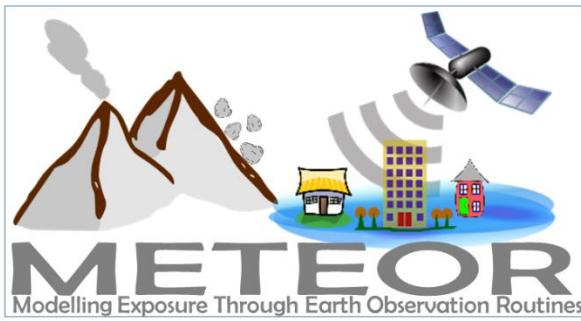


# METEOR: Ground Data Collection Using Protocols II

## Dar es Salaam, Tanzania

Report Number: 4.5/P

24 September 2020



### Contributors:

E. Kombe (HOT)

W. Evans (HOT)



Humanitarian  
OpenStreetMap  
Team



British  
Geological  
Survey



Oxford Policy  
Management



NSET  
Earthquake Safe Communities in Nepal



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## Document Verification

Project	METEOR: Modelling Exposure Through Earth Observation Routines	
Report Title	METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania	
Related Milestone	M4.5	
Reference as	E. Kombe & W. Evans (2020) METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania. Report 4.5/P	
Release Type	Public / Confidential / Confidential with Embargo Period	

Prepared by: Contributors		
Name(s):  E. Kombe  W. Evans	Signature(s):	Date(s):  24 September 2020
Approved by: Project Manager		
Name:  K. Smith	Signature:	Date:  28 September 2020
Approved by: UKSA IPP Project Officer		
Name:	Signature:	Date:

Date	Version	Alterations	Editor



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## Contents

<b>DOCUMENT VERIFICATION .....</b>	<b>I</b>
<b>ABBREVIATIONS .....</b>	<b>V</b>
<b>1. METEOR PROJECT INTRODUCTION .....</b>	<b>1</b>
1.1. PROJECT SUMMARY.....	1
1.2. PROJECT OVERVIEW .....	1
1.3. PROJECT OBJECTIVES .....	2
1.4. WORK PACKAGES.....	2
1.5. IN-SITU INPUTS AND VALIDATION.....	3
<b>2. OPENSTREETMAP .....</b>	<b>4</b>
2.1. WHAT IS IT? .....	4
2.2. OPEN DATA .....	4
2.3. DATA STRUCTURE .....	5
2.4. WHY USE IT?.....	6
<b>3. PARTICIPATORY MAPPING .....</b>	<b>7</b>
<b>4. GROUND DATA COLLECTION .....</b>	<b>8</b>
4.1. DATA MODEL.....	9
4.2. MAPPING TEAM .....	12
4.3. AREA OF INTEREST.....	16
4.4. SURVEY AUTHORISATION.....	19
4.5. TOOL PREPARATION .....	22
4.6. TEAM TRAINING.....	25
4.7. DATA MANAGEMENT.....	27
4.8. GROUND SURVEYING .....	28
4.9. DATA VALIDATION.....	31
<b>5. OPEN ACCESS .....</b>	<b>36</b>
5.1. DATA EXTRACTS.....	36
5.2. METHODOLOGY.....	39
<b>6. SUSTAINABILITY .....</b>	<b>42</b>
<b>7. CONCLUSION .....</b>	<b>43</b>
<b>APPENDIX 1: HOMOGENEOUS ZONE MAPS .....</b>	<b>45</b>
<b>APPENDIX 2: ATTRIBUTE VALUE BREAKDOWN .....</b>	<b>48</b>
<b>APPENDIX 3: SAMPLING PROCEDURE FOR BUILDINGS.....</b>	<b>53</b>
<b>APPENDIX 4: ONLINE INTERACTIVE MAP, DAR ES SALAAM METEOR BUILDINGS AND WARD BOUNDARIES ..</b>	<b>56</b>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## Figures

Figure 1: Tanzania Mapping Team with Civil Engineering Professor, at HOT/OMDTZ Office .....	13
Figure 2: Dar es Salaam Homogeneous Zones (link to high resolution version) .....	17
Figure 3: Dar es Salaam Sampling Points .....	19
Figure 4: Surveying Letter of Intention .....	21
Figure 5: OpenDataKit Spreadsheet.....	23
Figure 6: OpenDataKit Questionnaire (left), OpenMapKit (right).....	24
Figure 7: OpenMapKit Interface .....	25
Figure 8: Team Training at OpenMap Development Tanzania Office.....	26
Figure 9: OpenMapKit Server Table View .....	27
Figure 10: OpenMapKit Server Map View .....	28
Figure 11: OpenMapKit Diagram .....	29
Figure 12: JOSM Validation Tool .....	32
Figure 13: Building Photos .....	33
Figure 14: High Urban Homogeneous Zone on MapCampaigner.....	35
Figure 15: Attribute Information in MapCampaigner (Left) and OSM (Right).....	36
Figure 16: Describe Tab – Create Export Tool Function.....	37
Figure 17: Formats Tab – Create Export Tool Function .....	37
Figure 18: Data Tab – Create Export Tool Function .....	38
Figure 19: Summary Tab – Create Export Tool Function .....	38
Figure 20: Export Page – Completed Job .....	39
Figure 21: Tanzania METEOR Wiki Page .....	40
Figure 22: HOT Toolbox .....	41

## Tables

Table 1: METEOR Project Summary .....	1
Table 2: Overview of METEOR Work Packages.....	2
Table 3: Overview of HOT Deliverables .....	3
Table 4: Outline of OSM Ground Data Collection .....	8
Table 5: GED4ALL Simplified Building Taxonomy .....	10



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Table 6: Dar es Salaam Localised Building Taxonomy .....	11
Table 7: Digitisation Team .....	14
Table 8: Data Collection Team/Team on the Ground .....	15
Table 9: Dar es Salaam Homogeneous Zones .....	16
Table 10: Dar es Salaam Surveyed Buildings.....	19
Table 11: Before and After Survey Attribute Information .....	30
Table 12: Lateral Materials .....	30
Table 13: Lateral Systems .....	31
Table 14: Dar es Salaam MapCampaigner Projects .....	34
Table 15: Export Tool Data Extract (NOTE: export data older than 30 days is deleted).....	39
Table 16: Kathmandu METEOR Wiki Page Link.....	40
Table 17: HOT Toolbox Link .....	41



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## Abbreviations

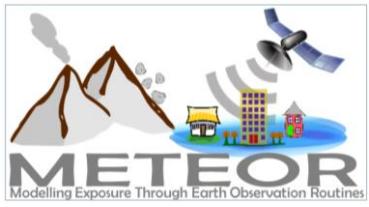
Acronym	Full Text and Description
<b>BGS</b>	British Geological Survey: An organisation providing expert advice in all areas of geoscience to the UK government and internationally
<b>COWI</b>	Danish international consulting group within engineering, economics and environmental science through which drone imagery was provided in 2016 to HOT
<b>DE</b>	Data Entry: The inputting of data or information into a computer, here it is often used to describe 'data entry specialists' who collect data in the field using mobile devices
<b>DMD</b>	Disaster Management Department: Prime Minister's Office of Tanzania focused on disaster risk
<b>DRM</b>	Disaster Risk Management: The application of disaster risk reduction policies and/or strategies
<b>DRR</b>	Disaster Risk Reduction: Disaster risk reduction is aimed at preventing new and reducing existing disaster risk and managing residual risk, all of which contribute to strengthening resilience and therefore to the achievement of sustainable development.
<b>EO</b>	Earth Observation: The gathering of information about Earth's physical, chemical and biological systems via remote sensing technologies, usually involving satellites carrying imaging devices
<b>FATHOM</b>	Provides innovative flood modelling and analytics, based on extensive flood risk research
<b>GCRF</b>	Global Challenges Research Fund
<b>GED4ALL</b>	Global Exposure Database for All: An open global database for hosting exposure information for multiple-hazard impact assessment
<b>GEM</b>	Global Earthquake Model: Non-profit organisation focused on the pursuit of earthquake resilience worldwide
<b>GFDRR</b>	Global Facility for Disaster Reduction and Recovery: A global partnership that helps



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



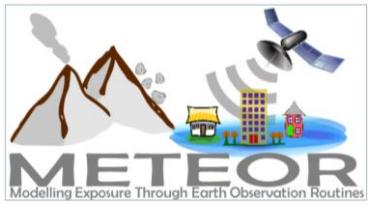
	developing countries better understand and reduce their vulnerability to natural hazards and climate change
<b>GIS</b>	Geographic Information System: A conceptualised framework that provides the ability to capture and analyse spatial and geographic data
<b>GNSS</b>	Global Navigation Satellite System: A satellite navigation system with global coverage; provides geolocation and time information to a receiver anywhere on or near the Earth where there is an unobstructed line of sight to four or more satellites
<b>GPS</b>	Global Positioning System: A satellite-based radio navigation system owned by the United States government and operated by the United States Space Force
<b>HOT</b>	Humanitarian OpenStreetMap Team: A global non-profit organisation that uses collaborative technology to create OSM maps for areas affected by disasters
<b>InaSAFE</b>	Free open software that produces realistic natural hazard impact scenarios for better planning, preparedness and response activities
<b>ImageCat</b>	International risk management innovation company supporting the global risk and catastrophe management needs of the insurance industry, governments and NGOs
<b>IPP</b>	International Partnership Programme: The UK Space Agency's International Partnership Programme (IPP) is a £30M per year programme, which uses expertise in space-based solutions, applications and capability to provide a sustainable economic or societal benefit to emerging nations and developing economies
<b>JOSM</b>	Java OpenStreetMap: Free Java software desktop editing tool for OSM data
<b>KLL</b>	Kathmandu Living Labs: A living lab and non-profit civic technology company based in Kathmandu, Nepal that primarily works on mobile technology and mapping
<b>M</b>	Milestone: Related to work package deliverables
<b>Mapathon</b>	Coordinated mapping event where individuals learn about OSM, and make edits to the map through remote digitising of satellite imagery
<b>METEOR</b>	Modelling Exposure Through Earth Observation Routines: A three-year project funded by the UK Space Agency to develop innovative application of Earth Observation (EO) technologies to improve understanding of exposure with a specific focus on the countries of Nepal and Tanzania



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



<b>NGO</b>	Non-Governmental Organisation: Organisations which are independent of government involvement
<b>NSET</b>	National Society for Earthquake Technology: Non-governmental organisation working on reducing earthquake risk in Nepal and abroad
<b>ODA</b>	Official Development Assistance: Government aid that promotes and specifically targets the economic development and welfare of developing countries
<b>ODbL</b>	Open Database License: A copyleft license agreement intended to allow users to freely share, modify, and use a database while maintaining this same freedom for others
<b>ODK</b>	OpenDataKit: Mobile application questionnaire form for collecting data
<b>OMDTZ</b>	OpenMap Development Tanzania: A registered NGO based in Dar es Salaam and operating across Tanzania. Since 2017, OMDTZ has promoted a number of community mapping projects
<b>OMK</b>	OpenMapKit: Mobile application that works with ODK to collect data in OSM
<b>OPM</b>	Oxford Policy Management: Organisation focused on sustainable project design and implementation for reducing social and economic disadvantage in low-income countries
<b>OSM</b>	OpenStreetMap: A collaborative project to create a free and open editable map database of the world
<b>OSMCha</b>	OpenStreetMap Changeset Analyzer: A web tool to help mappers analyse and review data changes to OSM. The objective of the tool is to help detect vandalism and act on bad changes to the map data
<b>OSMF</b>	OpenStreetMap Foundation: A not-for-profit organization that supports the OpenStreetMap Project
<b>Osmose</b>	OpenStreetMap Oversight Search Engine: One of many quality assurance tools available to detect issues in OpenStreetMap data. It detects a very wide range of issue types
<b>PDC</b>	Pacific Disaster Center: An applied research centre managed by the University of Hawaii, PDC is continuously developing new technologies and best practices to help our many global partners effectively mitigate, prepare for, respond to, and



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



	recover from disasters
<b>POSM</b>	Portable OSM: A project to bring together assorted OpenStreetMap tools onto a portable device which works without an internet connection
<b>QA</b>	Quality Assurance Specialist: People who ensure that the data is being collected correctly and to a high standard, often reviewing data after collection
<b>QGIS</b>	Free open source desktop geographic information system
<b>Ramani Huria</b>	A community mapping project based in Dar es Salaam, Tanzania; Swahili for “open map”
<b>SDGs</b>	Sustainable Development Goals
<b>TS</b>	Team Supervisor: Person responsible for managing and supervising teams, they possess high degrees of technical proficiency
<b>UAV</b>	Unmanned Aerial Vehicle: An aircraft without a human pilot on board, often simply referred to as ‘drones’
<b>UKSA</b>	United Kingdom Space Agency: An executive agency of the Government of the United Kingdom, responsible for the United Kingdom's civil space programme
<b>WASH</b>	Universal, affordable and sustainable access to WASH is a key public health issue within international development and is the focus of Sustainable Development Goal 6
<b>WP</b>	Work Package: Discrete sets of activities within the METEOR Project, each work package is led by a different partner and has specific objectives



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## 1. METEOR Project Introduction

### 1.1. Project Summary

Project Title	Modelling Exposure Through Earth Observation Routines (METEOR): EO-based Exposure, Nepal and Tanzania
Starting Date	08/02/2018
Duration	36 months
Partners	UK Partners: The British Geological Survey (BGS) (Lead), Oxford Policy Management Limited (OPM), SSBN Limited  International Partners: The Disaster Management Department, Office of the Prime Minister – Tanzania (DMD), The Global Earthquake Model (GEM) Foundation, The Humanitarian OpenStreetMap Team (HOT), ImageCat, National Society for Earthquake Technology (NSET) – Nepal
Target Countries	Nepal and Tanzania for “level 2” results and all 47 Least Developed ODA countries for “level 1” data
IPP Project	IPPC2_07_BGS_METEOR

Table 1: METEOR Project Summary

### 1.2. Project Overview

At present, there is a poor understanding of population exposure in some Official Development Assistance (ODA) countries, which causes major challenges when making Disaster Risk Management decisions. METEOR (Modelling Exposure Through Earth Observation Routines) takes a step-change in the application of Earth Observation exposure data by developing and delivering more accurate levels of population exposure to natural hazards.

METEOR is delivering calibrated exposure data for Nepal and Tanzania, plus “Level-1” exposure for the remaining Least developed Countries (LDCs) ODA countries. Moreover, we are: (i) developing and delivering national hazard footprints for Nepal and Tanzania; (ii) producing new vulnerability data for the impacts of hazards on exposure; and (iii) characterising how multi-hazards interact impact upon exposure. The provision of METEOR’s consistent data to governments, town planners and insurance providers will promote welfare and economic development and better enable them to respond to the hazards when they do occur.

METEOR is funded through the second iteration of the UK Space Agency’s (UKSA) International Partnership Programme (IPP), funded through the Global Challenges Research Fund (GCRF), which uses space expertise to deliver innovative solutions to real world problems across the globe. The funding helps to build sustainable development while building effective partnerships that can lead to growth opportunities for British companies.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



### 1.3. Project Objectives

METEOR aims to formulate an innovative methodology of creating exposure data through the use of EO-based imagery to identify development patterns throughout a country. Stratified sampling technique harnessing traditional land use interpretation methods modified to characterise building patterns can be combined with EO and in-field building characteristics to capture the distribution of building types. These protocols and standards will be developed for broad application to ODA countries and will be tested and validated for both Nepal and Tanzania to assure they are fit-for-purpose.

Detailed building data collected on the ground for the cities of Kathmandu (Nepal) and Dar es Salaam (Tanzania) will be used to compare and validate the EO generated exposure datasets. Objectives of the project look to: deliver exposure data for 47 of the least developed ODA countries, including Nepal and Tanzania; create hazard footprints for the specific countries; to develop critical exposure information from EO data; to create open protocols; and capacity-building of local decision makers to apply data and assess hazard exposure. The eight work packages (WP) that make up the METEOR project are outlined below in section 1.4.

### 1.4. Work Packages

Outlined below are the eight work packages that make up the METEOR project. These are led by various partners with a brief description of what each of the work packages cover. HOT is leading WP.4 Inputs and Validation, which focuses on the collection of exposure data in Kathmandu in Nepal and Dar es Salaam in Tanzania (Table 2). This data will assist with the validation and calibration of national exposure datasets created through the classification of building patterns from satellite imagery carried out by ImageCat in WP.3.

