool in QGIS (Figure 24). This means that in OSM exist 2674 km that do not appear in the KRB dataset and 321.62 km in the latter do not appear in OSM (Figure 25). Those numbers suggest a network of 7387km if the datasets were combined (Figure 26).

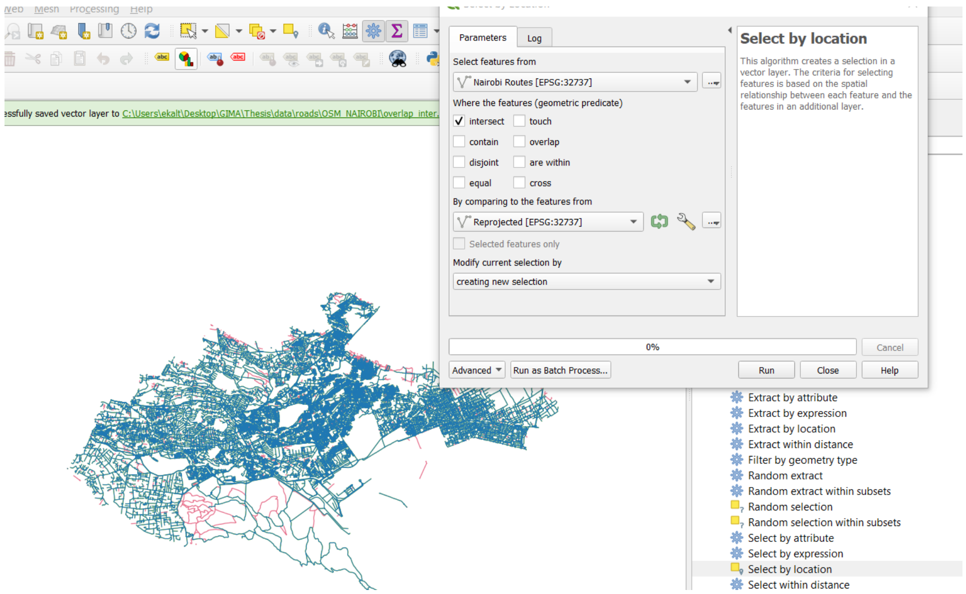


Figure 24. Select by location in QGIS to compare between KRB and OSM road network datasets for Nairobi County.

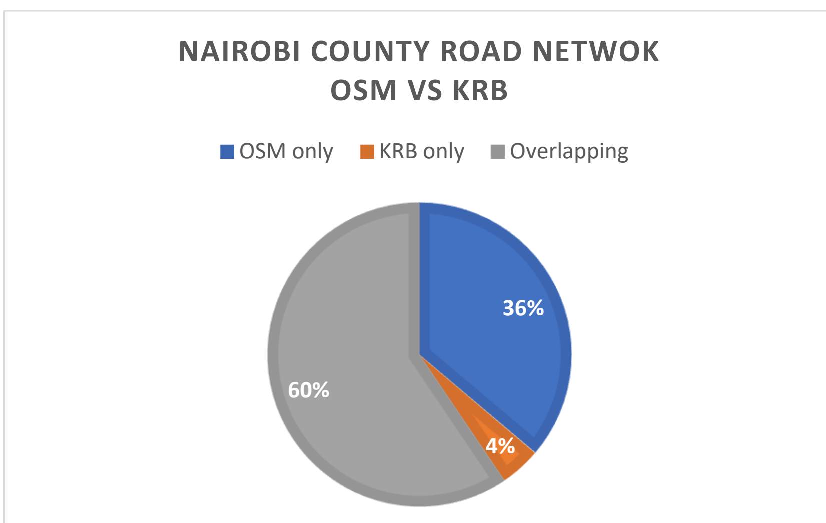
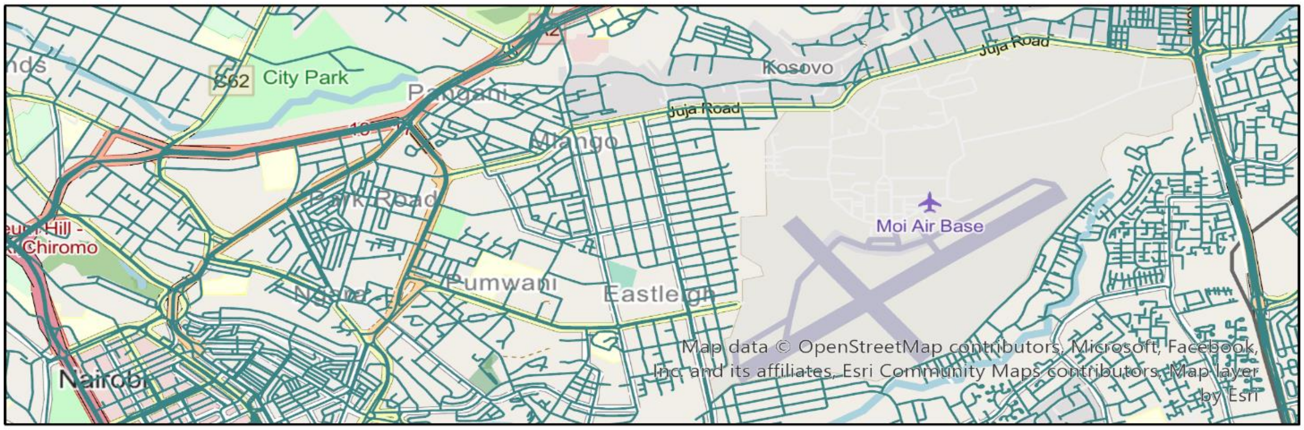


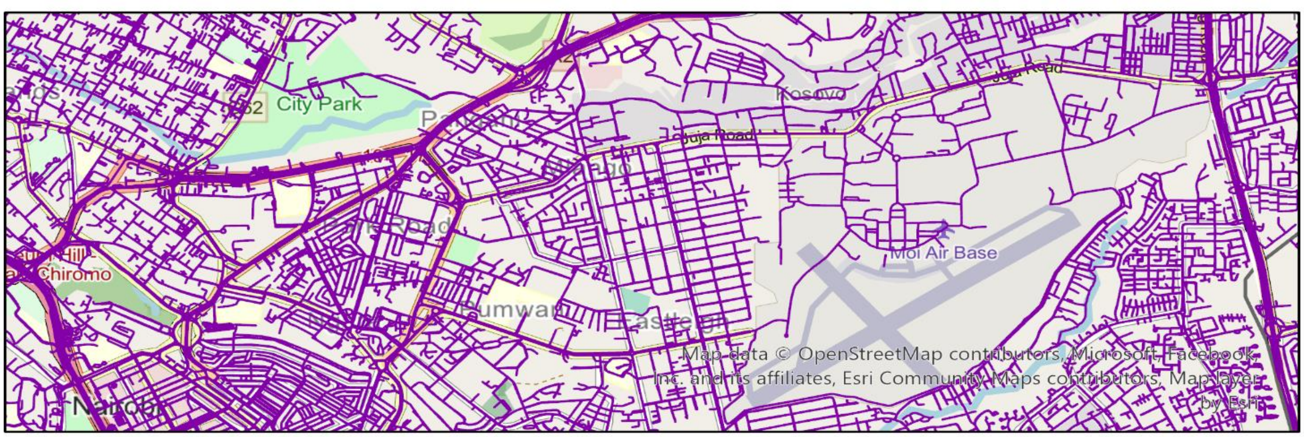
Figure 25. Chart showing the completeness of OSM road network when compared to the KRB authoritative data.

Nairobi County
KRB, OSM and combined Road Network



Kenya Roads Board

openstreetmap.org



Kenya Roads Board & openstreetmap.org

Figure 26. Maps giving a visual comparison for part of Nairobi's Road network from KRB, OSM and their combined datasets.

By looking closer into the two datasets, we can see that roads are classified differently. The Kenyan roads classification system has according to the authoritative data from the Kenya Roads Board 16 distinct classes. Those span from A to G and their respective urban versions (Au - Gu) plus NEW, NR, and S (Table 6 & Figure 27). Those letters indicate the functional role of the road segments which are explained in Table 8. In Nairobi County class F is absent. For OSM the values used for the key "highway" that represent roads and not objects are shown in Table 7 & Figure 28.

Table 6. Road classes of the authoritative data present in Nairobi County and their length

Kenya Roads Board Classification	Length in km
A + Au	204.40
B + Bu	331.47
C + Cu	8.57
D + Du	569.45
E + Eu	58.75
Fu	139.01
G + Gu	1591.50
NEW	428.87
NR	1318.31
S	63.58
Total	**4713.91**

Table 7. OSM highway tags of Nairobi County and their lengths

OSM highway tags	Length in km
motorway	55.81
trunk	192.67
primary	126.81
secondary	270.39
tertiary	344.61
residential	4131.30
unclassified	433.72
service	1510.97
Total	**7066.28**

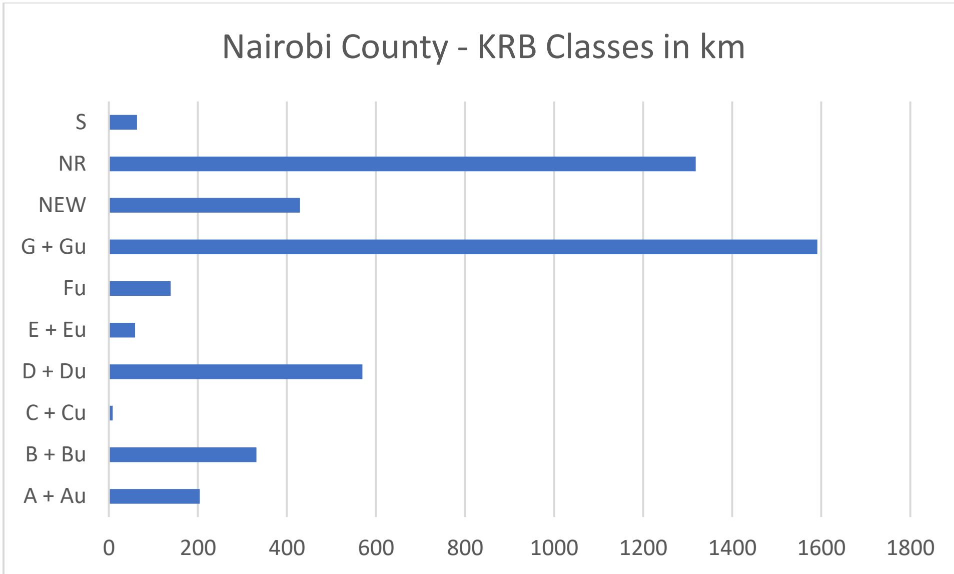


Figure 27. KRB for Nairobi - Total roads length of separate classes

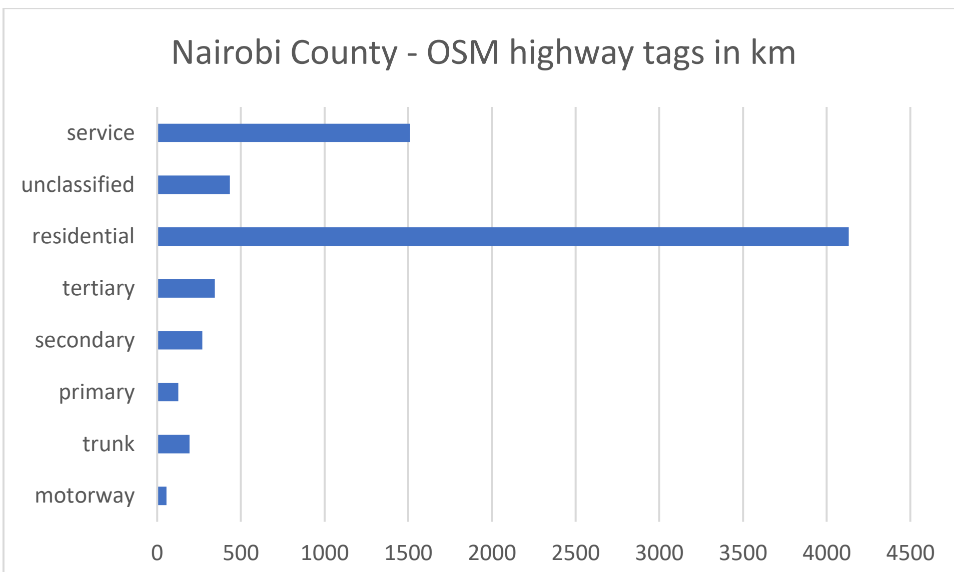


Figure 28. OSM Nairobi – Total roads length of different highway tags

To compare OSM road network data to the authoritative KRB data by type of road, there were several aggregations and assumptions to be made. First of all, for the KRB, each of the A-G classes were aggregated with their urban respective classes (Au-Gu). Then by following the traditional functional classification system (Forbes, 1999) we grouped those classes into:

- **Arterials or Trunks:** High-capacity roads that transfer the traffic.
- **Collectors:** Roads that collect the traffic and feed the arterials.
- **Local Roads:** Taking the traffic from properties to the collector.

To do the same for the OSM highway tags we had to make some compromises. The classes "tertiary" and "unclassified" were matching partially several of the authoritative classes. For simplifying the comparison, we classified them as "collectors" in the functional classification (Table 8) because most of their length matches classes C, D, Cu, and Du of the authoritative data. The comparison of the datasets by functional class shows again the that the OSM dataset the for Nairobi County roads network is more extensive in all classes (Figure 29).

Table 8. Reclassification of KRB and OSM road networks according to the traditional functional classification system

km	KRB classes	Description	Functional Class	OSM keys	km
599.45	S	Superhighway	Arterial Or Trunk	motorway, motorway_link	645.68
	A, Au	Major Arterial		trunk, trunk_link	
	B, Bu	Minor Arterial		primary, primary_link	
				secondary, secondary_link	
578.02	C, Cu	Primary	Collector	tertiary	778.33
	D, Du	Secondary		unclassified	
3536.44	E, Eu	Minor	Local Road	residential	5642.27
	Fu	Special Purpose		service	
	Gu	Local Access		residential	
	NEW			residential	
	NR			residential	

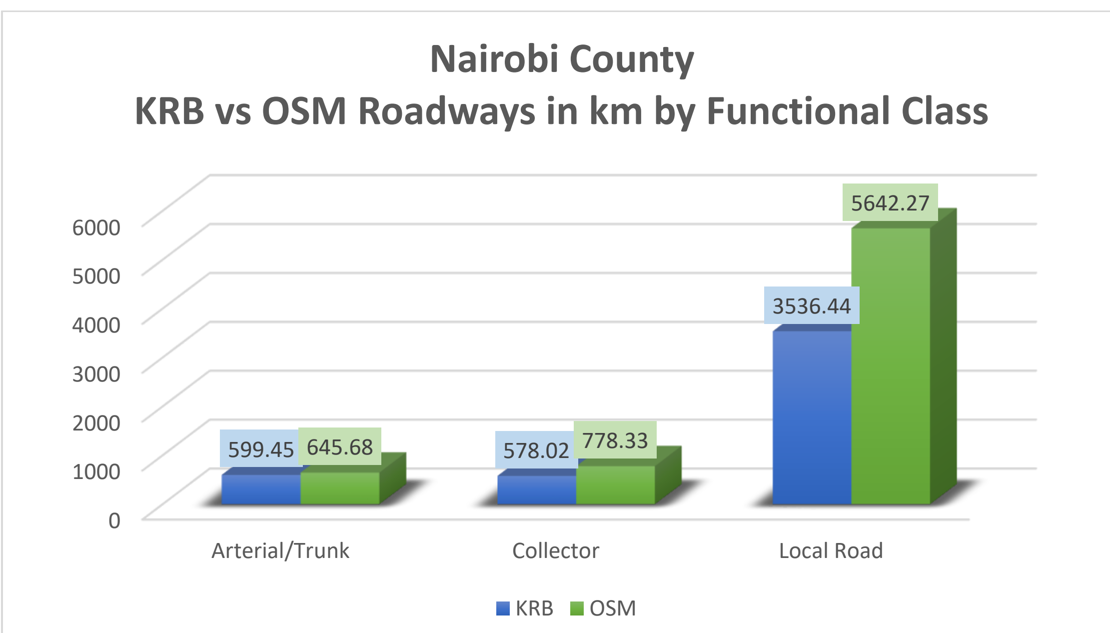


Figure 29. Comparison of Nairobi's KRB and OSM road networks according to the functional reclassification

5.1.4 Wajir County Health Facilities

Completeness

In Wajir County, the Authoritative data show 163 health facilities of all types while OSM data contain 89, which could suggest 54.6% completeness according to the relevant literature. As in the case of Nairobi, this can be far from the ground truth because of the OSM data that do not appear in the authoritative dataset. A comparison between OSM data and Authoritative was performed looking for matching (even partially) in names with similar coordinates. To facilitate the comparison this was made again at the constituencies level. What is interesting is that when the analysis for Wajir county began the triplified OSM data that we used did not include the mapping activity of the ongoing Humanitarian OpenStreetMap Team Nd Kenyan Red Cross project there. So, the results were as shown in Table 9. When the triplified dataset for Kenya was updated, including the contributions of the aforementioned mapping project the change was obvious for the Wajir East constituency with the completeness going as high as 55.81% from only 18.60% (Table 10 & Figure 31). In Figure 32 the results of the comparison are visualized on the map of Wajir county.

What is interesting to be mentioned here are cases where the administrative borderlines cross a settlement. This is the case of Bur Mayo which has two dispensaries, one on each side of the border. In the authoritative data, the distinction is done through the name given to the facility. For Wajir, it is called "Burmayow Dispensary" (The letter w at the end probably indicates that it is in Wajir's territory) while for Mandera it is called "Burmayo Dispensary". For OSM this is not necessary as the distinction derives from the enclosing OSM relations (administrative boundaries). What is maybe needed is an additional tag showing the administrating county for each object. Additional attention is needed for cases like the one of Korendile. In the query results, despite trying to eliminate all the duplicates (e.g. successful in the case of Wajir Referral Hospital), a point, as well as a polygon for the local health facility, are included in the results. The final clearing up of the data has also here as in the case of Nairobi been done manually in a CSV sheet by locating the duplicates of the names and then inspecting each one of them on the map

Table 9. Completeness of Health Facilities in OSM compared to KMH data for the County of Wajir (15/01/23)

	Authoritative	Auth. in OSM	% Completeness	OSM	Commission
Wajir North	20	10	50	11	1
Eldas	15	4	26.67	4	0
Tarbaj	23	11	47.83	12	1
Wajir West	25	10	40	12	2
Wajir South	34	8	23.53	8	0
Wajir East	43	8	18.60	11	3
Total	**160**	**51**	**31.87**	**58**	**7**

Table 10. Completeness of Health Facilities in OSM compared to KMH data for the County of Wajir (30/01/23)

	Authoritative	overlapping	% Completeness	OSM	Commission
Wajir North	20	10	50	11	1
Eldas	15	4	26.67	4	0
Tarbaj	23	11	47.83	12	1
Wajir West	25	10	40	12	2
Wajir South	34	8	23.53	8	0
Wajir East	43	24	55.81	42	18
Total	**160**	**67**	**41.87**	**89**	**22**

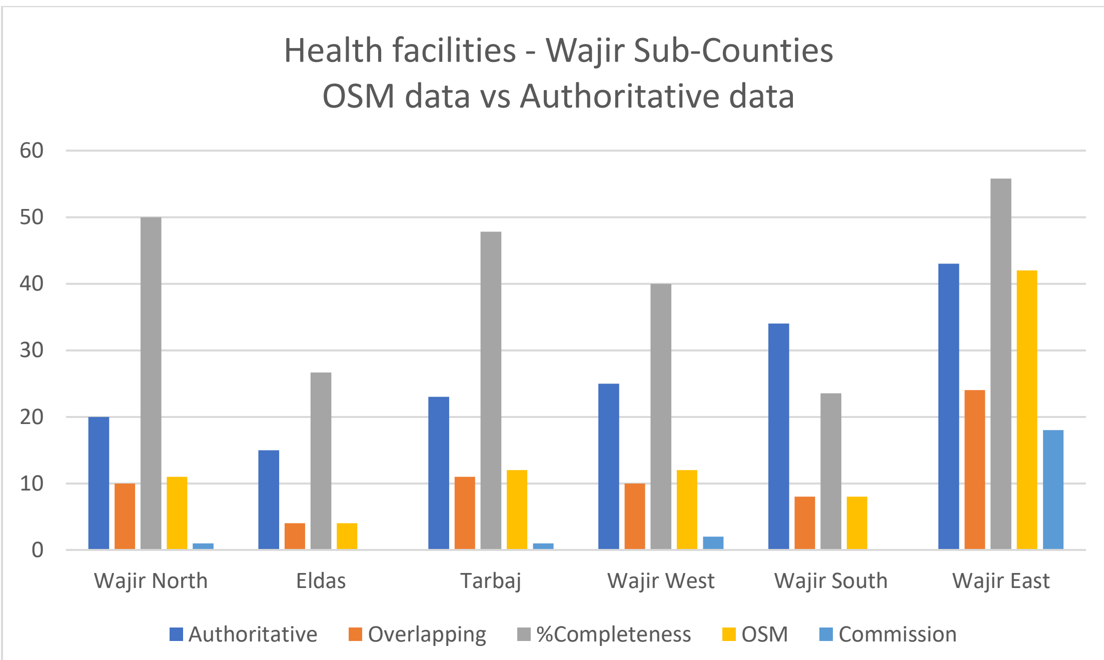
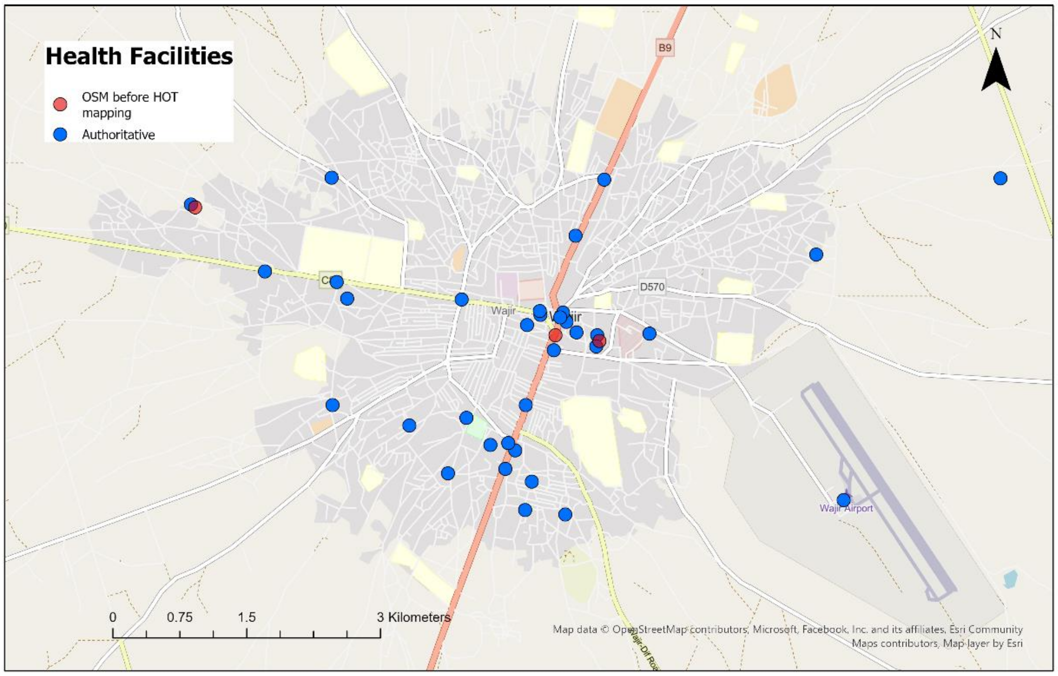


Figure 30. Chart showing four metrics of the completeness of OSM data compared to the authoritative dataset for Wajir's Health Facilities. The number of objects for both datasets, their overlap, the completeness of OSM based on the authoritative data, and the excess of data (commission) in OSM.

