METEOR: Import Existing Data into OSM

Report Number: WP4.1/P

31 August 2018

Contributors:

M. O'Hara

Document Verification

Project	METEOR: Modelling Exposure Through Earth Observation Routines
Report Title	METEOR: Import existing data into OSM
Related Milestone	MW4.1
Reference as	M. O'Hara (2018) METEOR: Import existing data into OSM. Report WP4.1/P
Release Type	Public / ~~Confidential~~ / ~~Confidential with Embargo Period~~

Prepared by: Contributors		
Name(s): M. O'Hara	Signature(s): DIGITAL SIGNATURE	Date(s): 31 August 2018
Approved by: Project Manager		
Name: K. Smith	Signature: Kay Smith	Date: 28 September 2018
Approved by: UKSA IPP Project Officer		
Name:	Signature:	Date:

Date	Version	Alterations	Editor

Contents

Figures

Tables

Acronyms

BGS	-	British Geological Survey: An organisation providing expert advice in all areas of geosciences to the UK government and internationally
DMD	-	Disaster Management Department, Prime Minister's Office of Tanzania, focused on disaster risk
DRM	-	Disaster Risk Management
DRR	-	Disaster Risk Reduction
EO	-	Earth Observation
FATHOM	-	Provides innovative flood modelling and analytics, based on extensive flood risk research
GEM	-	Global Earthquake Model, non-profit organisation focused on the pursuit of earthquake resilience worldwide
HOT	-	Humanitarian OpenStreetMap Team, a global non-profit organisation that uses collaborative technology to create OSM maps for areas affected by disasters
InaSAFE	-	Free open software that produces realistic natural hazard impact scenarios for better planning, preparedness and response activities
ImageCat	-	International risk management innovation company supporting the global risk and catastrophe management needs of the insurance industry, governments and NGOs
IPP	-	International Partnership Programme of UK Space Agency
KLL	-	Kathmandu Living Labs
Mapathon	-	Coordinated mapping event where individuals learn about OSM, and make edits to the map through remote digitising of satellite imagery
METEOR	-	Modelling Exposure Through Earth Observation Routines
NGO	-	Non-Governmental Organisation
NSET	-	National Society for Earthquake Technology, non-governmental organisation working on reducing earthquake risk in Nepal and abroad
ODA	-	Official Development Assistance
ODbl	-	Open Database License
OPM	-	Oxford Policy Management, organisation focused on sustainable project design and implementation for reducing social and economic disadvantage in low-income countries
OSM	-	OpenStreetMap, a collaborative project to create a free and open editable map database of the world
QGIS	-	Free open source desktop geographic information system
Ramani Huria	-	A community mapping project based in Dar Es Salaam, Tanzania
UKSA	-	United Kingdom Space Agency
WP	-	Work Package

1. 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, 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. METEOR Project Overview

At present, there is a poor understanding of population exposure in some 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. Providing new consistent data to governments, town planners and insurance providers will promote welfare and economic development in these countries and better enable them to respond to the hazards when they do occur.

METEOR is funded through the second iterations of the UK Space Agency’s International Partnership Programme, 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.

1.3. METEOR 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 and Dar es Salaam will be used to compare and validate the EO generated exposure datasets. Objectives of the project look to deliver exposure data or 47 of the least developed ODA countries, including Nepal and Tanzania. Create open protocols to develop critical exposure information from EO data 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 (WP) that make up the METEOR project. These are lead 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.

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.
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.

Table 2: Overview of METEOR Work Packages

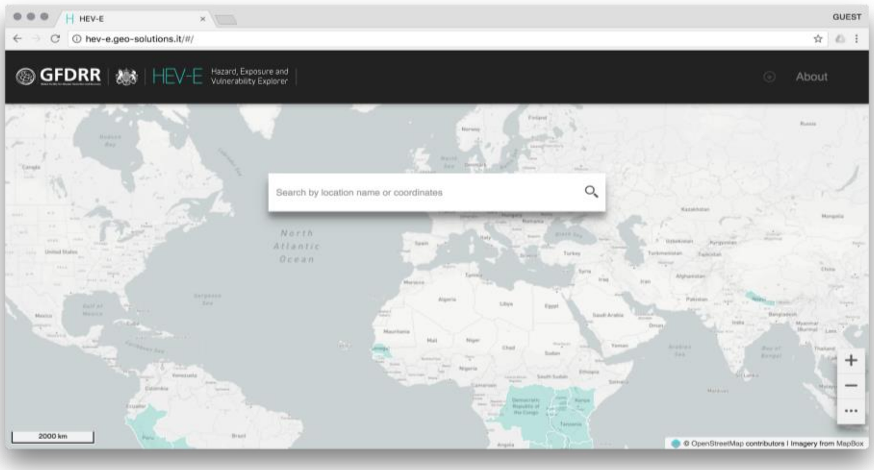


Figure 1: GFDRR Challenge Fund HEV-E Data Platform

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).

Work Package	Title	Overview
WP.4.1	Import Existing Data into OSM	Review and assess the suitability of existing open datasets for import into OSM for Kathmandu and Dar es Salaam.
WP.4.2	EO Mapping of Exposure Data	Remote mapping of building footprints and road networks in OSM for Kathmandu and Dar es Salaam.
WP.4.3	Protocols for Crowdsourcing Exposure Data	Draft protocols for the crowdsourcing of exposure data in OSM, covering data imports, remote mapping and ground data collection.
WP.4.4	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
WP.4.5	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
WP.4.6	Final Report	Deliver a final version of WP.4.3 along with the resulting data from WP.4.4. and WP.4.5.

Table 3: Overview of HOT Work Package

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)[1], 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[2], 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 (Open Data Commons, 2018):

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

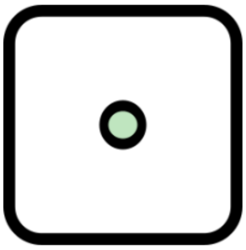
[1] https://opendatacommons.org/licenses/odbl/

[2] https://okfn.org/

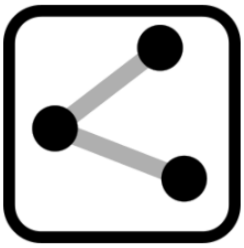
2.3. . Data Structure

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 by a way with the tags `building=yes` and `amenity=post_office`, to help identify the purpose of the building.

3. Importing Data

3.1. Overview

It is extremely important that the whole import process is planned and executed with more care and sensitivity than other edits in OSM, as it can have significant impacts on both existing data and the local mapping community. Therefore a methodology must be developed and strictly followed for each of the datasets that will be imported. There are Import Guidelines[3] created by the OSM community, which are outlined below and must be taken into account while developing the import methodology:

A. Prerequisites
B. Community Buy-in
C. License Approval
D. Documentation
E. Import Review
F. Uploading

3.1.1. Prerequisites

The first step of the methodology aims to cover the background knowledge leading up to the import, which includes having a basic understanding of OSM, the data structure and how it works. It is also important that the common import errors are properly reviewed (Table 4), and finally to identify the dataset that will be imported:

- ❖ Understand OSM basics
- ❖ Review import errors
- ❖ Identify data for import

General Import Errors
Not all data should be imported
Not all external data should be trusted
Imports might reduce mapper responsibility of data
Node tag duplication
Verifiability
Stale data
Data formats might not match
Fixing later can be difficult
Cross-referencing databases
Data density problems for editors

Table 4: Common Data Import Errors

[3] http://wiki.openstreetmap.org/wiki/Import/Guidelines

3.1.2. Community Buy-in

Having the import accepted by the OSM community is vital to process. Without the community buy-in, the import cannot take place. The plan for importing the data needs to be discussed with the community, which will initiate the process:

- Contact the local community to see if there is interest in importing the data
- Discuss the import plan with the OSM community
- Discuss the suitability of each layer for importing
- Complex large-scale imports should be reviewed by experienced OSM mappers
- Do not import the data without local community buy-in

Contact	Notes
imports@openstreetmap.org	Email Address
talk-(country)@openstreetmap.org	Email Address
https://lists.openstreetmap.org/	Mailing Lists
https://wiki.openstreetmap.org/wiki/User_groups	User Group
https://wiki.openstreetmap.org/wiki/Local_chapter	Local Chapters

Table 5: OSM Community Contacts

3.1.3. License Approval

Proper permissions and licenses to use the data in OSM must also be obtained. If the license of the data is not compatible with the ODbL license, then the data cannot be used. Some data policies are almost open, but may have conflicts with issues like prohibitions on commercial use, or the requirement for attribution. Getting permission to use data, even if the existing license might seem prohibitive, is as simple as asking the appropriate authority if they are willing to comply with the ODbL terms.

