

MONITORING SDG 9 WITH GLOBAL OPEN DATA AND OPEN SOFTWARE – A CASE STUDY FROM RURAL TANZANIA

Codrina Maria Ilie^{1,2}, Maria Antonia Brovelli³, Serena Coetzee⁴

¹Technical University of Civil Engineering of Bucharest, Bucharest, Romania,

²Terrasigna, Bucharest, Romania, codrina.ilie@terrasigna.ro

³Department of Civil and Environmental Engineering, Politecnico di Milano, Italy, maria.brovelli@polimi.it

⁴Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria,
South Africa, serena.coetzee@up.ac.za

Commission IV, WG IV/4

KEY WORDS: crowdsourcing, open geospatial data, open source software, rural accessibility, sustainable development goals (SDG), Tanzania

ABSTRACT:

The 17 goals adopted by the United Nations (UN) are aimed at achieving a better and more sustainable future for all. For each goal, a set of indicators has been defined. The indicators measure progress towards achieving the respective SDG. For the majority of these indicators, geospatial information is needed to evaluate the current state of the indicator. While geospatial information is largely available in developed countries, this is not the case in many developing countries of the world. Furthermore, skills and capacity for calculating indicator values are also limited in many developing countries. To address these shortcomings, the third challenge of the 2018 UN OSGeo Committee Educational Challenges called for the development of training material for using open source software together with freely available high resolution global geospatial datasets in support of monitoring SDG progress. The resulting training material provides a step-by-step guide for calculating the state of SDG indicator 9.1.1, *Proportion of the rural population who live within 2km of an all-season road*, using open software and open data with global coverage. Through the development of this training material, we showed that anyone can monitor progress towards achieving SDG indicator 9.1.1 for their specific part of the world. Because open source software and open data were used, the indicator calculation is cost effective and completely sustainable.

1. INTRODUCTION

The United Nations (UN) has adopted a set of 17 goals aimed at achieving a better and more sustainable future for all. The goals focus on global challenges, such as poverty, inequality, climate, environmental degradation, prosperity, and peace and justice for all. For each goal, a collection of targets and corresponding indicators are defined. The indicators measure progress towards achieving a specific target. For the majority of these indicators, geospatial information is required to determine the current state of the indicator. The SDGs are especially relevant for developing countries. Yet, while geospatial information is largely available in developed countries, this is not the case in many developing countries of the world. Furthermore, skills and capacity for calculating indicator values are also limited in many developing countries.

OSGeo (Open Source Geospatial Foundation) is a not-for-profit organization aimed at fostering the global adoption of open geospatial technology (www.osgeo.org). OSGeo and the UN joined forces in the UN OSGeo Committee whose work is aimed at identifying and developing open source geospatial software and services that meet the requirements of UN operations and UN aims (OSGeo, 2019). In 2018, the UN OSGeo Committee sent out a call for the development of educational material, comprising three challenges. Winners were guided by mentors to ensure that the material met the requirements of the target audience (OSGeo 2018).

The first two challenges were closely related to UN operations, and specifically the UN Open GIS Initiative, which is a project aiming at designing and developing an open source solution for peacekeeping missions (Eom, 2017). The third challenge was aimed at addressing the current lack of training material for using open source software together with freely available high resolution global geospatial datasets for environmental, social and economic analysis in support of UN SDGs. Such training material is especially relevant in developing countries where data is often scarce and resources for buying software are limited.

In this paper, we describe the material developed for the third challenge. In Section 2, the requirements and design of the training material are described. The training material is presented in Section 3 and Section 4 concludes.

2. DESIGN OF THE MATERIAL

2.1 Requirements

The following requirements were specified for the training material to be prepared for the third challenge:

- (1) the processing must be replicable anywhere on the globe, thus the datasets selected must have global coverage and be available under an open license;
- (2) the selected SDG indicator must have a significant geospatial component;
- (3) there must not be an ongoing or announced initiative to monitor the selected indicator, such

as the GEO Wetlands or GEO Human Planet initiatives.