Wajir City (Wajir East) Health Facilities before Humanitarian Mapping Project vs Authoritative Data



Wajir City (Wajir East) Health Facilities after Humanitarian Mapping Project vs Authoritative Data

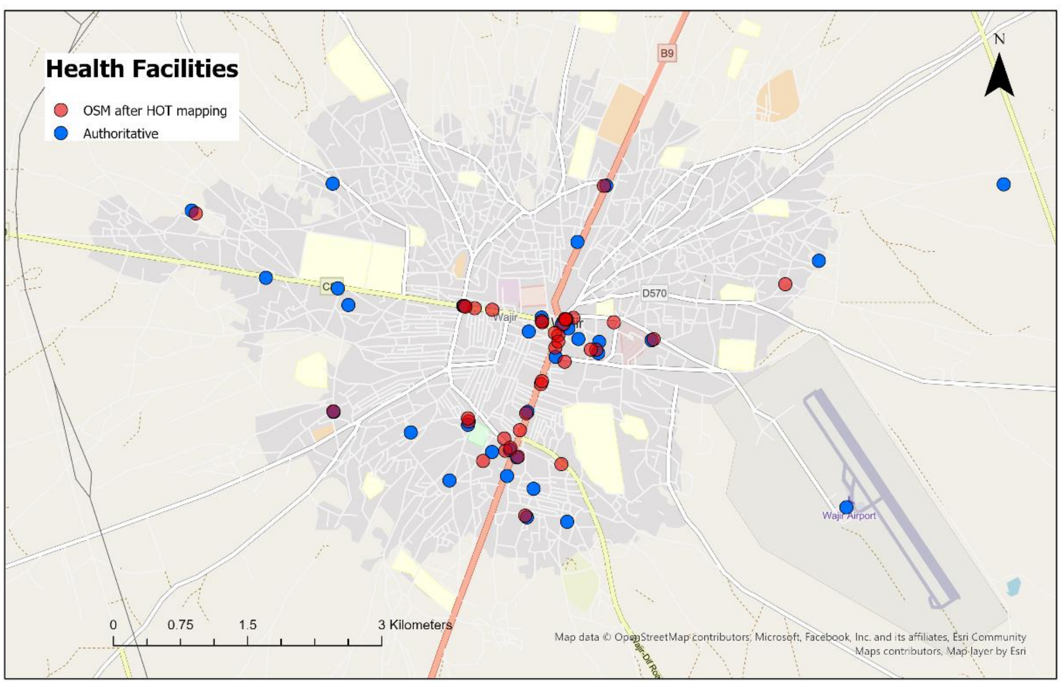


Figure 31. Maps of the city of Wajir (Wajir East) showing the change in OSM data during a Humanitarian mapping project

Wajir County - Health Facilities
OSM data vs Authoritative data

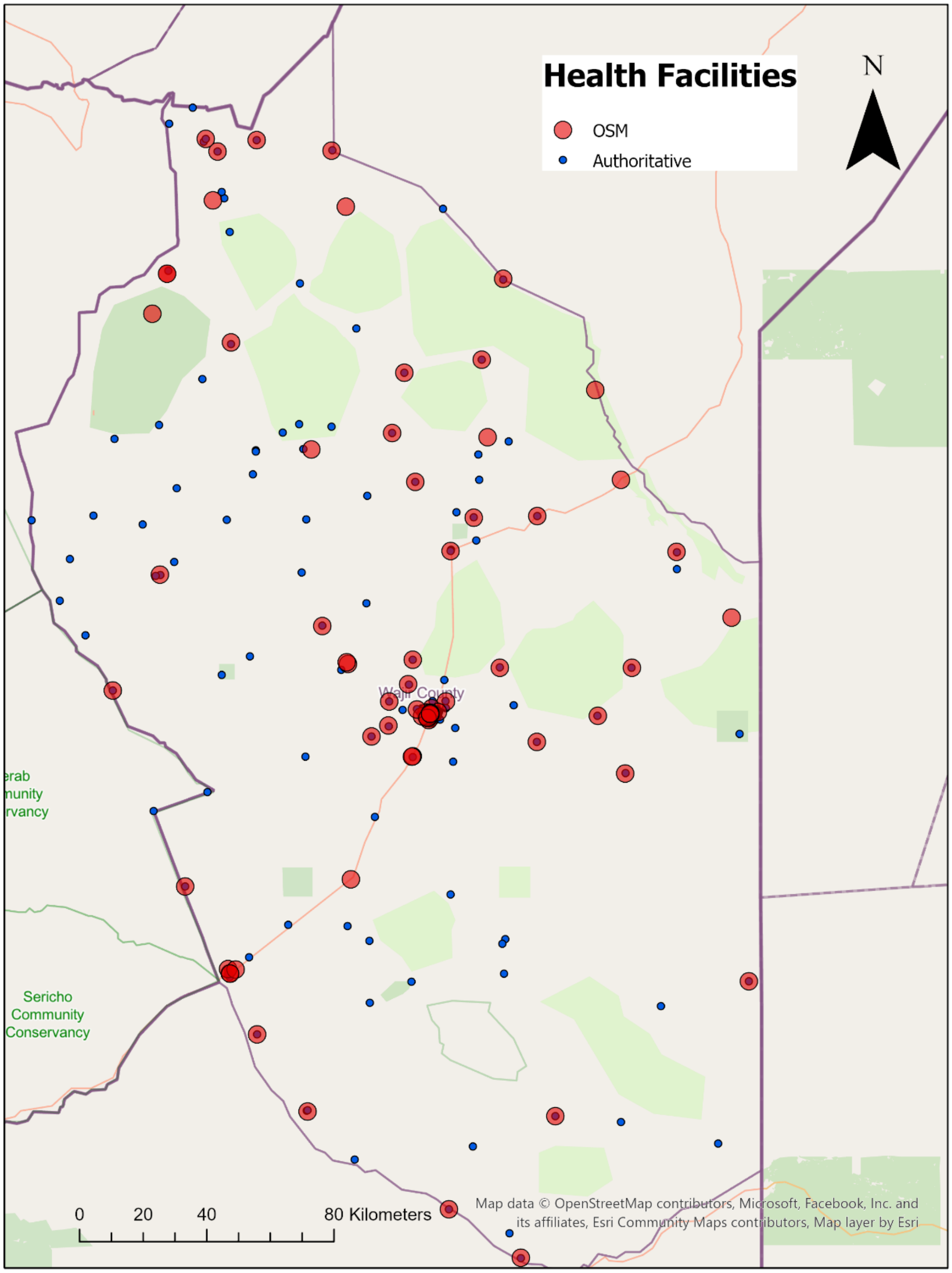


Figure 32. Map of Wajir County showing the health facilities according to OSM and the Kenyan Ministry of Health

Attribute completeness

The important attributes of the health facilities as identified by the users' needs analysis are the level according to the national classification system, the type of operator, the medical services offered, the contact details, and the equipment available. Table 11 shows the results of this type of quality assessment.

Classification Level

There are no tags in Wajir's OSM vocabulary implying the level of the health facility according to the national classification system.

In authoritative data, 88 out of 164 (**53.66%**) facilities give information about their classification according to the Kenyan Essential Package for Health (KEPH) levels 2-6.

Operator Type / Ownership

In OSM 6 out of 164 facilities in Nairobi indicate the type of operator which is **3.65%.** In the authoritative data of the Ministry of Health, this percentage is **100%**.

Contact

In OSM out of the 89 health facilities in Wajir County, none has a tag with the key "`contact`" or similar**.** For the Authoritative data of the Ministry of Health, this percentage is **100%**.

Medical services

The only key in OSM for Wajir county indicating medical services that is "`healthcare: specialty`" attributed only to 1 of the 89 objects (**1.1%**).

For the authoritative data of the Ministry of Health, 67 out of 164 facilities are accompanied by information concerning the offered services which is **40.85%**.

Available equipment

There are no tags in Nairobi's OSM vocabulary describing the available equipment in health facilities. Similarly, there are no such tags in the data of the Ministry of Health.

Table 11. Health facilities of Wajir County - Attribute completeness for the authoritative and OSM datasets measured in the percentage of objects that contain the attributes as described by the users' needs analysis.

	Authoritative %	OSM %
Classification	14.24	N/A
Operator	100	3.65
Contact	100	N/A
Services	55.46	1.1
Equipment	N/A	N/A

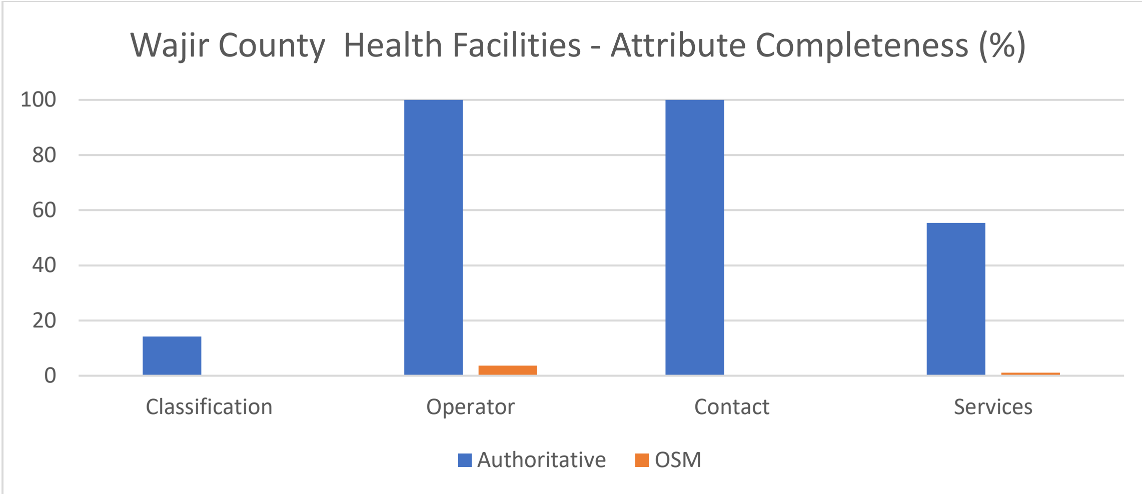


Figure 33. Chart showing the comparison of the attribute completeness of the health facilities in Wajir's County OSM and authoritative datasets.

Attribute Accuracy

This metric measures to what extent the attributes of OSM reflect in meaning those of the Authoritative dataset which are considered to be the ground truth. The attribute accuracy of the authoritative data is by default 100%.

Classification of Health Facilities

In the authoritative data the information about the class, according to the national classification system, is given by the tags "Level 2-6". In OSM Kenya the vocabulary has no tags indicating the Classification of the Health Facilities. Thus, OSM's data Attribute Accuracy for the health facilities classification, also in Wajir is not available.

Medical Services

In the authoritative data, there are 80 distinct tags describing the medical services offered by the Health Facilities while in OSM there is only 1 tag not matching any of the authoritative. This means **0%** attribute accuracy.

Operator Type / Ownership

In the authoritative data, 14 different tags are describing the ownership status. This is described in OSM for Wajir County by the key "`operator`". The matching to the authoritative data tag values of this key are 2 out of 2 ("private", "government") meaning 14.28%. However, the authoritative data tags of ownership can be grouped into four user-significant categories i.e., private, public, religious institutions, and NGOs. Thus, the OSM attribute accuracy for operator type in Wajir County is 50%.

Contact

There is no OSM tag in the health facilities of Wajir County indicating contact details. The attribute accuracy is not available.

Table 12. Health facilities of Wajir County - Attribute Accuracy for the authoritative and OSM datasets measured in the percentage of matching OSM values with the respective Authoritative tags.

	Authoritative %	**OSM %**
Classification	100	N/A
Operator	100	50
Contact	100	N/A
Services	100	0.0

5.1.5 Wajir County WATSAN

As in the case of Nairobi, authoritative data for WATSAN points of interest for Wajir County were not available by the time of writing. After a written application for the dataset to the County made by the Kenyan Red Cross, the data had not been received by the time of writing. The data found in OSM concerning drinking water from wells and ponds (used probably interchangeably for pans) seem to follow a specific vocabulary. In Figure 27 we can see a typical example of this. Although those edits could not be identified as part of an organized mapping project through hashtags or comments, the contributor is only one. After contacting the contributor through messaging in openstreetmap.org it was revealed that this was part of a mapping project of the University of Southampton. This information has not been able to be confirmed by any other source.

Tags

drinking_water	yes
irrigation_water	no
name	Anole Water Pan
natural	water
origin	manmade
water	pond
water_management	Community
water_quality	Soft
water_seasonality	Unreliable
water_source	Rain
watering_place_livestock	yes
watering_place_wildlife	no

Figure 34. Example of tagging a water point in Wajir County. The sophisticated keys indicate the involvement of an organized mapping project.

Completeness

For assessing the completeness of OSM water-related data (not hydrologically but human-related) of Wajir in absence of authoritative data, a map coming from Wajir's County official data found in a policy brief of 2019 (Bedelian, 2019) was digitized and used as reference data. To visually validate the digitization accuracy the data were overlayed on Wajir's official road network (Figure 35) and the results were satisfactory because usually, the major water points coincide with the road network. The overlapping of the water ponds data could not be calculated because there were no attributes in the reference dataset for comparison except for the type of the water point i.e., (borehole, pan, mega-pan, shallow well). In figure 35 a map of Wajir County with both datasets indicates a low overlapping. The query for the OSM water points can be found in Appendix C.

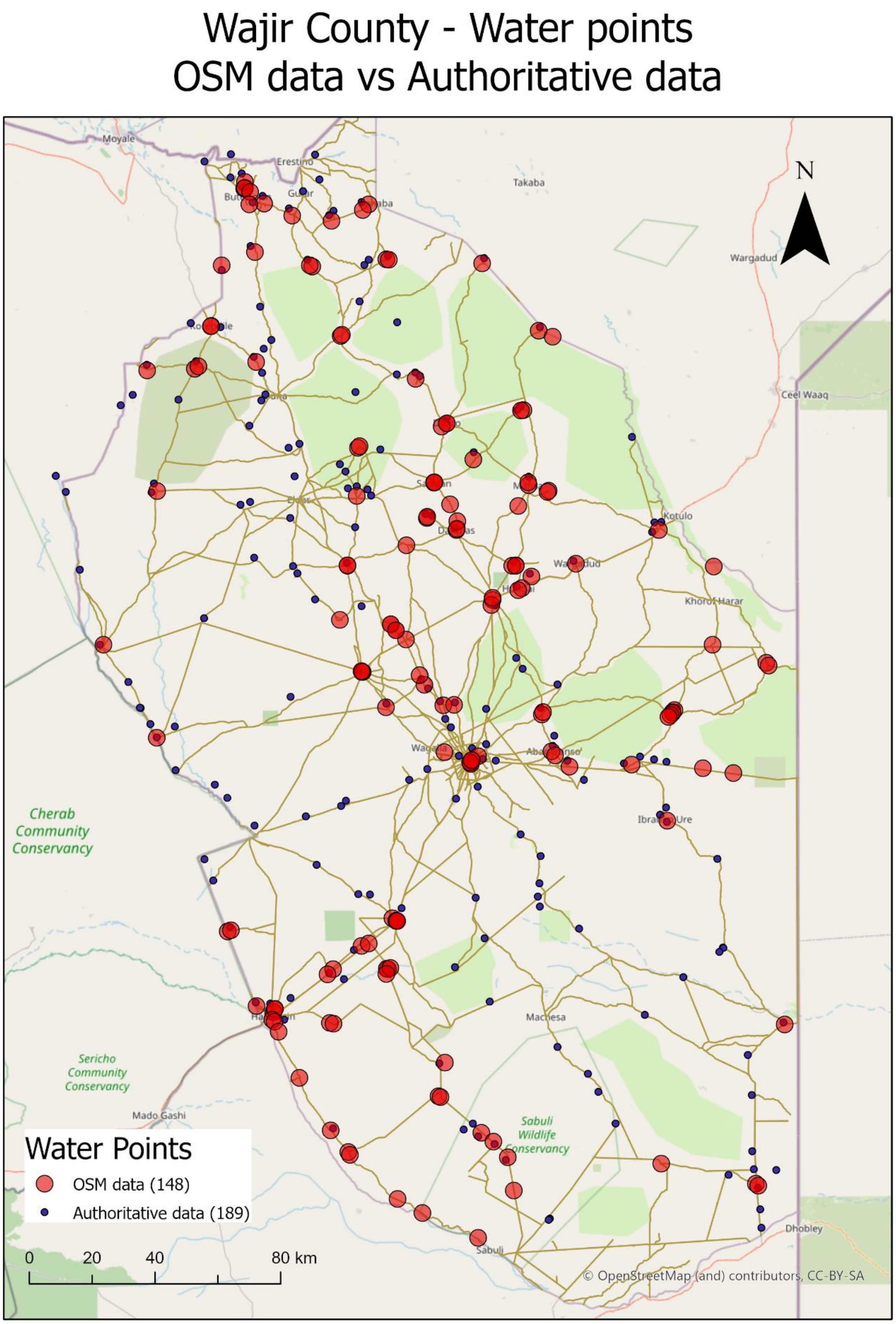


Figure 35. Map showing the water points in Wajir County. OSM data compared to the reference dataset which is the digitization result of a map published under the auspices of Wajir County.

5.1.6 Wajir County Road Network

The authoritative data for the road network of Wajir County, coming from the Kenya Roads Board, give a total length of 9431.72 Km. In OSM the whole road network of Wajir County accounts for 10822.42km. The overlapping of those two road networks was calculated using QGIS select by location processing tool to 7,892 km which means that 2930 km of OSM network are not in the KRB data and 1,539 km of KRB data that are not in OSM (Figure 35) with OSM completeness reaching **83.68%**. A combination of the two datasets would give a result closer to the ground truth with a length of 12,361 km (Figure 39).

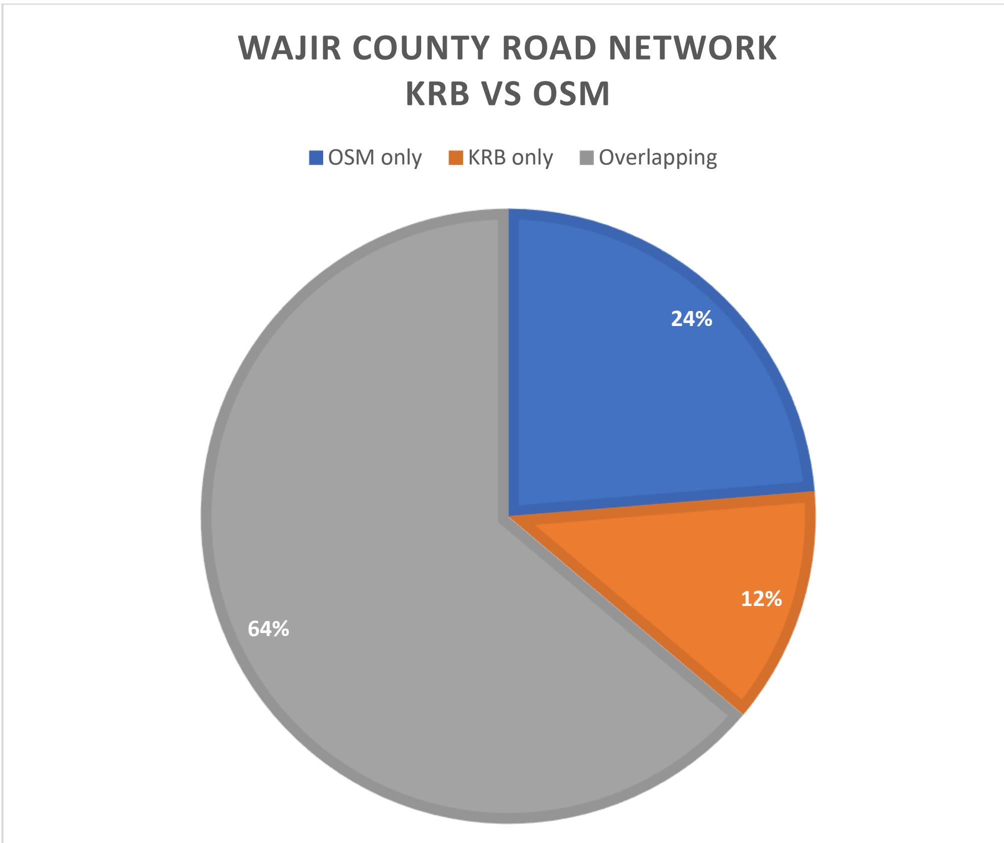


Figure 35. Chart showing the completeness of OSM road network in Wajir (omission - commission) when compared to the KRB authoritative data.

As similarly done for Nairobi County, tables 13 and 14 show the length for each class for the authoritative data of KRB (also in Figure 36) and the highway tags representing roads in OSM (also in Figure 37) both for Wajir County.

To compare the road networks from KRB and OSM for Wajir by type of road, a reclassification of both datasets (Tabble 15) was again necessary following the traditional functional classification as done for Nairobi County. This comparison is made again to show the attribute accuracy of the OSM highway tags. In all three functional classes, the comparison is far from similar to the overall comparison. It is out of scope to analyze further the attribute accuracy of the OSM road network as it has not been identified as a users' need. The only comment here is that the large number of unclassified road segments in Wajir County deteriorates the attribute accuracy.