License	ODbL
PD	Compatible
CC0	Compatible
ODBL v1.0	Compatible
UK Open Government Licence v3.0	Compatible
CC-BY v1.0-v4.0	Compatible

Table 6: Examples of ODbL Compatible Licenses

3.1.4. Documentation

There are three documentation requirements that must be carried out as part of the importing process, which are outlined below:

- ❖ Register the permissions and project to the Import/Catalogue[4] on the OSM Wiki
- ❖ Write a plan for the import on the OSM Wiki by adding a page outlining the details
- ❖ Add acknowledgement of the list to the Contributors[5] if this is required by data owners

Import Plan Outline			
Heading	**SubHeading**	**Sub SubHeading**	**Notes**
Goals			Identify goals of import
Schedule			List project timeframe
Import Data	Background	Data Source Site	Links to sources
		Data License	
		License Type	
		Permission Link	
		OSM Attribution	
		ODbL Compliant	
	OSM Data Files		Link to source data files
	Import Type		Identify if this is a one-time or recurring import, etc
Data Preparation	Data Reduction & Simplification		Describe plans, if any, to reduce data that will be imported
	Tagging Plans		Describe plan for mapping source attributes to OSM tags
	Changeset Tags		Describe how changeset tags will be used in import
	Data Transformation		Describe transformations and tools used

[4] https://wiki.openstreetmap.org/wiki/Import/Catalogue
[5] https://wiki.openstreetmap.org/wiki/Contributors

	Data Transformation Results		Link to OSM XML files
Data Merge Workflow	Team Approach		Describe team
	References		List all factors that will be evaluated in the import
	Workflow		Detail steps taken during import
	Conflation		Identify conflation approach
Q&A			Add a questions and answer plan

Table 7: OSM Wiki Import Plan Template

The page created on the OSM Wiki for the import of education facilities data from the Uganda Bureau of Statistics (UBOS)[6] (Figure 2), and the Mapbox[7] GitHub page for the import of county building data in Los Angeles (Figure 3), are great examples of how to document the intended workflow:

6 https://www.ubos.org/

7 https://www.mapbox.com/

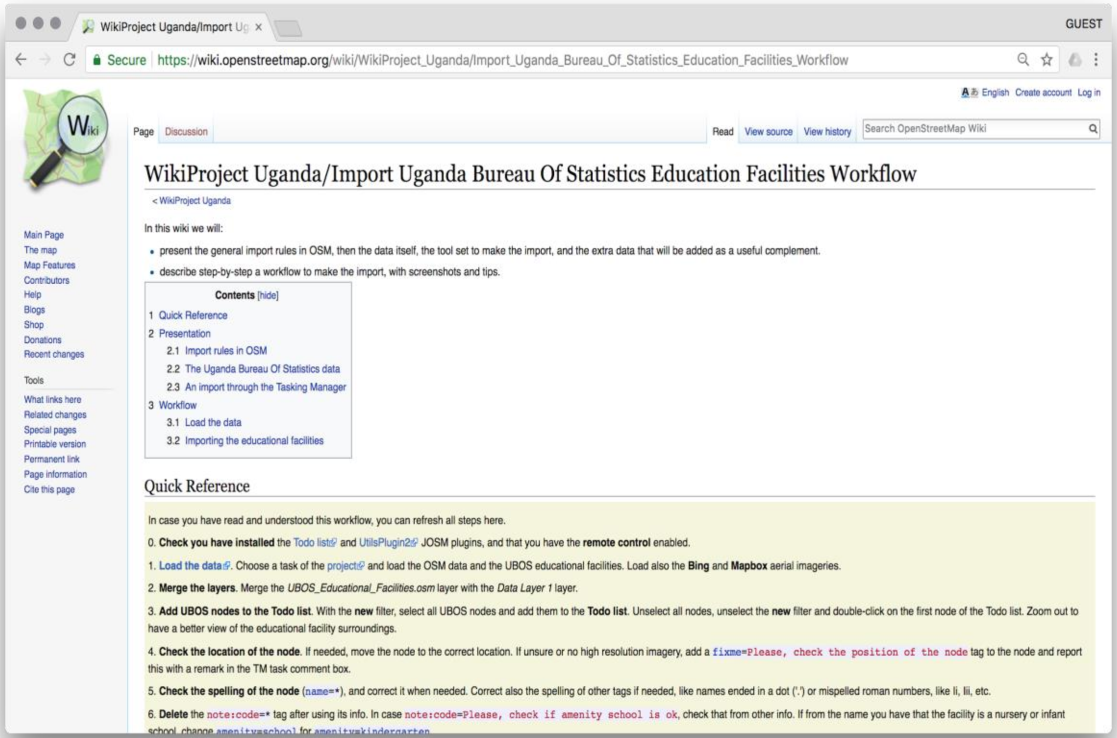


Figure 2: UBOS Education Facilities Import Workflow

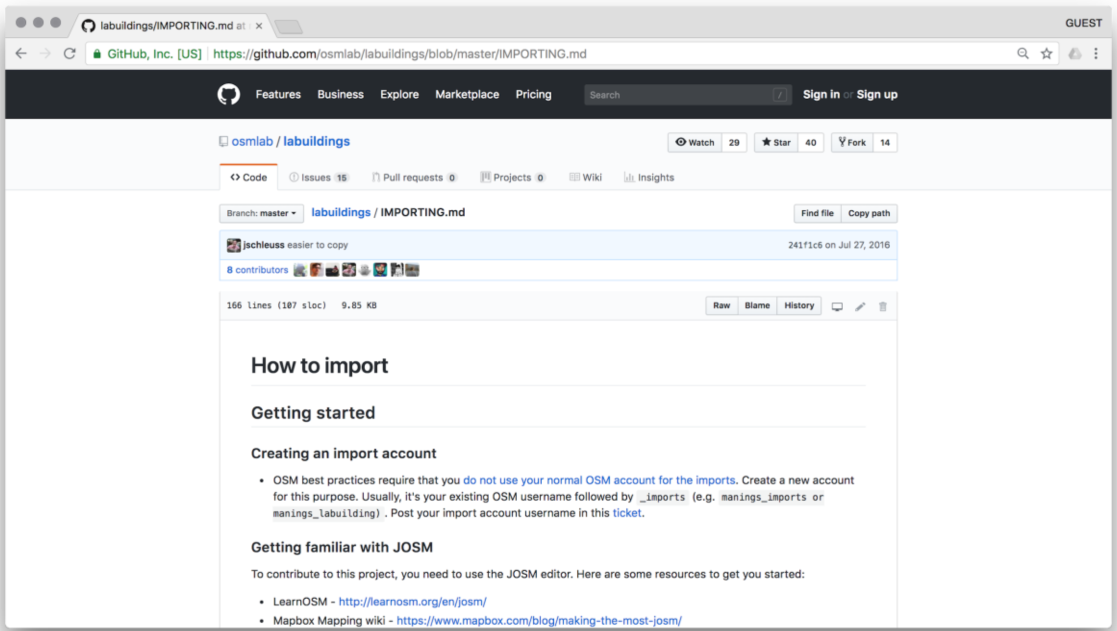


Figure 3: Los Angeles County Building Import Workflow

3.1.5. Import Review

The import must be reviewed by the appropriate authoritative. Outlined below are the steps that are required to ensure the review is carried out properly:

- Subscribe to the Imports Mailing List[8]
- Send a review of the Import Group[9]
- Prepare the data and make it available for review

3.1.6. Uploading

This section is focused on the actual importing of the data itself. The steps outlined below will help to guide the process to ensure that it runs smoothly:

- Follow the outlined plan
- Track the progress
- Provide updates to the community
- Let the community know when it is complete
- Use a dedicated OSM user account

Using a dedicated OSM user account specifically for imports is extremely important. Failure to do so can lead to the blocking of the account and the import process by the OSM Data Working Group (DWG)[10]. It is advised that a new account is registered using the following syntax:

"osm username"_import

[8] https://lists.openstreetmap.org/listinfo/imports

[9] imports@openstreetmap.org

[10] https://wiki.osmfoundation.org/wiki/Data_Working_Group

4. Focus Area

4.1. Nepal and Tanzania

Nepal and Tanzania were selected as case study countries for the METEOR project to assist with the development of innovative EO technologies, which is aimed at improving the understanding of exposure data. They were chosen due to the differences in the type of hazards and exposure present in both countries, with the primary natural hazards occurring in Nepal being earthquakes and landslides, while flooding makes up 95% of the natural hazard mortality in Tanzania (Guha-Sapir et al, 2016).