Table 2: Overview of METEOR Work Packages

Work Package	Title	Lead	Overview
WP.1	Project Management	BGS	Project management, meetings with UKSA, quarterly reporting and the provision of feedback on project deliverables and direction across primary stakeholders.
WP.2	Monitoring and Evaluation	OPM	Monitoring and evaluation of the project and its impact, using a theory of change approach to assess whether the associated activities are leading to the desired outcome.
WP.3	EO Data for Exposure Development	ImageCat	EO-based data for exposure development, methods and protocols of segmenting/classifying building patterns for stratified sampling of building characteristics.
WP.4	Inputs and Validation	HOT	Collect exposure data in Kathmandu and Dar es Salaam to help validate and calibrate the data derived from the classification of building patterns from EO-based imagery.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



WP.5	Vulnerability and Uncertainty	GEM	Investigate how assumptions, limitations, scale and accuracy of exposure data, as well as decisions in data development process lead to modelled uncertainty.
WP.6	Multiple Hazard Impact	BGS	Multiple hazard impacts on exposure and how they may be addressed in disaster risk management by a range of stakeholders.
WP.7	Knowledge Sharing	GEM	Disseminate to the wider space and development sectors through dedicated web-portals and use of the Challenge Fund open databases.
WP.8	Sustainability and Capacity-Building	ImageCat	Sustainability and capacity-building, with the launch of the databases for Nepal and Tanzania while working with in-country experts.

### 1.5. In-Situ Inputs and Validation

The project WP led by HOT is broken down into six deliverables, which are focused on the mapping of exposure data for the cities of Kathmandu and Dar es Salaam in OpenStreetMap (OSM). These involve importing existing data into OSM, the remote mapping of building footprints and road networks, drafting protocols for crowdsourcing exposure data, collecting detailed attribute information on the ground and producing a final report (Table 3).

Table 3: Overview of HOT Deliverables

Deliverable	Title	Overview
M4.1	Import Existing Data into OSM	Review and assess the suitability of existing open datasets for import into OSM for Kathmandu and Dar es Salaam.
M4.2	EO Mapping of Exposure Data	Remote mapping of building footprints and road networks in OSM for Kathmandu and Dar es Salaam.
M4.3	Protocols for Crowdsourcing Exposure Data	Draft protocols for the crowdsourcing of exposure data in OSM, covering data imports, remote mapping and ground data collection.
M4.4	Ground Data Collection using Protocols I	Collect exposure data on the ground for Kathmandu using a data model developed in line with the requirements for WP.3



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



M4.5	Ground Data Collection using Protocols II	Collect exposure data on the ground for Dar es Salaam using a data model developed in line with the requirements for WP.3
M4.6	Final Report	Deliver a final version of M4.3 along with the resulting data from Deliverables M4.4 and M4.5.

## 2. OpenStreetMap

### 2.1. What Is It?

OSM is a crowdsourced geospatial database of the world built largely by volunteers and professionals digitising aerial imagery, collecting attribute information on the ground and liberating existing public sources of geospatial data. Known as the ‘Wikipedia’ of maps, the data is freely accessible to all under the Open Database License (ODbL)<sup>1</sup>, meaning that it can be queried, used, manipulated, contributed to and redistributed in any form. OSM is the ideal database for humanitarian efforts and disaster management, as it is a great source of geographic baseline data for many cities around the globe, especially in countries with emerging economies that are not always on the map.

### 2.2. Open Data

The ODbL license is a copyleft ("share alike") agreement intended to allow users to freely share, modify, and use a database while maintaining this same freedom for others. Published by Open Data Commons, part of the Open Knowledge Foundation<sup>2</sup>, the ODbL license enables OSM to be a source of powerful geospatial data to make change, particularly in ODA countries where a potential lack of internal funds and skill sets make it difficult to create up-to-date exposure data themselves. Such data is necessary to conduct hazard impact risk assessments and carry out informed appropriate disaster management decisions. One project may focus on creating data in OSM for a specific purpose, but this data can then be used by anyone for their area of interest. This freedom of use is outlined on the ODbL website<sup>3</sup>.

<sup>1</sup> <https://opendatacommons.org/licenses/odbl/>

<sup>2</sup> <https://okfn.org/>

<sup>3</sup> <https://opendatacommons.org/licenses/odbl/summary/index.html>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



**You are free:**

- ❖ To Share: To copy, distribute and use the database
- ❖ To Create: To produce works from the database
- ❖ To Adapt: To modify, transform and build upon the database

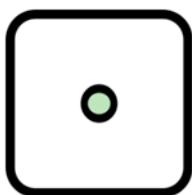
**As long as you:**

- ❖ Attribute: You must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database
- ❖ Share-Alike: If you publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL
- ❖ Keep open: If you redistribute the database, or an adapted version of it, then you may use technological measures that restrict the work (such as DRM) as long as you also redistribute a version without such measures

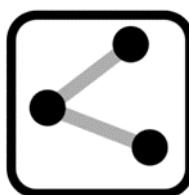
### 2.3. Data Structure

Geospatial data is stored in OSM as vectors, with three types of elements:

1. **Nodes**, which represent a point on the surface of the earth
2. **Ways**, which are sets of ordered nodes that can form lines or polygons
3. **Relations**, which are sets of nodes, ways and/or relations as members that are used to define logical or geographic relationships between other elements



Nodes



Ways



Relations

Each of these elements can have any number of key=value tags, that represents the attribute information for a given feature. For example, a post office may be represented with the tags building=yes and amenity=post\_office, to help identify the purpose of the building.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



#### 2.4. Why use it?

The most common question people ask is: Why would you use OpenStreetMap if there is Google Maps?<sup>4</sup> These platforms have many similarities and address the same basic human of spatially knowing where things are. In short, OSM represents an open approach to how data is collected and distributed, which makes it fast, free, flexible and widespread; an ideal combination for development application among others.

In the case of Haiti, the OSM community needed just a few hours to remotely digitise earthquake affected areas from satellite imagery, in contrast to commercial maps that had no way of responding in such a short period. The thousands of active volunteers around the globe are what made this possible, and the fact that every update is immediately visible to all other users, and is also version controlled. In many cases, the OSM community has been able to achieve even higher detail than any other map source.

The very idea behind OSM was to solve the problem of data restriction by using a Wikipedia model. Each edit made in OSM is owned by the community, whereas each change made in commercial map is owned by the providers. Data in commercial maps is copyrighted and so it can be subject to licensing fees and contractual restrictions. OSM is, and always will be, available for free. This open access to geospatial data makes it easy for researchers, governments, disaster management agencies and policy implementers to work collectively from the same database.

OSM is also ready for any kind of operation. Many major organisations are choosing OSM for their maps because it allows for customisation based on need. It has been used to collect functioning and non-functioning water facilities in northern Uganda, to mapping hundreds of thousands of shelters in Bangladesh, to planning logistics for a Yellow Fever campaign in the Democratic Republic of Congo and mapping the exposure information of infrastructure for hazard impact. It allows access to all of the map offline and can be formatted in local languages, transcending national boundaries.

Most importantly, OSM emphasises local knowledge across the globe. The barrier to entry is low, and there are many ways to contribute that do not require access to the latest technology. The result is a map made by local experts. OSM provides global map data in a unified tagging schema, although there are some local variations, this allows for consistency from country to country, for a truly global map.

---

<sup>4</sup> <https://welcome.openstreetmap.org/>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



### 3. Participatory Mapping

Participatory mapping is a phrase commonly used in the OSM community, which refers to the creation of maps by local communities. The maps are visual representations of what a community perceives as its environment and the significant features within it. Participatory mapping is based on the premise that local communities possess expert knowledge of their surroundings.

Often this process is supported by organisations, with a target goal and funds to assist with the research and direction, depending on the purpose of the data creation. The emphasis, however, is always on providing skills for the community to create the maps themselves in order to represent the spatial knowledge of local citizens. Participatory mapping, which is interchangeable with the term community mapping, is at the core of HOT as we believe that local people are the experts needed to create the necessary data in OSM. Frequently we utilise crowd-sourcing, by obtaining data contributions from large numbers of people, whether remote mappers digitising buildings and roads, or field mappers adding in local knowledge.

Participatory mapping relates to fostering a collaborative approach. The primary ingredient in the success of an open mapping project is the cultivation and care of a strong network of partners across the range of communities and institutions that have a stake in the project. These partnerships are critical for the implementation, uptake, and sustainability of open mapping projects. They can lead to shared investment of resources and technical expertise, innovative use cases for open data, increased legitimacy for the project, and stability over time. Because the open mapping ecosystem is a global phenomenon, it is likely that at the start of any project, there will be a pre-existing group of individuals and organisations already mapping in the target area. With proper outreach and engagement, these groups can be a powerful resource. Including them early in the planning and design phases of a project will help increase their sense of ownership and improve the likelihood of successful collaboration.

Government agencies responsible for areas as diverse as national statistics, urban planning, transportation, public health, environment and natural resource management, and disaster response have all benefited from working with open mapping and can bring valuable resources, expertise, and authority to a mapping project. The OSM community is active globally and will, in many cases have a local presence with experience mapping in the area, technical knowledge about mapping, and a group of passionate volunteers who may be willing to participate.

This is the case for crowd-sourcing the regional exposure data in the cities of Kathmandu and Dar es Salaam for the METEOR project. Kathmandu Living Labs (KLL) is a local company who has been involved with the OSM community for years, having worked on numerous projects with the Nepalese government. In Dar es Salaam, OpenMap Development Tanzania (OMDTZ) is a local organisation and partner of Humanitarian OpenStreetMap Team. OMDTZ has helped conduct the Ramani Huria Project, a community-based mapping initiative that began in Dar es Salaam training university students and local community members to create OSM maps of the most flood-prone areas of the city, in an effort to help combat the impact.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



#### 4. Ground Data Collection

Data collection in the city of Dar es Salaam and the neighbouring region of Pwani was conducted by HOT and OMDTZ following the outlined protocols in *Deliverable M4.3 Protocols for Crowd-Sourcing Regional Exposure Data*. The protocols focus on the addition of detailed attribute information of buildings observed from the ground to building footprints that were remotely digitised in OSM using aerial imagery. The methodology applied by HOT for the ground data collection is briefly outlined below and expanded through the following sections.

No.	Step	Description	Date
1	Digitisation	Remote mapping of the selected regions	Oct - Dec 2018
2	Data Model	Development of localised data model for Dar es Salaam	January 2019
3	Mapping Team	Recruitment of a local mapping team	January 2019
4	Area of Interest	Generation of randomised sampling points per homogeneous zone	January 2019
5	Survey Authorisation	Authorisation from district officials to survey areas of interest	January 2019
6	Tool Preparation	Preparation of data collection tools	January 2019
7	Team Training	Training with local structural engineers and use of data collection tools	February 2019
8	Data Management	Data management setup through dedicated online server	February 2019
9	Ground Survey	Collection of identified building attributes	February and March 2019
10	Data Validation and Quality	Data validation, spot checks and quality assurance	April 2019
11	Upload Data	Upload final data to OSM	April 2019

Table 4: Outline of OSM Ground Data Collection

The first step of any ground data collection is to create a data model, which was localised for the context of Dar es Salaam with the assistance of local structural engineers from Ardhi University who provided insight on types of commonly found buildings within the city. Once the data model was finalised, the mapping team was put together and the data collection tools prepared along with the creation of customised questionnaire forms following the localised data model. During this time, the team leaders determined the area of interest for collecting data through the generation of random sampling points within each of the homogeneous development zones, as well as ensuring access with the appropriate authorities.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Training for the team on how to use the surveying tools to capture the data on the ground was then carried out, along with the establishment of a data management system where the completed questionnaires could be organised, hosted and backed through an online server. The data collection commenced on the ground once everything was in order, which was then followed by a series of data validation rounds, as well as random spot checks by team leaders. A final round of data validation and assessment of the data quality assurance was conducted before uploading the information to OSM.

In total, the various steps involved with the ground data collection in Dar es Salaam took around 6 months, with the development of the data model starting in October 2018, and the ground surveying taking place from February to April 2019. The final phase of the project, which focuses on data validation, spot checking, and uploading the final data to OSM took place from March to April 2019.

#### 4.1. Data Model

The data model used for the collection of exposure information in Dar es Salaam was adopted from the model used by KLL in Kathmandu. This was derived from the Global Exposure Database for All (GED4ALL) data model, which was initially conceived through the second round of the GFDRR Challenge Fund.<sup>5</sup> The Challenge Fund aims to decrease disaster risk management costs and increase resilience by developing frameworks that facilitates a multi-hazard view of risk, taking into account the effects of volcanoes, floods, tsunamis, storms and cyclones on the structural integrity of buildings. HOT, in collaboration with the Global Earthquake Model (GEM) and ImageCat, assisted with the integration of OSM data into GED4ALL at the building level, by aligning the appropriate data tags in OSM to the identified taxonomy attributes<sup>6</sup>.

Two versions of the GED4ALL data model were created, a core data model containing an exhaustive list of structural building features that can be impacted by the hazards listed above and a simplified version. The simplified version is based on the most common, relevant and easy to define features of buildings, which were selected to help reduce the comprehensive multi-hazard risk assessment taxonomy for less experienced surveyors. This is the version that was adapted and used for the METEOR project. Data collection efforts are much more effective when the attributes required are kept to a minimum, where possible. This allows the surveyors to focus on the quality rather than the quantity of information. Table 5 below outlines the simplified version of the GED4ALL building taxonomy OSM data model.

---

<sup>5</sup> <https://www.gfdrr.org/en/challenge-fund>

<sup>6</sup> <https://wiki.openstreetmap.org/wiki/GED4ALL>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Table 5: GED4ALL Simplified Building Taxonomy

SIMPLIFIED GED4ALL BUILDING TAXONOMY			
No.	GED4ALL	OSM Key	OSM Description and Values
1	Material of the Lateral Load Resisting System	<a href="#">building:lateral:material=*</a>	Proposed lateral load resisting material tag (concrete_reinforced, concrete, concrete_steel, metal, masonry_reinforced, masonry, masonry_confined, earth, earth_reinforced, wood, other, unknown)
2	Lateral Load Resisting System	<a href="#">building:lateral:system=*</a>	Identify structural system of buildings (moment_frame, infilled_frame, braced_frame, post_beam, wall, dual_framewall, flat_slab, waffle_slab ,infill_flatslab, infill_waffleslab, hybrid, unknown)
3	Height	<a href="#">building:levels=*</a>	Number of above-ground levels of a building
		<a href="#">building:levels:underground=*</a>	Proposed tag for number of below-ground levels of a building
		<a href="#">building:slope=*</a>	Proposed tag for describing the slope of the building in degrees
4	Date of Construction or Retrofit	<a href="#">building:age=*</a>	Proposed age tag, associated with buildings (pre_2000, post_2000, unknown)
		<a href="#">building:condition=*</a>	Describe the condition of the building (good, average, poor, unknown)
5	Occupancy	<a href="#">building=*</a>	Describe the building purpose (residential, commercial, public, mixed_use, industrial, agriculture, assembly, government, educational, unknown)
		<a href="#">capacity:persons=*</a>	Describe the number of people a building can support
6	Ground Floor Hydrodynamics	<a href="#">ground_floor:openings=*</a>	Proposed openings tag, associated with building ground floors (yes, no, unknown)
		<a href="#">ground_floor:height=*</a>	Proposed height tag in meters, associated with building ground floors
7	Roof	<a href="#">roof:shape=*</a>	Well known roof shapes (flat, pitched, monopitch, sawtooth, curved, complex_regular, complex_irregular, unknown)
		<a href="#">roof:material=*</a>	Outer material for the building roof (masonry, earth, concrete, metal, wood, fabric, slate, stone, clay, unknown)



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



As the GED4ALL building taxonomy is designed to be used around the globe, the OSM values in the simplified table are merely guidelines to commonly used entries for the identified attributes. These values were localised for Dar es Salaam, with the support of a civil engineer expert from Ardh University. His input on structural aspects, such as commonly used building materials and lateral load resisting systems within Dar es Salaam were incorporated into the METEOR data model, while those generally not found within the area were removed. After discussion with the engineer we agreed to add some other proposed descriptions in each attribute, for example, we have added many local values in buildings:lateral:material, building:lateral:system, roof:shape and roof:materials. The final localised data model used for the ground data collection in Dar es Salaam can be seen in Table 6 below.