Table 13. Road classes of the authoritative data present in Wajir County and their length

Kenya Roads Board Classification	Length in km
A	1022.5
B	1513.1
C + Cu	1106.7
D	642.9
E	359.1
Fu	434.7
G	2144.01
NEW	1825.01
NR	383.7
Total	**9431.72**

Table 14. OSM highway tags of Nairobi County and their lengths

OSM highway classes	Length in km
track	3081.71
trunk	246.76
secondary	573.95
tertiary	2905.53
residential	426.52
unclassified	3547.93
service	30.97
road	9.05
Total	**10822.42**

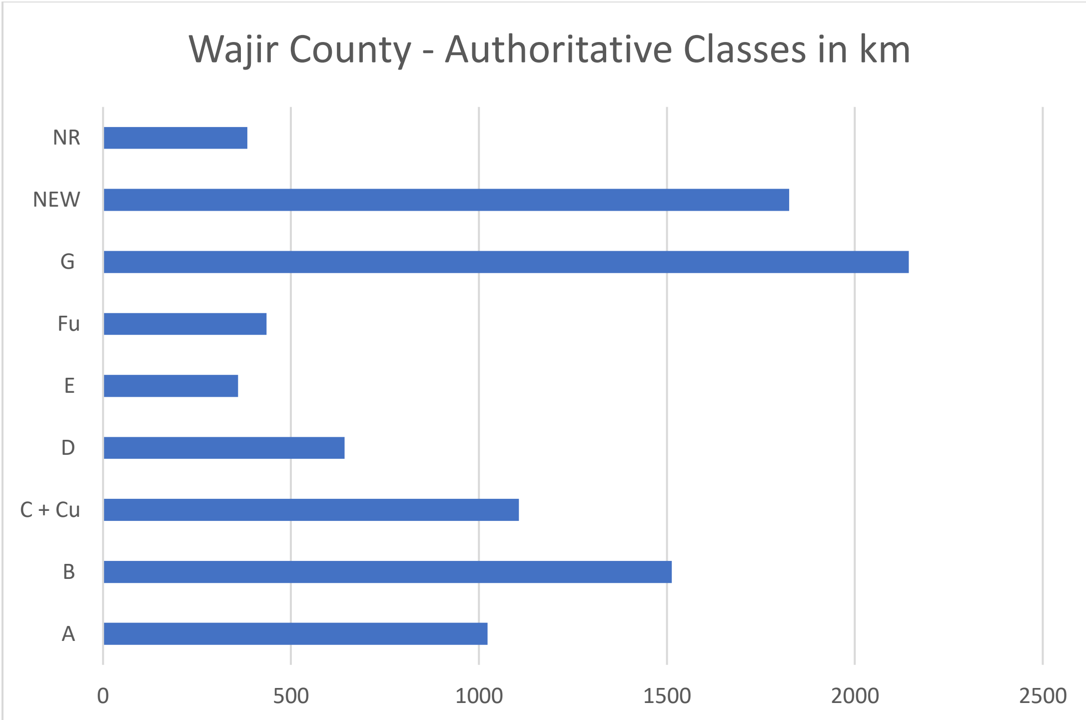


Figure 36. KRB for Wajir - Total roads length of separate classes

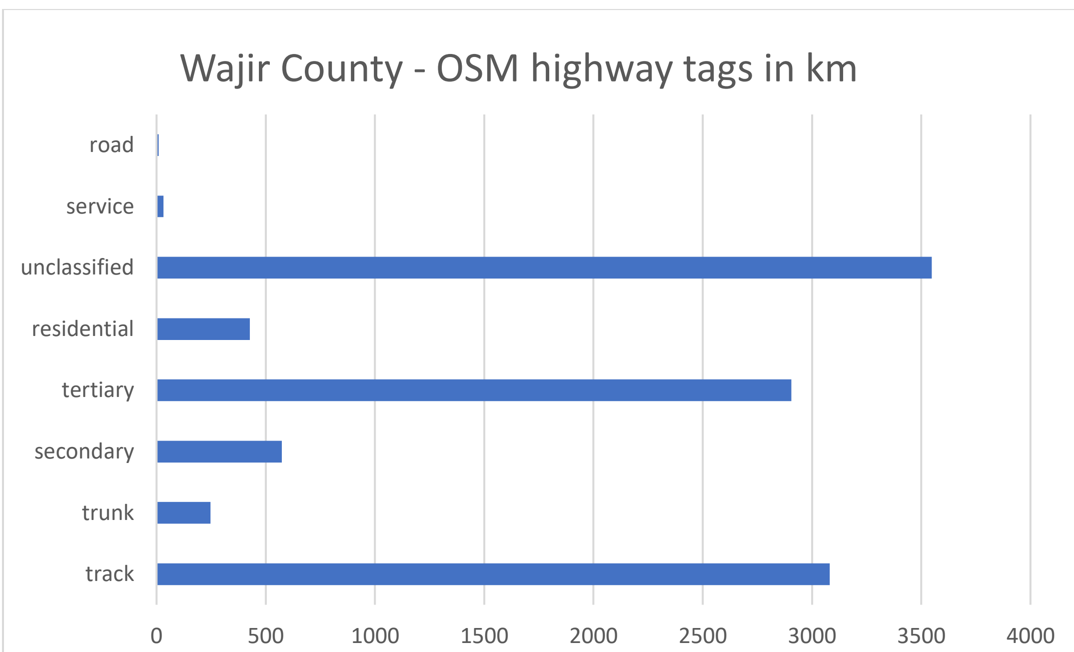


Figure 37. OSM Wajir – Total roads length of different highway tags

Table 15. Reclassification of KRB and OSM road networks according to the traditional functional classification system (Forbes, 1999)

km	KRB classes	Description	Functional Class	OSM keys	km
2535.6			Arterial Or Trunk		820.71
	A	Major Arterial		trunk, trunk_link	
	B	Minor Arterial		secondary, secondary_link	
1749.6	C, Cu	Primary	Collector	tertiary	6453.46
	D	Secondary		unclassified	
5146.52	E	Minor	Local Road	track	3548.25
	Fu	Special Purpose		service	
	G	Local Access		residential	
	NEW			residential	
	NR			residential	

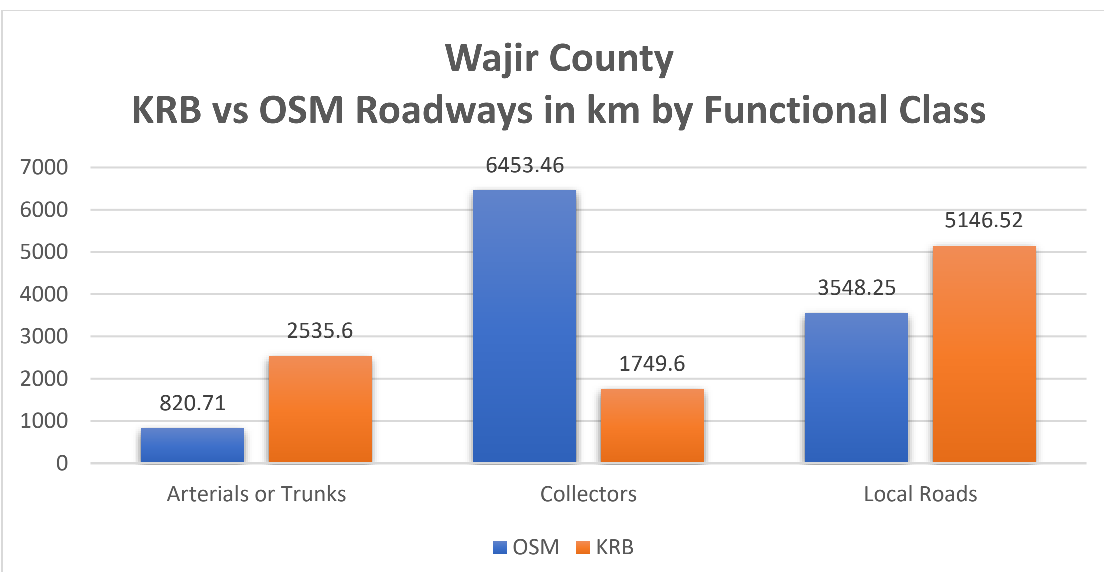
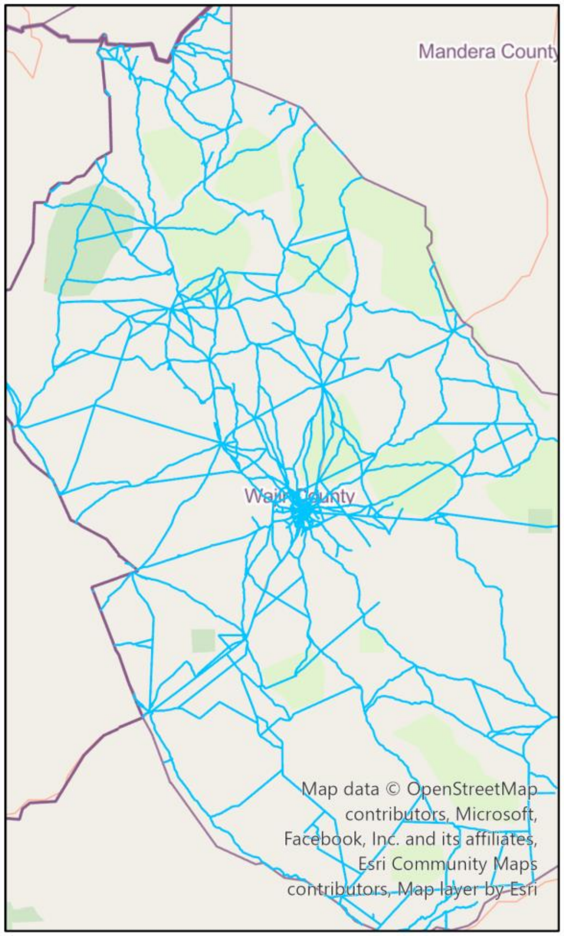
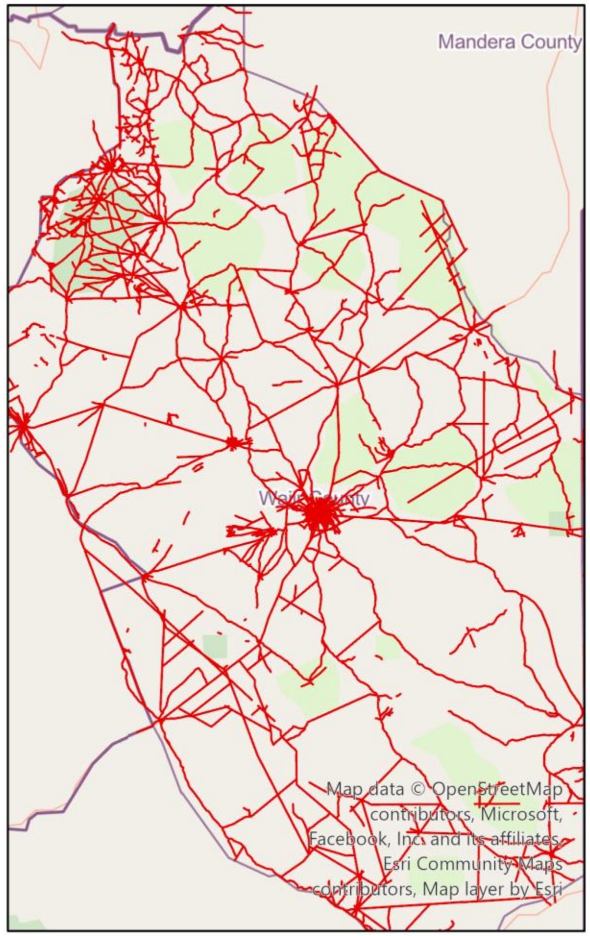


Figure 38. Comparison of Wajir's KRB and OSM road networks according to the functional reclassification

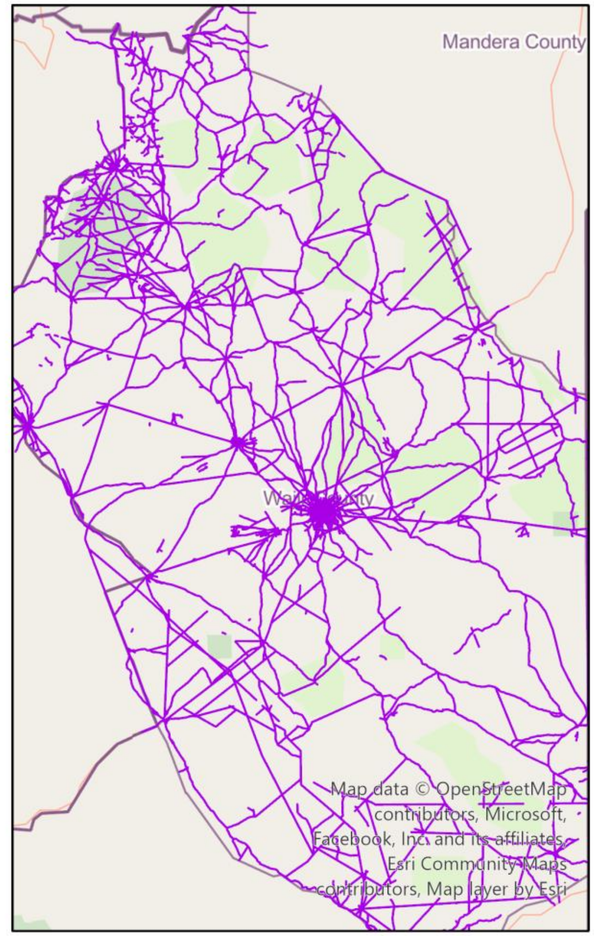
Wajir - KRB, OSM and combined Road Network



Kenya Roads Board



openstreetmap.org



Kenya Roads Board &
openstreetmap.org

Figure 39. Maps giving a visual comparison of Wajir's road network from KRB, OSM and their combined datasets.

5.2 Building an OSM ontology for the Kenyan Health facility classification system

One of the objectives of the present research is to take advantage of the existent vocabulary of OSM and turn it into an ontology. By taking advantage of as much information that is coming from the OSM data as possible it is attempted to satisfy the users of OSM and their needs as defined in the users' needs analysis. For example, the information missing from OSM in Kenya concerning the classifications of the health facilities could possibly be retrieved indirectly by other tags that imply the correspondent Level of each health facility. So instead of adding a new tag indicating the Level of the health facilities it is attempted to use and promote the current OSM file structure and tagging taxonomy. By analyzing the official classification of health facilities as described by the Kenyan Ministry of Health it is realistic to find some correspondence between OSM tags and one of the 6 official health facility classes (Figure 40).

Level 1 will not be analyzed as there are no authoritative nor OSM data for this volunteered body of the National Health System of Kenya.

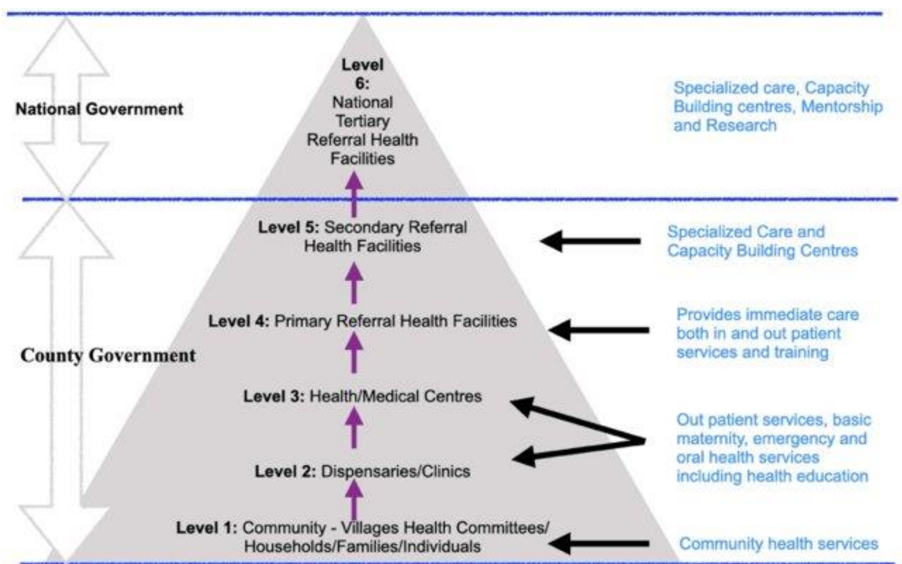


Figure 40. The official classification of Health facilities in Kenya (Toroitich, et al. 2022)

Level 2 is comprised of dispensaries and clinics. Dispensaries are public facilities while clinics are private, but they offer the same type of services. By building a query (Figure 43) for dispensaries and clinics with the tags "healthcare=clinic", "dispensing=yes", "health_facility:type=clinic" & "health_facility:type=dispensary" for the area of interest we can claim that we query for Level 2. For the county of Nairobi, this means 491 distinct objects in OSM. This number rises to 608 for the authoritative data. Nevertheless, in OSM the type of facility does not necessarily match the official classification. There are cases that a facility has the word clinic in its name, so the contributor draws the logical assumption that it is a clinic by type but in reality, it does not belong to Level 2. In other cases, the word clinic is in the name but not in the attributes. So, what is needed here is an attribute that the absence of which, clearly distinguishes from the next Level. This attribute is the existence of a medical officer. By adding to the query the tag "medical_staff:clinical_officer=0" facilities included that are not tagged as clinics or dispensaries but do not have a clinical officer. Then the results rise to 499. Two of the objects added were tagged as pharmacies and the other 6 were tagged as hospitals while their name tag stated clinic (Figure

42). In Figure 41 we can see that the results of the query used for Level 2 are clustered around the big informal settlements of Nairobi which are the grounds for numerous humanitarian mapping projects in the last decade. For the Health facilities of Nairobi, Level 2 can be described by the ontology in Figure 44.

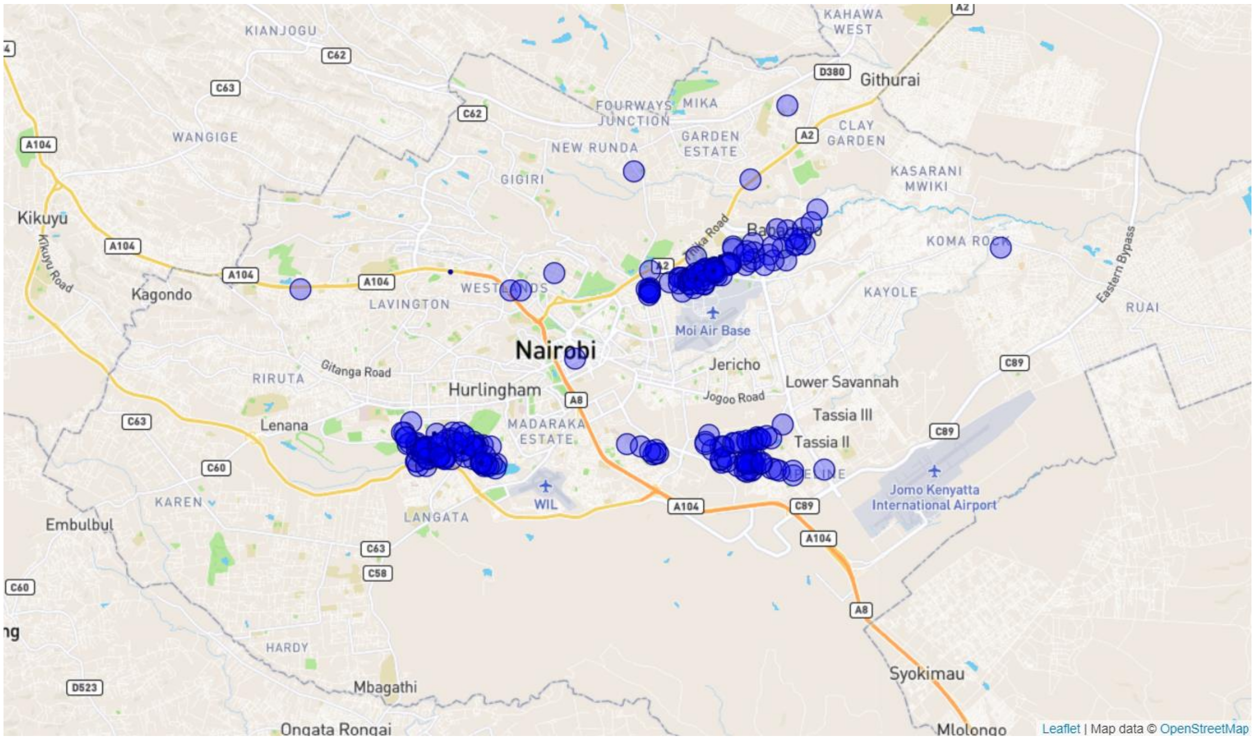
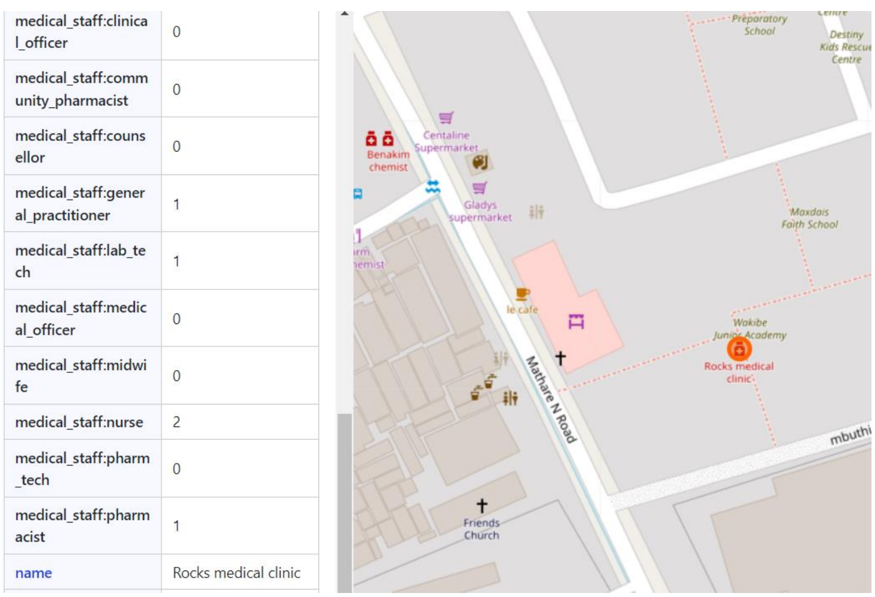


Figure 41. Health facilities in Nairobi County identified as Level 2 by the proposed ontology.



medical_staff:clinical_officer	0
medical_staff:community_pharmacist	0
medical_staff:counsellor	0
medical_staff:general_practitioner	1
medical_staff:lab_tech	1
medical_staff:medical_officer	0
medical_staff:midwife	0
medical_staff:nurse	2
medical_staff:pharm_tech	0
medical_staff:pharmacist	1
name	Rocks medical clinic

Figure 42. Example of a facility that is tagged as a hospital despite having the word “clinic in the official name”, but has no medical officer which shows that it is Level 2

```
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
 SELECT distinct * WHERE {{
SELECT ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare "clinic" .
 }}
 union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:dispensing "yes" .}}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:health_facility:type "dispensary" .
   }}
   union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:health_facility:type "clinic".
   }}
   union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "clinic" .
 }}
 union
 {SELECT distinct ?osm_id ?hasgeometry ?name  WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
      ?osm_id osmkey:name ?name .
  ?osm_id osmkey:medical_staff:clinical_officer "0" .
   } }}
```

Figure 43. SPARQL query in Qlever to capture Level 2 health facilities of Nairobi County from OSM

Level 3 is comprised of Health Centres, Sub-District Hospitals, and Maternity Centres. In Nairobi County, according to authoritative data, there are 30 such facilities of public ownership status and another 105 facilities of Level 3 of religious institutions, NGOs, or the private sector. In OSM by querying only for the relevant type of facility according to Nairobi OSM vocabulary i.e.,

"`health_facility:type=health_center;hospital`",
"`health_facility:type=health_center`",
"`health_facility:type=nursing_home_with_maternity`",
"`healthcare=nursing_home`", "`amenity=nursing_home`"

there are 15 objects. In the description of Level 3 by the Ministry of Health, Health Centres have a clinical officer in charge. By adding the tag of "`medical_staff:clinical_officer=1`" to the query (Figure 46) as an additional characteristic of Level 3 we get in return 48 facilities. In Figure 45 we can see an example of such a facility where the ontology (Figure 47) helps the identification of the correct class for an illtagged OSM object.

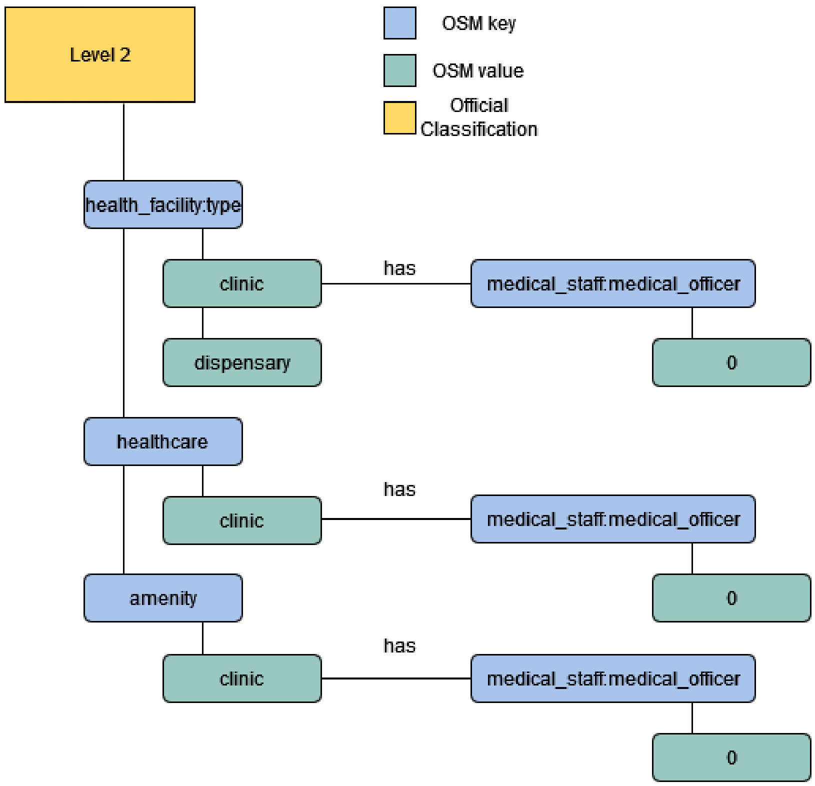


Figure 44. Proposed OSM ontology for Nairobi's Level 2 health facilities

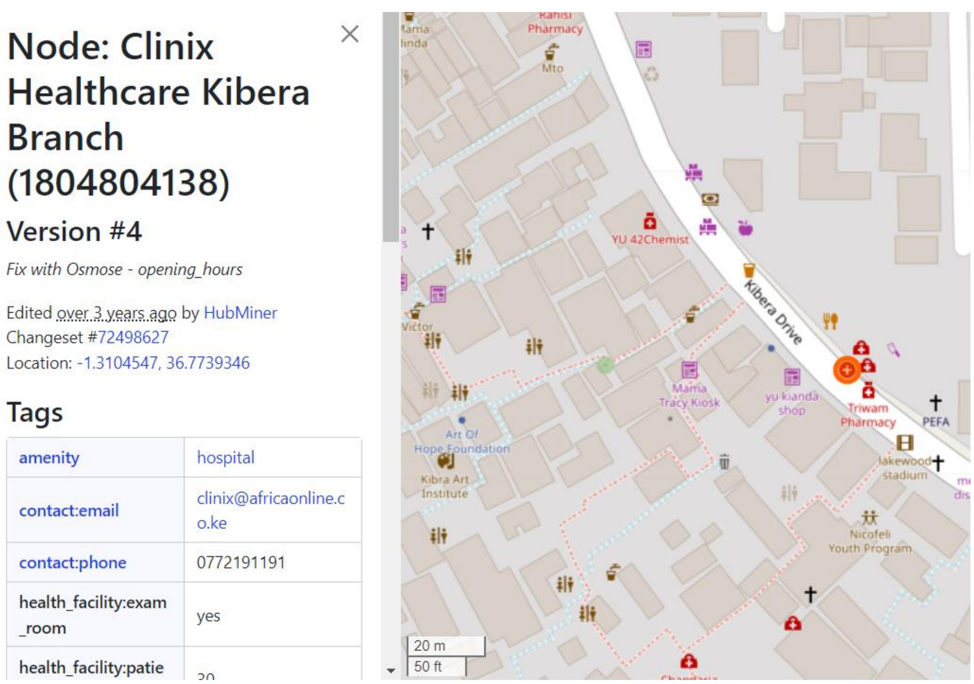


Figure 45. OSM object with the word Clinic in the name tagged as a hospital but falling under Level 3 - Health Center in the official classification. With the help of the tag **medical_staff:clinical_officer=1** is categorized also by the proposed ontology as Level 3

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
 SELECT distinct * WHERE {{
SELECT ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare "nursing_home" .
     # ?osm_id osmkey:medical_staff:clinical_officer "1" .
 }}
 union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "nursing_home" .}}

  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:health_facility:type "health_center" .
   }}
   union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:health_facility:type "health_center;hospital" .
   }}
   union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
   ?osm_id osmkey:health_facility:type "nursing_home_with_maternity" .