The scale of the development challenge in both Nepal and Tanzania is significant, for example the 2015 Gorkha Earthquake in Nepal was a major disaster with over 9,000 casualties, 22,300 injuries and the lives of 8 million people impacted. While Tanzania is one of the poorest countries in the world, with a gross domestic product (GDP) estimated at $43.8 billion, or $86.4 billion on a purchasing power parity (PPP) basis in 2014 (IMF, 2014). The work in Nepal and Tanzania will have significant application in pre-positioning, emergency response and prioritizing mitigation activities such as land use planning, land acquisition programmes, and building codes.

4.2. Kathmandu

Kathmandu, the capital of Nepal has been chosen to help verify the accuracy and ability of applying the development classification pattern to produce exposure datasets for the other 46 ODA countries. The population of Nepal reached 29,619,502 people in 2018, with 14,691,371 male (49.6%) and 14,928,131 female (50.4%), (CountryMeters, 2018). The city stands at an elevation of approximately 1,400 metres above sea level and covers an area of roughly 49.45 km squared. Kathmandu and adjacent cities are composed of neighbourhoods, which are utilised quite extensively and more familiar among locals. However, administratively the city is divided into 35 wards.

Kathmandu was chosen as the city to ground truth the exposure dataset created for Nepal due to the diversity of infrastructure and building types, which should provide a sufficient sample of information for validation. Also key to the decision is the close relationship HOT has with Kathmandu Living Labs (KLL)[11], a non-profit organisation focusing on mobile technology and mapping for humanitarian aims.

KLL has worked with the National Reconstruction Authority (NRA) and The Central Bureau of Statistics (CBS) of Nepal, as well as conducted a small scale data collection in collaboration with the National Society for Earthquake Technology (NSET)[12] of Nepal, mapping 10,000 buildings in Bharatpur (Republica, 2016). Following the 7.8 magnitude earthquake that struck in 2015, KLL were also key to

[11] http://www.kathmandulivinglabs.org/

[12] http://www.nset.org.np/

the successful global crowdsourced mapping effort of road networks and post disaster conditions in OSM, so that humanitarian aid teams on the ground could act quickly.

This pattern can be seen in the graphs below, with a lot of the Kathmandu data in OSM added in 2013, with a large spike in mapping activity following the earthquake in 2015 (Figure 4). There are currently 183,373 buildings mapped in OSM for the city of Kathmandu, 3,253 kilometers of roads and an estimated 7,079 amenities (Table 8). These statistics were obtained with the use of OSM Analytics online on August 3, 2018. Features determined as amenities are objects in OSM using an amenity tag as a key, so it should be kept in mind that there may be additional features present in the database that have not been correctly tagged and are therefore not counted.