The reasoning behind these requirements is straight forward. The UN Agenda that was defined to accomplish a better world under environmental, societal and economic aspects is *de facto* translated into the collection of 232 measurable indicators. As the scope of the UN SDGs spans the whole world, the indicators must be monitored globally. Yet, due to indisputable differences in development in different regions of the world, this is a very ambitious task. However, this task is becoming more and more feasible due to proven maturity of existing open source solutions for geospatial workflows and recent significant increases in the availability of open data, be it community driven or accessible as a result of an open data policy such as the Copernicus satellite data (<https://scihub.copernicus.eu/>).

Standing proof for this statement are the numerous international monitoring initiatives by various organizations, e.g. for monitoring SDG 2, Zero Hunger: the Group on Earth Observations (GEO) Global Agricultural Monitoring (GEOGLAM) (<http://geoglam.org/>), Sentinel-2

for Agriculture (<http://www.esa-sen2agri.org/>) and the Agricultural Market Information System (AMIS) (<http://www.amis-outlook.org/amis-about/en/>); for SDG 6, Clean Water and Sanitation: the European Space Agency's (ESA) Water Observation and Information System (WOIS) (<http://www.tiger-net.org/>), the GEO Wetlands Initiative (<https://geowetlands.org/>), Hydroweb (<http://ctoh.legos.obs-mip.fr/>) by the Center for Topographic studies of the Ocean and Hydrosphere (CTOH) and the UN's Integrated Water Resources Management (IWRM) Data Portal (<http://iwrmdataportal.unepdhi.org/>); for SDG 11, Sustainable cities and communities: the World Health Organization's (WHO) interactive air pollution maps (<http://maps.who.int/airpollution/>) and the GEO Human Planet Initiative's Global Human Settlement Layer (<https://ghsl.jrc.ec.europa.eu/>).

Unlike these initiatives, the training material described in this paper is not a service, but a step-by-step guide for determining the current state of one of the SDG 9 indicator that can easily be adopted as a solid starting point for an initiative that monitors SDG 9 at sub-national level and at minimal costs.

2.2 Selected SDG indicator 9.1.1 and study area

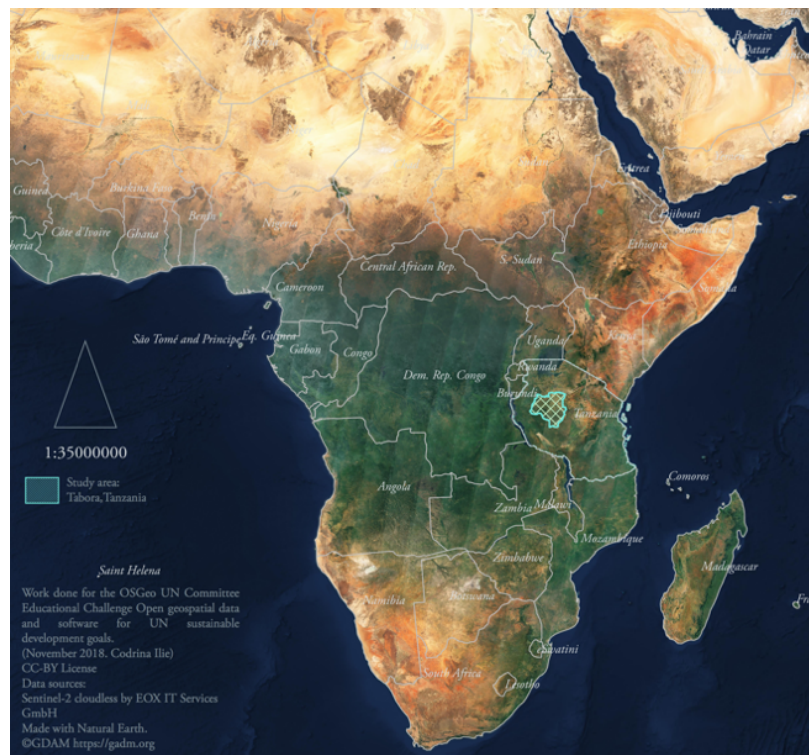


Figure 1. Tabora region in Tanzania

The training material was prepared for calculating indicator 9.1.1, *Proportion of the rural population who live within 2km of an all-season road*, with application to Tabora, a rural area of Tanzania (see Figure 1). This indicator measures progress towards SDG 9, which has the objective of building resilient infrastructure, promoting inclusive and sustainable industrialization and fostering innovation. Clear specifications for the terminology, methodology and possible data sources for calculating indicator 9.1.1 are available as work in progress published in December 2018 (UNSD, 2018). Monitoring this indicator relies heavily on

geospatial data, typically collected by road agencies and national statistics offices for their operational work. An initiative to monitor this indicator does not yet exist.