 }}
 union
{SELECT distinct ?osm_id ?hasgeometry ?name WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
       ?osm_id osmkey:name ?name .
  ?osm_id osmkey:medical_staff:clinical_officer "1".
       #?osm_id osmkey:medical_staff:medical_officer "0".
   } }}
```

Figure 46. SPARQL query in Qlever to capture Level 3 health facilities of Nairobi County from OSM

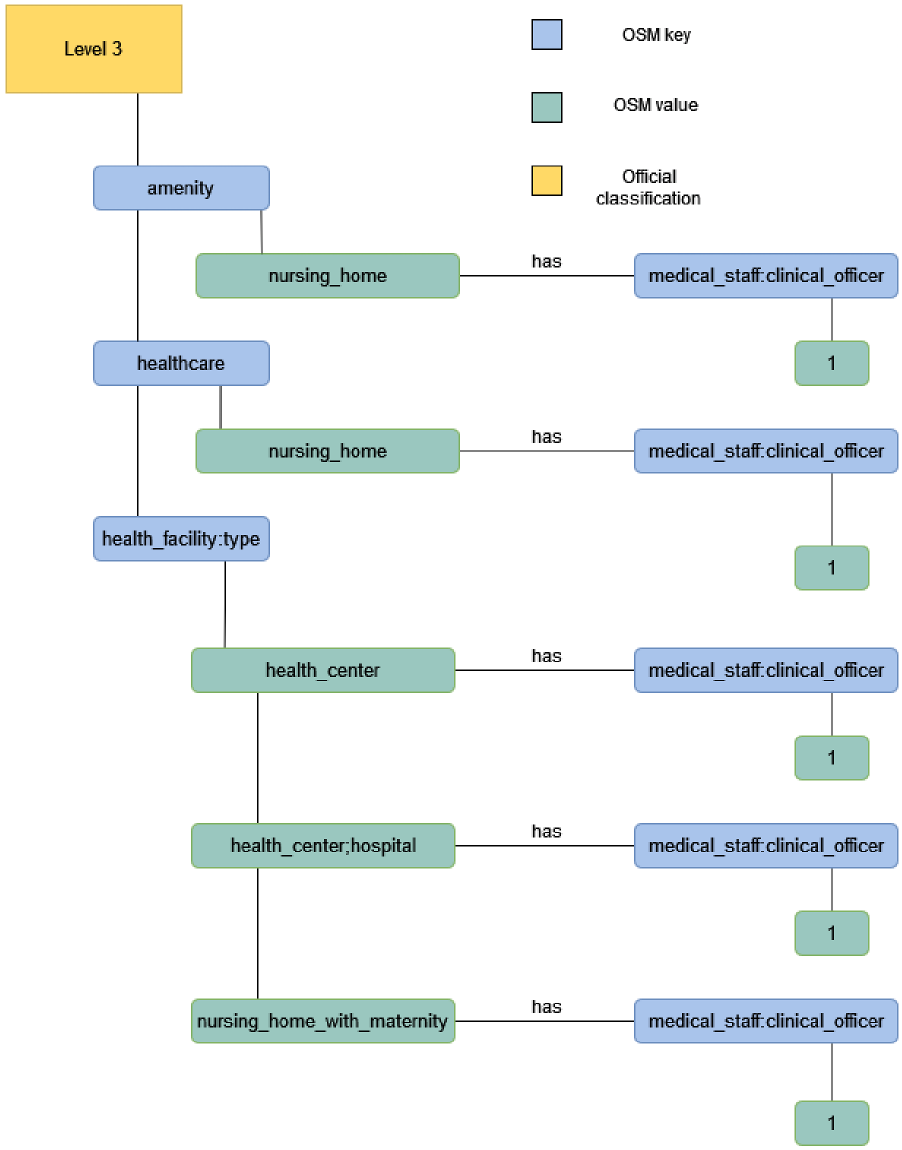


Figure 47. Proposed OSM ontology for Nairobi's Level 3 health facilities.

Level 4 comprises facilities that provide highly specialized services. The public facilities of this Level are the sub-county hospitals as well as referral laboratories. For Nairobi County, those are 10 (6 Hospitals and 4 Laboratories) in number. There are also 20 facilities owned by religious institutions and the private sector. Out of the 6 public hospitals of Level 4 in Nairobi County only 3 are found in OSM and they are poorly tagged. There are no tags that can be used to link them with their Level. The characteristic services that Level 4 offers and are different than the lower levels are surgeries and caesarian sections while at the same time, they have a medical officer in charge. A combination of those tags if they were available and also the absence of medical superintendents who manage Level 5 facilities could have helped relate OSM objects to the official Level 4. The type of facility from this Level and up has no significance because all are hospitals. Figure 48 shows a potentially helpful SPARQL query for Level 4 based on the existence of a medical officer. The tagging is also poor for the rest of the facilities.

```
PREFIX osm2rdf: <https://osm2rdf.cs.uni-freiburg.de/rdf#>PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct ?osm_id ?name ?hasgeometry WHERE {
  VALUES ?region { osmrel:3492709 osmrel:3492709 }
  ?region (osm2rdf:contains_area*/osm2rdf:contains_nonarea)|(osm2rdf:contains_area) ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:medical_staff:medical_officer ?medical_officer .
  #?osm_id osmkey:name ?name
  FILTER REGEX(?medical_officer, "1|5|8").
}

```

Figure 48. SPARQL query in Qlever to capture Level 4 health facilities of Nairobi County from OSM

Level 5 are facilities that provide specialized care including intensive care and life support. Those services, if tagged in OSM could help identify the Level 5 health facilities. In the case of Nairobi County, all three Level 5 hospitals that are part of the authoritative data exist also in OSM but again the tagging is so poor that they cannot be identified as level 5 through it but only from prior knowledge.

Level 6, the smallest class as of the number of objects, includes the National Referral Hospitals (NRH). Countrywide, in Kenya, there are 5 NRHs, all in Nairobi County. it would be reasonable to use the information coming from the name of each facility in OSM to classify them. All the health facility objects that include the words “referral” or “national” in their `name` tag can be considered candidates Level 6. In all their official names at least one of those two words is present. If those names are identical in OSM then it is possible to construct, by querying, a complete Level 6 for Nairobi County. The NRHs of Nairobi are:

1. National Spinal Injury Referral Hospital
2. Kenyatta National Hospital
3. Mathare National Teaching and Referral Hospital
4. Kenyatta University Teaching and Referral Hospital
5. Moi Teaching and Referral Hospital

While for the two Kenyatta hospitals, their official name is also in OSM this does not happen for the National Spinal Injury Referral Hospital which misses both keywords (Figure 50). Mathare National Teaching and Referral Hospital is called Mathare mental Hospital in OSM and Moi Teaching and Referral Hospital does not even exist in OSM. A second option would be to find distinct characteristics of the NRH that if identified in their OSM versions could indicate their class. Again, poor tagging prevents any attempt to identify Level 6 by the tags of the facilities.

The only but strong distinguishing characteristic of the NRHs in Kenya is that they operate under the authority of the national government contrary to all other public health facilities that operate under county governments. If query for OSM Health Facilities that use the tag "`operator:Government`" the result is one out of the five NRHs of Nairobi County (Kenyatta National Hospital). It should be noticed that the OSM contributors cannot be familiar with this detail about the operators of the health facilities but this is exactly why every single tag is of similar importance to any other. It is typical to see a facility of any level tagged as operated by the government but practically meaning that the facility is public. In the same manner, all NRH of Kenya have the tag `operator:type=public` with the Kenyatta National Hospital having also the correct one "`operator:Government`". For many users, the values public and government are interchangeable when referring to operator type as they do not have in mind the aforementioned distinction. Nevertheless, the SPARQL query for capturing Level 6 health facilities in OSM is shown in figure 48, and the respective ontology in figure 49.

```
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
 SELECT distinct * WHERE {{
SELECT ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare "hospital" .
     ?osm_id osmkey:operator "Government".
 }}
 union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:health_facility:type "hospital" .
  ?osm_id osmkey:operator "Government".
   }}
   union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "hospital" .
  ?osm_id osmkey:operator "Government".
 }}}
```

Figure 48. SPARQL query in Qlever to capture Level 6 health facilities of Nairobi County from OSM

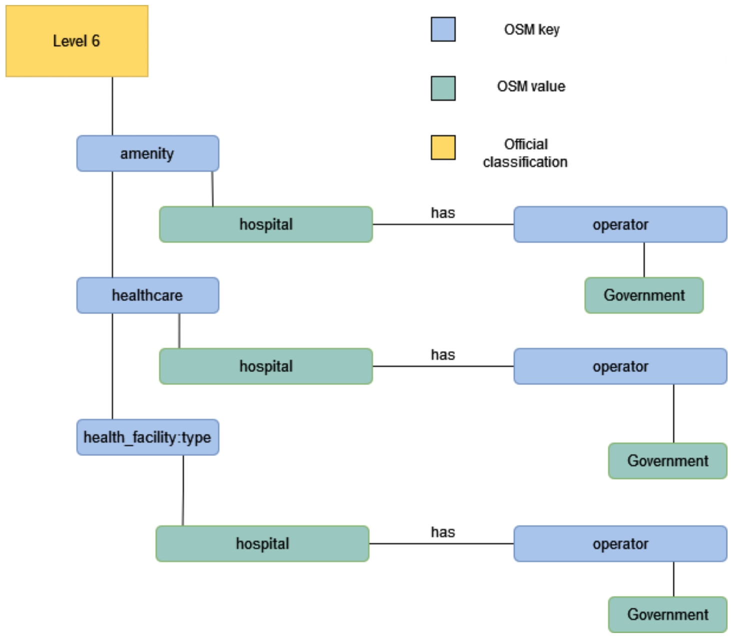


Figure 49. Proposed OSM ontology for Nairobi's Level 6 health facilities

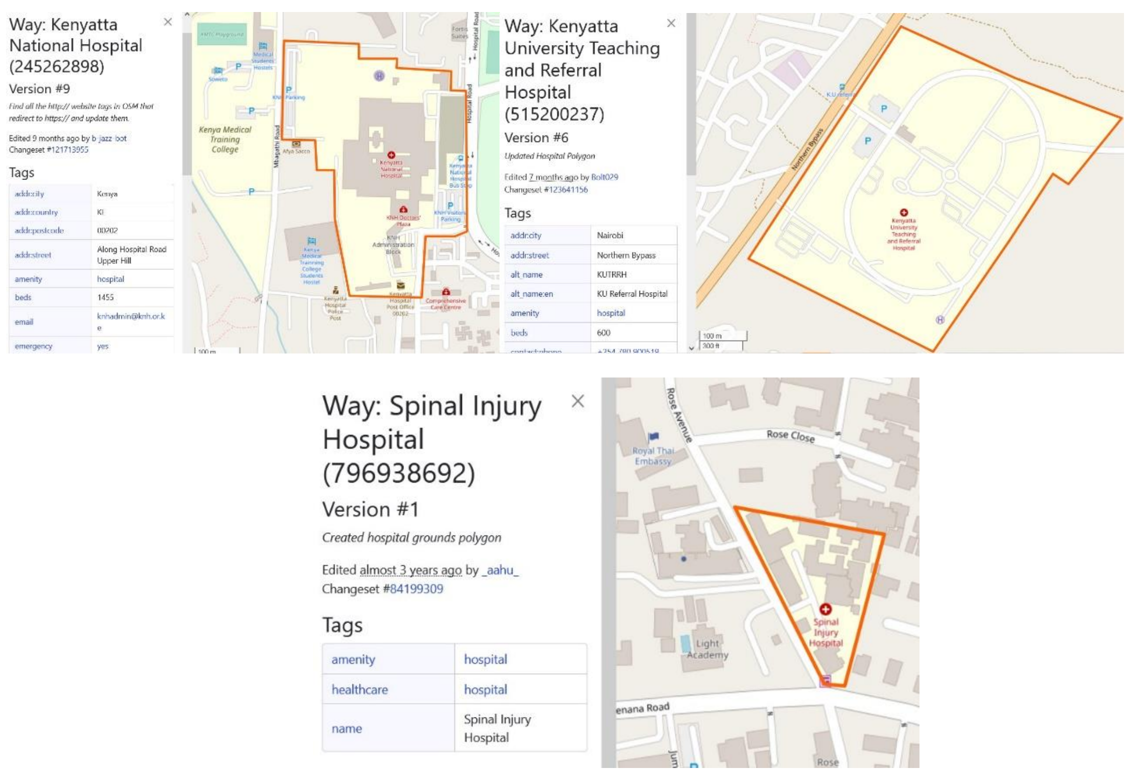


Figure 50. Three out of five National Referral Hospitals of Nairobi as represented in OSM.

5.3 Humanitarian OSM Mapping projects as a proxy for assessing OSM health facility data quality

Since the very beginning of the research, the visual examination of OSM data on the map concerning health services as they were described in the users' needs analysis showed significant clustering at the informal settlements of Nairobi County. What is common in those areas of low-income settlements in terms of OSM mapping are the numerous humanitarian mapping projects that have taken place there in the last decade. What humanitarian mapping seems to have achieved in those areas was to fill in the gap in the authoritative data concerning health facilities, adding more than 50km of footways for navigation inside only the major informal settlements and generating detailed data about water points, communal sanitation, and handwashing facilities. In the Analysis, we have already seen the improved quality of OSM data inside the informal settlements that most of the time are more complete than the authoritative counterparts. To prove the correlation of this increased level of quality to the humanitarian mapping projects spatial data proving their implementation is needed. It is common practice that in every organized mapping project, one or more hashtags are added in every changeset manually or automatically to indicate the source (mapping project name and/or organization involved). The triplified OSM data that is used throughout the research does not carry along the hashtags of the objects. Moreover, it is not possible to query in overpass turbo calling for hashtags. The most popular web application to perform such queries is OSMcha.

Object completeness

After experimenting with many hashtags like #hotosm, #missingmaps, #youthmappers which are mostly used in humanitarian OSM mapping projects it was obvious that OSMha could not filter for our specific area of interest. The only option was to query for local hashtags and comments. The most common for the entire Kenya that were occasionally coming along with the aforementioned international ones were, #mapkibera, #OSMkenya and #Unmappedkenya. We queried then using OSMcha in each constituency of Nairobi County for each one of those hashtags separately. Because of the high overlapping of those three tags in every constituency, we kept only the hashtags giving back the highest number of changesets. An extra column was created then in the completeness table of Nairobi County (Table 2) resulting in Table 16. Then a statistical analysis was performed to see if there is a positive and statistically significant correlation between the number of changesets coming from humanitarian mapping projects and the OSM object completeness of health facilities. From the 17 constituencies, one outlier (Kibra Constituency) and Embakasi South were removed. Embakasi South includes Mukuru one of the largest informal settlements in Nairobi. The OSM completeness of health facilities is relatively high (27.7%) but this does not reflect in the humanitarian mapping changesets. This is probably because most of the edits which were part of the initial Map Kibera project are done 12 years ago when the use of hashtags and comments was not typical.

In the scatter plot chart (Figure 51) the visual examination shows a strong positive correlation coefficient between the two variables. By performing a T-test analysis, we can confirm this as Pearson's Correlation is 0.918. This result is statistically significant as $p=0.0254$ at $p<0.05$ and $T_{stat}>T_{crit}$ (Table 17).

Table 16. Table showing the number of changesets containing the hashtags #mapkibera, #OSMkenya and #Unmappedkenya for the 17 constituencies of Nairobi County and their % completeness as measured in the quality assessment.

	Changesets of humanitarian mapping	% OSM health facilities completeness
Mathare (Mathare)	1326	86.67
Embakasi North (Korogocho)	184	30
Ruaraka	101	12.72
Kamukunji	34	8.33
Makadara (Mukuru)	34	25.45
Embakasi West	32	9.61
Embakasi Central	22	7.14
Embakasi East	32	15.38
Kasarani	253	26.67
Roysambu	114	9.91
Westlands (Various villages)	497	20.3
Dagoreti South	111	3.7
Dagoreti North (Kawangware)	236	26.92
Langata (Various villages)	203	18.68
Kibra (Kibera)	3371	52.94
Starehe (Mukuru)	164	21.70
Embakasi South (Mukuru)	29	27.77

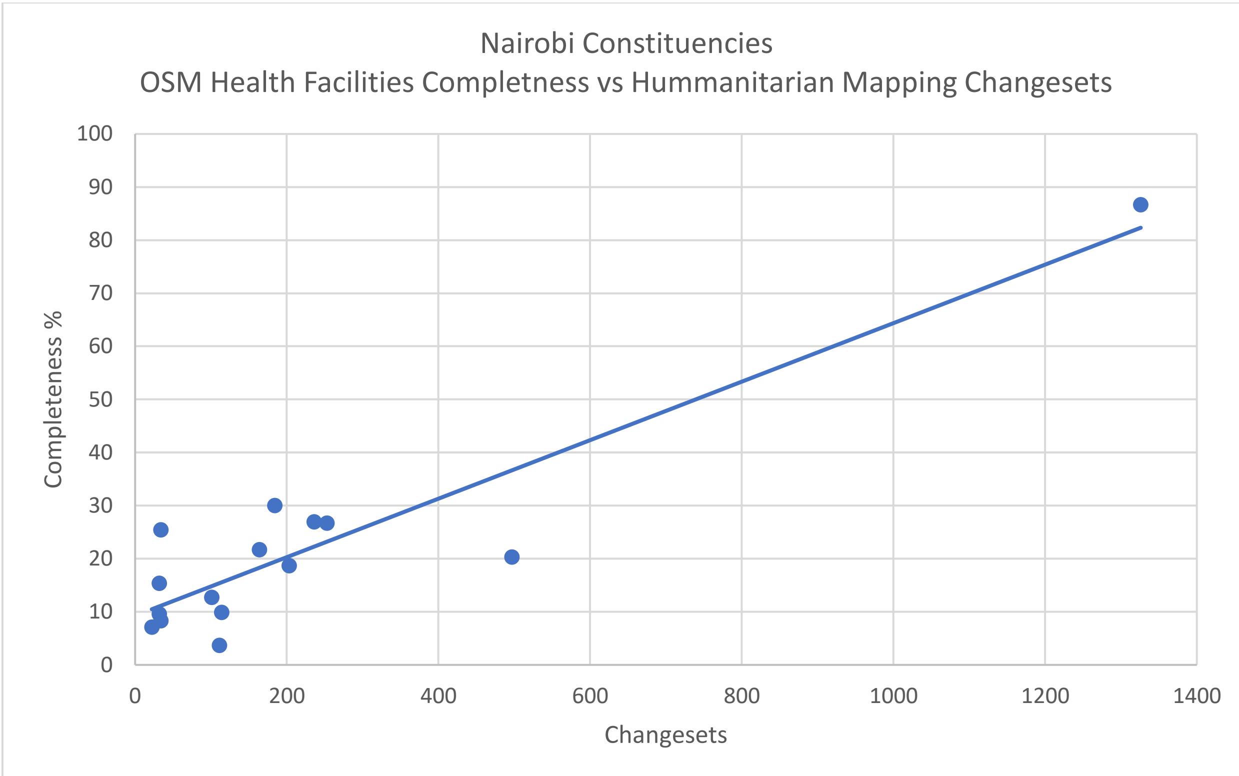


Figure 51. Scatter plot of Table 16 for visualization of the correlation between the two metrics

Table 17. Results of the statistical analysis of Table 16

t-Test: Paired Two Sample for Means

	changesets	*% Completeness*
Mean	222.8666667	21.54533333
Variance	108733.5524	391.9449838
Observations	15	15
Pearson Correlation	**0.917877256**	
Hypothesized Mean Difference	0	
df	14	
t Stat	**2.50168988**	
P(T<=t) one-tail	0.012691826	
t Critical one-tail	1.761310136	
P(T<=t) two-tail	**0.025383653**	
t Critical two-tail	**2.144786688**	

Attribute completeness

To find out if there is a correlation between humanitarian mapping projects and the attribute completeness of the health facilities in Nairobi County "`medical_service`" key was chosen to be examined because it is dedicated to the health domain. The SPARQL query used for finding all OSM objects in Nairobi containing this key is shown in figure 52. An overview of the results can be seen on the map in figure 53. It is obvious from the map that only in the informal settlements of Kibera and Mathare (with a few objects in Kawangware) are the medical services attributed. In that case, a statistical analysis such as the one performed for object completeness is not relevant. The medica_service tag key in Nairobi County is only present where humanitarian mapping projects have taken place.

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct ?osm_id ?hasgeometry WHERE {
 osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:medical_service:outpatient | osmkey:medical_service:family_planning
  | osmkey:medical_service:pregnancy_test | osmkey:medical_service:malaria|
  osmkey:medical_service:general_medical_services |osmkey:medical_service:condom_distribution
  |osmkey:medical_service:vct_hiv_counselling_test
  |osmkey:medical_service:circumcision
  |osmkey:medical_service:immunizations|osmkey:medical_service:sexually_transmitted_infections_management
  |osmkey:medical_service:antenatal_care|osmkey:medical_service:dental_services
  | osmkey:medical_service:minor_surgery|osmkey:medical_service:psychosocial_support_counselling

 |osmkey:medical_service:integrated_management_of_childhood_illness|osmkey:medical_service:growth_and_nutrition
al_support
  |osmkey:medical_service:inpatient| osmkey:medical_service:home_based_care
  |osmkey:medical_service:youth_friendly_services| osmkey:medical_service:ear_nose_throat
  |osmkey:medical_service:pmtct |osmkey:medical_service:tb_diagnosis
  |osmkey:medical_service:gender_based_violenve_services|osmkey:medical_service:tb_treatment
  |osmkey:medical_service:tb_lab|osmkey:medical_service:antiretroviral_therapy
  |osmkey:medical_service:basic_emergency_obsteric_care|osmkey:medical_service:palliative_car
  |osmkey:medical_service:comprehensive_essential_obsteric_care
  |osmkey:medical_service:disease_support_groups|osmkey:medical_service:prevention
  | osmkey:medical_service:maternity|osmkey:medical_service:blood_sugar_test
  |osmkey:medical_service:cancer_screening| osmkey:medical_service:blood_transfusion "yes" .
  }
```

Figure 52. SPARQL query for finding all OSM objects in Nairobi County containing tags with the key medical_service.