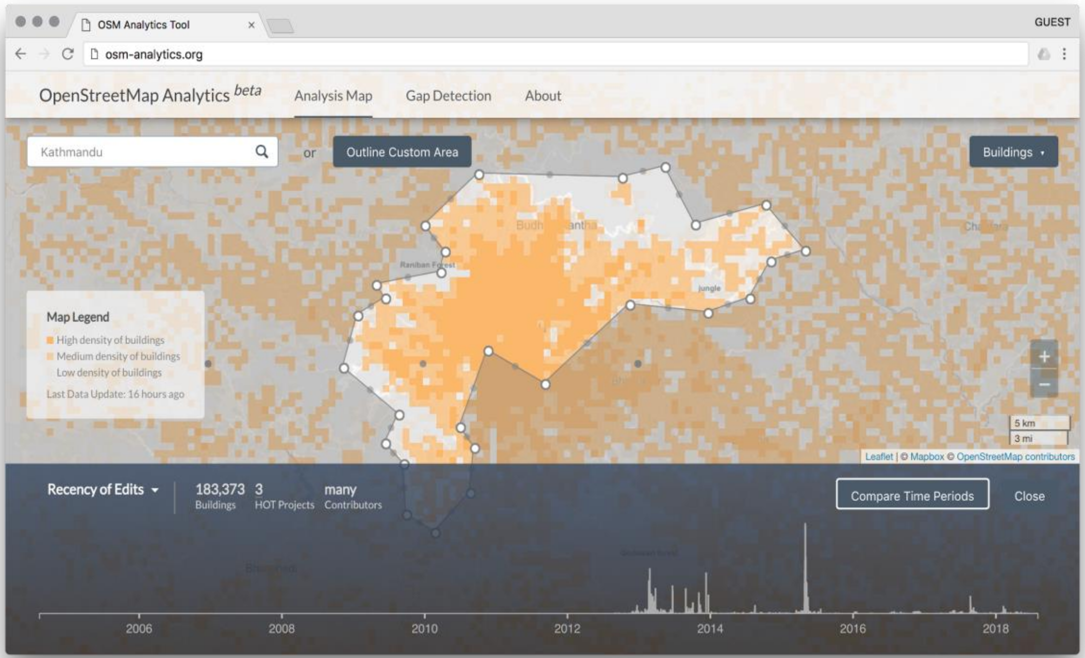


Figure 4: Kathmandu OSM Building Count

Kathmandu	
OSM	**Statistics**
Buildings	183,373
Roads	3,253
Amenities	7,079

Table 8: Kathmandu OSM Statistics

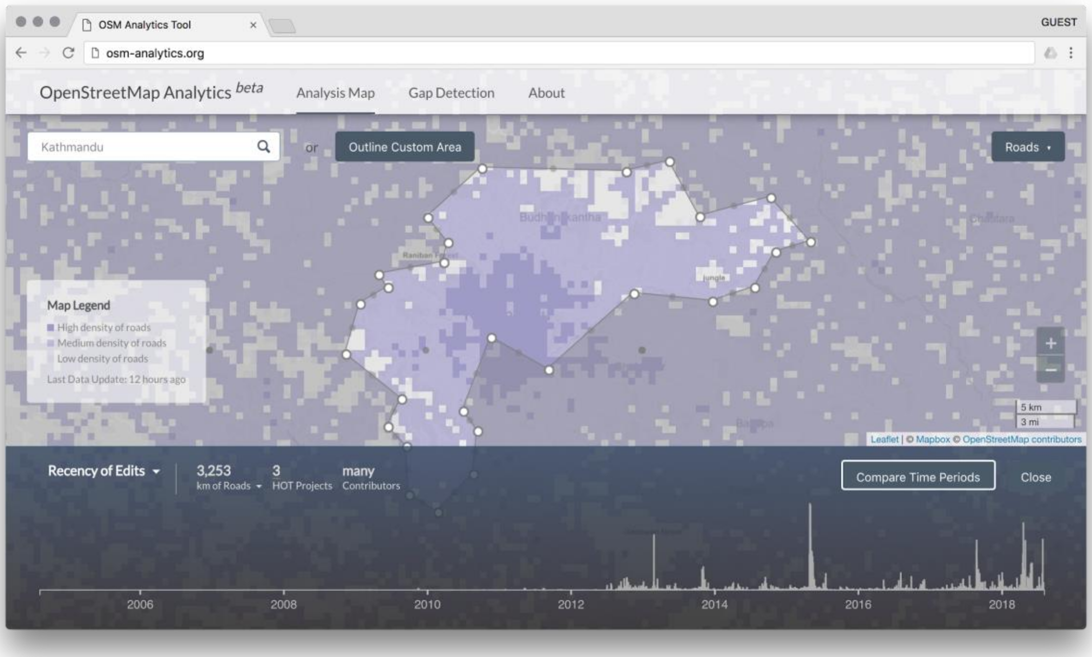


Figure 5: Kathmandu OSM Road Length

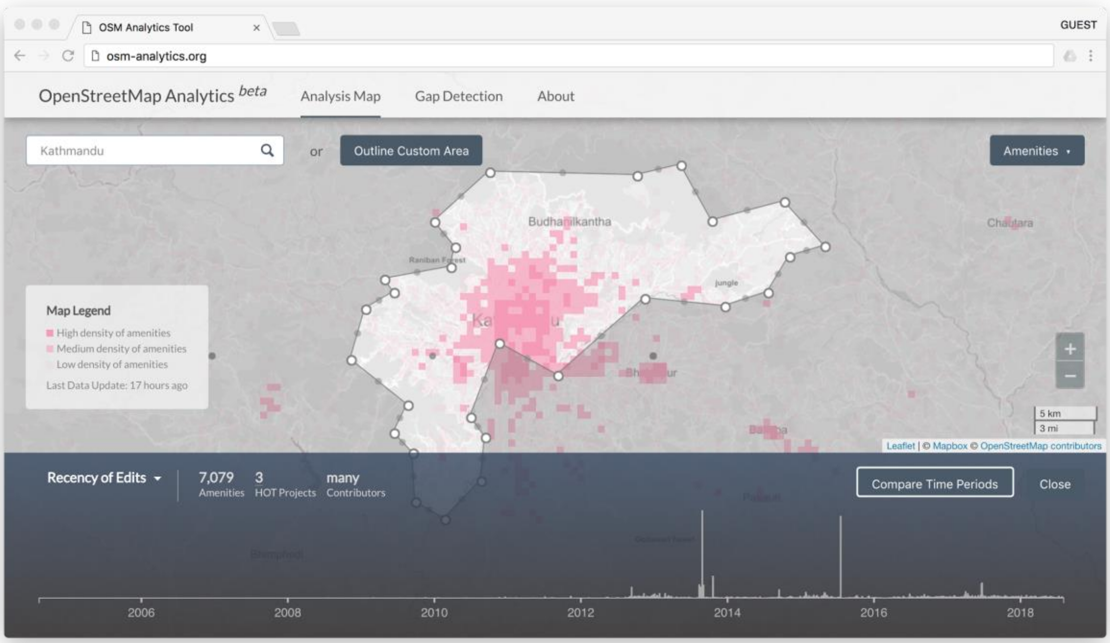


Figure 6: Kathmandu OSM Amenities Count

4.3. Dar es Salaam

The city of Dar es Salaam in Tanzania has also been chosen to help validate the applicability of the classification pattern for developing exposure datasets. Dar es Salaam is especially important as a benchmark city because it is the largest and fastest growing east African metropolitan area with a population of 5 million, which is expected to grow by 85% through 2025, and infrastructure and assets estimated at US$5 billion (ReliefWeb, 2017). It is one of the 31 administrative regions that make up the country and is situated towards the eastern side, sitting along the coast. The city consists of 90 wards, spread across 5 districts: the Northern Kinondoni; Central Ilala, Ubungo, Southern Temeke and Kigamboni.

Dar es Salaam was chosen as the city to ground truth the exposure dataset created for Tanzania due to the diversity of infrastructure and building types, which should again provide a sufficient sample of information for validation. Also key to the decision is the presence HOT has in Dar es Salaam already with the Ramani Huria[13] project, which focuses on the mapping of drainage infrastructure in OSM. This data is being collected in an effort to help reduce the impacts caused by the large scale flooding experienced in Dar es Salaam annually. Health care services, which are key to reducing illnesses and mortality during flooding events, as well as toilets, water sources, and building infrastructure data is also being collected.

Currently there are 905,177 buildings mapped in OSM for the city of Dar es Salaam, 14,672 kilometers of roads and an estimated 13,010 amenities (Table 9). These statistics were also obtained on August 3, 2018. A lot of the data added to OSM for the city has been carried out by Ramani Huria, with their efforts almost doubling the amount in 2017 (Figure 7), when they went from 454,587 in August, to 897,934 by December 2017. This difference can be explored with OSM Analytics[14], an online tool developed by HOT to help assess the number of OSM edits for buildings and roads between different time periods.

[13] http://ramanihuria.org/

[14] http://osm-analytics.org/

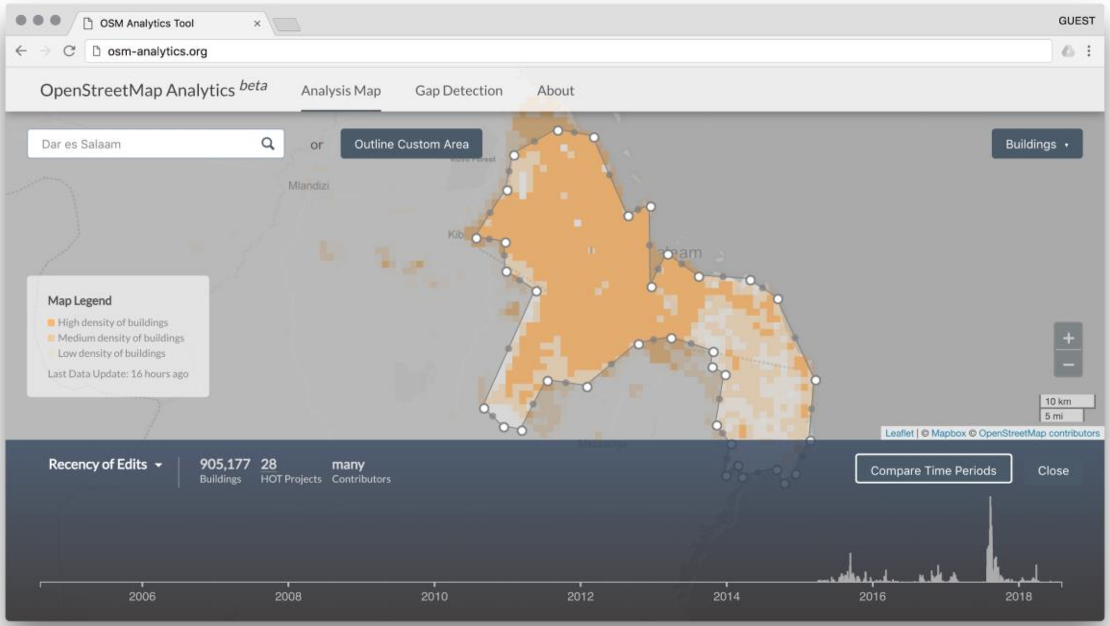


Figure 7: Dar es Salaam OSM Building Count

Dar es Salaam	
OSM	**Statistics**
Buildings	905,177
Roads	14,672
Amenities	13,010