2.3 Selected open source software

The latest version of QGIS (www.qgis.org), an open source GIS, was selected for the training material due to its intuitive and robust functionalities, and its wide and intensely active community of developers and users. Apart from QGIS, a variety of alternative open source geospatial

solutions are available for processing and visualization of complete geospatial workflows. A good starting point for understanding this ecosystem is by exploring OSGeo activities, projects and initiatives on its website (<https://www.osgeo.org/>).

2.4 Selected open datasets with global coverage

To calculate SDG 9.1.1 based on the World Bank methodology (World Bank, 2016) requires the following geospatial information: spatial distribution of the population; urban areas (so that they can be excluded); the road network; and road conditions (to identify all-season roads).

For the spatial distribution of the population, the WorldPop Alpha version 2015 estimates of numbers of people per grid square, with national totals adjusted to match UN population division estimates (WorldPop 2013) was used. The dataset estimates the number of people per grid square, with national totals adjusted to match UNSD population estimates. The spatial resolution is 100m and the units are estimated number of persons per grid square.

For urban areas, data prepared by the Global Rural-Urban Mapping Project (GRUMP) v1 was used. This vector dataset is produced by the NASA Socioeconomic Data and Applications Center (SEDAC) and consists of polygons based on night-time lights and approximated urban extents (circles) based on buffered settlement points (CIESIN et al. 2017). As the prepared exercise had to be globally replicable, nuanced country-specific or regional differences in urban and rural definitions were not considered.

Global coverage geospatial data of a road network, especially with road conditions, is the most problematic aspect of the SDG 9.1.1 assessment. Even though there are relevant datasets, such as the Global Roads Open Access Data Set, Version 1 (gROADSv1) from SEDAC (CIESIN and ITOS, 2013), the data granularity (only major roads and coarse geometry) is a serious limitation for usage in the SDG 9.1.1 calculation. Thus, the training material was developed with OpenStreetMap (OSM) as the data source for geospatial road data (www.openstreetmap.org).

OSM is a collaborative project, therefore the data may vary in quality (completeness, accuracy, precision, etc.) over different regions. A significant advantage of the collaborative approach to gathering geospatial data worldwide, is the well documented framework that allows a clear understanding of the data structure and semantics.

While the quality may differ from one region to another, the data structure and semantics are described consistently across the whole dataset. This consistency supports initiatives such as the one presented in this paper because the quality can be understood and can be improved before calculating the SDG indicator. Furthermore, active and constant use of OSM data in a specific region can trigger the inception of an OSM contributing community and further on, help to consolidate the community so that data gaps can be filled.

2.5 Calculating SDG indicator 9.1.1

Generally, accessibility is based on distances between locations, either straight line distances or distances travelled via a network. The method proposed by the World Bank for measuring SDG 9.1.1. relies on straight line walking distances to all season roads (World Bank, 2016). There are many ways in which the method could be improved, e.g. by varying the walking distance based on topography or population density. In line with the requirements specified for the third challenge, this paper focuses on applying the method proposed by the World Bank, not on improving or evaluating the pros and cons of this and other methods. That is, the training material was designed to test the efficiency of open source solutions and open data for assessing an SDG indicator in regions that would otherwise lack capacity to monitor the SDG with traditional data sources.

3. TRAINING MATERIAL

The training material is composed of two main segments. Firstly, it presents the workflow for calculating the rural access index (RAI) for the Tabora Region in Tanzania, pointing out some particularities in calculating the RAI; secondly, it introduces the model built using the same open source solution to automate the SDG 9.1.1 computation.

3.1 Workflow for calculating the RAI

Calculation of the RAI is divided into three stages: (1) preparing the geospatial data; (2) computing the rural access index; and (3) presenting the results. The conceptual schema of the workflow designed to calculate SDG 9.1.1 *Proportion of the rural population who live within 2km of an all-season road*, using the World Bank methodology is presented in Figure 2.