Nairobi County
OSM health facilities having tag key(s) of the type medical_service:*

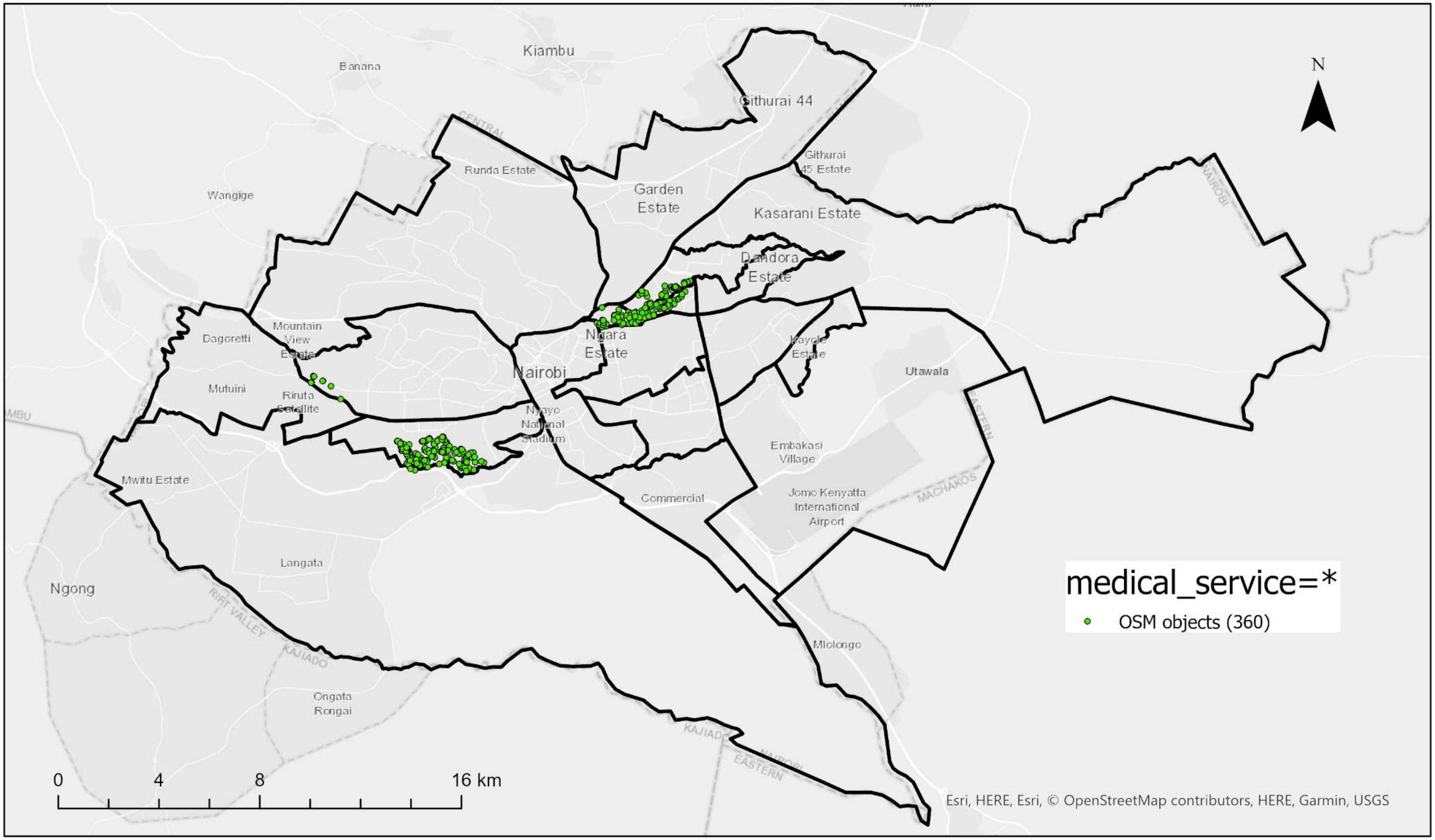


Figure 53. Map of Nairobi County showing the OSM objects that have medical_service key in one or more of their tags. The clustering of data around the informal settlements of Mathare and Kibera is obvious.

6 Towards an OSM-based web application for health services

The last objective of the research is to find out to what extent can we use semantic web technology to build a web application that based on triplified OSM data and by taking into account external RDF sources (e.g. wikidata) can give back to the user (as identified by the present research) the most of the information available.

The triplified OSM data are available in the osm2rdf library provided by the University of Freiburg as ttl files. As we have seen in practice, the data for Kenya were not updated weekly as stated by the provider but were up to date after a period of almost 5 months. We were able to check that by confirming the presence of edits that happened in Wajir city in September 2022 and were available in the turtle file in the middle of January 2023. For the health facilities we have already identified which OSM tags can satisfy users' needs. The next step would be to express those needs in RDF queries (Appendix C) that when called can give back the desired results from the turtle dataset and potentially be combined with external RDF sources from the web. This has already been done during the quality assessment. The OSM triplified dataset used has geometry data for every object so the queries give back location information.

Classification

This type of information is not available in Kenya OSM either with a dedicated tag or with a generic tag. For this reason, we have built ontologies that can easily be translated to SPARQL queries. This step has also been completed for Levels 2,3 and 6 of the official health facilities classification of Kenya as described in paragraph 5.2 containing the queries.

Services

The services of the health facilities have been quite comprehensively described in OSM with 35 valid dedicated keys of the type "`medical_service:*`" from which 21 represent services identical to the authoritative data. The dilemma here is including all tags or just those that are aligned with the ministry of health. This is a question that is out of the scope of our research so we will use all 35 keys. Those keys can be categorized into smaller groups for interface accommodation. Figure 54 shows the SPARQL query to be used for the services provided by the facilities. The name of the service (*****) can be added by the user with autocomplete help fed by the list of the available options.

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
SELECT ?osm_id ?hasgeometry ?type ?name WHERE {
osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:medical_service:***** "yes" .
  }

```

Figure 54. SPARQL query for all medical services in OSM for Nairobi County

Operator

The operator or the ownership as it is found in the authoritative data is easily queried for Nairobi with the key “`oerator_type`”. The 14 values that exist for health services can be categorized into 4 groups to match the users’ needs: public, private, NGO, and religious institutions. The information about the type of operator is important only after the desired service or level is found in one facility. The user can be informed by clicking on the results and be redirected to direct to the respective OSM object and see all other available information as shown in figure 55. In case the user wants to see all public health facilities within the bounding box of choice the query would be as in figure 56.

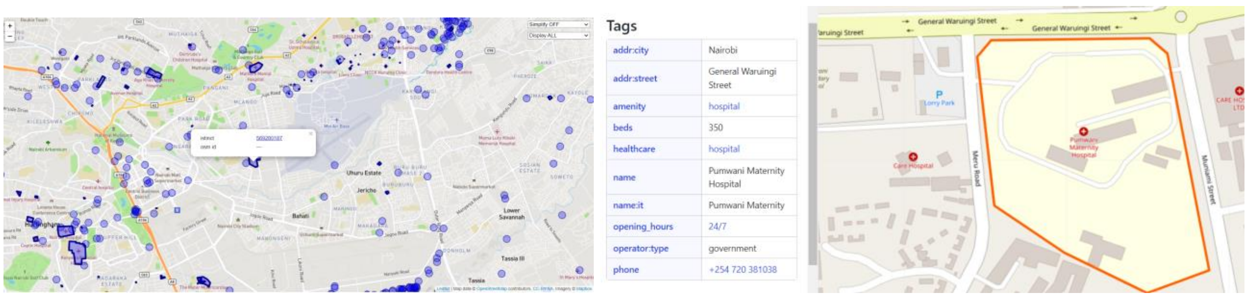


Figure 55. Example of using the proposed web application as a dedicated health facility query for OSM objects

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct * WHERE {
{SELECT distinct ?osm_id ?hasgeometry WHERE {
   osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:operator:type "government" .
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
   osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:operator:type "ministry_of_health" .
  }}
  union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
   osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:operator:type "public" .
  }}}
```

Figure 56. SPARQL query to select only the public from all the health facilities for Nairobi County

Contact

The contact attribute is important information but only when the user has found what they need. So there will not be a dedicated query and interface button for it.

To run those queries against OSM data it is necessary to download the latest version of the turtle file for our area of interest (Kenya, for the present research) from https://osm2rdf.cs.uni-freiburg.de/ and store it in a local disc. Of course, the dataset should be uploaded on a server for a web application to be built. The automatic regular update of the dataset should be scripted in any case for the data to be up to date. This dataset contains all OSM data for Kenya. With the help of a SPARQL engine instead of selecting the data we construct sub-graphs that are narrowing down the dataset. Using the operator CONSTRUCT instead of SELECT we can construct for example a sub-graph for Nairobi (Figure 57) and then a sub-graph of the health facilities in Nairobi (Figure 58). By doing so, generic keys such as “phone” or “operator_type” can be queried only for the features of interest.

```
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
Construct
{osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .}
WHERE
{osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
}
```

Figure 57. SPARQL Constructing sub-graph of Nairobi from the original graph of Kenya.

```
PREFIX osm2rdf: <https://osm2rdf.cs.uni-freiburg.de/rdf#>
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
Construct
{
  VALUES ?region { osmrel:3492709 osmrel:3492709 }
  ?region (osm2rdf:contains_area*/osm2rdf:contains_nonarea)|(osm2rdf:contains_area) ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity ?amenity .
  FILTER REGEX(?amenity, "hospital|clinic|pharmacy")
}
 WHERE {
  VALUES ?region { osmrel:3492709 osmrel:3492709 }
  ?region (osm2rdf:contains_area*/osm2rdf:contains_nonarea)|(osm2rdf:contains_area) ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity ?amenity .
  FILTER REGEX(?amenity, "hospital|clinic|pharmacy")
}
```

Figure 58. Short example of creating a sub-graph for specific facilities from the original OSM ttl file of Kenya

Finally, the return of the queries which will contain geometry information (coordinates) will be visualized through “Leaflet” or “GeoDjango” if a pure python environment is desired on an interactive map that pops up the OSM id as a direct link to the actual OSM object (Figure 55). In Figure 59 there is a diagram describing the basic setup of the proposed web application.

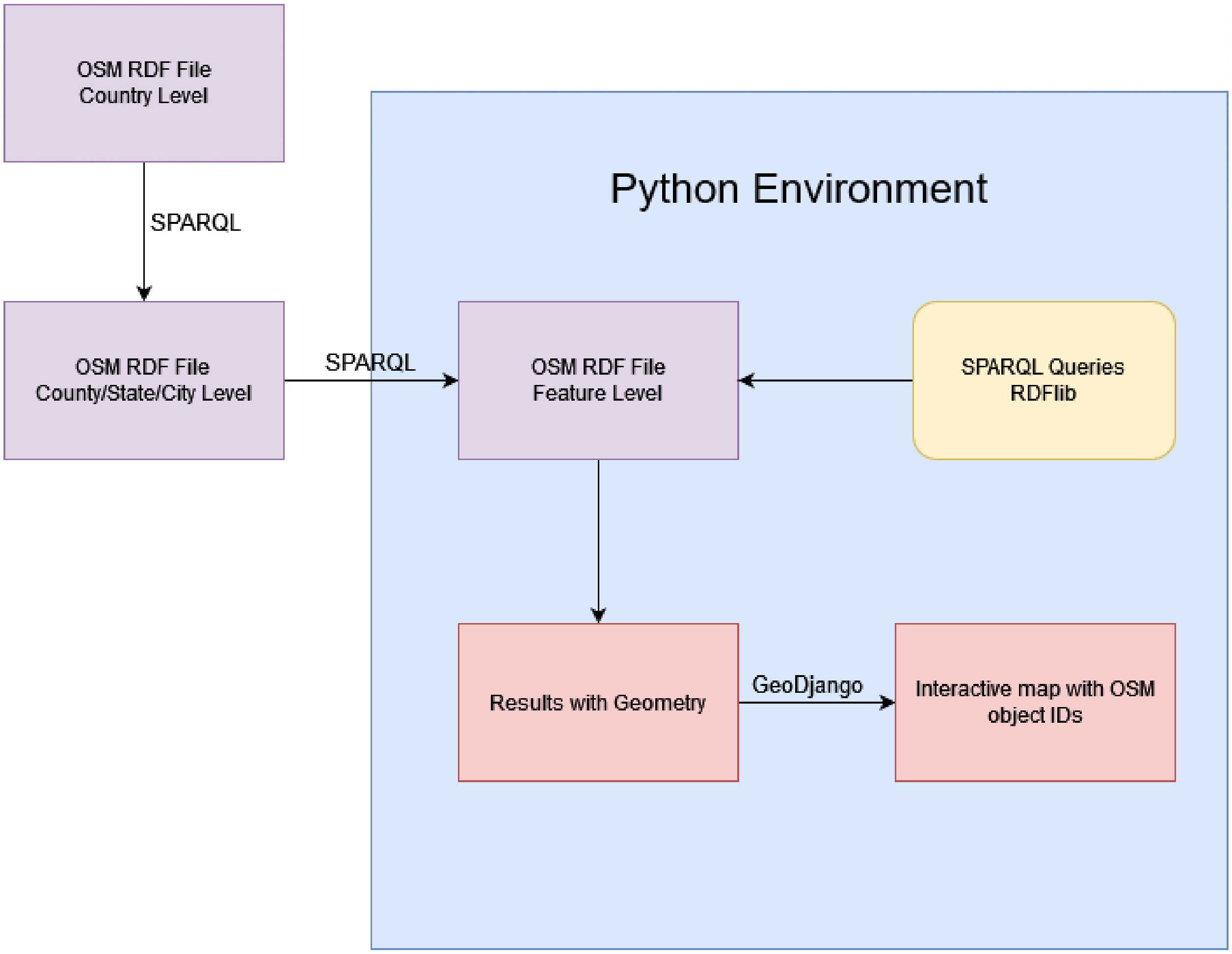


Figure 59. Diagram showing the basic structure of a web application for health facilities.

It should be mentioned here that there are geographical limitations in the use of the application. If the application is fed with the dataset of another area, then the queries are not valid anymore because they are based on the vocabulary of a specific area. The solution to this problem is to build queries based on the whole OSM vocabulary.

7 Discussion

In the course of assessing OSM data for supporting Health Services in the Global South, the present research looked into the cases of two Kenyan Counties, Nairobi and Wajir. The first contains the country's capital city with a large, dense population including huge informal settlements with very low income. The second, is mostly rural in terms of population, with low density and one of the lowest average incomes countrywide.

Fitness of Use Framework

The methods that are found in the relevant literature for assessing OSM data as a whole for an AOI lack domain-level sensitivity. For example, an OSM attribute completeness assessment with a moderate result can easily ignore that there may be a specific attribute category (e.g., health facilities) that is attributed significantly better or another like water points much worse. For this reason, our assessment was guided by a users' needs analysis for the specific type of OSM data. What was proven to be important for health field workers of the Kenyan Red Cross and their office-based colleagues was the actual completeness of the health facilities objects in OSM (but not their accurate position), their level according to Kenya Essential Package for Health (KEPH) which dictates the properties of the official health facilities classification's properties, the medical services provided, and equipment offered in each facility and their contact details. Apart from the health facilities, knowing the location of water and sanitation points (including handwashing) was considered important as well as a complete road network for navigation.

Analyzing the Vocabularies

The needs of the users had to be translated into OSM data to analyze the capacity of the OSM database to support health services. For this, a vocabulary analysis for the areas of the use cases was performed. The results showed that there are no tags in OSM in Kenya used to represent the classification of the health facilities or the equipment they use for two different reasons. The first is that in OSM, there is no key explicitly stating the classification level of a health facility. The second is that while for the equipment there is a dedicated key in OSM (`healthcare:equipment`) it is not used in the areas of the use cases. For Nairobi in OSM, there is an extensive vocabulary for the services offered by the health facilities and their contact information which is not used in Wajir. Similarly in Nairobi, there is an extensive WATSAN vocabulary that is missing from Wajir. In any case, the capacity of OSM in representing Health Services related data as those defined by the users is found by the present research adequate. In places like the informal settlements of Nairobi where humanitarian mapping has taken place, it is evident that extensive vocabulary has been used for the attribution of health and WATSAN facilities. In places that lack this organized mapping, the vocabulary is much poorer like in the case of Wajir. In the case of Nairobi County, 63 keys were found related to health facilities with 280 corresponding values while for Wajir there are only 3 keys and 8 different values. This finding when combined with the spatial distribution of the specialized health facilities related tags (tags using underscores and colons for detailed attributes, e.g., `health_facility:type=hospital`) gave the first indication that mapping projects can improve OSM data quality. This is because the sophisticated and organized OSM vocabulary was found in objects clustered around the informal settlements of Nairobi which are the grounds of most humanitarian mapping projects. Those mapping projects have built a taxonomy inside OSM that can easily be followed by future contributors.

Quality assessments

To prove that humanitarian mapping projects contribute largely to the quality of health services-related data in OSM the first step was to perform a quality assessment for the data defined in the users' needs analysis. For the health facilities, the first assessment was on the completeness of objects. In Nairobi County, the OSM data was compared to the authoritative data which is the conventional ground truth. The results showed 21.38% completeness but this percentage when zooming in to the informal settlements climbed to 64.64 %. An interesting result was the commission of 686 objects in OSM that are not present in the authoritative data with 485 (70%) of them being in the informal settlements. Those numbers indicate the importance of humanitarian mapping in adding new data not only for OSM but also by generating data that did not exist anywhere else before. One of the main objectives of mapping projects like the monumental Map Kibera, which is still in progress contributing tens of thousands of changesets since 2009, is to put the informal settlements on the map where there was only a blank before. The fact that those settlements in Nairobi are informal coincides with the total neglect from the side of the authorities. This is obvious from a data perspective as the authoritative data include 99 health facilities while in OSM there are 549. In all other areas of the county, the authoritative data include 1187 health facilities and OSM 412.

In Wajir County the completeness of health facilities in OSM was found at the beginning of the research to be 32.87% but because there was a humanitarian mapping project of the Kenyan Red Cross in progress between September and December of 2022 this percentage is at the time of writing 41.87%. Especially for the Wajir East sub-county where most parts of the fieldwork of the project were carried out the percentage of completeness raised from 18.6% to 55.8% giving us another strong indication of a positive correlation between humanitarian mapping projects and the completeness of OSM health-related data.

The attribute completeness of OSM data for the health facilities in Nairobi County was then assessed. Again, the overall percentage for attributing the type of available services in each facility is quite low reaching just 33.19% for OSM when for the authoritative data is 55%. The same percentage in OSM for Mathare informal settlement is more than 87%. The information about the ownership status or the operator type of the facilities is in almost 50% of OSM objects for the whole of Nairobi and is almost 75% of Mathare informal settlement. The contact details are found only in 7.8% of the Nairobi objects With Mathare having 0% meaning that this attribute was not defined as important during the mapping projects there. Nevertheless, this assessment provided more indications of increased OSM data quality in the healthcare domain due to humanitarian mapping projects. In Wajir, there are practically no attribute tags for the health facilities so far as shown in the vocabulary analysis.

For the assessment of OSM WATSAN data and especially those matching the users' needs analysis i.e., water points, communal sanitation points, and handwashing facilities the comparison was made with UN-Habitat data of the major informal settlements of Nairobi. This was due to the absence of authoritative data. The WATSAN data in Nairobi County are almost exclusively in the informal settlements areas which again indicates a higher quality of data produced during humanitarian mapping projects. In comparison, the data of OSM seem to be more complete when about water points and communal sanitation while the UN-Habitat data seem more complete when about handwashing facilities probably because the mapping was made as a response to covid-19 pandemic. Interestingly, UN-Habitat did not seem to use for their project the datasets that already existed in OSM which are proven to be more extensive in the cases of water points and communal sanitation facilities. For Wajir County it was not possible to find reference data set for WATSAN. Data for water points across the county came from digitizing a map coming from the county government. There were no data to assist in measuring overlap, omission, and commission of data. When examined visually on the map, the overlap of OSM and authoritative data was moderate. In net numbers, authoritative water point data surpassed moderately those of OSM

The road networks both in Nairobi and Wajir counties are more extensive in OSM than in the authoritative data of Kenya Roads Board. For Nairobi, there are 56.7% more roads in length in OSM compared to the authoritative data, at 93% completeness and this percentage for Wajir is 31% at 83% completeness. Those numbers of high completeness percentage combined with extensive road segments commission give strong indications that the OSM network is sufficient enough to support health services by assisting navigation.

It is obvious, from the numbers describing the assessment, that OSM data concerning health facilities in Nairobi or Wajir counties is far from being useful for health services at the County level but can be valuable when used for smaller areas like the informal settlements of Nairobi where authoritative data are almost absent. The percentages of completeness of health facilities in the informal settlements show explicitly that the humanitarian mapping projects have fulfilled at least one of their primary objectives. The assessment was made through comparisons between authoritative and OSM datasets, following the relevant literature.

Any comparison has inherently the characteristic of competition which is obvious in many scientific papers that try to prove that OSM is a valuable or useless source of information. But when talking about health services data, competition should be out of the picture. In both the cases of OSM data in Nairobi and Wajir where the overlapping with authoritative data is low, a combination of the datasets would give the health services a valuable tool for addressing all kinds of challenges. It is not of the slightest importance if OSM data is better or not compared to the authoritative data which is in any case a question very difficult to answer. What is of real importance is the need to use all available data sources for a common humanitarian cause in the vulnerable environments of the Global South. A complimentary to the authoritative data role for OSM could be a gamechanger for the health services.

Humanitarian Mapping Projects: A proxy of quality

After the assessment of the OSM data, several hints came to the surface about the correlation between the humanitarian mapping and the quality in terms mainly of completeness but also attributes of OSM health facilities data. For this reason, a statistical analysis of the produced data was performed. The results showed a strong positive correlation coefficient between the number of changesets with a hashtag of a humanitarian mapping project and the object completeness in the same area. This finding means that a tool could be developed that identifies the need for health facility mapping in an area, based on the absence or the small number of changesets that come from humanitarian projects. This correlation can also be used as a proxy for quality assessment and an applied method for supporting part of the relevant literature that shares the idea of the credibility of the contributor as an indicator of quality. Credibility is defined as trustworthiness combined with expertise which are characteristics that can be attributed to organized mapping projects due mostly to the presence of high supervision and guidance throughout their implementation. This could similarly be extended to other metrics than completeness. For attribute completeness concerning the `medical_service` key inside tags, the overview of the map visualization was enough to confirm the correlation because those tags are present only in the informal settlements and are produced by humanitarian mapping projects.

Building Ontologies using SPARQL

Throughout the OSM data assessment, we used the triplified version of the OSM dataset for Kenya. This was done for two reasons. The first reason is that by using SPARQL querying language it is easy to construct graphs that can work as ontologies. This way the use of an existing vocabulary in OSM is promoted rather than adding new tags. From the users' needs analysis results, there was one type of information that did not exist in OSM; the classification of the health facilities. After studying carefully, the Kenya Essential Package for Health (KEPH) where the official classification and its properties are described in detail, we were able to translate some distinct characteristics of Levels 2-6 into OSM tags. This way when querying for those tags we essentially query for the respective Level. It was thus, possible to construct an ontology for Levels 2-6. Not all identified OSM health facility objects fall under a specific KEPH Level because of the oftentimes poor tagging of objects. In some other cases, errors in the attribute tagging by the mappers can attribute the wrong classification to objects. In any case, the ontology is described so that when followed correctly by the contributors can increase the amount of information available. The aim here was not to find all facilities of each Level using a single query but to show a way that if followed correctly can substitute the need for a new tag indicating explicitly the classification. In some cases when designing an ontology, along with the respective queries that utilize it, a choice should be made between tolerance to overlapping or omitted data. The first choice is inclusive but inaccurate while the second exclusive but accurate. All this process was done for Nairobi County which has an extensive OSM vocabulary in use. Wajir on the other hand has a so limited vocabulary that it was not possible to do the same. There are two options in that case. The first one is for the contributors to follow Nairobi's vocabulary so that the proposed ontology is valid or follow OSM features suggestions something that was not done in any of the humanitarian mappings of Nairobi as seen in the vocabulary analysis. In any case, the promotion of an ontology highly depends on consistency from the side of the contributors. A change in editing habits can bring a change in pattern and make an ontology useless. Moreover, the use of words that are interchangeable in life (e.g., clinic, hospital, etc.) cannot be so in OSM. Every tag matter and the decision to use it or not and in what form can play a very important role if OSM is ever going to be a reliable tool for supporting any kind of humanitarian or other official services or just be a very good basemap.