Table 9: Dar es Salaam OSM Statistics

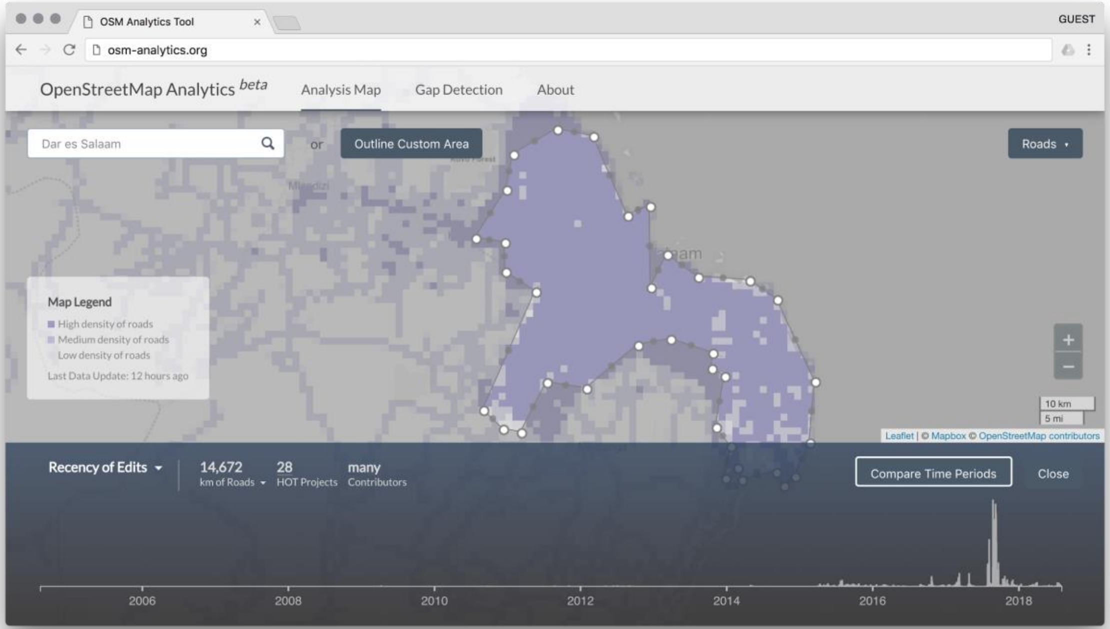


Figure 8: Dar es Salaam OSM Road Length

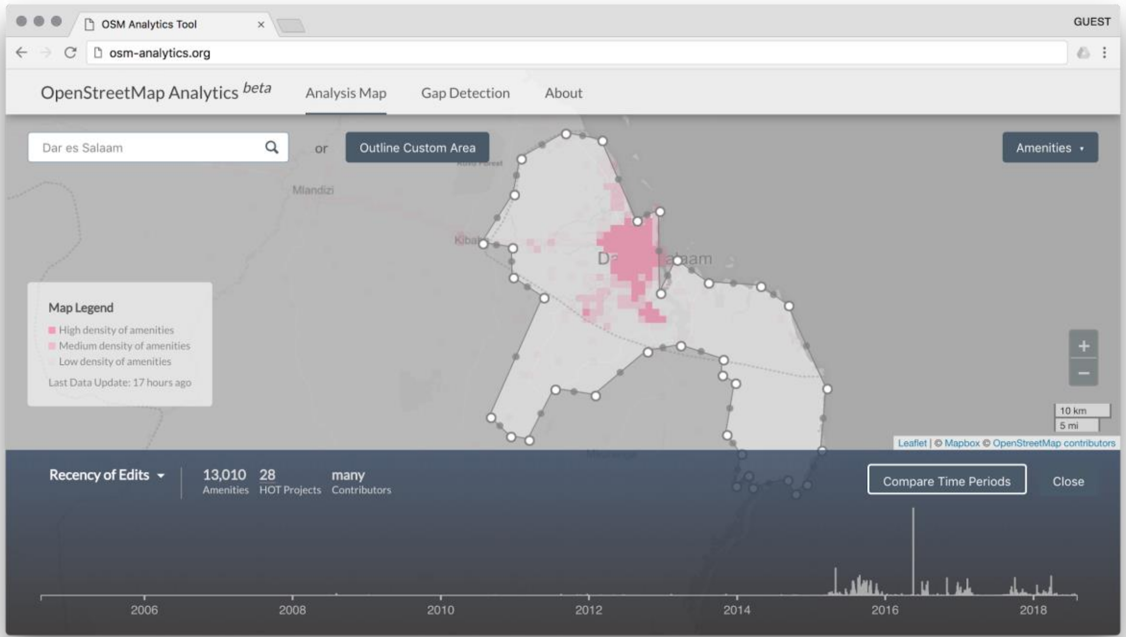


Figure 9: Dar es Salaam OSM Amenities Count

5. Datasets

5.1. Power Grid

The overall aim of the METEOR project is to establish a methodology and produce a protocol for the development of exposure data that includes both the multi-hazard taxonomy of structural and functional characteristics of the built environment and future population projections.

An extensive multi-hazard taxonomy was created as part of the second round of the Global Facility for Disaster Reduction and Recovery (GFDRR) Challenge Fund[15] in 2018, through a collaboration between the Global Earthquake Model (GEM)[16], ImageCat[17] and HOT. This exposure taxonomy and the accompanying database schema created is known as the Global Exposure Database for All (GED4ALL).

Numerous Nepal and Tanzania datasets were assessed for their suitability to import into OSM as part of the project. In the end, the datasets deemed most appropriate were those containing power grid information. The taxonomy for power grids in GED4ALL follows the HAZUS[18] classification system, which outlines four main components: substations, distribution circuits, transmission towers and energy generation facilities. The attributes identified as key for assessing the structural integrity of power grid infrastructures for the GED4ALL and the supporting OSM tags are outlined in the tables below:

POWER GRID: SUBSTATION			
No.	**GED4ALL**	**OSM Key**	**OSM Description**
		power=substation	Place dedicated to step-up or step-down the voltage=* within an electrical power grid, generally linked to the rest of the electrical network by one or more sets of power=lines and which may contains one or more power=transformers
		substation=transmission	Substation whose main function is to connect and switch transmission lines transmitting power between areas
		power=transformer	Static device that converts a given power voltage to another power voltage. Usually located within a substation
1	Component	voltage=*	The highest voltage of operation within the facility
2	Anchorage	anchored=*	Proposed OSM anchored key, associated with storage tanks (yes, no, unknown)
3	Code Provisions	substation:code=*	Proposed OSM provision code key, associated with storage tanks (none, low, moderate, high, unknown)

15 https://www.gfdrr.org/en/challenge-fund

16 https://www.globalquakemodel.org/

17 http://www.imagecatinc.com/

18 https://www.fema.gov/media-library/assets/documents/24609

POWER GRID: TOWER			
No.	**GED4ALL**	**OSM Key**	**OSM Description**
		power=tower	For big towers or pylons carrying electricity cables. Normally constructed from steel latticework and carry high voltage electricity cables
		material=*	Default value is steel (steel, wood, concrete, aluminium, composite)
		structure=*	Default value for steel towers is lattice (lattice, tubular, solid)
		height=*	Height in meter of the tower

POWER GRID: LINES			
No.	**GED4ALL**	**OSM Key**	**OSM Description**
		power=line	High-voltage power lines used for power transmission, supported by towers/pylons, or in some places or situations, by poles
		voltage=*	Voltage at which the line is operated (operating voltage)
		operator=*	Name of the company which operates this power line section (cable operator)
		cables=*	Number of different phase conductors for this power line section (number of cables)
		circuits=*	Number of different and separated electrical circuits built within this power line section (number of circuits)

Table 10: GED4ALL Power Grid Taxonomy Tags

Photo	Tagging
	power=substation voltage=380000;150000 substation=transmission name=380kV Bleiswijk operator=TenneT
	power=substation voltage=25000;10000 substation=distribution location=indoor
	power=substation voltage=10000;400 substation=minor_distribution location=kiosk operator=Stedin

Figure 10: OSM Substation Tag examples

ENERGY GENERATION FACILITIES			
No.	**GED4ALL**	**OSM Key**	**OSM Description**
		power=plant	An industrial facility for the generation of energy
1	Energy Source	plant:source=*	The source of the energy generated by a power plant (oil, geothermal, nuclear, hydroelectric, wind, solar, tidal wave, gas, biomass)
2	Power Capacity	plant:output:electricity=*	Power plants can output multiple forms of energy (yes / x W / x kW / x MW / x GW)
		plant:output:hot_water=*	
		plant:output:hot_air=*	
		plant:output:cold_water=*	
		plant:output:cold_air=*	
		plant:output:compressed_air=*	
		plant:output:steam=*	
		plant:output:vacuum=*	

Table 11: GED4ALL Energy Generation Taxonomy Tags

Power plant	Tagging	Descritption	OSM ref
Verrerie solar power plant, France	• type=multipolygon • power=**plant** • plant:output:electricity=24 MW • operator=Akuo Energy • name=Centrale solaire de Verrerie • start_date=2013-06-14	A 34ha solar farm producing 24 MW of electricity when sun comes up in south of France. With a little aerial imagery help, generators are taken as groups of solar cells for sake of simplicity. It is actually possible to use for each photovoltaic module but it's really expensive in mapping time. The relation accepts two closed and fenced perimeters, all generators and a transmission substation connecting to the power grid	Relation 3501886

Figure 11: OSM Renewable Energy Tag examples

5.2. Energy Data Info

The power grid datasets selected were obtained from the ENERGYDATA.INFO[19] (EDI) online platform, which is an open data service providing access to datasets and data analytics that are relevant to the energy sector. It has been developed as a public good available to governments, development organizations, private sector, non-governmental organizations, academia, civil society and individuals to share data and analytics that can help achieving the UN Sustainable Development Goal 7 of ensuring access to affordable, reliable, sustainable and modern energy for all (EnergyData.Info, 2018). The online service is a collaboration between 15 organisations such as Facebook, Georgia Tech, Berkeley Lab, World Bank (WB) Group, WorldPop and World Resources Institute. The datasets selected were the *Nepal - Electricity Transmission Network (2013)*[20] and the *Tanzania - Electricity Transmission Network (2014)*[21]. The Nepal dataset was collected and prepared for a project of the WB in October 2013, which includes transmission lines, substations, as well as power stations for both existing as well as planned projects (Table 12). The dataset for Tanzania was created from a combination of data sources, including data collected and prepared for a project of the WB in April 2014, the Africa Infrastructure Country Diagnostic (AICD) study and data from OSM (Table 13). The datasets were initially found on the UN OCHA Humanitarian Data Exchange (HDX) [22] platform which hosts open datasets for humanitarian use.