Table 6: Dar es Salaam Localised Building Taxonomy

KATHMANDU BUILDING TAXONOMY			
No.	Attribute	OSM Key	OSM Description and Values
1	Material of the Lateral Load Resisting System	building:lateral:material=*	Proposed lateral load resisting material tag (steel, steel-concrete_composite, block, brick, stone, corrugated_galvanized_iron, earth, wood, bamboo, light_gauge_steel or cold_formed_steel, iron_sheet, industrial_troughed_sheet)
2	Lateral Load Resisting System	building:lateral:system=*	Identify structural system of buildings (moment_resisting_frame, shear_wall, braced_frame, framed_concrete_structure, Dual system (frame wall), masonry_wall, confined_masonry, hybrid)
3	Height	building:levels=*	Number of above-ground levels of a building
		building:levels:underground=*	Number of under-ground levels of a building
4	Date of Construction or Retrofit	building:age=*	Proposed age tag, associated with buildings (pre_2000, post_2000, unknown)
		building:condition=*	Describe the condition of the building (good, average, poor)
5	Occupancy	building=*	Describe the building purpose (residential, commercial, public, mixed_use, industrial, agriculture, assembly, government, educational, health, construction)
		capacity:persons=*	Describe the number of people a building can support
6	Roof	roof:shape=*	Well known roof shapes (flat, pitched, monopitch,



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



			double_pitch, sawtooth, curved, complex_regular, complex_irregular, domed, unknown)
		roof:material=*	Outer material for the building roof (clay tiles, earth, concrete, metal, wood, fabric, slate, stone, composite_tiles, cgi, bamboo, thatch, mixed, unknown)
7	Neighbouring Condition	building:neighbour=*	Describe the neighbouring condition of the building (attached, free_standing)
8	Geological Site Condition	building:geological_site=*	Describe the geological site the building is built upon (slopy_land, flood-prone_area, river_bank)

#### 4.2. Mapping Team

Once the data model was finalised, a team to conduct the data collection on the ground was put together. The structure of the team greatly influences the management, efficiency, quality and overall success of surveying on the ground, especially when working with large amounts of information. A tiered team structure ensures that the data collection can be carried out systematically, so that the line of responsibility and guidance is clear. As data collection is crowdsourced through OSM and focuses on the structural engineering of buildings, the surveying team for Dar es Salaam consisted of HOT and OMDTZ mapping supervisors, field mappers, and some structural engineering students to assist with spot checking and validation. This allowed for a range of specialised expertise combined with the local knowledge of surveyors and residents from Dar es Salaam.

Recruitment of the team began in October 2018, and focused mainly on assembling a team of surveyors who had already engaged with various HOT Tanzania Projects and had a basic understanding of data collection tools and processes. Therefore, we started digitising buildings footprints by using the digitisation team that had been created and has a wide experience in digitisation processes. Next, we constructed a field team by February 2019, this team is different from those who used in digitisation. Most of the people in this team consisted of recent graduates from Ardi University<sup>7</sup> and the University of Dar es Salaam<sup>8</sup>.

An added advantage of this recruitment was that surveyors had been involved in many of the activities within HOT Tanzania projects such as Ramani Huria, and also had a background of the planning activities and how buildings are classified in OSM as well as having great knowledge of the Dar es Salaam areas and the ward locations. After making sure everyone was available in Dar es Salaam we conducted training both in the classroom and with practical exercises the same month followed with field mapping by March 2019.

---

<sup>7</sup> <http://www.aru.ac.tz/>

<sup>8</sup> <https://www.udsm.ac.tz/>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Figure 1: Tanzania Mapping Team with Civil Engineering Professor, at HOT/OMDTZ Office

The team project manager was in charge of planning the whole mapping approach, from developing the workflow, setting up the tools, hiring the team, conducting the training and strategically dividing up the city, helping in reporting as well as overseeing all the data collected and validated. Once the work was delegated down the line, the quality assurance specialists could focus on ensuring that the data was being collected correctly and to a high standard, while providing feedback both down the line to the data entry specialists, and to the project manager.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



No.	Name	Team Position	Connected Institute
1.	Dorica Mugusi	Team Lead	HOT/OMDTZ
2.	Gwantwa Andrew	Quantity and Assurance	OpenMap Development Tanzania
3.	Dorice Lucas	Quantity and Assurance	OpenMap Development Tanzania
4.	Jamal Amani	Editor	OpenMap Development Tanzania
5.	Mariamu Crispoo	Editor	OpenMap Development Tanzania
6.	Nuru Athumani	Editor	OpenMap Development Tanzania
7.	Samwel Kyando	Editor	OpenMap Development Tanzania
8.	Frank Charles	Editor	OpenMap Development Tanzania
9.	Abubakari R Kachongo	Editor	OpenMap Development Tanzania
10.	Christina Reuben	Editor	OpenMap Development Tanzania
11.	Edger Mahali	Editor	OpenMap Development Tanzania
12.	Aron Martin	Quantity and Assurance	OpenMap Development Tanzania
13.	Erasto Eliud	Editor	OpenMap Development Tanzania
14.	Asibwene Kalulunga	Editor	OpenMap Development Tanzania
15.	Godfrey Faustine	Editor	OpenMap Development Tanzania
16.	Janeth Mlinga	Editor	OpenMap Development Tanzania
17.	Neema Kombe	Editor	OpenMap Development Tanzania
18.	Damon Monyo	Editor	OpenMap Development Tanzania
19.	Furaha Mrema	Editor	OpenMap Development Tanzania
20.	Godfrey Hussein	Editor	OpenMap Development Tanzania
21.	Neema Alphonse	Quantity and Assurance	OpenMap Development Tanzania

Table 7: Digitisation Team



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Table 8: Data Collection Team/Team on the Ground

No.	Name	Team Position	Connected Institute
1.	William Evans	Global Project Manager	Humanitarian OpenStreetMap Team
2.	Emanuel Kombe	Project Manager	HOT/ OMDTZ
3.	David Mwasaka	Quantity & Assurance	OpenMap Development Tanzania
4.	Neema Elias	Quantity & Assurance	OpenMap Development Tanzania
5.	Zaituni Rashid	Quantity & Assurance	OpenMap Development Tanzania
6.	Paschal Mapera	Quantity & Assurance	OpenMap Development Tanzania
7.	Evarist Isdory	Surveyor	OpenMap Development Tanzania
8.	Glory Emanuel	Surveyor	OpenMap Development Tanzania
9.	Gloria Elevan	Surveyor	OpenMap Development Tanzania
10.	Asia Abdallah	Surveyor	OpenMap Development Tanzania
11.	Jackson Hombo	Surveyor	OpenMap Development Tanzania
12.	Gisella John	Surveyor	OpenMap Development Tanzania
13.	Debora Chamaku	Surveyor	OpenMap Development Tanzania
14.	Rashid Iddi	Surveyor	OpenMap Development Tanzania
15.	Joan Kamwela	Surveyor	OpenMap Development Tanzania
16.	William Mponeja	Surveyor	OpenMap Development Tanzania
17.	Hawa Magongo	Surveyor	OpenMap Development Tanzania
18.	Felister Rwehumbiza	Surveyor	OpenMap Development Tanzania
19.	Jeremiah Baptister	Surveyor	OpenMap Development Tanzania
20.	Josephina Tinuga	Surveyor	OpenMap Development Tanzania
21.	Hashiru Mutaremwa	Surveyor	OpenMap Development Tanzania
22.	Ibrahim Alphonse	Surveyor	OpenMap Development Tanzania
23.	Grace Makala	Validator	Ardhi University
24.	Halima Rajun Kisavul	Validator	Ardhi University



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



25.	Joseph Asantemungu	Validator	Ardhi University
26.	Josephat Mhanga	Validator	Ardhi University
26.	Magret Naimani	Validator	Ardhi University

#### 4.3. Area of Interest

Dividing up and delegating target areas to the mapping team also plays a large role in successfully collecting data on the ground. How the area is split up is dependent on the type of mapping initiative, whether it is looking at the city or town as a whole, or whether it is isolated sections. For the METEOR project, Dar es Salaam was divided into homogeneous development type zones making the area of interest targeted portions of the city.

These areas were classified into homogenous zones by ImageCat, based on structural development patterns identified remotely. There were originally 7 zone types identified in Dar es Salaam, which include High Urban Development, Industrial Development, Rural Settlement, Single Family Residential, Informal Zones, Urban Development, and Natural Forest Land. The Natural Forest Land development pattern was eventually discarded for the purposes of this project, due to the fact that there were no buildings of any kind to be found in this zone, which left 6 homogeneous zones where buildings would be surveyed (Figure 2 and Table 9).

No.	Homogenous Zone Type	No. Zones Identified
1	Single Family Residential	11
2	High Urban Development	4
3	Industrial Development	11
4	Informal Settlement	10
5	Urban Development	4
6	Rural Settlement	4
	<b>Total</b>	<b>44</b>

Table 9: Dar es Salaam Homogeneous Zones

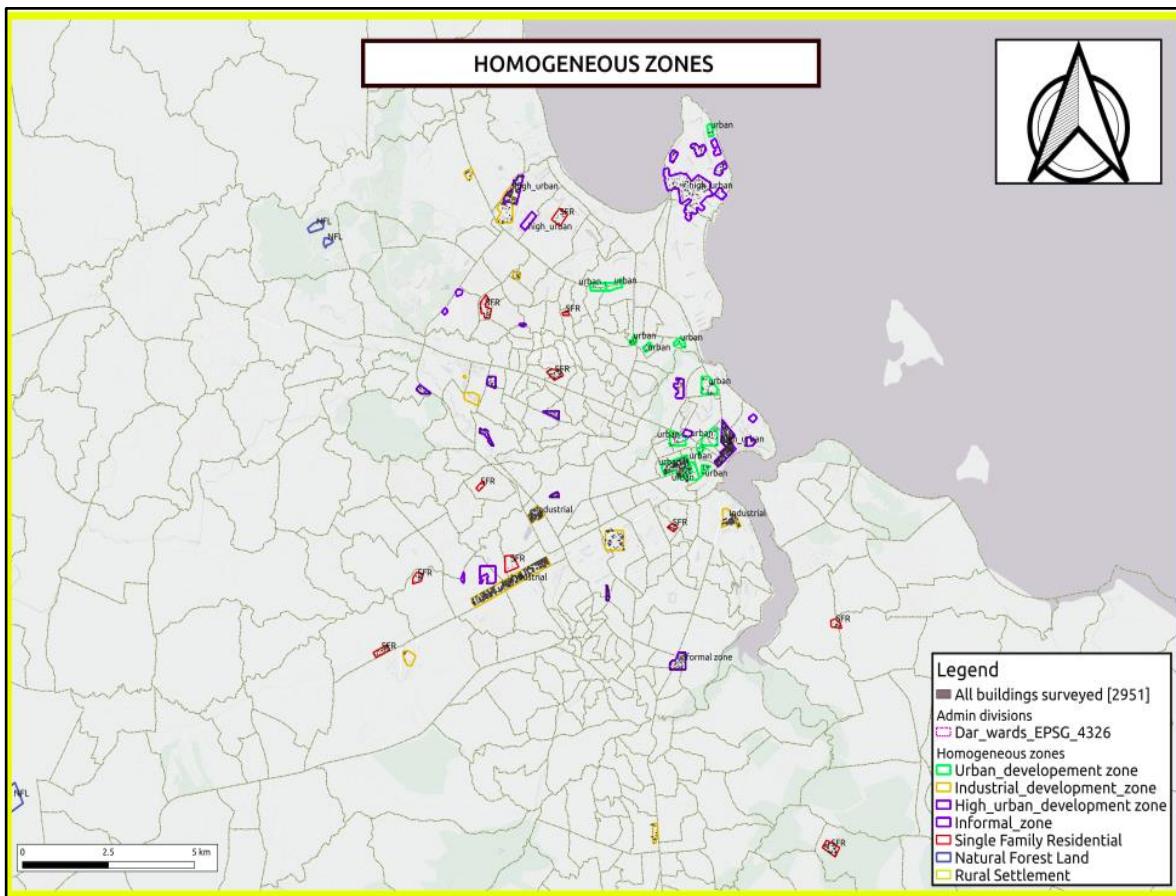


Figure 2: Dar es Salaam Homogeneous Zones ([link to high resolution version](#))



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



ImageCat requested that a randomised sampling strategy be applied to the homogenous zone types, so that there are 40 randomised points per development type, with 10 buildings surveyed around each of the points. These sampling points were used as coordinates on the ground, from which the mappers randomly selected 10 buildings to survey and collect data for, following the project data model. As there are 6 homogeneous zone types in Dar es Salaam and Pwani, ideally a total of 2,400 buildings were to be surveyed keeping the error margin at 5%.

The randomised points were created according to the development zones and a Homogeneous Zone Samples file shared by ImageCat. As in Deliverable M4.4 for Nepal, the best approach appeared to be cluster sampling. However, in dense areas with only a few more buildings than in our sample selection, there was little point in clustering as virtually all clustered buildings would overlap.

Therefore, for three of the zones (Rural Settlement, Single Family Residential, and Urban Development) the Tanzania team used random points, following the clustered approach. For the other three zones (High Urban Development, Industrial Development, and Informal Settlements) they used the QGIS Field Calculator to generate the random buildings. These are representative because of the large proportion of chosen buildings to total population (i.e. surveying 400 out of 434 buildings in the Industrial Zone).

This means that roughly 2,400 buildings were to be surveyed, though the team added in additional buildings to help with the error margin. However, as for the sampling in Nepal, several of the sampling points fell within military and government complexes, where ground surveying is prohibited and, in such cases, where possible alternative points were chosen. When it was not possible to select an alternative point, they were ignored. In the end, a total of 2951 buildings were surveyed around 44 sampling areas across the 6 homogeneous zones. In comparison to the data collected in Nepal, the points in Dar es Salaam were much further distributed, below are a few examples of some of the sampling points. (Figure 3 and Table 10).

For more details regarding the sampling procedure for buildings, the team has prepared a "[Sampling procedure for buildings METEOR TZ](#)" tutorial attached in this report as Appendix 3, that explains how they approached clustered and non-clustered sampling with open source tools.

**Note:** The team in Tanzania has in fact digitised *all* of the buildings in the areas provided by ImageCat, therefore, there shouldn't have been bias when randomly selecting. This is one of the benefits of working in a city that has been thoroughly digitised by past open mapping projects.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*

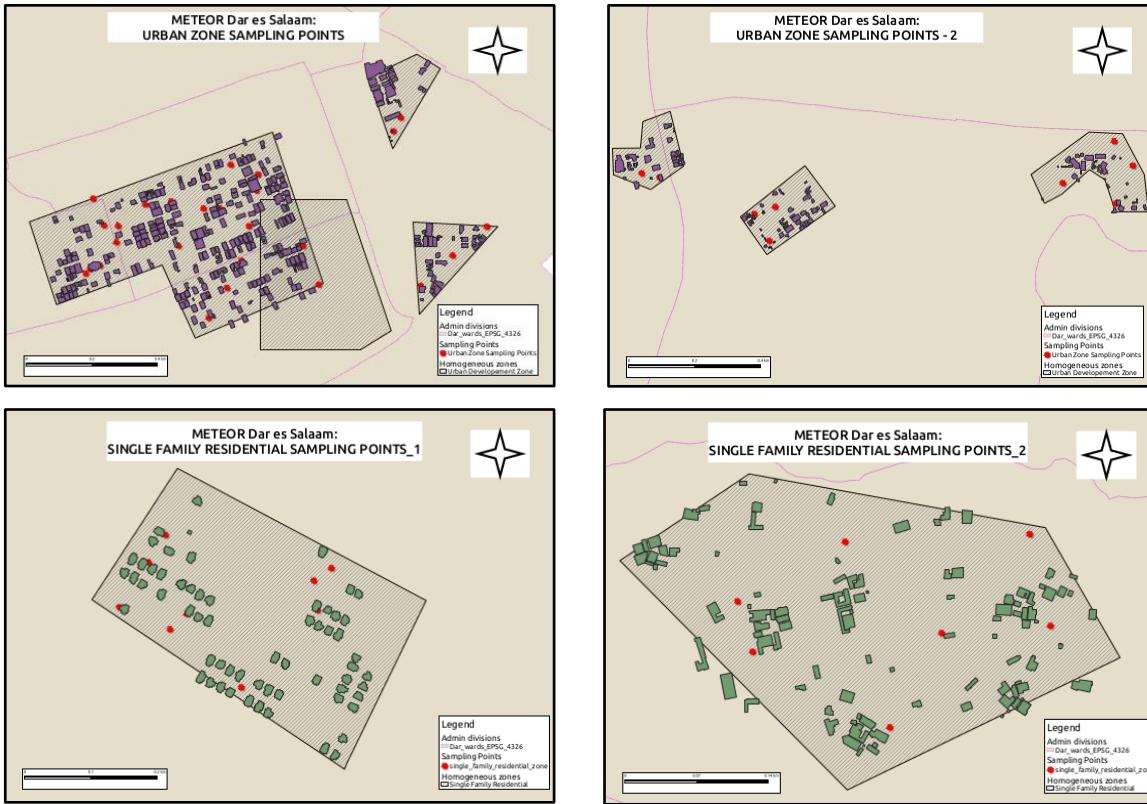


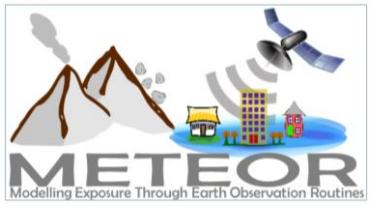
Figure 3: Dar es Salaam Sampling Points

No.	Homogenous Zone Type	No. Buildings Surveyed
1	Single Family Residential	467
2	High Urban Development	584
3	Urban Development	505
4	Industrial Development	423
5	Informal Settlement	538
6	Rural Settlement	434
	<b>Total</b>	<b>2951</b>

Table 10: Dar es Salaam Surveyed Buildings

#### 4.4. Survey Authorisation

No official authorisation was required as the surveying teams did not enter any property or restricted areas, as well as interview anyone. However, there was a formal request sent to the four Municipal Offices in Dar es Salaam, and afterwards the teams informed the local government offices of the survey activities that were going to be done. The surveyors informed the local government of their intended activities and delivered the permission letter before carrying out the ground data collection.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



This was done at the ward level, and was accompanied by a letter of intention composed by OMDTZ/HOT, Municipal (district) and local government. This process is vital to the safety and success of data collection in Tanzania. A copy of the authorisation letter can be found below (Figure 4).