A web application for OSM health services

The second reason for using semantic web technologies is to have the option of building a lightweight web application for finding health facilities based on their attributes. It would use a relatively small ttl file as a database containing all OSM health facilities and WATSAN objects along with their geometry and tags. Semantic web technologies can potentially enrich the results of a query with external RDF sources (e.g., Wikidata). That could mean for example querying for an object, a value, or a tag against OSM but asking in parallel for related media from Wikidata. After constructing all the queries for the assessment and the ontologies that promote the use of the existing tags the setting up of the application is possible. The turtle file for Kenya is provided by the University of Freiburg, and it is regularly updated even if not as often as promised. For extracting smaller graphs, it is possible to construct graphs from this turtle file for specific OSM relations which in our case are Nairobi and Wajir Counties. Then by utilizing the queries already built for the OSM data quality assessment (object completeness, attribute completeness, attribute accuracy), and also the ontology of the facility classification Level, the potential user could with one click run them against the database. The whole application can be developed in Django using Python with RDFlib for the SPARQL queries and GeoDjango or Leaflet for the visualizations.

Limitations

One of the major limitations of the present research is that it is largely based on a third party. The Department of Algorithms of the University of Freiburg is supplying the triplified dataset that we have been using throughout the research, but there is no guarantee that this will be continued. Of course, there are several tools available that triplify OSM data but not automatically which means that for all the procedures described, we should every time or regularly triplify the data again. This is because OSM is organic in nature and data can rapidly and massively change especially after humanitarian mapping in an AOI. But change is not only concerning the number of objects mapped but also their tagging.

In that sense, another limitation that is inherent in OSM becomes obvious. The vocabulary analysis that was performed during the research can be outdated if the tagging patterns change or if new mapping projects occur that follow the proposed vocabulary by the OSM community which has not been happening until now.

Another major limitation of the proposed framework is the locality of the OSM vocabulary. All of the present research was based on the vocabulary analysis made for Wajir and Nairobi Counties. Even within the same country, Kenya, the vocabulary varies from county to county. This means that the queries that have been built based on the vocabulary analysis are only valid for the areas of interest.

Regarding the proxy of quality assessment for OSM health facilities the limitation is that there is not any tool available, until the moment of writing, that can relate hashtags of changesets to coordinates. This means that producing a thematic map of this proxy can be done only manually and only within administrative boundaries as done with OSMCha.

Further research

Using the existence of humanitarian mapping projects as a proxy for quality could be extended to more domains than health facilities. Humanitarian mapping through OSM has contributed to mapping buildings, road networks, amenities, public services, and more features. Research on if this could be a proxy for the OSM data quality in those domains might help finding new methods to perform some quick preliminary quality assessments of OSM data.

During the limited time of the research, the time spent on building the SPARQL queries was not enough to come to more sophisticated and thus shorter versions. Limiting the multiple UNION use by employing instead the VALUES clause can be an option in this direction. Some tests done during the research were not successful, leaving out some of the desired results. Making the queries shorter can make them more adequate for a potential web application for reducing response time and coding lines.

The development of a tool that could handle hashtags and comments from OSM objects as real data with coordinates that can be queried for, could help group data according to their source mapping project. An insight into every mapping project separately and their quality assessment could highlight best practices for future use or gaps that should be filled. An idea of achieving that in another way would be trying to scrape data from OSMCha and check whether the results of the queries contain the geometry of the changesets.

8 Conclusions

In this research the main objective was, since the beginning, to support the vulnerable communities of the Global South by giving access to as more as possible reliable data resources with zero cost. The Volunteered Geographic Information platforms and especially OpenStreetMap seemed an ideal source of information but at the same time underutilized. The first step was to choose two use cases to work with that have different socioeconomic characteristics: the Nairobi and Wajir counties of Kenya. Through an extensive literature review on the quality assessment of OSM, it was obvious that the methods used were too generic to capture the quality level of specific domains. This problem was addressed by the present research through a users' needs analysis that identified what is important information for the field health workers, their coordinators, and the individuals. The analysis showed that the important information is the completeness of the dataset of health facilities, the services they provide, their ownership status, and their contact details along with data about the water and sanitation of an area of interest and the road network. Answering the first research question, OpenStreetMap has an unlimited capacity of representing any kind of object and phenomena on the map. The totally free tagging system is guaranteeing that. This was obvious in the case of Nairobi County where even if the contributors did not use the suggested by the OSM community vocabulary they could describe to detail services and operational details of the health facilities. The divergence though from the information offered by the authoritative data was significant. The needs identified were very specific and really close to what the authoritative data have to offer.

Those needs were then translated into OSM tags but only after studying in detail the relevant OSM vocabulary of each use case. By combining the users' needs analysis results with the vocabularies, a fitness-of-use framework was constructed in which the domain-specific quality assessment was later made. During the assessment, the biggest challenge was to find all possible tags that indicate a health facility or a medical service. The vocabulary analysis gave the answers to this problem. By using triplified OSM data for the use cases and extensive SPARQL queries it was possible to unify the heterogenous and fragmented vocabulary of OSM in any data category indicated by the user's needs analysis. Those queries containing all tags of a specific feature, or all the values of a specific key could potentially lead to an ontology that gives information that is not represented by tags. This was attempted in this research by building an ontology for the classification of the health facilities in Nairobi according to the national official system. Querying for the tags of the ontology to find the health facilities of specific levels gave back some positive results but also highlighted the inherent weakness of every VGI which has to do with wrong tagging practices by the users. The validation of the proposed ontologies was not done in-depth, but it is something that can be investigated in the future. Concerning the second research question, the use of SPARQL querying language and RDF version of OSM data made it possible to construct sophisticated queries combining any kind of keys and values to locate exactly the information needed. This process of constructing queries and sub-graphs was then easily turned into an ontology for describing a phenomenon that is not even tagged in OSM.

Afterward, the quality assessment was made using mainly authoritative datasets. The metrics used were chosen from the relevant literature to measure the object completeness of the health facilities, WATSAN points, and road networks, the attribute completeness, and the attribute accuracy of the facilities. The results for Nairobi showed that the OSM health facilities and WATSAN data were clustered around the informal settlements where most of the humanitarian mapping has taken place. For Wajir county, the tagging is relatively poor and the completeness low, but it increased dramatically by a humanitarian mapping project that was taking place there during the research. This spatial heterogenicity of OSM data indicated two things. The first was that OSM data at the moment is not able to support health services at the County level for Kenya but also that humanitarian mapping can significantly improve OSM data quality.

A statistical analysis was then performed to prove if there is a correlation between the humanitarian mapping activity in the informal settlements of Nairobi and the data quality in terms of completeness. The correlation proved to be positive and strong, and this partially replies positively to the question of whether humanitarian mapping activity can be a proxy of quality for OSM health facilities in terms of completeness. To be absolutely positive more research is needed with larger samples for the statistical analysis. There are also indications that this correlation can be extended to attribute completeness quality, but this also needs further research. As for the third research question, organized mapping projects were found to be correlated to the completeness of health facilities in OSM as well as to their attribute completeness. This phenomenon was present mainly in the places where the state is absent i.e., the informal settlements of Nairobi County. This is no surprise as humanitarian mapping has this very objective.

The last objective of the present research was to pave the way toward a web application based on OSM data that can support health services in the Global South. Having already described the fitness-of-use framework and the queries for finding data from triplified OSM datasets and by building ontologies to capture information that is not tagged but exists in OSM the theoretical base is available. The triplified dataset is provided on a regularly updated basis by the University of Freiburg and a python environment is adequate for dealing with RDF data and web GIS applications. The basic components are available as well as the tools and the concept so a web application that is dedicated to health services using OSM data is possible with a scale limitation because of the heterogenous OSM vocabulary. All the work done for answering the three first research questions produces the basic material for the components of a proposed web application that could do the simple thing of querying for information that assists health services against a chaotic OSM database without the need for deep vocabulary knowledge by the user.

Throughout this research, it was made clear that no dataset from those used was complete and accurate. The authoritative data were much more structured and provided information that is much closer to the needs of the users. This is probably the result of accumulated experience and the fact that those datasets are produced by or in collaboration with people already working in the domain. The main drawback of the authoritative datasets as seen in the use cases was not the currentness as someone could imagine but the neglect of whole areas that are called informal settlements. Explaining the reasons for this is out of the scope of this research and extends to the grounds of social geography and the theories of social exclusion. The real value of OSM is that it can provide free valuable data, especially in areas where even the state is absent. Areas that are blank on the map. This is the story of Kibera claiming to be the largest informal settlement in Africa, at least until some years ago. In 2009 the young Kiberans decided to put their homeland on the map. And they made it with OpenStreetMap. Now the information is there, and no one can claim they do not know. It is up to the Kenyan authorities to use those data for enriching their official datasets, and it is up to the OSM communities and mapping teams to add authoritative data to OSM. The aim should not be only to make OSM more complete but to also improve the authoritative data through crowdsourcing. It is not about which option is better but about how can the world be better through joint datasets and efforts. On the other side uploading bulk authoritative data is probably taboo for OSM communities but this is maybe the way to create the most complete geographic database in the world. The small overlapping of health facilities in the case of Nairobi shows that a combined effort from both sides could bring wonderful results. What is needed is consistency during the editing because all tags matter. Hopefully, the various OSM mapping teams can bring in that characteristic and spread it throughout the OSM global community.

Reference List

Anderson, J., Sarkar, D., Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *8*(5), 232. https://doi.org/10.3390/ijgi8050232

Anelli, V. W., Calì, A., di Noia, T., Palmonari, M., & Ragone, A. (2016). Exposing Open Street Map in the Linked Data Cloud. In H. Fujita, M. Ali, A. Selamat, J. Sasaki, & M. Kurematsu (Eds.), Trends in Applied Knowledge-Based Systems and DataScience (pp. 344–355). Springer International Publishing.

Antoniou, V., & Skopeliti, A. (2015). MEASURES AND INDICATORS OF VGI QUALITY: AN OVERVIEW. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *2*.

Athanasiou, S., Hladký, D., Giannopoulos, G., García-Rojas, A., & Lehmann, J. (2014). GeoKnow: Making the Web an Exploratory Place for Geospatial Knowledge. ERCIM News, 2014.

Ballatore A., Bertolotto M., Wilson, D. (2013). Geographic Knowledge Extraction and Semantic Similarity in OpenStreetMap. Knowledge and Information Systems. 37. 10.1007/s10115-012-0571-0.

Barron C., Neis P., Zipf A., (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS. 18. 10.1111/tgis.12073.

Bast, H., Brosi, P., Kalmbach, J., & Lehmann, A. (2021). An Efficient RDF Converter and SPARQL Endpoint for the Complete OpenStreetMap Data. Proceedings of the 29th International Conference on Advances in Geographic Information Systems, 536–539. https://doi.org/10.1145/3474717.3484256

Bedelian C. (2019). Improving the governance of water points in Wajir County. 10.13140/RG.2.2.25319.27043

Behrens J., van Elzakker C. P. J. M., Schmidt M. (2015). Testing the Usability of OpenStreetMap's iD Tool. *The Cartographic Journal*, *52*(2), 177–184. https://doi.org/10.1080/00087041.2015.1119457

Borkowska S., Pokonieczny K. (2022). Analysis of OpenStreetMap Data Quality for Selected Counties in Poland in Terms of Sustainable Development. *Sustainability*. 14(7):3728. https://doi.org/10.3390/su14073728

Chen, J., Deng, S., & Chen, H. (2017). CrowdGeoKG: Crowdsourced Geo-Knowledge Graph (pp. 165–172). https://doi.org/10.1007/978-981-10-7359-5_17

Davidovic N., Mooney P., Stoimenov L., Minghini M. Tagging in Volunteered Geographic Information: An Analysis of Tagging Practices for Cities and Urban Regions in OpenStreetMap. *ISPRS International Journal of Geo-Information*. 2016; 5(12):232. https://doi.org/10.3390/ijgi5120232

Devillers R., Bédard Y., Jeansoulin R., Moulin B., (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. International Journal of Geographical Information Science 21, 261–282. DOI: https://doi.org/10.1080/13658810600911879

Flanagin A., Metzger M. (2008). The Credibility of Volunteered Geographic Information. GeoJournal. 72. 137-148. 10.1007/s10708-008-9188-y.

Fonte C., Antoniou V., Bastin L., Estima J., Jokar Arsanjani J., Laso Bayas J., See L., Vatseva R. (2017). Assessing VGI Data Quality. 10.5334/bbf.g.

Forbes G. (1999). Urban roadway classification. In *Urban Street Symposium*. Washington, DC: Dallas, Texas.

Girres J.F., Touya G. (2010). Quality Assessment of the French OpenStreetMap Dataset. T. GIS. 14. 435-459. 10.1111/j.1467-9671.2010.01203.x.

Golder S., Huberman B. (2006). Usage Patterns of Collaborative Tagging Systems. J. Information Science. 32. 198-208. 10.1177/0165551506062337.

Goodchild M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69(4):211–221.

Haklay M., Weber P. (2008). OpenStreetMap: User-Generated Street Maps. *Haklay, M. and Weber, P. (2008) OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing, 7 (4). Pp. 12-18. ISSN 15361268*, *7*. https://doi.org/10.1109/MPRV.2008.80

Haklay M. (2010). How good is volunteered geographical information? A comparative study of Open StreetMap and ordnance survey datasets. Environment and Planning. B, Planning & Design, 37(4), 682–703. doi:10.1068/b35097

Haklay M., Basiouka S., Antoniou V., Ather A. (2010) How many volunteers does it take to map an area well? the validity of Linus' law to volunteered geographic information. Cartogr. J. 47(4), 315–322 (2010)

Hashemi P., Abbaspour R. (2015). Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept. 10.1007/978-3-319-14280-7_2.

Helbich M., Amelunxen C., Neis P. (2012). Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. Angewandte Geoinformatik.

Herfort B., Lautenbach S., Porto de Albuquerque J. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Sci Rep* **11**, 3037. https://doi.org/10.1038/s41598-021-82404-z

International Organization for Standardization. ISO 19157:2013 Geographic Information—Data Quality; ISO:Geneva, Switzerland, 2013.

Jackson S., Mullen W., Agouris P., Crooks A., Croitoru A., Stefanidis A. (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. ISPRS International Journal of Geo-Information. 2. 507-530. 10.3390/ijgi2020507.

Jokar Arsanjani J. & Barron C., Bakillah M., Helbich M. (2013). Assessing the Quality of OpenStreetMap Contributors together with their Contributions. 16th AGILE Conference. Leuven, May 14-17, 2013

Jokar Arsanjani J., Zipf A., Mooney P., Helbich M. (2015a). An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. 10.1007/978-3-319-14280-7_1.

Jokar Arsanjani J., Zipf A., Mooney P., Schauss A. (2015b). Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets. 10.1007/978-3-319-14280-7_3.

Keßler C., Trame J., Kauppinen, T. (2011). Tracking Editing Processes in Volunteered Geographic Information: The Case of OpenStreetMap. Conference On Spatial Information Theory.

Kaur J., Singh J., Sehra S., Rai H. (2017). Systematic Literature Review of Data Quality Within OpenStreetMap. 177-182. 10.1109/ICNGCIS.2017.35.

Ludwig I., Voss A., Krause-Traudes M. (2011). A Comparison of the Street Networks of Navteq and OSM in Germany. Advancing Geoinformation Science for a Changing World, 65–84. doi:10.1007/978-3-642-19789-5_4

McGough A., Kavak H., Mahabir R. (2022). Revisiting Linus' Law in OpenStreetMap: An Agent-Based Approach. In: Thomson, R., Dancy, C., Pyke, A. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2022. Lecture Notes in Computer Science, vol 13558. Springer, Cham. https://doi-org.proxy.library.uu.nl/10.1007/978-3-031-17114-7_12

Mooney P., Corcoran P., & Winstanley A. C. (2010). Towards Quality Metrics for OpenStreetMap. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 514–517. https://doi.org/10.1145/1869790.1869875

Mooney P., Corcoran P. (2012a). The Annotation Process in OpenStreetMap. , 16(4), 0–0. doi:10.1111/j.1467-9671.2012.01306.x

Mooney P., Corcoran P. (2012b). Characteristics of Heavily Edited Objects in OpenStreetMap. Future Internet. 4. 285-305. 10.3390/fi4010285.

Qian X, Di L, Li D, Li P, Shi L, and Cai L 2009 Data cleaning approaches in Web2.0 VGI application. In Proceedings of the Seventeenth International Conference on Geomatics, Fairfax, Virginia: 1–4

Raymond E. S. (2001). The Cathedral and the Bazaar, p. 19, O'Reilly Media, Sebastopol, CA.

Sehra S. S., Singh J., Rai H. (2014). [IEEE 2014 Eleventh International Conference on Information Technology: New Generations (ITNG) - Las Vegas, NV, USA (2014.04.7-2014.04.9)] 2014 11th International Conference on Information Technology: New Generations - A Systematic Study of OpenStreetMap Data Quality Assessment. , (), 377–381. doi:10.1109/itng.2014.115

Senaratne H., Mobasheri A., Loai Ali A., & Capineri C. & Haklay, Muki. (2016). A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science. 31. 10.1080/13658816.2016.1189556.

Simon R., Haslhofer B., Jung J. (2011). Annotations, Tags and Linked Data. Metadata Enrichment in Online Map Collections Through Volunteer-Contributed Information. e-Perimetron. 6. 129-137.

Smith B., Mark D. (2001). Do Mountains Exist? Towards an Ontology of Landforms. Environment and Planning B Planning and Design. 30. 10.1068/b12821.

Špiranec S., Ivanjko T. (2013). Experts vs. Novices Tagging Behavior: An Exploratory Analysis. Procedia - Social and Behavioral Sciences. 73. 456–459. 10.1016/j.sbspro.2013.02.076.

Stadler C., Lehmann J., Höffner K., Auer S. (2012). LinkedGeoData: A Core for a Web of Spatial Open Data. Semantic Web Journal, 3, 333–354. https://doi.org/10.3233/SW-2011-0052

Sui D.Z. (2008) The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS, Computers Environment and Urban Systems, 2008, 32, 1–5

Toroitich A. & Dunford L. Armitage R., Tanna S. (2022). Patients Access to Medicines – A Critical Review of the Healthcare System in Kenya. Risk Management and Healthcare Policy. Volume 15. 361-374. 10.2147/RMHP.S348816.

Vandecasteele A., Devillers R. (2015). Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. 10.1007/978-3-319-14280-7_4.

Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., & Zipf, A. (2020). *Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. International Journal of Geographical Information Science, 1–27.* doi:10.1080/13658816.2020.1730848

Zhang H., Malczewski J. (2017). Quality Evaluation of Volunteered Geographic Information: The Case of OpenStreetMap. 10.4018/978-1-5225-2446-5.ch002.

Appendix A

Health facilities and WATSAN OSM vocabulary for Wajir and Nairobi Counties

1. Nairobi

1.1 Nairobi Health Facilities

OSM key	Number of values
amenity (only related to health facilities)	9
healthcare	13
healthcare_speciality	18
health_facility	1
health_facility:type	24
health_facility:referals	67
health_facility:description	9
health_facility:patients_per_day	39
health_facility:exam_room	2
health_facility:bed	10
health_facility:cot	5
despensing	26
medical_service	1
medical_service:outpatient	2
medical_service:family_planning	2
medical_service:pregnancy_test	2
medical_service:malaria	2
medical_service:general_medical_services	2
medical_service:condom_distribution	2
medical_service:vct_hiv_counselling_test	2
medical_service:circumcision	2
medical_service:immunizations	2
medical_service:sexually_transmitted_infections_management	2
medical_service:antenatal_care	2
medical_service:dental_services	2
medical_service:minor_surgery	2
medical_service:psychosocial_support_counselling	1
medical_service:integrated_management_of_childhood_illness	1
medical_service:growth_and_nutritional_support	2
medical_service:inpatient	2
medical_service:home_based_care	2
medical_service:youth_friendly_services	2
medical_service:ear_nose_throat	2
medical_service:pmtct	2
medical_service:tb_diagnosis	2
medical_service:gender_based_violenve_services	2
medical_service:tb_treatment	1
medical_service:tb_lab	2

medical_service:antiretroviral_therapy	2
medical_service:other	2
medical_service:basic_emergency_obsteric_care	2
medical_service:palliative_care	2
medical_service:comprehensive_essential_obsteric_care	1
medical_service:disease_support_groups	4
medical_service:prevention	8
medical_service:health_insurance	2
medical_service:maternity	1
medical_service:blood_sugar_test	1
medical_service:cancer_screening	1
medical_service:blood_transfusion	1
medical_staff:pharmacist	6
medical_staff:nurse	10
medical_staff:clinical_officer	6
medical_staff:counsellor	7
medical_staff:lab_tech	5
medical_staff:pharm_tech	5
general_practitioner	4
medical_staff:community_pharmacist	3
medical_staff:medical_officer	4
medical_staff:midwife	6
medical_staff:other	16
medical_staff:description	6

amenity (10/108)

```
"pharmacy"
"hospital"
"clinic"
"dentist"
"laboratory"
"doctors"
"medical transport service"
"traditional health centre"
"nursing_home"
```

healthcare (13)

```
"hospital" .
"clinic" .
"pharmacy" .
"alternative" .
"laboratory" .
"dentist" .
"yes" .
"*" .
"centre" .
"clinic;doctor;physiotherapist" .
"doctor" .
"hospice" .
"nursing_home" .
```

healthcare:speciality (18)

```
"gynaecology" .
"gynaecology;paediatrics;radiology" .
"paediatrics" .
"general" .
"gynaecology;paediatrics" .
"radiology" .
"general;dentist" .
"community;general" .
"fertility" .
"gynaecology;paediatrics;surgery" .
"gynaecology;radiology" .
"oncology;general;cardiology;ophthalmology" .
"biological_haematology;blood_check;dental_oral_maxillo_facial_surgery;dermatology;
                                  emergency;gynaecology;occupational;vaccination"
"clinic;community" .
"dentist" .
"general;gynaecology;paediatrics;dentist;orthopaedics;radiology;urology;
           physiotherapy;surgery;cardiology;neurology;plastic_surgery" .
"gynaecology;paediatric_surgery" .
"paediatrics;radiology" .
```

health_facility:type (24)

"chemist_dispensing" .
"medical_clinic" .
"chemist_over_the_counter" .
"pharmacy" .
"clinic" .
"alternative" .
"herbalist" .
"laboratory" .
"health_center" .
"hospital" .
"other" .
"dispensary" .
"dentist" .
"dental_clinic" .
"nursing_home_with_maternity" .
"health_programme" .
"centre" .
"vct_center" .
"chemist_over_the_counter;chemist_dispensing" .
"health_center;hospital" .
"Community based organization" .
"Dispensing drugs" .
"Health clinic" .
"rehabilitation_center" .

health_facility:referals (67)

"KNH" .
"Mbagathi District Hospital,KNH" .
"KNH,Mbagathi" .
"IDH,KNH" .
"yes" .
"Mbagathi,Kenyatta National Hospital" .
"st mary langata" .
"GENERAL" .
"IDH,Mbagathi,IDEWES" .
"Kenyatta National Hospital, Mbagathi Hospital" .
"Kenyatta,Mbagadthi and St Mary,s Lang'ata" .
"KHN" .
"Kibera South,Shofco" .
"KNH, Mbagathi" .
"KNH, Mbagathi," .
"KNH,IDH" .
"KNH,Magqathi,St Marys" .
"KNH,Mbagathi,TB-refer to AMREF" .
"KNH,St Mary's Langata,mbagathi" .
"Mbagathi" .
"Mbagathi, KNH" .
"Mbangathi,KNH" .
"MSF Belgium Kwa wanga" .
"MSF,St Marys,Langata,KNH,mbagathi" .
"Nairobi Women,FIDA,CRADO" .
"their other branch in Kawangware" .