[19] https://energydata.info/

[20] https://energydata.info/dataset/nepal-electricity-transmission-network-2013

[21] https://energydata.info/dataset/tanzania-electricity-transmission-network-2014

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

Field	Value
Source	http://projects.worldbank.org/P115767/nepal-india-electricity-transmission-trade-project?lang=en
Author	Christopher Arderne
Last Updated	March 29, 2018, 10:39 PM (UTC+01:00)
Created	February 23, 2017, 9:27 PM (UTC+00:00)
Topic	Transmission and distribution
Country	Nepal
Published Year	2013
Datasets	Transmission Network, Power Stations, Substations
Link	https://energydata.info/dataset/nepal-electricity-transmission-network-2013
License	Creative Commons Attribution 4.0

Table 12: Metadata for EDI Nepal Power Grid Data

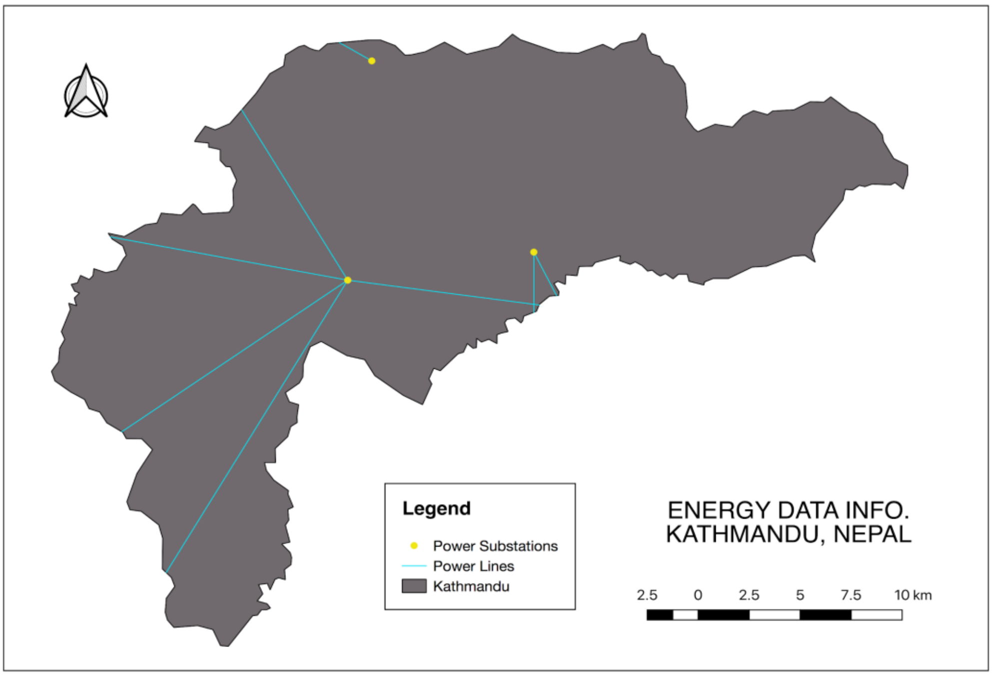


Figure 12: Kathmandu Map of EDI Power Grid Dataset

Field	Value
Source	https://datacatalog.worldbank.org/dataset/africas-infrastructure-electricity
Author	Christopher Arderne
Last Updated	April 10, 2018, 9:41 PM (UTC+01:00)
Created	March 9, 2017, 9:17 PM (UTC+00:00)
Topic	Transmission and distribution
Country	Tanzania, United Republic of
Published Year	2014
Datasets	Transmission Network, Power Stations
Link	https://energydata.info/dataset/tanzania-electricity-transmission-network-2014
License	Open Data Commons Open Database License 1.0

Table 13: Metadata for EDI Dar es Salaam Power Grid Data

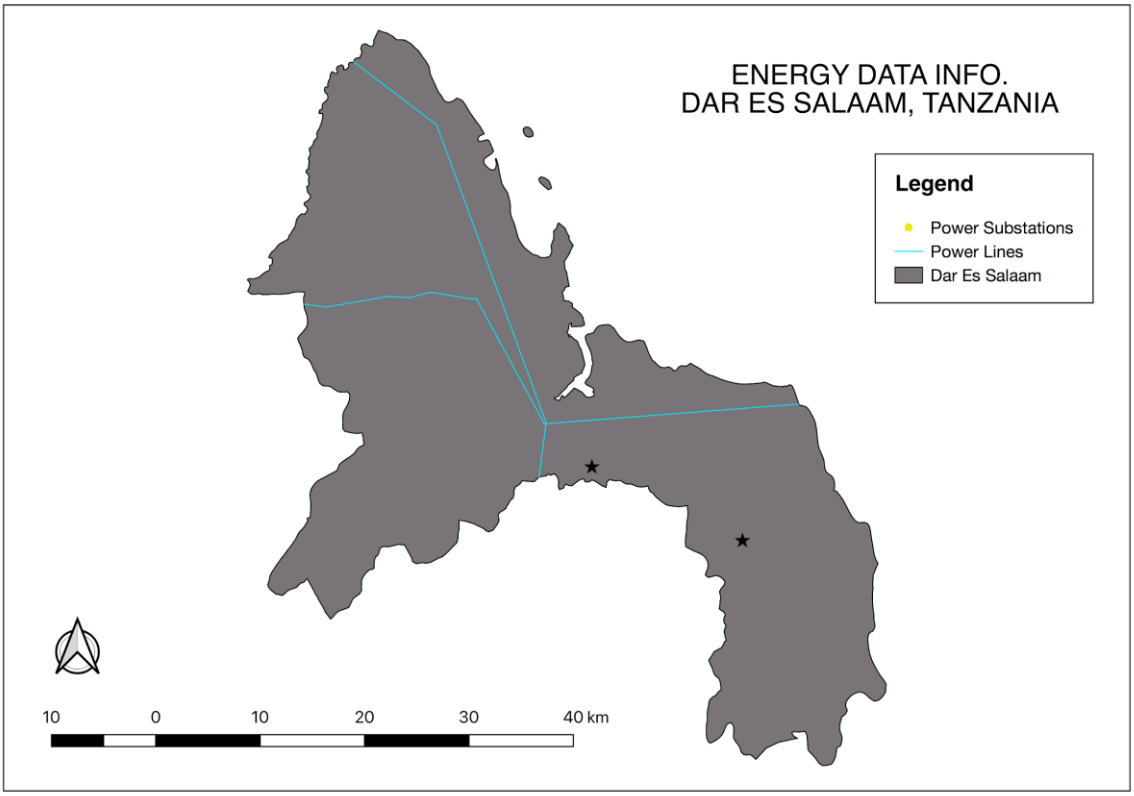


Figure 13: Dar es Salaam Map of EDI Power Grid Dataset

5.3. OpenStreetMap

The power grid data for Kathmandu and Dar es Salaam was extracted from OSM with the use of the HOT Export Tool [23]. The online open service allows users to easily create customised extracts of up-to-date OSM data in various file formats. Data was extracted using the following YAML syntax:

```yaml
power_grid:
  types:
    - points
  select:
    - name
    - power
  where:
    - power='plant'
    - power='substation'
    - power='tower'


power_lines:
  types:
    - lines
  select:
    - name
    - power
  where:
    - power='line'
```