**OpenMap  
Development  
Tanzania**

OpenMap Development Tanzania (OMDTZ)  
NGO Registration Number: 00NGO/0009412  
Plot 228, House 15, Lukuledi Street  
Regent Estate, Mikocheni  
Dar es Salaam, Tanzania  
Email: [omdtanzania@gmail.com](mailto:omdtanzania@gmail.com)  
Twitter: <https://twitter.com/HOTTanzania>

January 28<sup>th</sup>, 2019

Municipal Director  
Ilala Municipal Council  
1 Mission Street  
11883 Dar es Salaam

Municipal Director  
Kinondoni Municipal Council  
2 Morogoro Road  
14883 Dar es Salaam

Municipal Director  
Temeke Municipal Council  
P.O. Box 45085  
Dar es Salaam

Municipal Director  
Ubungo Municipal Council  
Kibamba Area, Morogoro Road  
P.O. Box 55068  
Dar es Salaam

Municipal Director  
Kigamboni Municipal Council  
P.O. Box 36009  
Dar es Salaam



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



RE: METEOR PROJECT

SUB: GROUND DATA COLLECTION FOR BUILDINGS IN DAR ES SALAAM

Reference is made to the above subject.

OpenMap Development Tanzania (OMDTZ) is working together with the Humanitarian OpenStreetMap Team (HOT) on the Modelling Exposure through Earth Observation Routines (METEOR) project. This project is focused on developing innovative applications of earth observation (EO) technologies to improve understanding of exposure to help minimize risks from natural hazards (specifically flooding). As part of this project, recent urban planning graduates are conducting research work on visually assessing building structures. This assessment work will begin from February 4, 2019 and ends on April 4, 2019.

We are also working with the Disaster Management Department (DMD) from the Prime Minister's Office of Tanzania, which acts as the central planning, coordinating and monitoring institution for prevention, mitigation, preparedness, response and post-disaster recovery for all disaster risks. It is responsible for promoting education, knowledge and IT in disaster management for public awareness. DMD works with several METEOR partners and specialists in Ardh University and the University of Dar es Salaam to develop open databases that store exposure and hazard information. DMD will ensure that METEOR outcomes are fit-for-purpose and used in Disaster Risk Management in Tanzania.

The data we will be collecting is: shape of the building, type of the building structure (RCC, masonry etc.), type of roof, number of stories and other related information. No sensitive data will be collected. We hope that any information and technique on collecting such information will lead to better flooding resiliency.

If you have any concerns, please reach out to me on my mobile +255 714 166 665. Thank you for your cooperation.

Regards,

Executive Director OMDTZ

Innocent Maholi

*Figure 4: Surveying Letter of Intention*

The team was also very conscious of their behaviour and overall code of conduct, carrying out their data collection in a respectfully appropriate manner. For example, the surveyors did not enquire about the details of building attributes from individuals who did not independently volunteer to share this information.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



#### 4.5. Tool Preparation

Surveyors can choose from a number of free and open OSM tools that facilitate on-the-ground data collection in both online and offline settings. The tools used by the mapping team in Dar es Salaam included the mobile applications OpenDataKit (ODK), OpenMapKit (OMK) and Maps.me. ODK<sup>9</sup> is an Android application that replaces paper forms used in survey-based data gathering, allowing users to easily author, field, and manage mobile data collection solutions. It supports a wide range of question and answer types, which can be managed by a set of rules in order to help reduce the occurrence of errors during data collection. It is also designed to work well without network connectivity. Once forms have been downloaded, it does not depend on the internet again until the collected data needs to be uploaded to a server, or not at all if the data is transmitted directly to a desktop.

The biggest portion of preparing ODK for use, is the creation of the customised questionnaire that will be read by the tool and presented to the surveyors through their mobile application. The questionnaire was created by HOT with the use of a spreadsheet and excel sheet (Figure 5) and the localised OSM data model outlined in the previous section. Once the spreadsheet is complete, it can be converted into a form for consumption by ODK using XLSForm<sup>10</sup> or by loading it onto OMK Server, which is detailed in the 4.7 Data Management section below.

---

<sup>9</sup> <https://opendatakit.org>

<sup>10</sup> <https://opendatakit.org/xlsform/>



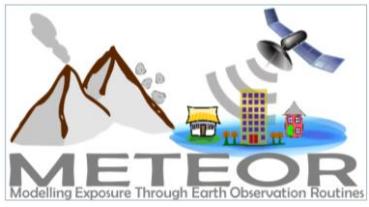
# METEOR: Ground Data Collection Using Protocols II:

## Dar es Salaam, Tanzania



start	start		
end	end		
today	today		
username	username		
phononenumber	phononenumber		
deviceid	deviceid		
note	title	MODELLING EXPOSURE THROUGH EARTH OBSERVATION ROUTINES: METEOR DAR ES SALAAM AND PWANI	
osm_building_tags	osm_buildings	Building	Tag attributes about this building
begin_group	building_photos		
image	image_1	Building Image 1	Building must best fit the image viewpo view]
image	image_2	Building Image 2	side view[right or left]
end_group			
note	end_001	You have reached the end of this form.	
<hr/>			
list name	name	label	
building_tags	addr:city	Name of city	
building_tags	addr:municipal	Name of Municipality	
building_tags	addr:ward	Name of the ward	
building_tags	addr:subward	Subward Name	
building_tags	addr:street	Street Name	
building_tags	building:lateral:material	Lateral load resisting material	
building_tags	building:lateral:system	Structural system of building	
building_tags	building:levels	Number of above-ground levels of a building	
building_tags	building:levels:underground	Number of below-ground levels of building	
building_tags	building:age	Age of building	
building_tags	building:condition	Describe the condition of the building	
building_tags	building	Describe the building purpose	
building_tags	capacity:persons	Describe the number of people a building can support	
building_tags	roof:shape	Shape of roof	
building_tags	roof:material	Outer material for the building roof	
building_tags	building:neighbor	Neighbor condition	
building_tags	building:geological_site	Geological/Site condition	
<hr/>			
building:lateral:material	concrete_reinforced	Concrete reinforced	
building:lateral:material	steel	Steel	
building:lateral:material	steel-concrete_composite	Steel-concrete composite	
building:lateral:material	block	Block	
building:lateral:material	brick	Brick	
building:lateral:material	stone	Stone	
building:lateral:material	corrugated_galvanized_iron	Corrugated Galvanized Iron	
building:lateral:material	earth	Adobe (earth)	
building:lateral:material	wood	Timber/Wood	
building:lateral:material	bamboo	Bamboo	
building:lateral:material	light_gauge_steel or cold_formed_steel	Light Gauge Steel/Cold Formed Steel	
building:lateral:material	iron_sheet	Iron Sheet	
building:lateral:material	industrial_troughed_sheet	Industrial Troughed Sheet	
building:lateral:material	plywood	Plywood	
building:lateral:material	mixed		
building:lateral:system	moment_resisting_frame	Moment Resisting Frame	
building:lateral:system	shear_wall	Shear Wall	
<hr/>			
form_id	form_title	default_language	
meteor_dsm	meteor_dsm	english	

Figure 5: OpenDataKit Spreadsheet



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Form ID	Description	Status
meteor_dsm	ID: meteor_dsm_v1	Selected
meteor_dsm_v2	ID: meteor_dsm_v2	Selected
meteor_dsm_v3	ID: meteor_dsm_v3	Selected
meteor_pwani_v1	ID: meteor_pwani_v1	Selected
meteor_v4	ID: meteor_v4	Selected
meteor_validation_v1	ID: meteor_validation_v1	Selected
meteor_validation_V2	ID: meteor_validation_V2	Selected

**Tags** SAVE TO ODK COLLECT

**building:lateral:material**

Lateral load resisting material  
*building:lateral:material*

- Concrete reinforced  
*concrete\_reinforced*
- Steel  
*steel*
- Steel-concrete composite  
*steel-concrete\_composite*
- Brick  
*brick*
- Stone  
*stone*
- Adobe (earth)

Figure 6: OpenDataKit Questionnaire (left), OpenMapKit (right)

OMK<sup>11</sup> (OMK) is a free and an open source tool that is used to create professional quality mobile data collection surveys on the ground in OSM. OMK launches directly from ODK and allows users to add points of interests and attribute information in the form of tags for OSM, following the customised survey questionnaire (Figure 6). This allows detailed information to be securely collected, saved, and uploaded to the server of choice. OMK is an extension that launches directly from within ODK, when the OSM question type is enabled in a standard survey. It allows the user to browse OSM features, and to create and edit OSM tags. Therefore, ODK must also be installed on the Android phone in order to use OMK to collect and edit data in OSM. Preparation for this tool included the creation of an OSM base layer, containing the footprints of the buildings that were selected for surveying on the ground (Figure 7). This allows the mappers to easily identify and select a specific building geometry to add structural details. Other useful base layers created by HOT Tanzania for OMK included the boundaries of the homogenous zones identified across Dar es Salaam and Pwani.

<sup>11</sup> <http://openmapkit.org>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*

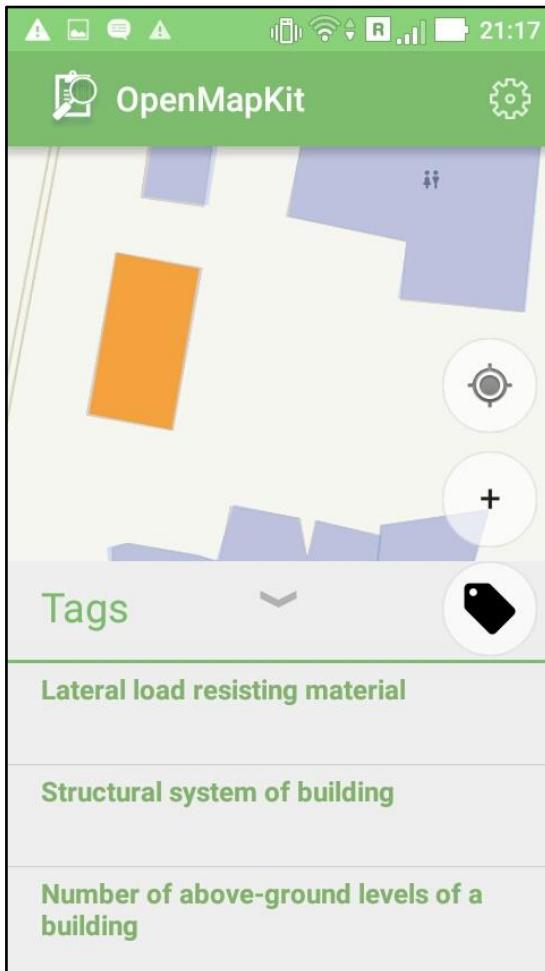


Figure 7: OpenMapKit Interface

#### 4.6. Team Training

A full day training session was conducted at the OMDTZ office by one senior civil engineer from Ardhi University, Mr. Frank Mushi, that focused on the team becoming experts in the building taxonomy that had been created in the data model. The training needed only one full day because most of the surveyors have been involved in various projects by using free and open applications and how to classify buildings as well. During the training the primary focus was training surveyors on details of building structures, materials commonly used in Dar es Salaam and the challenges of visually inspecting buildings from the ground. With this training, the team could identify the different attributes necessary to complete the building taxonomy in their survey forms. Some specific training materials included: structural building features<sup>12</sup>, loading of structures<sup>13</sup>, and classification of

<sup>12</sup> [https://docs.google.com/presentation/d/1KErOR2Ylxey1zdcM-svT\\_qryhWTzgpnD2yctLP0FbR8/edit](https://docs.google.com/presentation/d/1KErOR2Ylxey1zdcM-svT_qryhWTzgpnD2yctLP0FbR8/edit)

<sup>13</sup> <https://drive.google.com/drive/u/0/folders/1cdchSMWzj6vpouvg6hjkI5N2oaA1Vm>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



structures<sup>14</sup>, which we are including as training resources. On the same day, the application and the forms were tested. This training covered the use of the ODK, OMK and Maps.me applications on the mobile phones of the surveyors. For this portion of the training, the surveyors were taken out to the field where they collected real world data on the ground to simulate the actual data collection process (Figure 8). The data that was collected during training, was then evaluated to provide feedback to the surveyors, as part of the learning process.



Figure 8: Team Training at OpenMap Development Tanzania Office

---

<sup>14</sup> <https://drive.google.com/drive/u/0/folders/1cdchSMWzj6vlpoouvg6hjkI5N2oaA1Vm>



# METEOR: Ground Data Collection Using Protocols II:

## Dar es Salaam, Tanzania



### 4.7. Data Management

After surveying, the data needs to be retrieved from the devices. This can be done simply by manually copying the data directly from the mobile phone devices over to a computer, however, this does not scale well when there are large numbers of people collecting data. It also means the data is generally not backed up until the surveyors return to the office, and if the devices are lost, then all the collected data is also lost. HOT Tanzania decided to use a data collection server as their central repository for managing the hosting of their questionnaire forms, deployments, collection and aggregation of responses.

HOT Tanzania setup an online OpenMapKit (OMK) Server<sup>15</sup> to help manage their data flow and backup their submissions (Figure 9). A surveyor can retrieve blank questionnaire forms directly from the server, and as long as there is internet connectivity, the completed questionnaires can be submitted back to the server from the field. The submitted data is then immediately backed up by the server, greatly reducing the chances of loss. In addition to the deployment of forms to ODK, overall data management and automatic backing up of submitted forms, OMK Server also provides the ability to view all the surveyed features on a map (Figure 10), as well as filter and export selected data.