"Ushirika Medica Center,Coptic" .
"YES" .
"Amref" .
"AMREF,Tabitha Kwawanga" .
"General" .
"Kawangware Herbal" .
"Kenyatta and Mbagathi District Hospital" .
"Kenyatta/Mbagathi" .
"KHN,Nairobi west hospital" .
"KNH & Patient's preference" .
"KNH, COPTIC, Nairobi West, Nairobi Women" .
"KNH, Mbagathi, Nairobi Women, Coptic" .
"KNH, ST MARYS OTIENDE" .
"KNH,Mbagathi,St Mary" .
"KNH,Mbangathi" .
"KNH,Pumwani" .
"KNH,St Marys Mission" .
"KNH,Ushirika, Mbagathi" .
"Kwa Wanga,Mbagathi District Hospital" .
"Lea toto,PAG,KICOSHEP,Huduma" .
"Main Shifco branch,KNH,Mbagathi,Ndama Place" .
"major hospitals" .
"Marie stopes Eastliegh" .
"Mbagadhi district hospital/Vipawa medical clinic" .
"Mbagathi District Hospital" .
"Mbagathi District Hospital,KNH,Nairobi Women Hospital" .
"Mbagathi District Hospital,St Mary's Mission Hospital,KNH" .
"Mbagathi,Pathocare" .
"Mbagathi,St Marys" .
"MSF,KNH,Mbagathi" .
"Otiende Health Center" .
"PAG,AMREF Kibera South" .
"PAG,MSF Belgium,CDC,CFK" .
"SENYE, ST. MARYS, USHIRIKA" .
"St Mary's hospital" .
"ST. MARY'S, KNH, Mbagathi" .
"Tabitha,AMREF" .
"Trans-chem" .
"Ushirika Medica Center" .
"Vipawa" .
"Woodly Health Center,KNH" .

health_facility:description (9)

```
"CHURCH" .
"clinic" .
"Counselling and Youth Friendly Centre for Mental Health" .
"Exam Room,Over the Counter and Prescription" .
"Home-Based Care Programme forb HIV/AIDS Positive Women" .
"offer talks on sexually reproductive health" .
"Physiotherapy and Psychotherapy Home" .
"Reseach institution for tropical infections, administrative center of CDC"
"some prescription" .
```

health_facility:patients_per_day (39)

```
"20" .
"10" .
"10-20" .
"30" .
"15" .
"50" .
"12" .
"25" .
"120" .
"13" .
"20-30" .
"100" .
"14" .
"17" .
"20-40" .
"20-60" .
"200" .
"31" .
"32-35" .
"90-100" .
"above 10" .
"Above 50" .
"Average 15" .
"Average 18" .
"10 and above" .
"10- 15" .
```

"10-15" .
"16" .
"3-5" .
"30-60" .
"30-70" .
"5" .
"5-10" .
"60" .
"7" .
"80" .
"Average 20" .
"Average 30" .
"varies" .

health_facility:exam_room (2)

"yes" .
"no" .

health_facility:bed (10)

"7" .
"5" .
"10" .
"2" .
"3" .
"6" .
"0" .
"14" .
"4" .
"8" .

health_facility:cot (5)

"1" .
"2" .
"0" .
"8" .
"14" .

health_facility (1)

"Clinic with a laboratory" .

dispensing (26)

```
"yes" .
"no" .
"trust and rejoice medical clinic" .
"Pharmacy" .
"chemist" .
"medical clinic" .
"awambo chemist" .
"Baba jomba clinic" .
"Bayland pharmacy" .
"Candy Phamacy" .
"Caros chemist and cosmestics" .
"chemist (not functional)" .
"clinic" .

"Grace chemist (not functional)" .
"Imara centre chemist" .
"Kayaba medical chemist" .
"Lengo medical clinic" .
"Lucky summer pharmacy" .
"Mandazi road chemist" .
"Mauzuri Chemist" .
"Mlembe chemist" .
"Mwema dispensary" .
"Romiera medial centre" .
"St. maranatha clinic" .
"ST.JAMES MEDICAL CLINIC" .
"Tiriki chemist" .
```

medical_service (1)

```
"family_planning" .
```

medical_service:* “yes”, “no”

osmkey:medical_service:outpatient
osmkey:medical_service:family_planning
osmkey:medical_service:pregnancy_test
osmkey:medical_service:malaria
osmkey:medical_service:general_medical_services
osmkey:medical_service:condom_distribution
osmkey:medical_service:vct_hiv_counselling_test
osmkey:medical_service:circumcision
osmkey:medical_service:immunizations
osmkey:medical_service:sexually_transmitted_infections_management
osmkey:medical_service:antenatal_care
osmkey:medical_service:dental_services
osmkey:medical_service:minor_surgery
osmkey:medical_service:psychosocial_support_counselling
osmkey:medical_service:integrated_management_of_childhood_illness
osmkey:medical_service:growth_and_nutritional_support
osmkey:medical_service:inpatient
osmkey:medical_service:home_based_care
osmkey:medical_service:youth_friendly_services
osmkey:medical_service:ear_nose_throat
osmkey:medical_service:pmtct
osmkey:medical_service:basic_emergency_obsteric_care
osmkey:medical_service:tb_diagnosis
osmkey:medical_service:gender_based_violence_services
osmkey:medical_service:tb_treatment
osmkey:medical_service:tb_labs
osmkey:medical_service:antiretroviral_therapy
osmkey:medical_service:other
osmkey:medical_service:palliative_care
osmkey:medical_service:comprehensive_essential_obsteric_care
osmkey:medical_service:disease_support_groups
osmkey:medical_service:prevention
osmkey:medical_service:health_insurance
osmkey:medical_service:maternity
osmkey:medical_service:blood_sugar_test
osmkey:medical_service:cancer_screening
osmkey:medical_service:blood_transfusion

medical_service:prevention (8)

```
"Bilhazia,Other water born dieases" .
"CHOLERA" .
"HIV" .
"Malaria,Diarrhea,STIs" .
"none" .
"Polio,Measles" .
"STIs" .
"Tetanus,Malaria" .
```

medical_service:disease_support_groups (4)

```
"HIV" .
"HIV/AIDS" .
"Hiv/Aids,STI" .
"none" .
```

medical_service:other

```
"Maternity" .
"Blood pressure test" .
"Blood sugar test" .
"also offer gynaecological service" .
"chemist" .
"community cleaning" .
"curative, laboratory" .
"HIV1 and 2 screening" .
"Injections,Dresswounds" .
"lab testing,vaccination" .
"offer income generating programmes,training and loaning scheme,awareness and training in hiv/aids,
                                    recreation,sports,drama,video and community library services." .
"pregnancy crisis control and gynaecologist" .
"ultra sound scanning" .
```

medical_staff:* "number" (10)

```
osmkey:medical_staff:pharmacist
osmkey:medical_staff:nurse
osmkey:medical_staff:clinical_officer
osmkey:medical_staff:counsellor
osmkey:medical_staff:lab_tech
osmkey:medical_staff:pharm_tech
osmkey:medical_staff:general_practitioner
osmkey:medical_staff:community_pharmacist
osmkey:medical_staff:medical_officer
osmkey:medical_staff:midwife
```

osmkey:medical_staff:other

```
"0" .
"Herbalist" .
"2" .
"1 Lawyer, 5 Paralegals" .
"4" .
"Nutritionist" .
"One attendence" .
"1" .
"2 attendance" .
```

osmkey:medical_staff:description

```
"community health volunteers" .
"nurse aid" .
"Have community health workers" .
"Physiotherapist" .
"Social workers" .
"Teachers 3." .
```

operator:type

```
"private" .
"ngo_cbo" .
"ngo_international" .
"faith_based_organization" .
"ministry_of_health" .
"ngo_national" .
"ngo_faith_based" .
"government" .
"ngo" .
"private;individual" .
"community" .
"Naomi Nkaambi" .
"ngo_self_help" .
```

1.2 Nairobi WATSAN (water and sanitation)-related features

OSM key	Number of values
amenity (watsan related)	4
drinking_water	2
water_point	2
water_source	6
water	8
water:availability_status	2
water:source_of_water	1
water:hand_wash_facility	1
Water_Point:operational_status	1
water_tank	1
watsan:type	24
watsan:type_description	17
watsan:users_per_day	147
watsan:user_per_day	2
watsan:water_availability	13
watsan:water_availability_description	6
watsan:connection	9
watsan:connection_description	5
watsan:location	11
watsan:location_description	11
watsan:cleaner	32
watsan:cleaner_description	2
watsan:handwashing	10
watsan:hand_washing	8
watsan:handwashing_description	2
watsan_handwashing:no_point	1
watsan_handwashing:water	2
watsan_handwashing:soap	1
watsan_handwashing:without_water	1
watsan:bathroom	3
watsan:sanitary_bin	2
watsan:cost	174
watsan:cost_per_20l	3
watsan:availability	3
watsan:toilet_type	14
watsan:toilet_type_description	5
watsan:operational_status	8
watsan:operator	7

watsan:toilet_status	47
watsan:toilet_condition	2
watsan:type_status	1
watsan:water_public	2
watsan:dumping_site	2
watsan:features_per_location	20
watsan:open_defecation_area	1
watsan:storage_tank	2
watsan:opening_hours	21
watsan:urban_agriculture	2
watsan:other	2
watsan:other_description	12
watsan:disabled_access	2
watsan:light	2
watsan:pee_point	2
watsan:kwaho_pet_bottless	2
watsan:kwaho_sodis	1
watsan:kwaho_description	5
watsan:kwaho_hwts	1
watsan:kwaho_health_products	2
watsan:kwaho_health_products_description	1
watsan:funding	3
watsan:recycling	1
watsan:litres_per_day	4
watsan:menstrual_facility	2
watsan:biocentre	2
watsan:owner	1
watsan:private	1
watsan:toilet_operator	1
watsan:operartor	1
watsan:operetor	1
watsan:operartional_status	1
watsan:operatonal_status	1
watsan:operators	1
watsan_toilet:operational_hours	8
watsan_toilet:type_description	6
watsan_toilet:owner	1
watsan_toilet:type	3
watsan_toilet:status	6
watsan:toilet:status	3
watsan:toilet:type	1
watsan_goal:type	6

watsan_goal:operation_hours	3
watsan_goal:type_description	2

amenity (4/108)

```
“drinking_water”
“water_point”
“community_centre;drinking_water”
“school”
```

drinking_water (2)

```
"yes" .
"no" .
```

water_point (2)

```
"operational" .
"non operational" .
```

water_source (6)

```
"tap" .
"well" .
"water_tank" .
"main" .
"truck" .
"groundwater" .
```

water:source (6)

```
"tap" .
"water_tank" .
"water" .
"unknown" .
"none" .
"None" .
```

water (8)

"pond" .
"basin" .
"river" .
"reservoir" .
"lake" .
"wastewater" .
"lake;pond" .
"drain" .

water:availability_status (2)

"always" .
"often" .

water:source_of_water (1)

"tap" .

water:hand_wash_facility (1)

"yes" .

Water_Point:operational_status (1)

"operational" .

water_tank (1)

"yes" .

2. Wajir County

2.1 Wajir Health Facilities

OSM key	Number of values
healthcare	3
amenity (healthcare related)	4
dispensing	2

2.2 Wajir WATSAN (water and sanitation)-related features

OSM key	Number of values
amenity(watsan related)	2
drinking_water	2
water	4
water_source	4
water_seasonality	7
Waterway	7

Appendix B

Python script for scraping data from the Kenya Master Health Facility List

```python
import csv
import sys
import requests


# Set the API token and the maximum number of rows to fetch
# The API token can be found by inspecting the headers of the network requests to the API in the browser
# and looks like this: Authorization: Bearer 5qLDOCEIGzRovWF6uGuzS7eUsGf8DQ
api_token="5qLDOCEIGzRovWF6uGuzS7eUsGf8DQ"
max_rows = 5000

headers = {
    "Authorization": "Bearer {}".format(api_token),
    'Accept': 'application/json',
    'Content-Type': 'application/json',
    }

# Get the data from the http://kmhfl.health.go.ke/ API
def get_counties():
    # Get the data from the API
    r = requests.get('https://api.kmhfl.health.go.ke/api/common/filtering_summaries/?fields=county',headers=headers)
    # Return the data as a JSON object
    return r.json().get('county')

def get_county_data(county_id):
    r = requests.get('https://api.kmhfl.health.go.ke/api/facilities/material/?fields=id,code,name,&page_size={}&county={}'.format(max_rows,county_id),headers=headers)
    return r.json().get('results')

def implode_services(service_list):
    return ",".join(service_list)

def write_to_csv(data, county):
    with open('{}.csv'.format(county), 'w', newline='') as csvfile:
        fieldnames = ['name', 'sub_county_name', 'lat','long','service_names']
        writer = csv.DictWriter(csvfile, fieldnames=fieldnames,extrasaction='ignore')
        writer.writeheader()
        for row in data:
            writer.writerow(row)

if __name__ == "__main__":
    # get the parameter that was passed to the script
    try:
        parameter = sys.argv[1]
    except IndexError:
        exit("No parameter passed to the script")

    #get all counties from the API and filter the one that matches the parameter
    counties = get_counties()
    county = list(filter(lambda c: c.get('name') == parameter, counties))

    #check if the county was found
    if not len(county):
        print("County not found. Please check the spelling and try again. Counties are case sensitive.")
        exit()


    county_id = county[0].get('id')
    print("{} county ID: {}".format(parameter,county_id))

    #request the data for the county from the API
    print("Getting data for {} county".format(parameter))
    county_data = get_county_data(county_id)
    for row in county_data:
        row.update({'service_names': implode_services(row.get('service_names'))})

    write_to_csv(county_data, parameter)
    print("Done")
```

Appendix C

SPARQL queries for OSM data – Nairobi and Wajir Counties

1. Health Facilities Nairobi

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct ?osm_id ?hasgeometry ?name WHERE {
{SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
 ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare ?healthcare .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "hospital" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "clinic" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "clinic" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "pharmacy" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "pharmacy" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "laboratory" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
```

```
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "laboratory" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "multipolygon", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "doctors" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "multipolygon", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "nursing_home" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "point", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "nursing_home" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "multipolygon", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "medical transport service" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "point", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "medical transport service" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "multipolygon", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "dentist" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "point", "i")
}}
union
{SELECT distinct ?osm_id ?hasgeometry WHERE {
osmrel:3492709 ogc:contains ?osm_id .
?osm_id geo:hasGeometry ?hasgeometry .
?osm_id osmkey:amenity "dentist" .
?osm_id osmkey:name ?name
filter regex (?hasgeometry, "multipolygon", "i")
}}
}
```

2. Health facilities Wajir

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT * WHERE {
{SELECT distinct ?osm_id ?hasgeometry ?name WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare ?healthcare .
      ?osm_id osmkey:name ?name .
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
 ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:healthcare ?healthcare .
  ?osm_id osmkey:name ?name
    filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:dispensing ?dispensing .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:dispensing ?dispensing .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "hospital" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "hospital" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "clinic" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "point", "i")
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "clinic" .
  ?osm_id osmkey:name ?name
  filter regex (?hasgeometry, "multipolygon", "i")
  }}
```

3. WATSAN points Nairobi

3.1 Water points

```
PREFIX osmway: <https://www.openstreetmap.org/way/>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX ogc: <http://www.opengis.net/rdf#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
 SELECT ?osm_id ?hasgeometry ?name WHERE {
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
 osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "drinking_water" .
  }}
  union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "water_point" .
  }}
   union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:drinking_water "yes" .
      }}
       union
  {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "community_centre;drinking_water" .
      }}
  }
```

3.2 Sanitation points

```
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "toilets" .
   }
```

3.3 Handwashing points

```
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct * WHERE {
{SELECT distinct ?osm_id ?hasgeometry ?handwashing WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:watsan:handwashing ?handwashing .
  filter REGEX(?handwashing, "sink|water_tank|leak_tin|leak_tins|yes|soap|tank|other")
   }}
   union
   {SELECT distinct ?osm_id ?hasgeometry ?handwashing WHERE {
  osmrel:3492709 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:water:hand_wash_facility "yes"
  }}}
```

4. Water points Wajir

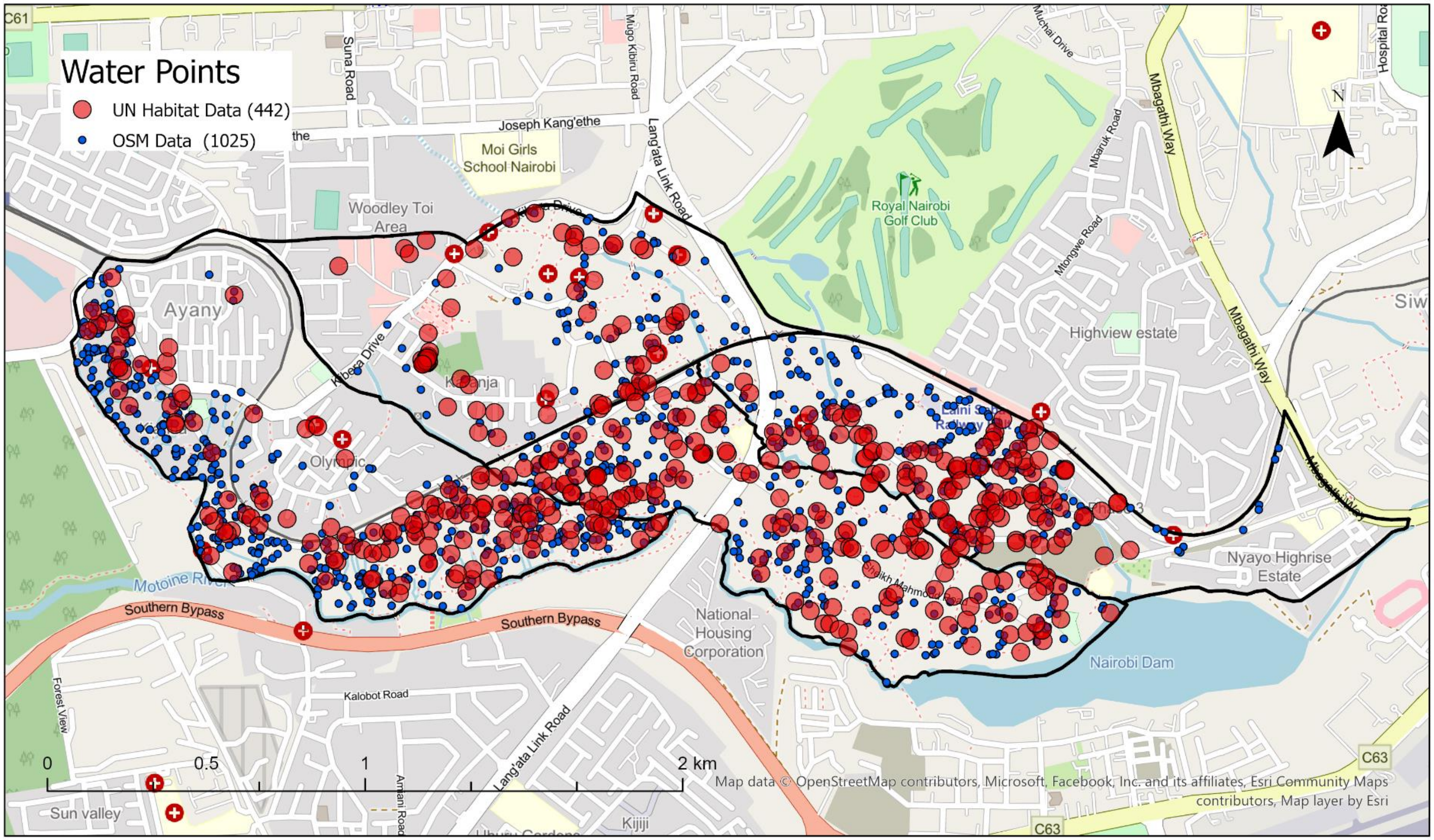
```
PREFIX osm: <https://www.openstreetmap.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX osmrel: <https://www.openstreetmap.org/relation/>
PREFIX ogc: <http://www.opengis.net/rdf#>
SELECT distinct * WHERE {
{SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:drinking_water "yes" .
   }} union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "drinking_water" .
   }}  union
   {SELECT distinct ?osm_id ?hasgeometry WHERE {
  osmrel:3495566 ogc:contains ?osm_id .
  ?osm_id geo:hasGeometry ?hasgeometry .
  ?osm_id osmkey:amenity "water_point" .
   }}}
```

Appendix D

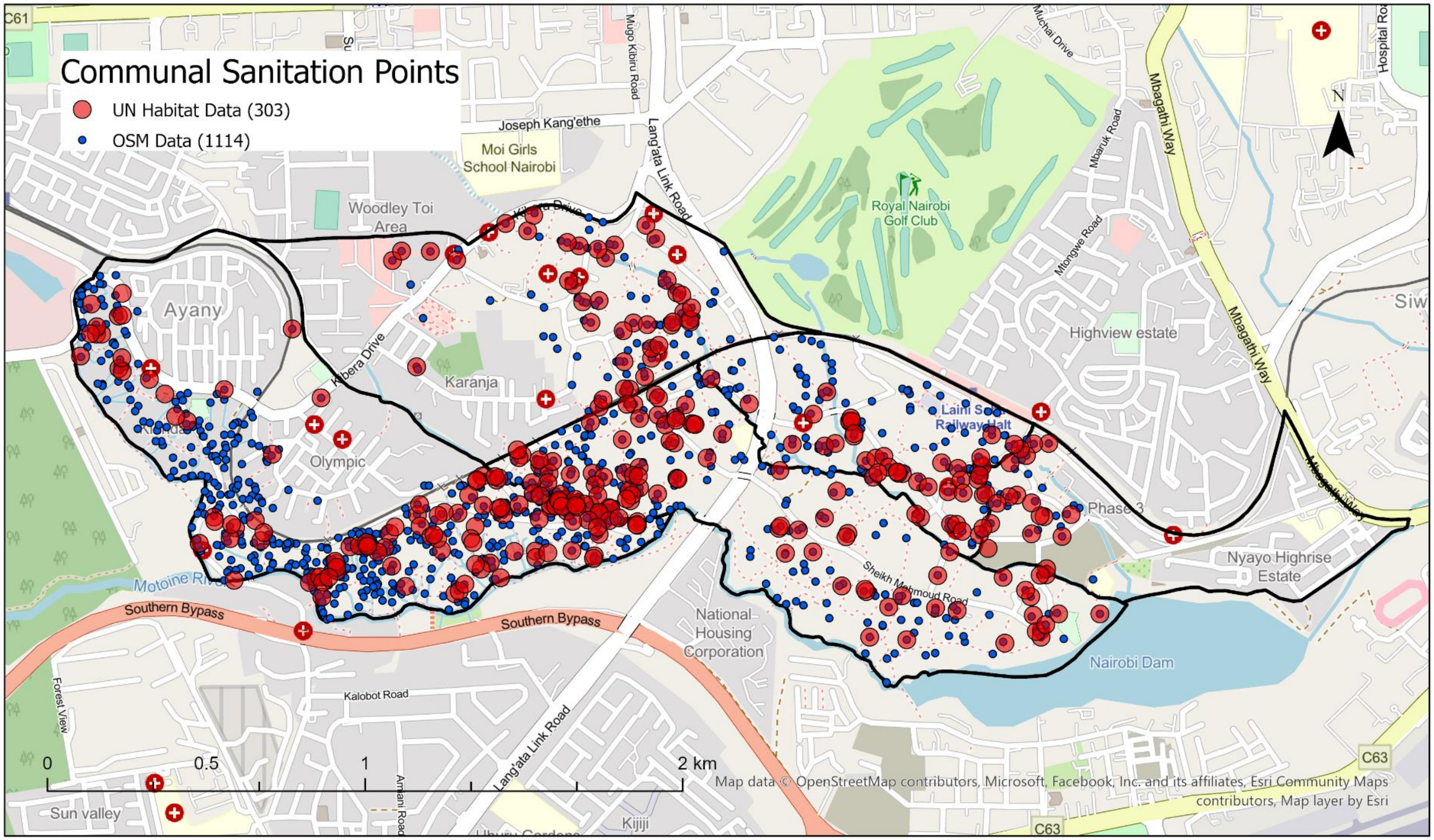
WATSAN Points in the informal settlements of Nairobi County