[23] https://export.hotosm.org/

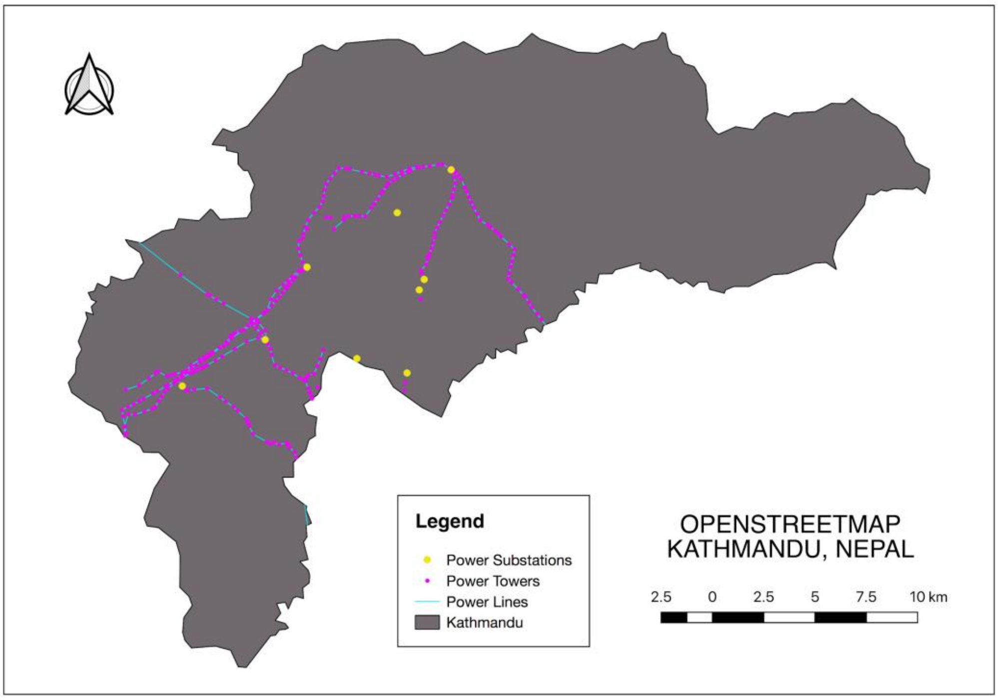


Figure 14: Nepal map of OSM Power Grid Dataset

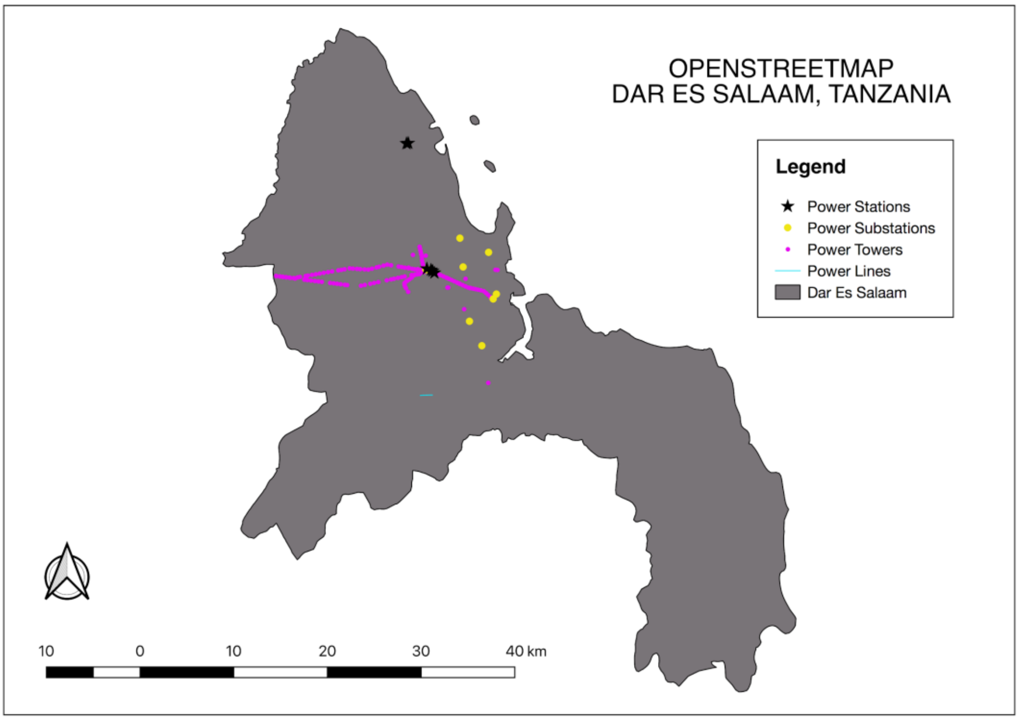


Figure 15: Dar es Salaam Map of OSM Power Grid Dataset

5.4. Discussion

A comparison of the power grid features available in the EDI datasets against the OSM features for both Kathmandu and Dar es Salaam easily show that the current data in OSM is more complete. The OSM power grid data for Kathmandu has more power substations, power towers and power lines contained within it (Table 14). Similarly the OSM power grid data for Dar es Salaam has more power plants, power substation and power towers contained within it. The EDI dataset however contains more power line information (Table 15).

Not only does OSM overall contain more power grid features, but the accuracy of the data is of higher quality. Looking at a zoomed-in snapshot of the EDI data for Kathmandu, it can be seen that the power plant does not appear on the satellite imagery in the background (Figure 16). This is a stark difference compared to the power substations and power towers in the OSM data, which can easily be seen against the satellite imagery base layer (Figure 17). The satellite imagery chosen to help assess the accuracy of the power grid features presents in the EDI and OSM datasets is the DigitalGlobe Premium[24], which was released in May 2017 to assist with the development of open data. This imagery product is made up of the most recent imagery, with crowdsourced fillers for areas identified with cloud cover, alignment errors or low resolution issues present, to provide that highest quality imagery.

Similarly, the EDI power grid data for Dar es Salaam contains less features than the OSM data (Table 15). The power plants contained within the EDI data cannot be seen in the satellite imagery, while the power plants in the OSM data can easily be detected in the baselayer (Figure 18 and Figure 19). The EDI power grid data contains much more information for the power lines compared to what is currently in OSM, but upon closer inspection it appears that the data is not accurate with the vector lines not following any present power lines in the satellite imagery (Figure 19).

[24] https://platform.digitalglobe.com/premium-imagery-maps-api/

Kathmandu		
Feature	**EDI**	**OSM**
Power Plants	0	0
Power Substation	3	9
Power Towers	0	297
Power Lines	66,047 km	86,619 km

Table 14: Kathmandu EDI vs OSM Power Grid Statistics

Dar es Salaam		
Feature	**EDI**	**OSM**
Power Plants	2	6
Power Substation	0	9
Power Towers	0	145
Power Lines	100,229 km	13,141 km