OpenMapKit Server						
		Home	Deployments	Upload Form	Archived Forms	admin
<b>meteor_dsm_v3</b>	Total submissions: 1082	Start	End	Username	Submission Time	Image
<a href="#">ODK Data</a>	<a href="#">OSM Data</a>					
<a href="#">Manage</a>						
From <a href="#">?</a>	To <a href="#">?</a>					
<input type="text" value="YYYY-MM-DD"/>	<input type="text" value="YYYY-MM-DD"/>					
Username	Choose an item...					
<a href="#">Filter</a>	<a href="#">Clear</a>					
Page Size	Page					
<input type="button" value="5000"/>	<input type="button" value="1"/>					
<a href="#">Switch to Map view</a>						
Mar 15, 2019 12:52 PM	Mar 15, 2019 1:20 PM	Josephine A. Tinuga	Mar 15, 2019 6:44 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 7, 2019 1:04 PM	Mar 7, 2019 2:13 PM	Neema Elias	Mar 7, 2019 2:41 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 13, 2019 1:22 PM	Mar 13, 2019 1:30 PM	William J. Mponeja	Mar 13, 2019 10:23 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 13, 2019 1:34 PM	Mar 13, 2019 1:36 PM	Neema Elias	Mar 13, 2019 11:04 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 11, 2019 11:47 AM	Mar 11, 2019 11:48 AM	Paschal SM	Mar 12, 2019 10:08 AM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 1, 2019 12:56 PM	Mar 1, 2019 12:58 PM	Evarist Isdory	Mar 1, 2019 1:45 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 8, 2019 3:21 PM	Mar 8, 2019 3:25 PM	Joan kamwela	Mar 8, 2019 3:30 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 6, 2019 11:13 AM	Mar 6, 2019 11:16 AM	Glory Emanuel	Mar 6, 2019 1:50 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 8, 2019 12:52 PM	Mar 8, 2019 4:17 PM	Asia Komba	Mar 8, 2019 4:19 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 1, 2019 12:20 PM	Mar 1, 2019 12:30 PM	Hashiru Hashim	Mar 1, 2019 2:38 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 1, 2019 11:05 AM	Mar 1, 2019 11:49 AM	Gloria Elevan	Mar 1, 2019 12:36 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 11, 2019 2:11 PM	Mar 11, 2019 2:22 PM	Gisela John	Mar 12, 2019 8:03 AM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 8, 2019 12:24 PM	Mar 8, 2019 12:28 PM	Joan kamwela	Mar 8, 2019 3:29 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	
Mar 4, 2019 11:58 AM	Mar 4, 2019 12:00 PM	Joan kamwela	Mar 4, 2019 2:12 PM	<a href="#">Download image</a>	<a href="#">OSM File</a>	

Figure 9: OpenMapKit Server Table View

<sup>15</sup> <https://github.com/hotosm/OpenMapKitServer>



# METEOR: Ground Data Collection Using Protocols II:

## Dar es Salaam, Tanzania



Figure 10: OpenMapKit Server Map View

### 4.8. Ground Surveying

Once the data management system was set up through the OMK Server, the team was trained on how to identify the location of points using OMK with aerial imagery basemap mbtiles on their mobile devices. MBTiles is a file format used for storing map tiles as a single file - the most common use case as base layers for mobile mapping applications. The imagery used was taken by a fixed-wing drone with 10cm resolution, originally collected by COWI in 2016 and provided by Tanzania's Ministry of Lands, Housing and Human Settlements. In addition to the imagery the surveyors could see the OSM building footprints in their allocated areas of interest, so that they were able to complete the attribute information following the localised data model observed from the ground, and then submit it back to the OMK server once the survey was completed (Figure 11).

This facilitated easy identification of the buildings, where the high contrast of the images allowed the team to identify the roof colour, and navigate between buildings to be surveyed. In addition, the team was composed of young Tanzanians who were familiar with the areas, making the data collection more efficient. For safety reasons, surveyors were divided into 9 pairs so that they would be working together in the field.

The team used both a clustering method as well as a non-clustering method to allocate the sampling points in the six identified zones. These were spread out across the city of Dar es Salaam, and some required two hours of transportation time to reach, especially rural settlements and some single-family residential zones. To reach these, surveyors used boda bodas (motorcycles) to be efficient in terms of time and money.

The houses around each sampling point were chosen and visually inspected by the surveyors. The field surveyors were given discretion on how far or near to specific coordinates they could move and were



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



advised not to infringe on the 20-metre buffer of another point. Surveyors were asked to ignore any points that fell inside restricted zones, such as military compounds. Any surveyors encountering difficulties with accessing particular directions of a point were asked to collect data in another direction where possible.

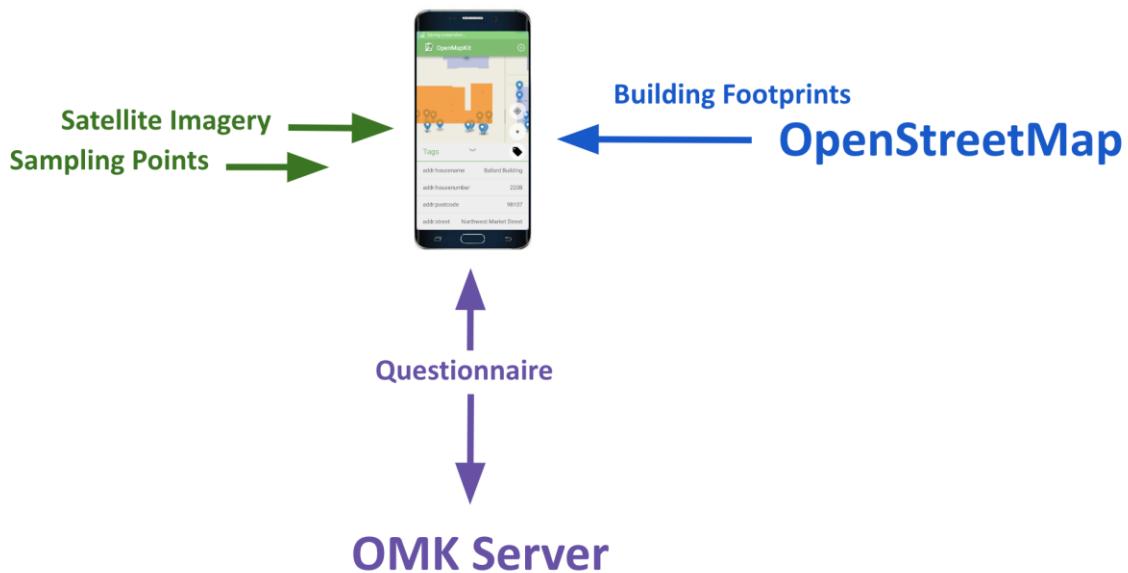


Figure 11: OpenMapKit Diagram

Throughout the data collection period, weekly meetings were held and used as a place for the surveyors to discuss and address any challenges faced. This provided them with the opportunity to ask any specific technical questions, learn more about OSM, OMK and OMK Server, as well as get feedback on the quality of the data being collected to help iron out any issues before continuing on with the data collection. At the end of the data collection a total of 2,951 buildings were surveyed. Some of these buildings had pre-existing attribute information following the tags identified in the localised data model such as `building`, `building:neighbor`, `building:levels`, `roof:material` and `roof:shape`, while the other tags were not present at all. Table 11 shows how the data surveyed through this project has added to and enriched the pre-existing attribute information in OSM for the selected buildings in Dar es Salaam.

Looking at the breakdown of the data values collected for the key attributes, it can be seen that the lateral structure materials used for buildings in the surveyed areas of Dar es Salaam are roughly 73.7% concrete blocks, 14.2% reinforced concrete, 4% steel, 2% brick and 1% steel-concrete composite and wood (Table 12). The values for the lateral system are approximately 62% masonry wall, 14.8% confined masonry, 9% framed concrete structure and 7.5% moment resisting frame (Table 13).



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Before going to the field, the team gathered statistics on the building attributes based on the digitisation of the buildings. The buildings selected to be surveyed had some attribute information before the METEOR Project. After downloading the data using the HOT Export Tool, these sample buildings were imported into QGIS and reviewed. By clipping all of the buildings according to the boundaries provided by ImageCat, the team was able to later compare the attribute information before and after data collection.

No.	Tags	Before	After	% Change
1	building	1548	2951	47.5%
2	building:age	206	2951	93%
3	building:condition	105	2951	96.4%
4	building:geological_site	0	2951	100%
5	building:lateral:material	0	2951	100%
6	building:lateral:system	0	2951	100%
7	building:levels	1461	2951	50.5%
8	building:levels:underground	0	2951	100%
9	capacity:persons	0	2951	100%
10	roof:material	222	2951	92.5%
11	roof:shape	0	2951	100%
12	source	0	2951	100%

Table 11: Before and After Survey Attribute Information

No.	Lateral Materials	Count	Percentage
1	concrete_reinforced	419	14.2%
2	brick	56	1.9%
3	steel	119	4%
4	earth	22	-
5	bamboo	3	-
6	wood	40	1%
7	steel-concrete_composite	40	1%
8	stone	8	-
9	cgi	14	-
10	hybrid	2	-
11	light_gauge_steel	11	-
12	block	2175	73.7%
13	mixed	11	-
14	Industrial_troughed_sheet	7	-

Table 12: Lateral Materials



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



No.	Lateral System	Count	Percentage
1	moment_resisting_frame	221	7.5%
2	masonry_wall	1835	62%
3	hybrid	24	1%
4	braced_frame	17	1%
5	dual_system (frame wall)	44	1.5%
6	confined_masonry	437	14.8%
7	framed_steel_structure	47	1.6%
8	framed_timber_structure	14	-
9	framed_concrete_structure	261	9%
10	shear_wall	51	1.7%

Table 13: Lateral Systems

#### 4.9. Data Validation

Validating the data to ensure that it is complete and correct is a key part of the ground data collection process. The main objective is to catch any mistakes in the data before uploading it to OSM, so that anyone using the information can be confident of the data accuracy. Data validation was carried out at four stages by the Tanzania team during the ground data collection process. The first phase of data validation is when the data quality specialists are cleaning up the information they collected on the ground. There are several tools that can be used to assist with data validation, but the tool generally used the most which was also applied by HOT/OMDTZ is the use of JOSM (Figure 12).

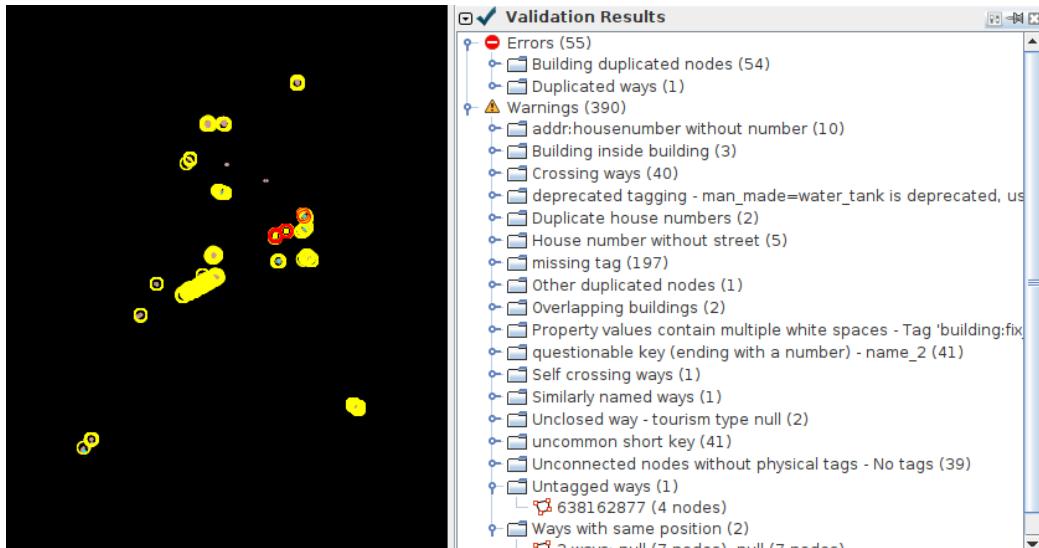


Figure 12: JOSM Validation Tool

This desktop GIS tool is specifically designed for OSM data and can help highlight any errors and warnings in the data with the ‘Validation’ tool. Geospatial errors detected include intersecting geometries and unconnected nodes, while attribute information errors include missing tags, misspelled tags, and incorrectly assigned tags. The ‘TODO’ tool, which is also available in JOSM, allows the user to select all the features in a layer and organise them so that they can be reviewed systematically one by one, so that none of them are missed.

The second phase of data validation took place when 50% of the overall data was collected, while the third data validation phase took place when 90% of the data was collected. Both of these phases were conducted by the QA specialists. To do this, a CSV file of all the surveyed information was downloaded, which was then divided into a few groups based on the mappers name or the mobile device ID through which the data was uploaded. Once the data was sorted into groups, every tenth row was validated by a QA specialist based on the images gathered by the surveyors of the building in question. This process was carried out, so that 10% of all data surveyed was validated at the office.

Spot checking was also carried out by visiting the locations of the buildings that fell within the 10% of the data that had been validated. The building information was checked by the QA specialists to ensure that the information on the ground matched the images and data collected by the field surveyors (Figure 13). The engineering students from Ardhi University were specifically brought on to assist with this validation process.



Figure 13: Building Photos

The fourth and final data validation phase took place after 100% of the data was collected through both peer review and GIS experts checking the data in the office. This process involved the surveyors reviewing the data of one another in the office and also revisiting some of the areas to check if the data was collected well. Therefore, this final validation was done in certain zones that had complicated attributes such as Industrial, High Urban and Urban zones. The aim of this was to ensure that the information captured is correct and followed the data model.

Data quality assurance goes hand in hand with the data validation process, and ensures that the data reaches a minimal level of quality. The higher the quality your data is, the more useful it is to end users, so it is essential to manage the process along the way. A tool recently developed by HOT aims to support this data quality monitoring process, so that surveying teams and their supervisors can easily ensure a certain level of quality. MapCampaigner<sup>16</sup> assists with the crowdsourcing of data collection and data quality assessment of detailed attribute information in OSM.

Campaigns can be set up on the tool by specifying the area of interest and the features to collect detailed information for. The features and their associated attribute information are specified through a data model. This tool allows the surveying process and collected data to be evaluated, monitoring the overall quality by providing statistics on the number of features collected, how much of the area has been covered, attribute completeness, features by type, potential errors, as well as information on user engagement. HOT/OMDTZ set up ten projects through the MapCampaigner tool for

---

<sup>16</sup> <https://campaigns.hotosm.org/>



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



monitoring the quality assurance of the data collected in Dar es Salaam. These campaigns are split up by the homogeneous zone types (Table 14).

No.	Homogeneous Zone Type	MapCampaigner Project
1	Single Family Residential	<a href="https://campaigns.hotosm.org/campaign/90c4cc568a5a47d49db343a08d658d1e">https://campaigns.hotosm.org/campaign/90c4cc568a5a47d49db343a08d658d1e</a>
2	Rural Settlement	<a href="https://campaigns.hotosm.org/campaign/9323035dde544c868793a92844b37c53">https://campaigns.hotosm.org/campaign/9323035dde544c868793a92844b37c53</a>
3	Urban Development	<a href="https://campaigns.hotosm.org/campaign/203467dc917f4f39ad25cace439555b3">https://campaigns.hotosm.org/campaign/203467dc917f4f39ad25cace439555b3</a>
4	Urban_Mixed	<a href="https://campaigns.hotosm.org/campaign/ab1e11ac372a4a418e3b78e7b568d61c">https://campaigns.hotosm.org/campaign/ab1e11ac372a4a418e3b78e7b568d61c</a>
5	Urban Zone 1	<a href="https://campaigns.hotosm.org/campaign/473705af7e88403986bd910ab5694b29">https://campaigns.hotosm.org/campaign/473705af7e88403986bd910ab5694b29</a>
6	Industrial Development Zone 1	<a href="https://campaigns.hotosm.org/campaign/6e03da6b7af54bf48d44d5adb0ddc17e">https://campaigns.hotosm.org/campaign/6e03da6b7af54bf48d44d5adb0ddc17e</a>
7	Industrial Development Zone 2	<a href="https://campaigns.hotosm.org/campaign/6e03da6b7af54bf48d44d5adb0ddc17e">https://campaigns.hotosm.org/campaign/6e03da6b7af54bf48d44d5adb0ddc17e</a>
8	Informal	<a href="https://campaigns.hotosm.org/campaign/7316fb4466ce4393879ab9501556c9c1">https://campaigns.hotosm.org/campaign/7316fb4466ce4393879ab9501556c9c1</a>
9	High Urban Zone 3	<a href="https://campaigns.hotosm.org/campaign/ce18d14f52ba45b8a96d0a2e4a11fb27">https://campaigns.hotosm.org/campaign/ce18d14f52ba45b8a96d0a2e4a11fb27</a>
10	High Urban Zone 4	<a href="https://campaigns.hotosm.org/campaign/35bb6252b9e948ffa53c04a5445c2ef8">https://campaigns.hotosm.org/campaign/35bb6252b9e948ffa53c04a5445c2ef8</a>

Table 14: Dar es Salaam MapCampaigner Projects

The Tanzania team had some technical challenges with MapCampaigner early in the project, specifically it was difficult to add boundaries that contained multi-polygons that are distantly dispersed, and this led to initial errors. The team troubleshooted these issues, and was able to resolve them so that MapCampaigner was used later in the data collection process. Something to note is that the campaign links are no longer active (also true of the Nepal data collection), as they may have been deactivated with the latest update of the software, but the screenshots in Figure 14 and Figure 15 show how MapCampaigner was used during the project. One of the benefits of using MapCampaigner was the ability to assess the completeness of the attributes collected by the team.