OSM data vs UN-Habitat data

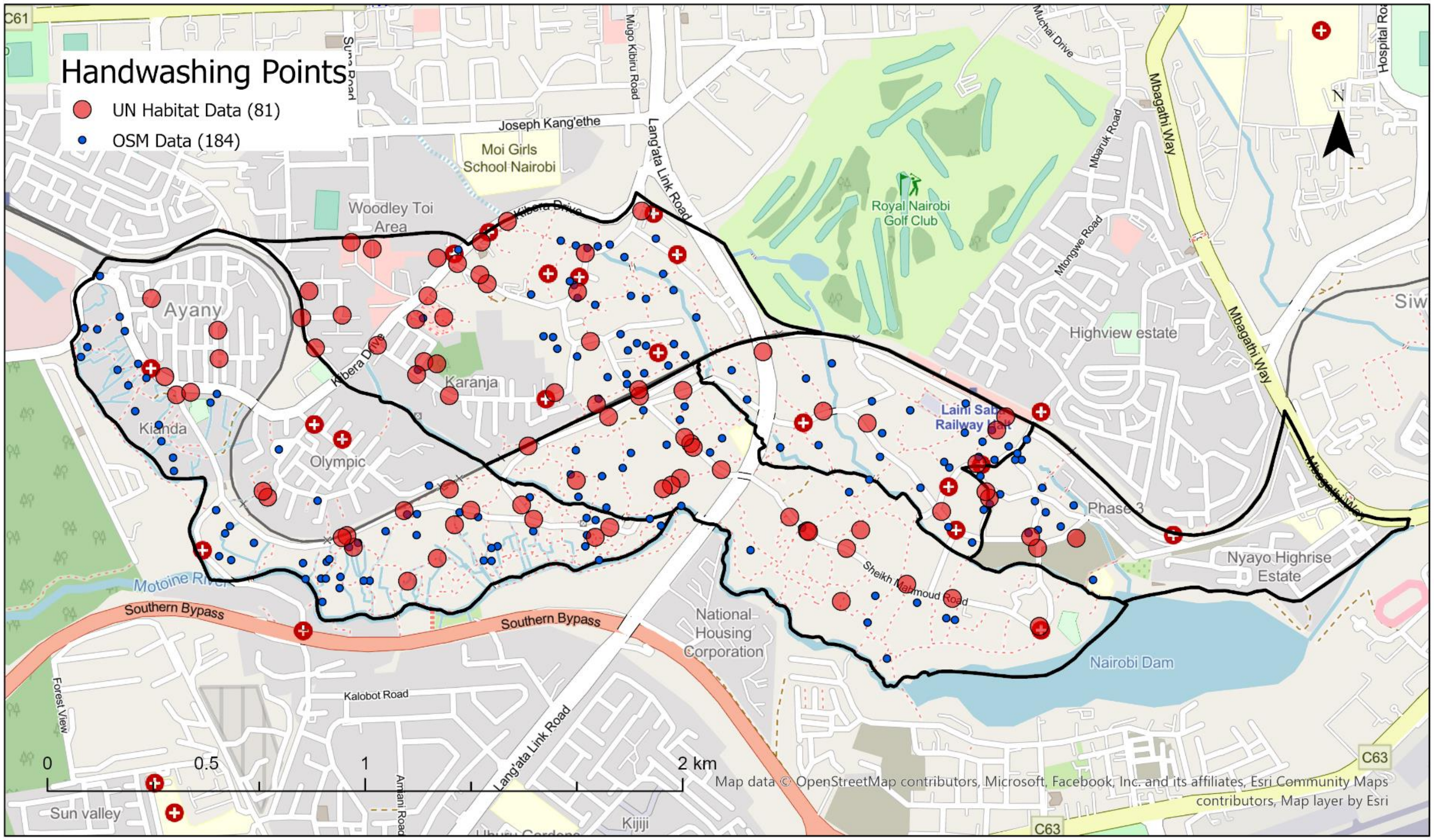
Kibera Informal Settlement - Water Points
OSM Data vs UN Habitat Data



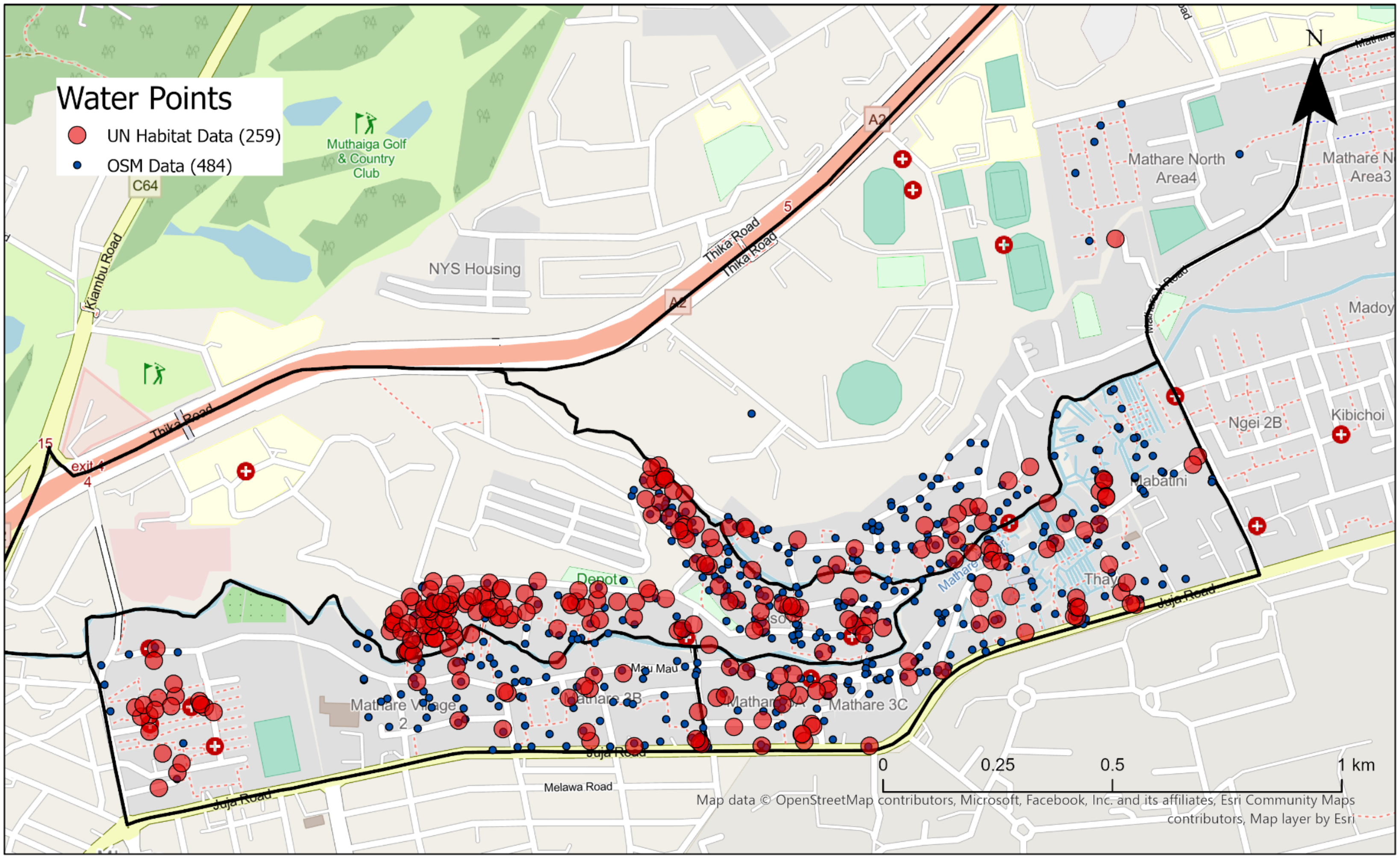
Kibera Informal Settlement - Communal Sanitation Points
OSM Data vs UN Habitat Data



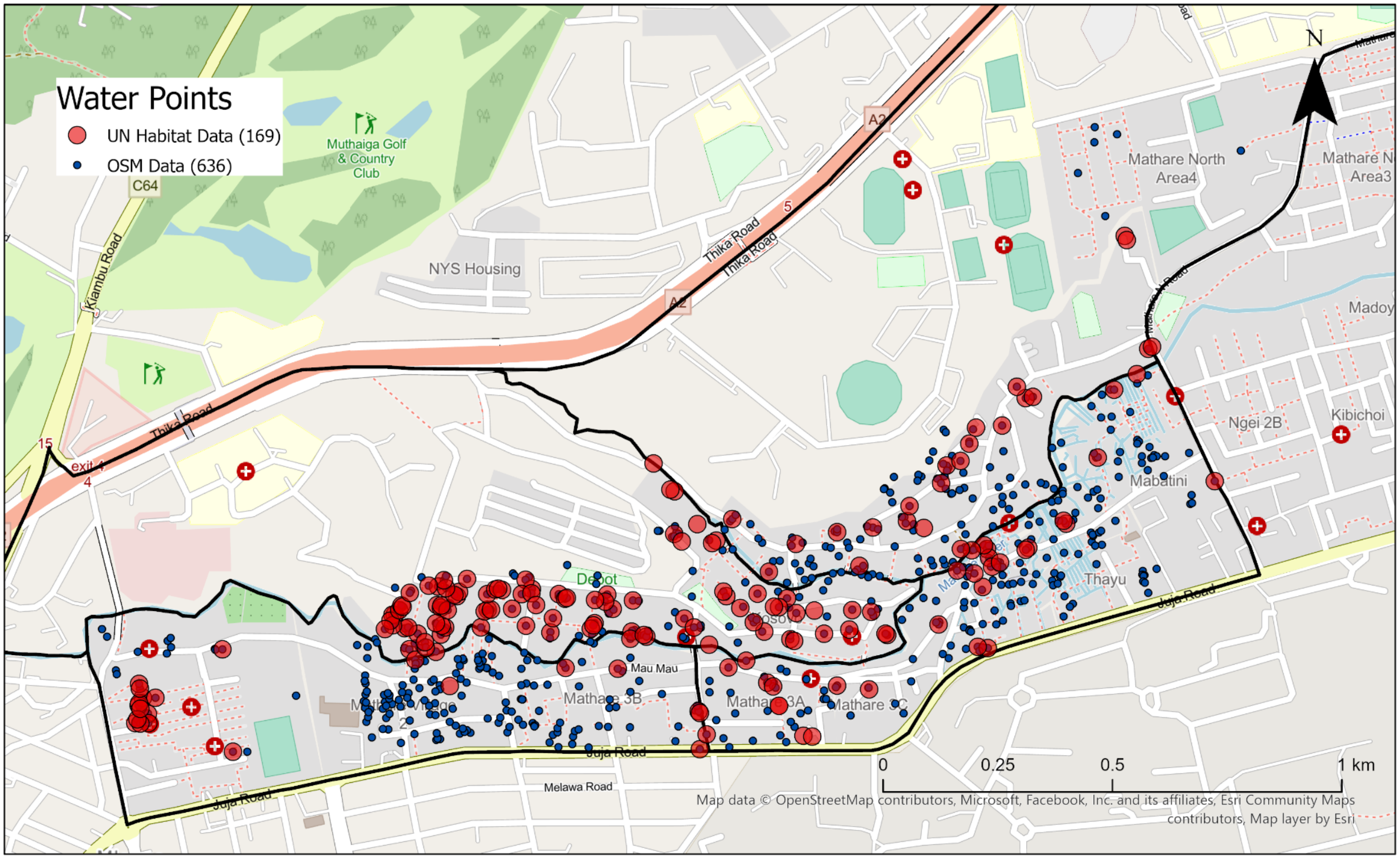
Kibera Informal Settlement - Handwashing Points
OSM Data vs UN Habitat Data



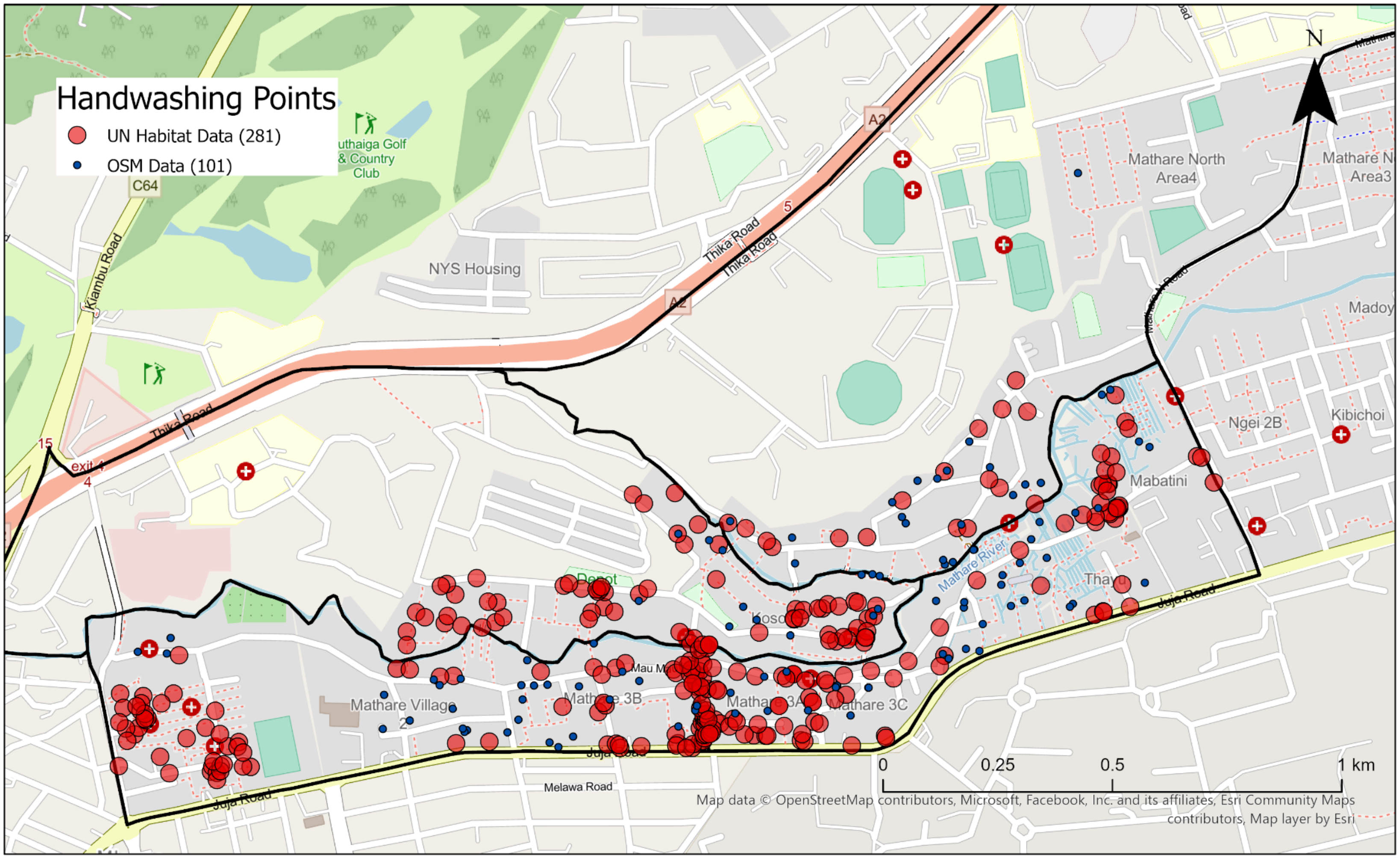
Mathare Informal Settlement - Water Points
UN Habitat Data vs OSM Data



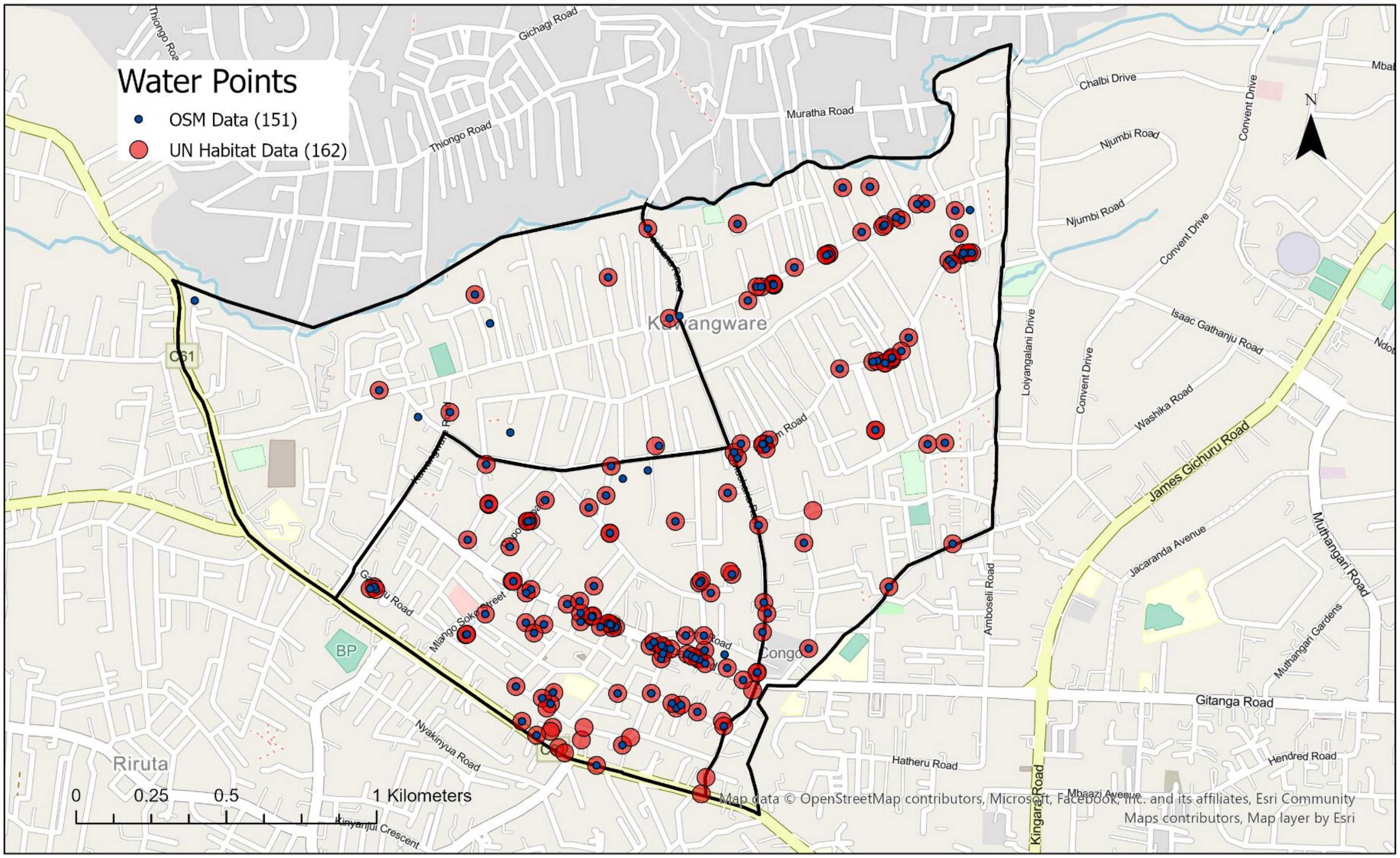
Mathare Informal Settlement - Communal Sanitation Points
UN Habitat Data vs OSM Data



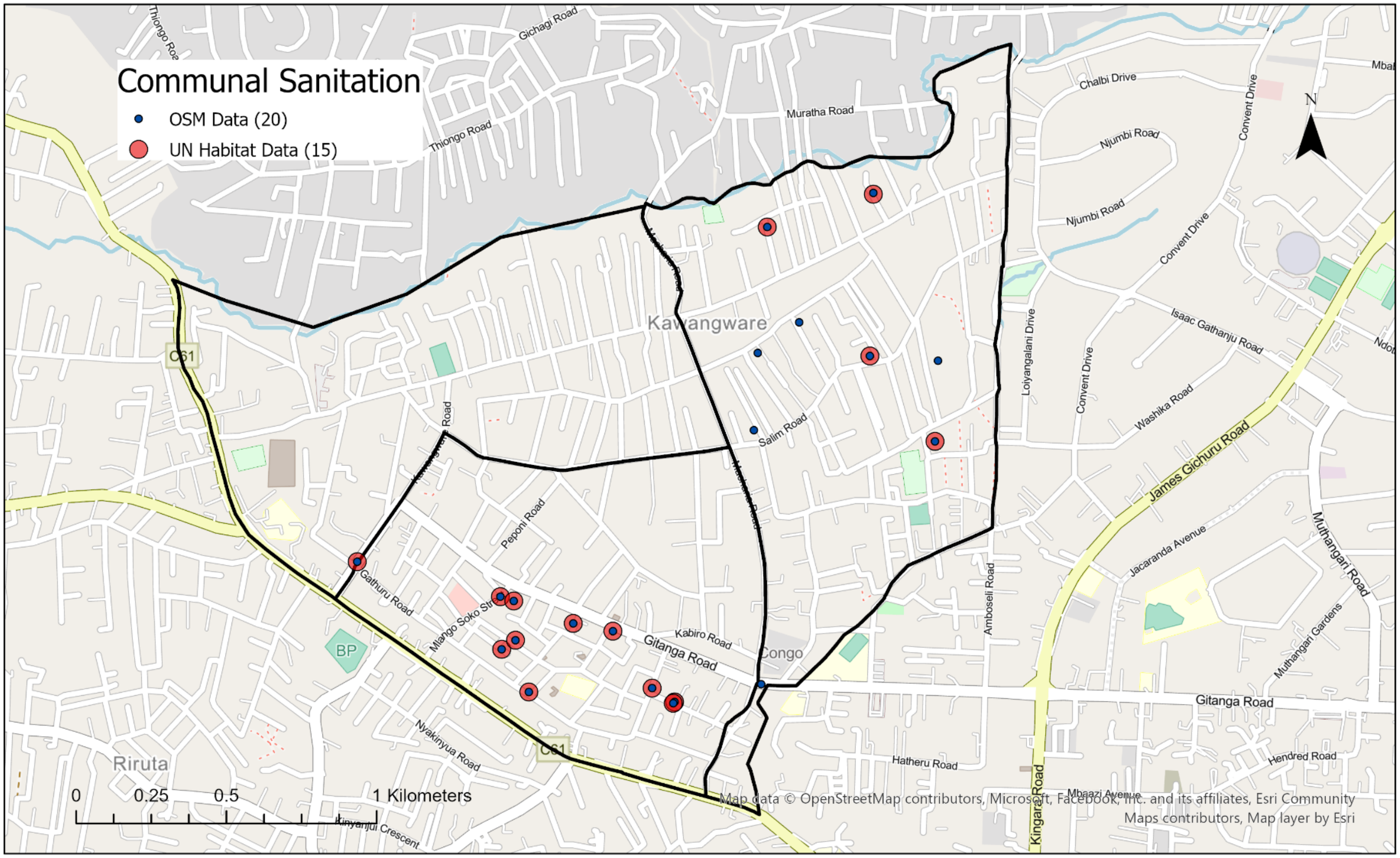
Mathare Informal Settlement - Handwashing Points
UN Habitat Data vs OSM Data



Kawangware Informal Settlement - Water Points
OSM Data vs UN Habitat Data



Kawangware Informal Settlement - Communal Sanitation Points
OSM Data vs UN Habitat Data



Kawangware Informal Settlement - Handwashing Points
OSM Data vs UN Habitat Data