Table 15: Dar es Salaam EDI vs OSM Power Grid Statistics

Figure 16: Kathmandu Map of EDI Power Grid Dataset – Satellite Baselayer

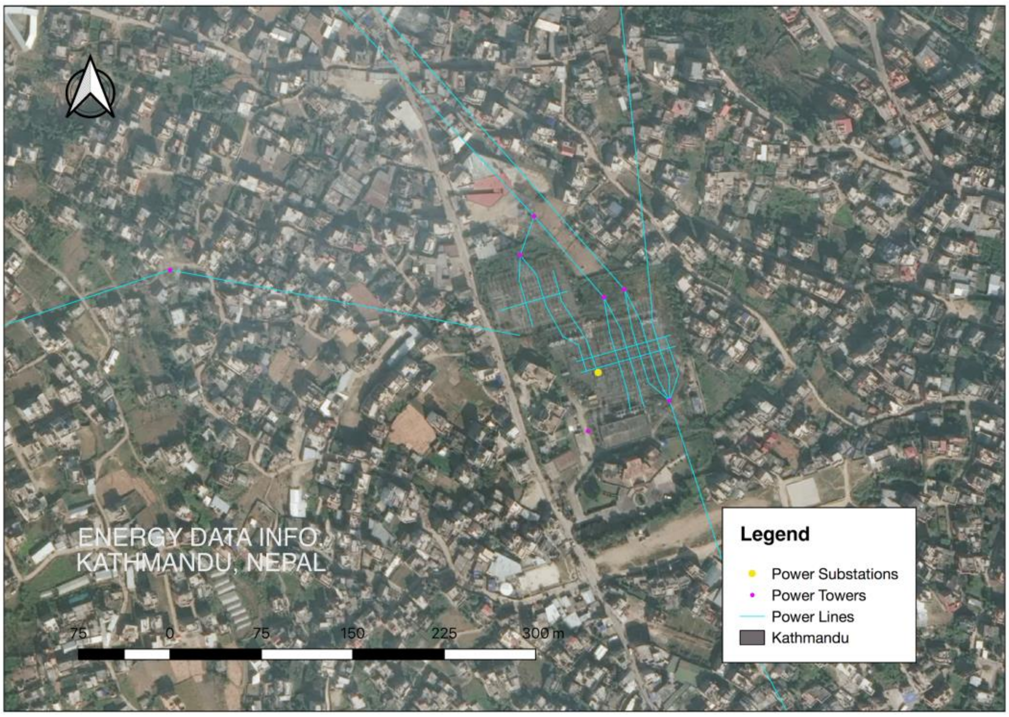


Figure 17: Kathmandu Map of OSM Power Grid Dataset – Satellite Baselayer

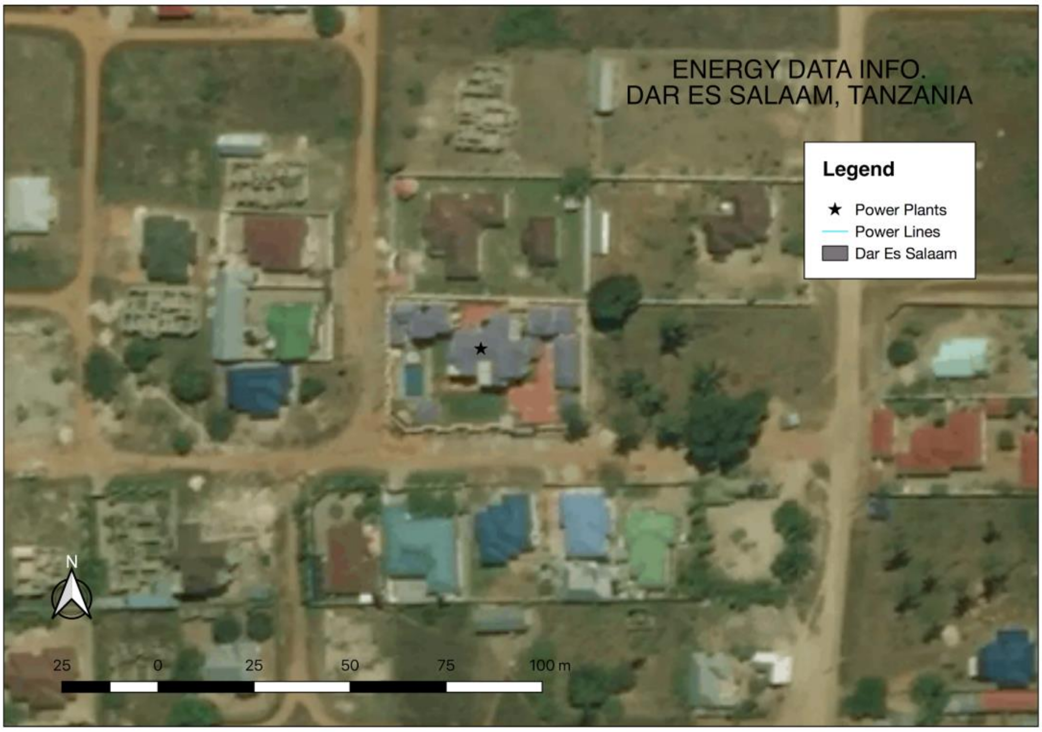


Figure 18: Dar es Salaam Map of EDI Power Grid Dataset – Satellite Baselayer

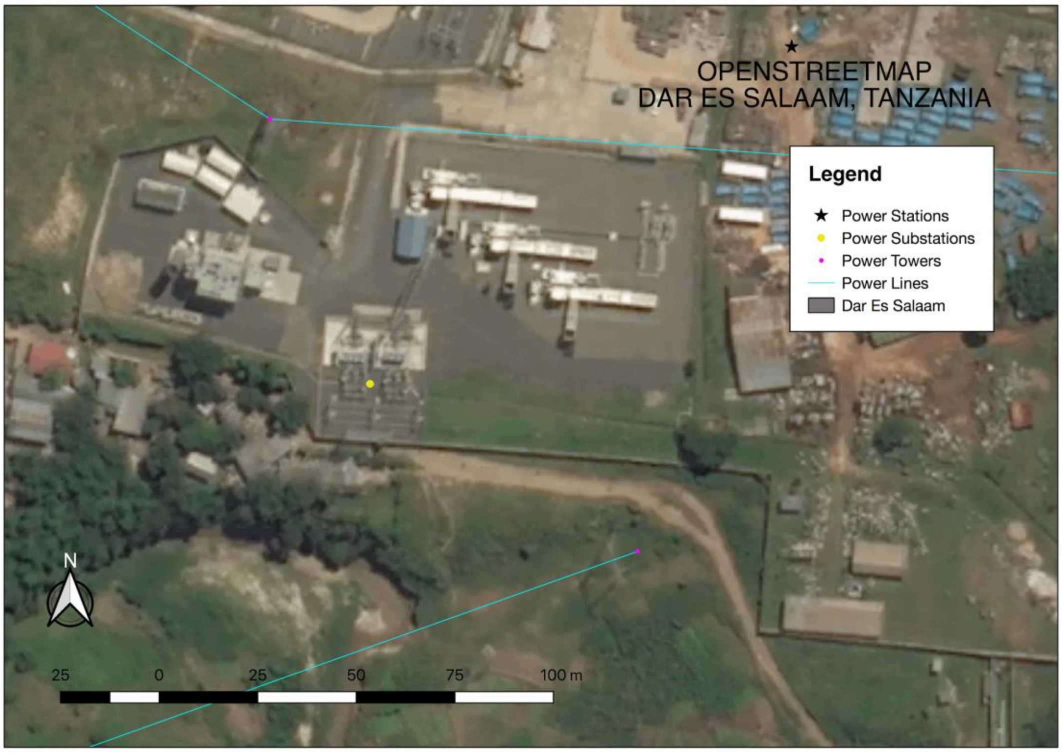


Figure 19: Dar es Salaam Map of OSM Power Grid Dataset – Satellite Baselayer

There can be several reasons why the quality of the EDI datasets are not as high as we would have hoped for. It was noted on the EDI website that both datasets are based on digitised PDF maps, and are intended as a schematic of rough locations of the power network and are not suitable for applications requiring high accuracy. This could be one of the reasons, along with other data processing errors, such as the use of poor satellite imagery for digitisation, or inaccurate GPS capturing of the location of features on the ground. Another contributing factor could also be that the data is simply out of date with the Nepal project for the data collection taking place in 2013 and the Tanzania work published in 2014. Although this is only five years ago, both Kathmandu and Dar es Salaam have gone through many structural changes since then with natural hazards and urban development occurring in the cities.

6. Conclusion

Several datasets were assessed for suitability to import into OSM. Discussions with NSET introduced the possibility of using data from a 2011 building census carried out in 2011. The granularity of the data was captured at the municipal level, with attributes such as the building materials, floor type, wall type, roof type, number of floors, year of constructions and number of occupancy, all of which is the right type of information hosted in OSM. Unfortunately, however, there were no GPS coordinates for the data making it a very labour intensive task to properly import and difficult to verify remotely. Open datasets available online from the Government of Nepal Ministry of Home Affairs[25], but none of these were suitable for import into OSM, as they were either the wrong type of information, incomplete, out of date, did not have the appropriate license or the source was simple not reliable enough. The same was the case for Dar es Salaam datasets assessed for import.

One of the strengths of OSM data is the fact that it is open, and used by numerous individuals and organisations who are checking the quality of the information. Data that is closed or not accessed that often is much more likely to never be validated by external sources. Open data is a difficult topic in Tanzania. There is a lot of data around, but it is not really in a usable state, being out of date, incomplete and lacks the appropriate licensing. Conversations with Ramani Huria, highlighted that their project also attempted to import existing open datasets, with the licensing and restrictive media laws[26] making it difficult. The WB has helped the Tanzania government set up a platform for open data[27], however the licensing is not very clear as to what or how the data can be used, with the current text not compatible enough with the ODbL license [28]. For example the National Bureau of Statistics (NBS) open data portal requests for the user to sign in[29], then an authentication message pops up

[25] http://drm.moha.gov.np/layers/

[26] https://freedomhouse.org/report/freedom-press/2016/tanzania

[27] http://opendata.go.tz

[28] http://opendata.go.tz/en/pages/kanuni-na-masharti

[29] http://nbs.go.tz/nbstz/index.php/english/geographical-information-system-gis/910-gis-open-data-portal-for-nbs

stating that "The site you are attempting to access is not public. Please sign in using your ArcGIS organization credentials", but then there is no where to sign up[30].

The power grid datasets for Kathmandu and Dar es Salaam from EDI seemed like great candidates for importing into OSM, given that it was the right type of structural information for the project, it is from reputable sources, the data was hosted on several platforms and the appropriate license was available. However after further analysis of the data, it turned out that the quality of the data was not appropriate for import into OSM, and after reviewing all the steps involved with the import process, it would make sense to channel that effort into the next stages of the project that focus on the creation of the exposure data through the remote digitisation of satellite imagery, followed by the collection of associated attribute information on the ground. This will ensure that the data created in OSM for use by the METEOR is accurate and of high quality for the best results, and to follow the guidelines required by the community.

7. References

CountryMeters (2018) https://countrymeters.info/en/Nepal [Online] Accessed: 1 August 2018.

EnergyData.Info (2018) https://energydata.info/ [Online] Accessed: 25 June 2018.

Guha-Sapir D, Hoyois Ph., Wallemacq P. Below. R. (2016) Annual Disaster Statistical Review 2016: The Numbers and Trends. Brussels: CRED.

International Monetary Fund (2014) World Economic Outlook Database, April 2014 [Online] Available: http://www.imf.org/external/pubs/ft/weo/2014/01/weodata/weorept.aspx?pr.x=66&pr.y=7&sy=2014&ey=2014&scsm=1&ssd=1&sort=country&ds=.&br=1&c=738&s=NGDPD%2CNGDPDPC%2CPPPGDP%2CPPPPC&grp=0&a=

My Republica (2016) Cover Story: Data Revolution [Online] Available: http://admin.myrepublica.com/the-week/story/43132/banking-on-data.html Accessed: 12 July 2018.

Open Data Commons (2018) https://opendatacommons.org/ [Online] Accessed: 1 July 2018.

ReliefWeb (2017) Problems of Population Growth and Climate Change Converge in Dar-es-Salaam [Online] Available: https://reliefweb.int/report/united-republic-tanzania/problems-population-growth-and-climate-change-converge-dar-es-salaam Accessed: 12 July 2018.

WorldPop (2013) Tanzania 100m Population, Version 2. University of Southampton. DOI: 10.5258/SOTON/WP00288.

WorldPop (2017) Nepal 100m Population, Version 2. University of Southampton. DOI: 10.5258/SOTON/WP00531.

[30] http://portal-nbs-tz.opendata.arcgis.com/