# METEOR: Ground Data Collection Using Protocols II:

## Dar es Salaam, Tanzania



MAP CAMPAIGNER About Resources Create Campaign kombe1207 ▾

### Meteor/Dar es Salaam/High-Urban-Development-Zone\_1

● Inactive

[Buildings](#)

Started : 18 Feb 2019 Ends : 18 Mar 2019

Modelling Exposure Through Earth Observation Routines[METEOR], is the project aim to cover building taxonomy around selected homogeneous zones. The project will help to improve the understanding of the distribution and character of building and infrastructure's exposure to natural hazards. This High Urban Development zone in Dar es Salaam consists all of the buildings available in high urban area, so the project will capture the full details of the buildings for the selected zone

Campaign Manager : [kombe1207](#), [Nuru\\_Kalulunga](#), [dorica\\_mugusi](#), [William Perry Evans](#), [nsmith](#), [mataharimhairi](#) Creator : [kombe1207](#) Map: Default

Participate

[OpenMapKit](#)

[Maps.me](#)

Share

Features Collected	Users Engaged	Areas Covered
233	0	0%

Feature details

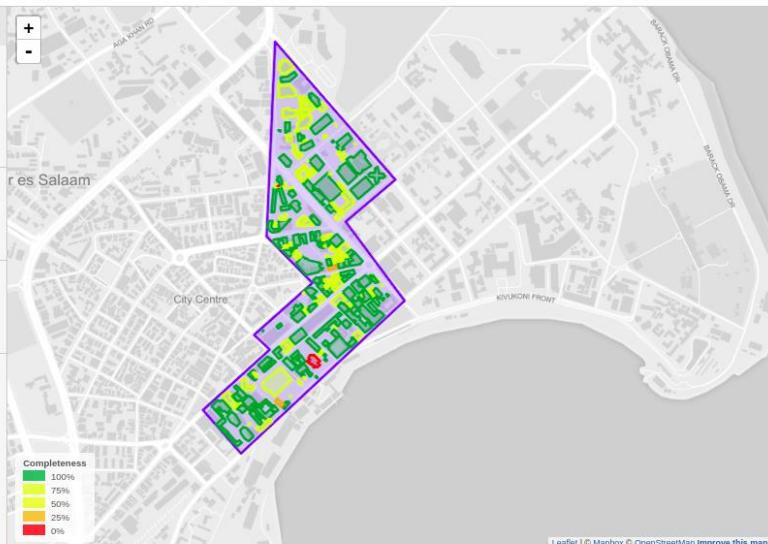
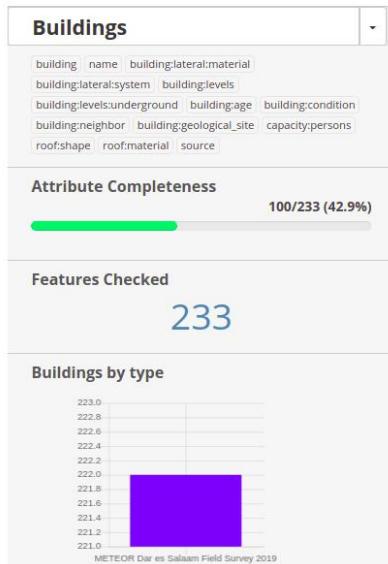


Figure 14: High Urban Homogeneous Zone on MapCampaigner



# METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania

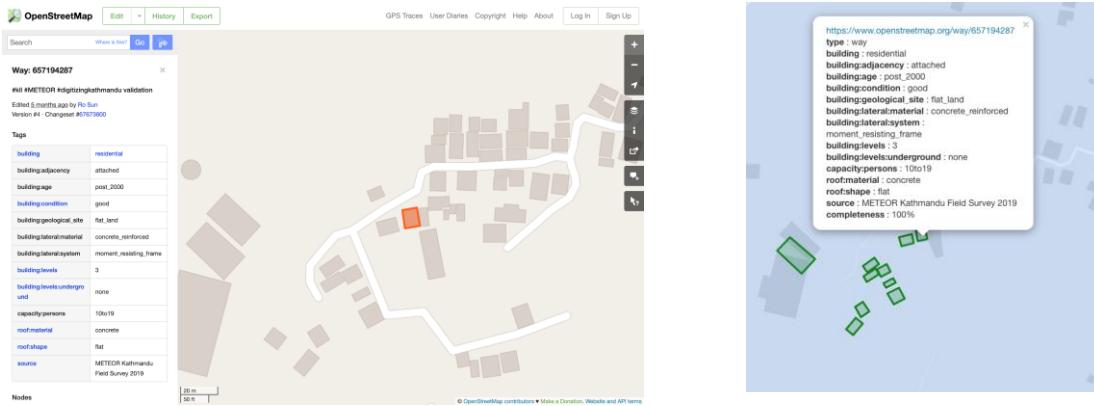


Figure 15: Attribute Information in MapCampaigner (Left) and OSM (Right)

## 5. Open Access

### 5.1. Data Extracts

The final collected and validated Dar es Salaam data, is now hosted on OSM and available for anyone to access and use. There are multiple ways to extract data from OSM such as the use of Overpass Turbo or JOSM, but for those with little to no experience with OSM, the Export Tool<sup>17</sup> is recommended. The Export Tool was created by HOT as a way to provide individuals with low technical skills to easily extract data from OSM through a user-friendly interface and straightforward process.

The Export Tool allows users to create custom OSM data extracts for anywhere in the world, simply by selecting an area of interest (Figure 16), map features (Figure 18) and file formats to convert the data into (Figure 17). Within minutes, up-to-date OSM data is exported (Figure 19).

<sup>17</sup> <https://export.hotosm.org/>



# METEOR: Ground Data Collection Using Protocols II:

## Dar es Salaam, Tanzania



Figure 16: Describe Tab – Create Export Tool Function

Figure 17: Formats Tab – Create Export Tool Function



# METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania



The screenshot shows the 'Data Tab – Create Export Tool Function'. On the left, there's a 'Feature Selection' section with a YAML configuration:

```

Buildings:
  types:
    - polygons
  select:
    - building:lateral:material
    - building:lateral:system
    - building:levels
    - building:levels:underground
    - building:age
    - building:condition
    - building
    - capacity:persons
    - roof:shape
    - roof:material
    - building:neighbor
    - building:geological_site
    - source
  where: source = "METEOR Dar es Salaam Field Survey 2019"
  
```

Below this is a 'Load from JOSM Preset XML' button and a 'Next' button. On the right, there's a map titled 'Dar es Salaam Region' showing the coastline and inland areas. A dashed blue rectangle highlights the area of interest (AOI). Labels on the map include 'Kibaha', 'Dar es Salaam', and 'Mkuranga'. At the bottom of the map interface, it says 'Area Of Interest (AOI)' and 'Dar es Salaam Region Tanzania, Dar es Salaam'. There are also 'ZOOM TO SELECT' and 'Draw Screen' buttons.

Figure 18: Data Tab – Create Export Tool Function

The screenshot shows the 'Summary Tab – Create Export Tool Function'. On the left, there are fields for 'Name' (METEOR Dar es Salaam Filed Survey 2019), 'Description' (METEOR), and 'Export Formats' (GeoPackage (.gpkg) and OSM (.pbf)). There are also checkboxes for 'Buffer AOI - expand an uploaded boundary by 0.02 degrees', 'Publish this Export', 'Unfiltered files - all OSM data for PBF, MWM, OBF, IMG formats. Use only for small areas.', and 'Bundle for POSM'. Below these is a large red 'Create Export' button. A note at the bottom says 'OpenStreetMap database last updated 14 minutes ago'. On the right, the map interface is identical to Figure 18, showing the Dar es Salaam Region with the highlighted AOI and various tools available on the right side.

Figure 19: Summary Tab – Create Export Tool Function

The only requirement to use the Export Tool is to sign in with an OSM username and provide an email address, which is used to send a link of completed jobs in case the connection is lost, the export will still run and notified the user once it is ready.

A job has already been set up on the Export Tool for the Dar es Salaam data (Figure 20), which can be found in Table 15. The data is extracted from OSM using the source = "METEOR Dar es Salaam Field Survey 2019" tag, which was applied to each of the buildings surveyed by the team. This selects the specific buildings within the area of interest, and then filters the attributes based on those identified in the YAML syntax in the 'Feature Selection' tab of the Tool.



## METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania



There are multiple file formats to choose from, but only the GeoPackage and pbf formats have been selected. If other formats are desired, the job can be ‘Cloned’ and the settings changed before running the job again. The file can be downloaded by clicking on the highlighted link on the export page once it has finished running.

METEOR Dar es Salaam Filed Survey 2019	
Description:	Export data from METEOR Dar es Salaam field Survey 2019
Id:	7607cd74-5864-4923-b7fc-88ccbad77aa3
Project:	METEOR
Area:	4119 sq km
Created at:	Thursday, May 21st 2020, 2:37 pm
Created by:	kombie1207
Published:	Yes
Export formats:	<a href="#">GeoPackage (.gpkg)</a> <a href="#">OSM (.pbf)</a>
OSM Analytics:	<a href="#">View this area</a>
<a href="#">Features</a> <a href="#">Re-Run</a> <a href="#">Clone</a> <a href="#">Delete</a>	

Thursday, May 21st 2020, 2:37 pm	
Status:	COMPLETED
ID:	76a30ef9-a787-43c5-a799-05da03dd5a66
Finished:	Thursday, May 21st 2020, 2:37 pm
Duration:	a few seconds
GeoPackage	<a href="#">METEOR_Dar_es_Salaam_Filed_Survey_2019.gpkg.zip (.gpkg)</a> (398 kB)
OSM	<a href="#">METEOR_Dar_es_Salaam_Filed_Survey_2019.osm.pbf (.pbf)</a> (184 kB)

Figure 20: Export Page – Completed Job

Dar es Salaam Data Extract
<a href="https://export.hotosm.org/en/v3/exports/c1f46025-6f93-4957-925d-83fed720a51b">https://export.hotosm.org/en/v3/exports/c1f46025-6f93-4957-925d-83fed720a51b</a>

Table 15: Export Tool Data Extract (NOTE: export data older than 30 days is deleted)

### 5.2. Methodology

Creating data in OSM is not the only aim of HOT-led projects, but providing open access to the methodology and tools applied is also at the core. Sharing knowledge is the best way to increase awareness of the issues and goals at hand, which also allows for the sustainability of the data in the years to come.

One of the ways in which the project methodology is shared with the greater community is through the OSM wiki<sup>18</sup> (Figure 21, Table 16). The wiki is a common place for documenting projects related to creating data in OSM. It is also the first place that should be checked for similar projects conducted in the targeted area of interest. This way the processes and lessons learned can be reviewed and built upon. It is also common to provide the details of the project manager who can be contacted for further discussions regarding the work and provide feedback.

<sup>18</sup> <https://wiki.openstreetmap.org/>



# METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania



Wiki English Create account

Page Discussion Read View source View history Search OpenStreetMap Wiki

## METEOR Project in Tanzania

Contents [hide]

- 1 METEOR Project in Tanzania
- 2 Project Overview
- 3 Goals and Deliverable
- 4 Mapping Timeline
- 5 Area of Interest
- 6 Remote Digitization
  - 6.1 Mapping Approach
  - 6.2 Methodology
    - 6.2.1 Tasking Manager Projects
      - 6.2.1.1 List of Tasks Mapped from HOT Tasking Manager
      - 6.2.2 JOSM Application
    - 6.3 Quality Assurance
      - 6.3.1 Quality Assurance (Before and After digitization)
        - 6.3.1.1 Building Alignment
        - 6.3.1.2 Building Shape
        - 6.3.1.3 Undigitized Buildings
        - 6.3.1.4 Quality of image used
      - 6.3.2 Challenges During Building Digitization
  - 7 Data Collection
    - 7.1 Surveying Approach

Figure 21: Tanzania METEOR Wiki Page

Local Partner Organisation	Link
OpenMap Development Tanzania/Ramani Huria	<a href="https://wiki.openstreetmap.org/wiki/METEOR_Project_in_Tanzania#METEOR_Project_in_Tanzania">https://wiki.openstreetmap.org/wiki/METEOR_Project_in_Tanzania#METEOR_Project_in_Tanzania</a>

Table 16: Kathmandu METEOR Wiki Page Link

The OpenStreetMap Foundation provides an overview on the expectations for organised editing efforts in OSM through their ‘Organized Editing Guidelines’ page.<sup>19</sup> Their goal is to provide a framework to both organised mapping initiatives and the communities to encourage good organised mapping, which will assist with the sustainability and usability of the data created. The organised editing guidelines apply to any edits that involve more than one person and can be grouped under one or more sizeable, substantial, coordinated editing initiatives. One of the key sections of the guidelines focus on the creation of a wiki page, to assist with the recording and sharing of activities with the rest of the OSM community for constructive feedback.

A page for the METEOR project has been set up on the HOT website<sup>20</sup>, with contact details for the project manager, as well as links to news update blogs and other key pages related to the work. In addition, OpenMap Development Tanzania have set up a project page on the OSM wiki (Table 16) for the data created in Tanzania, as part of the METEOR project following the OSM Foundation ‘Organized Editing Guidelines’. The main sections of the wiki page include an overview of the project (manager, contact details, time frame, etc), the purpose and goal, as well as details of the data creation process.

<sup>19</sup> [https://wiki.osmfoundation.org/wiki/Organised\\_Editing\\_Guidelines](https://wiki.osmfoundation.org/wiki/Organised_Editing_Guidelines)

<sup>20</sup> <https://www.hotosm.org/projects/modelling-exposure-through-earth-observation-meteor/>



# METEOR: Ground Data Collection Using Protocols II:

## Dar es Salaam, Tanzania



This includes links to the Tasking Manager projects, imagery used, building statistics before and after the digitisation stage (quantity and quality), as well as before and after maps of the homogeneous zones for the remote digitisation phase. The localised OSM data model derived from the GED4ALL schema, tools used for the data collection, statistics of attribute completeness before and after data collection, statistics of data quantity and quality before and after surveying are also provided through the wiki page.

While the wiki page provides more of an overview to the METEOR project aim, methodology, tools and results, it does not go in depth on how to set up a mapping project and apply the appropriate tools. Outlining the steps in *METEOR Deliverable M4.3: Protocols for Crowd-Sourcing Regional Exposure Data* report is not the most effective way to reach a larger audience, as HOT has seen in the past and are currently developing an open online 'HOT Toolbox' that will host all of the methodologies applied for setting up and executing mapping projects (Figure 22, Table 17). The site includes presentation, videos and the ability to export materials as PDFs. It covers a range of topics spanning from designing and coordinating a mapping project, to data use and analysis. This will continually evolve as the methodologies applied by HOT also change to reflect the lessons learned and the emergence of new better equipped tools are developed.

The screenshot shows the HOT Toolbox website at <https://hotosm.github.io/toolbox/>. The page has a header with the HOT Toolbox logo, a search bar, and a link to 'hotosm.org'. The main content area is titled 'HOT Toolbox' and contains two columns of text. The left column lists categories such as 'Introduction', 'Designing and Coordinating a Mapping Project', '1.1 Mapping Project Workflow', '1.1.1 LEGIT Project Workflow', '1.3 Hardware', '1.4 Safety and Security for Field Operations', '1.5.1 Setting up phones and servers', '1.8 Trainings and Workshops', 'Core Technology', '2.1 OpenStreetMap', '2.1.1 Opening OSM accounts', 'Digitization and Editing', '3.1 Working with the HOT Tasking Manager', '3.2 Working With Imagery', and '3.3 Editing with iD and JOSM'. The right column provides descriptions for these items. At the bottom, there is a note about the Humanitarian OpenStreetMap Team being a 501(c)(3) organization and a 'Contribute' button.

Figure 22: HOT Toolbox

Open Access Training Material	Link
HOT Toolbox	<a href="https://hotosm.github.io/toolbox/">https://hotosm.github.io/toolbox/</a>

Table 17: HOT Toolbox Link



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## 6. Sustainability

A common concern about crowd-sourcing mapping initiatives is their ability to have a sustainable impact after the end of the project cycle. In this case, sustainability can take on multiple meanings, including continuing use and maintenance of the data, continued activity of the local mapping community, or ongoing investment in the project by new participants. Defining at the start of the process, what goals are the most relevant to help the sustainability of the data will more likely achieve them. It is also important to take stock at the end of a project of what was learned during the implementation, in order to guide continued work or new efforts in other locations. Key goals for the sustainability of the METEOR ground data collection was the creation of open access information on the project itself, including the aim, methodology, tools and results through the OSM wiki page, as well as the HOT Toolbox where interested parties can access more detailed steps on how to set up their own mapping projects.

Other aspects of the data collection focused on the sustainability of the project, includes the use of the ‘source=METEOR Dar es Salaam Field Survey 2019’ tag, which allows users to see and reference how the data was created. Reading the open access documents and protocols on the mapping project can work as a guide to support similar mapping activities in the area to enhance the data or to simply keep it up-to-date. Similarly, the ‘fixme=’ tag was also applied to the data collected. This tag was used to describe any mapping errors discovered by the surveyors on the ground. For example, the “north side of the building is attached to a neighbouring wall” or “the building complex is actually 4 separate buildings evenly divided”. This tag was included in the ground data collection model to maximise the improvement of OSM. It has also made it easier for the OSM Tanzania community to continue improving OSM beyond the project period.

Since HOT’s partner OMDTZ is a local organisation based in Dar es Salaam, Tanzania, they plan to monitor the data uploaded to OSM. They work on several OSM-related projects throughout the year, and will continue to do so in the future as they are an integral part of the OSM community across all of Tanzania, and act as a point of contact and reference for related mapping projects in the country. MapCampaigner was a new tool introduced to OMDTZ on the METEOR project, which they hope to continue to utilise going forward as it helps the teams to monitor specific data in OSM.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## 7. Conclusion

Overall the ground data collection effort was successful, providing OSM data at the building level for 2951 randomly selected structures within the 6 homogeneous development type zones in Dar es Salaam and the neighbouring region of Pwani. The OSM data will then be used by ImageCat to aggregate and smear for the rest of Tanzania to create exposure datasets at higher levels.

One of the qualities that makes the METEOR Project unique is its effort to calibrate the Earth Observation exposure data with testing and validation via on the ground data collection in Nepal and Tanzania. In many ways, the data collection in Dar es Salaam has followed closely along the same lines as in Kathmandu, by using the Protocols for Crowdsourcing Exposure Data (METEOR Deliverable M4.3). At the same time, everything from the data model to the data collection has been tailored to the context of Tanzania, ensuring that the outputs are fit-for-purpose. However, no data collection project is ever perfect, and this one is no exception. Building on the lessons from Nepal, some of the work in Tanzania was improved, while new insights were learned. Moving forward, these will be helpful for communities to improve the way data is created in OSM to meet its intended purpose.

Among the takeaways HOT would like to share is the idea that certain building classifications may differ country to country, for instance ‘informal settlement’. In Nepal, informal settlements often refer to slums, and their physical form (structure of the building) as well as vulnerability is different than in Tanzania. In Tanzania, and Dar es Salaam specifically, it is known that up to 70% of the city is informal, which indicates a high level of socioeconomic vulnerability as many live in unplanned, untitled plots. Yet in terms of building structure, these are usually built with concrete blocks and corrugated metal roofs, reducing their vulnerability when compared with actual slums.

Another takeaway that HOT would like to emphasise is the finding that the profile of rural areas is significantly different than that of Dar es Salaam. Thus, extrapolating from the city to wider Tanzania may produce a less representative sample. Buildings that are made of mud are considered to be more vulnerable, and these are found primarily in remote areas outside of cities, examples being the Arusha or Manyara Regions. Though these buildings are often more vulnerable, the risk depends on where they exist and whether they are exposed to potential natural hazards. The Tanzania team had even planned on doing supplementary fieldwork to enhance the model with additional exposure sets, but with the onset of the COVID-19 pandemic, HOT confirmed with ImageCat that the data already provided would be sufficient to the modelling.

It should also be kept in mind that the data collection methods applied for the METEOR project are by no means an exhaustive and comprehensive reference for all the processes and tools that can be used for collecting data in OSM. No one solution fits all, and an approach should be formulated based on the data collection focus, the local community and the targeted area of interest. Approaches that work in one country might not work in another, and newer more appropriate tools for data collection may become available. However, the completion of ground data collection in Dar es Salaam does give evidence that the established protocols and tools will be relevant for other ODA nations.



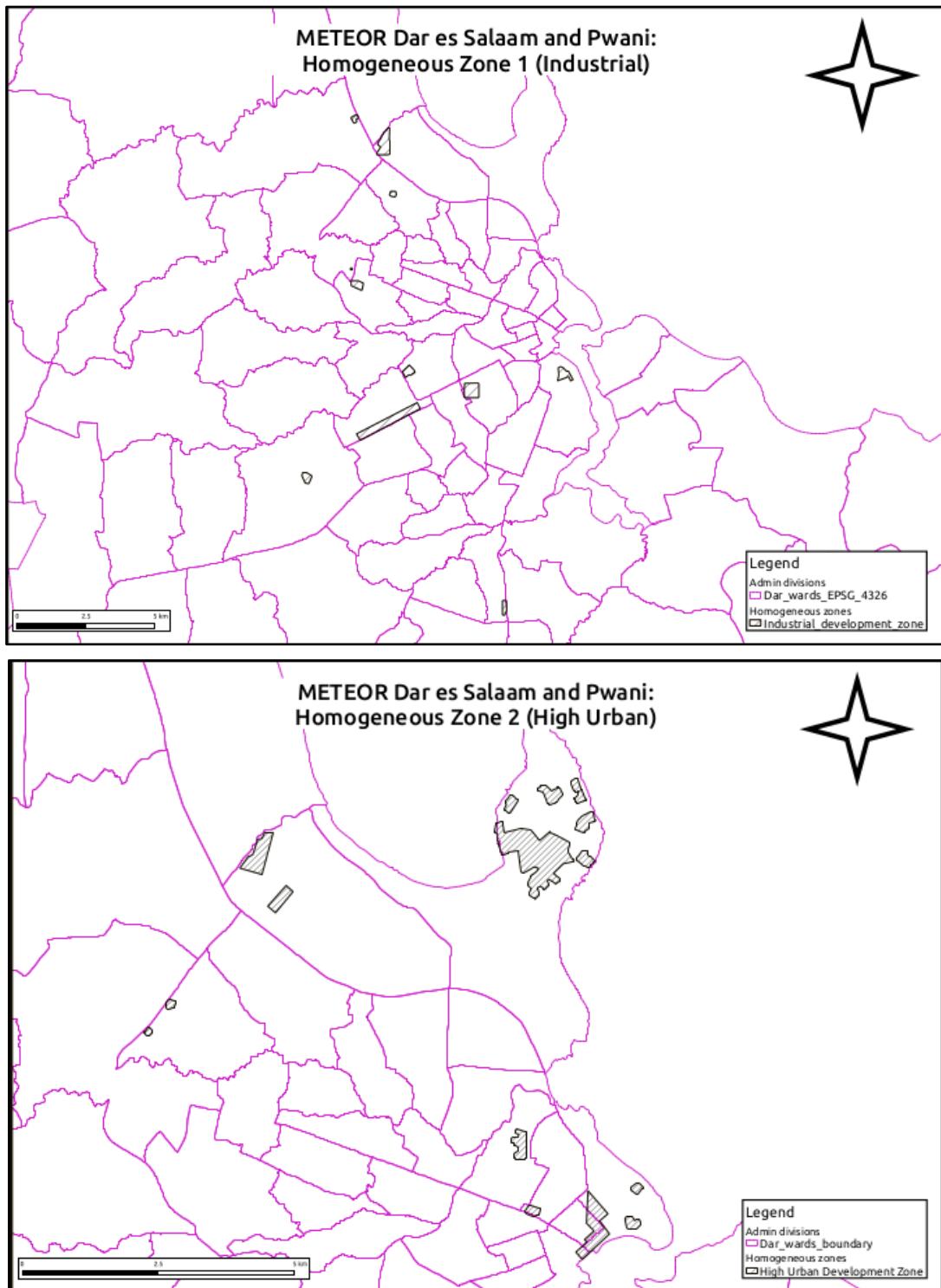
*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



HOT has been conducting OSM data collection for over seven years, and the methodology is continuously evolving as lessons are learned, and processes become more streamlined. Many of our software development projects are focused on assisting the creation, extraction and analysis of OSM data. Helping provide the tools to empower local citizens with the capacity to place themselves on the map, is one of the biggest contributions we can make towards the OSM community as we work towards the Sustainable Development Goals (SDGs) together.

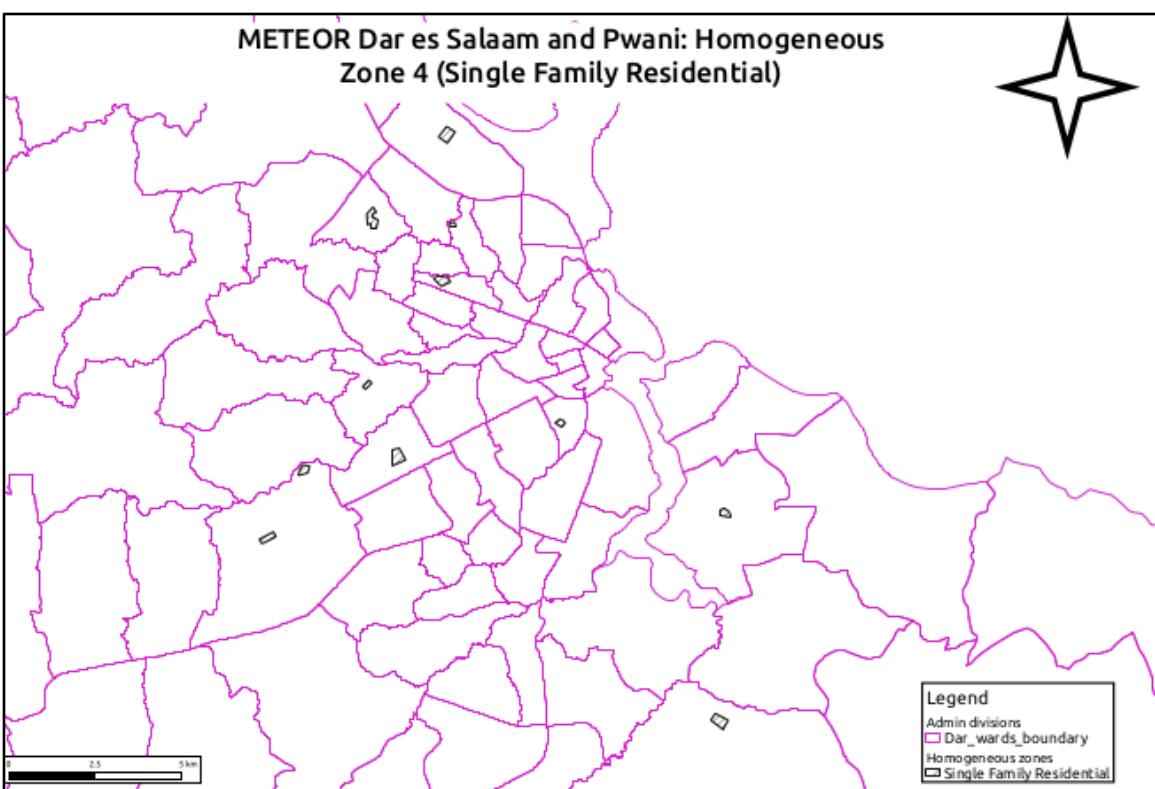
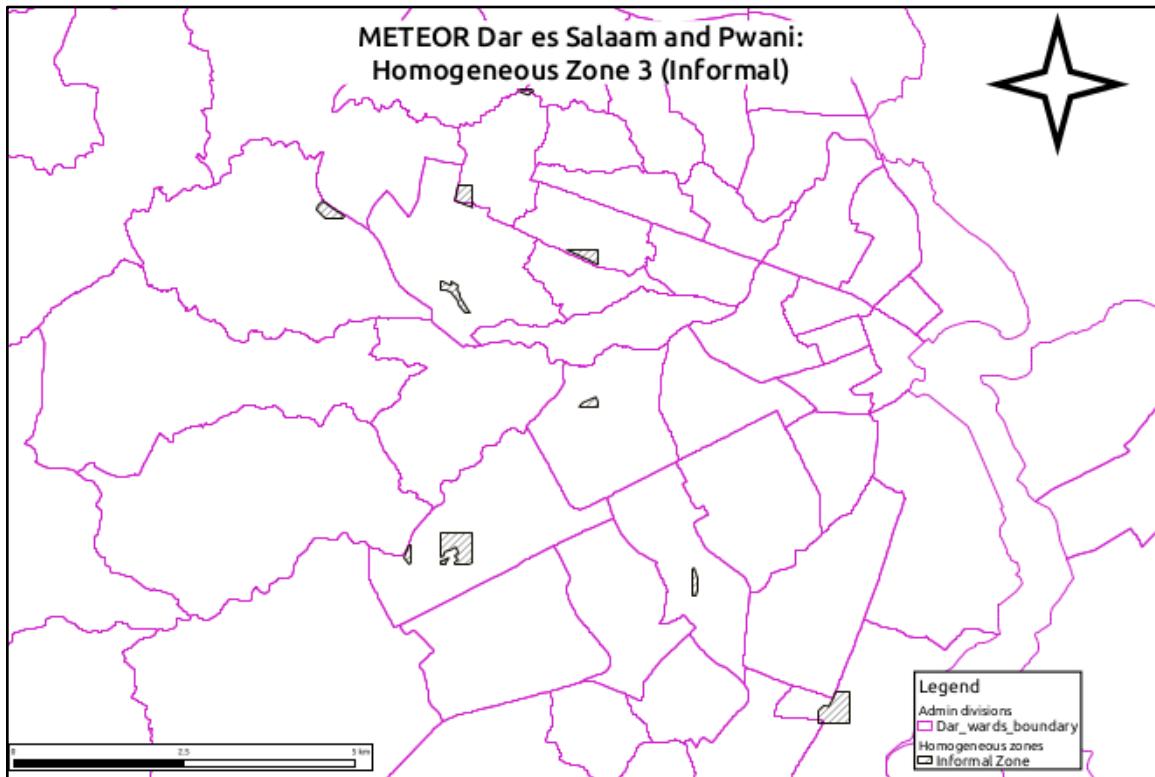
Accomplishing the key aims of the METEOR Project still necessitates further engagement so that the exposure data can connect with resilience efforts, risk assessments, and Disaster Risk Management plans. As the project shifts from the data collection and modelling phase towards knowledge sharing, sustainability, and capacity-building, HOT and our local partner OMDTZ remain ready to facilitate and achieve the broader goals of the project.

## Appendix 1: Homogeneous Zone Maps



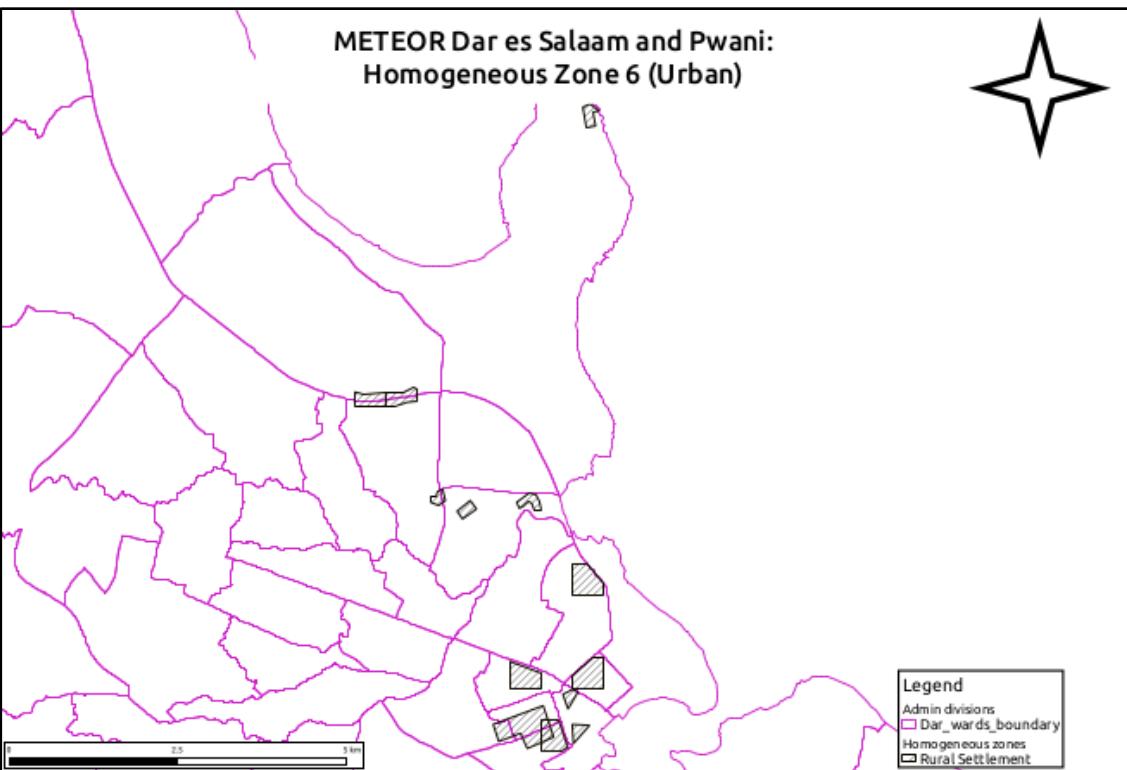
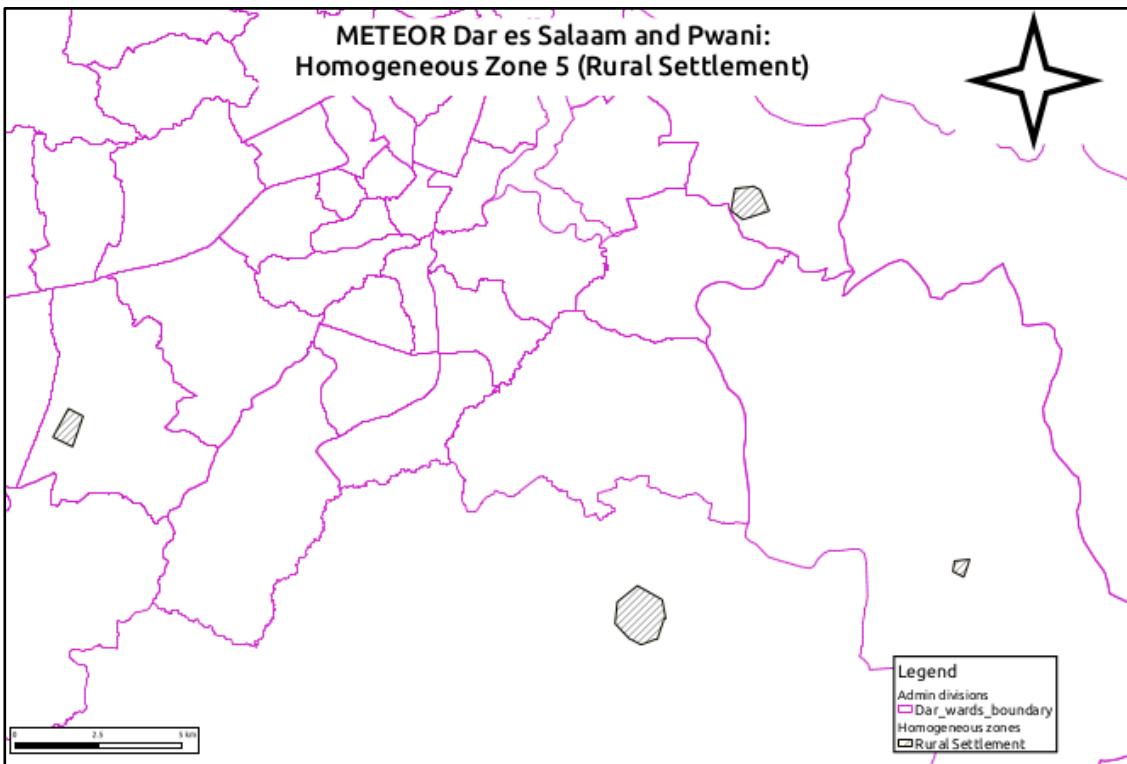


*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*





*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*





*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## Appendix 2: Attribute Value Breakdown

No.	Building	Count	Percentage
<b>1</b>	residential	1578	53.5%
<b>2</b>	mixed_use	343	11.6%
<b>3</b>	commercial	479	16%
<b>4</b>	government	29	1%
<b>5</b>	educational	12	-
<b>6</b>	industrial	364	12%
<b>7</b>	agriculture	2	-
<b>8</b>	public	120	4%
<b>9</b>	construction	16	1%
<b>10</b>	toilet	1	-
<b>11</b>	assembly	6	-
<b>12</b>	health	1	-

No.	Building Neighbor	Count	Percentage
<b>1</b>	attached	967	32.8%
<b>2</b>	free_standing	1984	67.2%

No.	Building Age	Count	Percentage
<b>1</b>	post_2000	1886	63.9%
<b>2</b>	pre_2000	1065	36.1%

No.	Building Condition	Count	Percentage
<b>1</b>	good	1452	49.2%
<b>2</b>	average	1224	41.5%
<b>3</b>	poor	275	9.3%



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



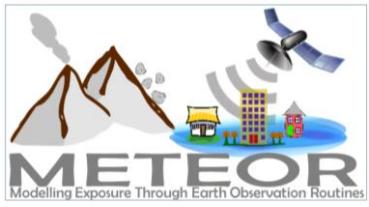
No.	Geological Site	Count	Percentage
<b>1</b>	flat_land	2485	84.2%
<b>2</b>	slopy_land	331	11.2%
<b>3</b>	river_bank	17	1%
<b>4</b>	flood_prone_area	118	4%



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



No.	Building Levels	Count	Percentage
<b>1</b>	1	2596	88%
<b>2</b>	2	136	4.6%
<b>3</b>	3	57	1.9%
<b>4</b>	4	41	1.4%
<b>5</b>	5	21	1%
<b>6</b>	6	29	1%
<b>7</b>	7	13	-
<b>8</b>	8	17	1%
<b>9</b>	9	4	-
<b>10</b>	10	6	-
<b>11</b>	11	6	-
<b>12</b>	12	7	-
<b>13</b>	13	3	-
<b>14</b>	14	4	-
<b>15</b>	15	2	-
<b>16</b>	19	1	-
<b>17</b>	20	1	-
<b>18</b>	21	1	-
<b>19</b>	23	1	-
<b>20</b>	25	1	-
<b>21</b>	26	1	-
<b>22</b>	28	1	-
<b>23</b>	31	1	-
<b>24</b>	35	1	-

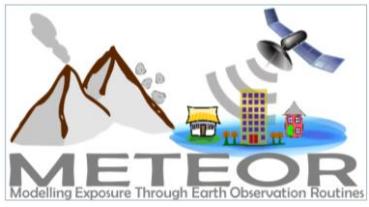


*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



No.	Building Levels Underground	Count	Percentage
<b>1</b>	none	2923	99%
<b>2</b>	1	27	1%
<b>3</b>	2	1	-

No.	Capacity Person	Count	Percentage
<b>1</b>	10 to 19	687	23%
<b>2</b>	20 to 49	303	10%
<b>3</b>	5 to 9	924	31%
<b>4</b>	50 to 100	146	5%
<b>5</b>	500 and above	2	-
<b>6</b>	300 to 500	11	-
<b>7</b>	200 to 300	58	2%
<b>8</b>	100 to 200	277	9%
<b>9</b>	Less than 5	542	18%
<b>10</b>	Less than 9	1	-



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



No.	Roof Material	Count	Percentage
<b>1</b>	concrete	375	12.7%
<b>2</b>	cgi	2095	71%
<b>3</b>	mixed	15	1%
<b>4</b>	wood	1	-
<b>5</b>	unknown	60	2%
<b>6</b>	slate	1	-
<b>7</b>	composite_tiles	40	1%
<b>8</b>	thatch	11	-
<b>9</b>	earth	1	-
<b>10</b>	fabric	20	1%
<b>11</b>	metal	165	5.6%
<b>12</b>	clay_tiles	140	4.7%
<b>13</b>	stone	1	-
<b>14</b>	Industrial_troughed sheet	25	1%
<b>155</b>	glass	1	-

No.	Roof Shape	Count	Percentage
<b>1</b>	flat	486	16.5%
<b>2</b>	monopitch	607	20.6%
<b>3</b>	double_pitch	312	10.6%
<b>4</b>	complex_regular	112	3.8%
<b>5</b>	pitched	1350	45.7%
<b>6</b>	unknown	56	1.9%
<b>7</b>	complex_irregular	9	-
<b>8</b>	curved	13	-
<b>9</b>	sawtooth	3	-
<b>10</b>	domed	3	-



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



## Appendix 3: Sampling Procedure for Buildings

### Goal

The goal is a statistically valid random and systematic procedure to select a subset of buildings that generates clusters (groups of buildings in close proximity) so that field mapping is efficient.

If the sampling rate is so high as to generate a large number of overlaps (for example, if the desired sampling rate is 400 out of 434 buildings, as is the case in the Industrial zone of Dar es Salaam in our current project), clustering does not make sense; most buildings are part of more than one cluster.

Therefore, we have applied the following rule: if the desired sample is more than 1/3 of the total buildings, we select a random subset of those buildings without clustering. If the desired sample is less than 1/3 of the total structures, we use the clustered random sampling procedure described below.

### Procedures

#### Non-clustered sampling

This procedure simply generates a random sample of the structures without taking proximity into account. We use the randomly generated ranking number to order the buildings and select the required number from the ordered ranking.

The random ranking number is generated using the QGIS Field Calculator (under the hood this is the Python random number generator). We add a new column to the attribute table of the building layer, in which the function `rand({min}, {max})` generates a random integer. We use zero and one billion as our min and max—this large number is to avoid any potential duplicates—and select the appropriate number of buildings (in our current project 400) with the *lowest* random number.

- Use the QGIS Field Calculator, create a new column in the building layer with *Output field name* `random`, choose *Whole number (integer)* as the type, and the expression `rand (0, 1000000000)` to populate the new column. Save the changes and exit editing mode.
- In the Attribute Table of the building layer, click the column header of the Random column. This will order the entire dataset by the random number, creating a sequential ranking. Select the first  $n$  elements (in our case 400), and in the Layers panel select `Export -> Save Selected Features As` and create a new GeoPackage with an appropriate name. This GeoPackage will be your random sample.

#### Clustered sampling

First, prepare your files and data. Keep things organized—name your files, put them in appropriately named folders, and name your layers in QGIS!

- Load up the zones and buildings as layers in QGIS. Split the zones into multiple files using `Vector -> Data Management Tools -> Split Vector Layer` selecting `typeclean` as the Unique ID Field. That will create a new shapefile for each type of area. Load the resulting shapefiles into the project in a single folder.
  - NOTE: the QGIS Split Vector Layer tool only outputs shapefiles, not GeoPackages. This is OK for the zones in this procedure, but should not be used for buildings (or any other OpenStreetMap layer from which we need the attributes, as the Shapefile



## METEOR: Ground Data Collection Using Protocols II: Dar es Salaam, Tanzania



format truncates all attribute table headers to 10 characters, which results in a lot of mysterious columns).

Now we need to create random points within the zones. Since each zone may be multiple polygons, we cannot simply choose a number of points, as we need a consistent density of points in all polygons within one zone. To do this, we calculate a required density of points per square meter for each zone and apply this across all the polygons in that zone. First, we calculate the areas of all zones.

- Go to each new layer, open the Attribute Table, and use the Field Calculator to create new field called *area*, select type *decimal number (real)* selecting Geometry -> \$area.
- Run the Vector -> Analysis Tools -> Basic Statistics for Fields tool for each layer. we need the SUM of the area column, so choose area as the *Field to calculate statistics on*, and copy the SUM from the resulting log (this will be an area in square meters) into another document. Keep this handy.

Now we can generate random points in the zone polygons to use as the centres for our clusters. In the example here, we wish to select 400 buildings in clusters of 10. However, if we create only 40 random points, some will be sufficiently close to one another that their clusters overlap (some buildings will be within the nearest 10 of more than one cluster centre). Therefore, we will create more points than seem necessary to select more buildings than we need (overshoot rather than undershoot), and later remove the excess buildings in a systematic fashion. Let's create 60 random points per zone rather than 40 (the expression below uses 60 as the number of points to divide by the number of square meters in each zone).

- Go to Vector -> Research Tools -> Random Points in Polygon. Choose *Points Density* as the sampling strategy, and use the expression  $60 / \{SUM\}$  where SUM is the number we got from the Basic Statistics for Fields (use that number, don't type out the word SUM). This expression gives us a very tiny number, something like 0.00005, that represents the number of points per square meter. when we press *Run*, the chosen number of points (in this case 60) will be created at a consistent density across all polygons in the zone.

Before going further, we need to add a random number attribute into the random point layer, as well as to the building layer (just as we do for non-clustered sampling). This random ranking attribute is used to remove the buildings exceeding the number needed (in our case 400). To retain the clustering, we will first remove the buildings associated with the highest-scoring random points. Since this probably won't get us to our exact target (since in our case each point has around 10 buildings associated with it, so there will be zero to nine excess buildings left after removing the excess clusters), we'll trim off the remaining excess buildings using the random scores of the buildings themselves.

- Use the QGIS Field Calculator, create a new column in the building layer with *Output field name* random, choose *Whole number (integer)* as the type, and the expression `rand(0, 1000000000)` to populate the new column.
- Then do the same thing with the random points layer.

Now we wish to select the 10 closest buildings around each random point. The QGIS Distance Matrix tool will do this, but it only works on points, not polygons. So, we will generate centroids for the building polygons.



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



- Create centroid points for the building layer using Vector -> Geometry Tools -> Centroids. The resulting layer should have all the same attributes as the buildings, including the OSM ID (probably a column called osm\_id).
- Now go to Vector -> Analysis Tools -> Distance Matrix. Your input layer is Random Points, and your target layer is Centroids. The "Target unique ID field" should be the osm ID. Output matrix can be linear and in the field for "Use only the nearest (k) target points," put 10. This should give us a new layer called Distance Matrix that contains only the 10 nearest building centroids from each random point.

Some building centroids may be among the 10 nearest to more than one random point, in which case the total number of points in the Distance Matrix will not be 10X the number of random points. This is why we used more points than 1/10 the number of buildings needed (in this case 60 rather than 40).

Now we need to link the chosen centroids back to the original building layer. While we're at it, we'll add the random ranking number from the random points to the Attribute Table of the building layer—we'll use that to remove the excess structures first by removing those from the excess random points (and therefore clusters), then by trimming off the remaining excess buildings.

- Open the Properties of the buildings, go to Joins, and push the green plus button on the bottom to add a new join. The Join layer is Distance matrix, the join field is TargetID, and the Target field is osm\_id (or whatever the column with the OSM ID is called). Custom Field Name Prefix should be DM\_ (or something equally short), and Joined Fields should be InputID only (this will create a column in the building layer with the random ranking number of the random point it is associated with).
- Open the Attribute Table of the building layer, and look at the last column on the right (called DM\_InputID if we followed these instructions carefully). It should be blank for some buildings (the ones that don't fall into the selected clusters), and contain a number from zero to one billion in a few rows. Sort the column so that all of the buildings with a DM\_InputID are on top, select them all, close the Attribute Table, and right-click on the building layer.
- Choose Export -> Save Selected Features As and save this as a new GeoPackage -- it will contain only the buildings within the clusters for that particular sampling zone.
- However, this file still contains too many buildings! So we sort again in ascending order, first by the Random column, then by the DM\_InputID column. Now we can select the first  $n^*$  buildings (for us 400) and delete the rest (by entering editing mode, pressing the \*Invert selection button, and pressing the delete button. Because of the sort order (first by the building random number, then by the cluster random number which is in the DM\_InputID column), this will retain the clustering by trimming first the buildings belonging to the excess clusters, then the excess buildings in the final cluster (so the last cluster may have as few as 1 building in it, but all of the others will include the nearest 10 to their random point).



*METEOR: Ground Data  
Collection Using  
Protocols II:  
Dar es Salaam, Tanzania*



Appendix 4: Online Interactive Map, Dar es Salaam METEOR Buildings and Ward Boundaries



uMap: [https://umap.openstreetmap.fr/en/map/meteor\\_tz\\_buildings\\_417725#17-6.81326/39.28848](https://umap.openstreetmap.fr/en/map/meteor_tz_buildings_417725#17-6.81326/39.28848)

\*Note: Images were previously linked for data validation, currently are no longer being hosted. Images taken during data collection are available, and have been provided as “TZ METEOR Data with Photos” including a GeoPackage of the data with a field for image ID, along with a folder containing all images.