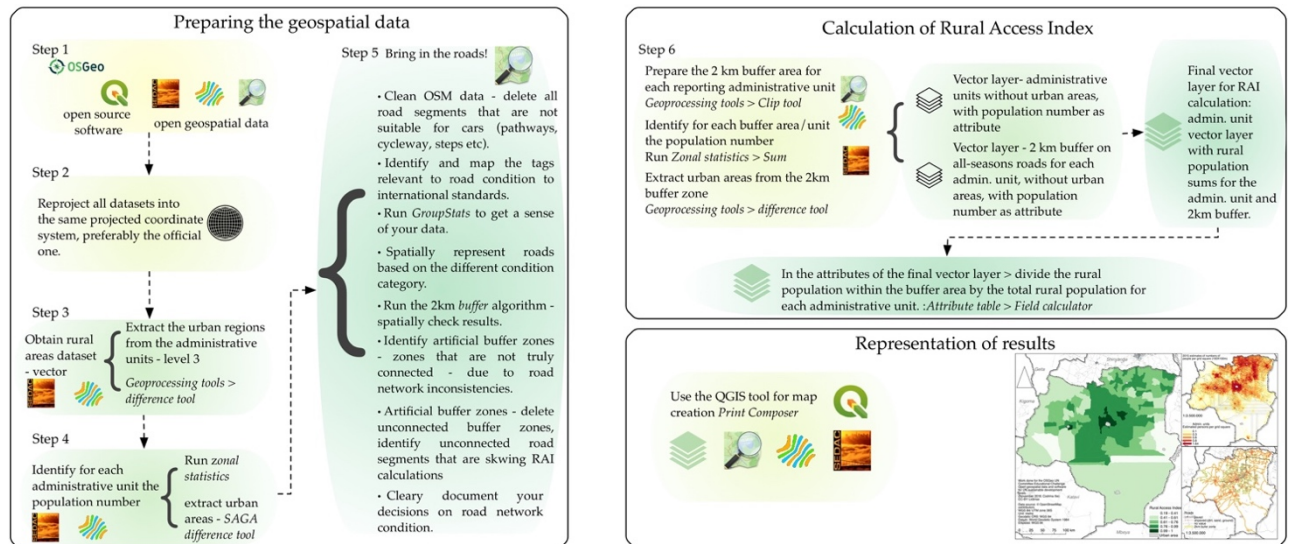


Figure 2. Workflow for calculating SDG 9.1.1

Data preparation is the most time consuming, as it implies a thorough analysis of the data, spanning from the formal structure (i.e. format) to aspects related to consistency, precision, scale etc., to the cleaning of the datasets. For this first stage, there are five steps in the workflow. The introductory step consists of importing all selected datasets into the selected open source software, QGIS3¹.

The second step focuses on reprojecting these datasets to the same projected coordinate system. This is necessary because the RAI calculation requires processing of road geometries. To identify the appropriate coordinate system, one can use a web service such as epsg.io². For the Tabora use case, the appropriate coordinate reference system is WGS 84 / UTM zone 36S - EPSG: 32736.

Step 3 is designated to preparing the rural areas of the region analysed, using the geoprocessing tool difference. As stated in the United Nations dedicated RAI metadata document, “the RAI can be calculated at subnational levels down to the level of granularity of the underlying datasets.” Thus, given the scale of our datasets and the purpose of the use case, RAI has been calculated at administrative level 3, according to terminology of the Database of Global Administrative Areas corresponding to wards in the Tanzanian administrative subdivisions.

The purpose of Step 4 is to identify for the rural areas of each administrative unit - ward - the spatial distribution of the population. Using the Zonal Statistics tool, the population number is extracted from the WorldPop dataset.

Step 5 is dedicated to preparing the road network and road conditions. As evident from the workflow, this is the most extensive processing step. For the current use case, only OSM data was used. However, if data from other sources is available, it could be used without affecting the workflow. Because topology is not significant in the RAI calculation, the use case does not address this aspect. As only all-season

roads suitable for any kind of vehicle (average modern automobile) are relevant for the RAI, all roads that are tagged as "highway" = 'cycleway', "highway" = 'pedestrian' or "highway" = 'path' and also roads under construction were removed. Next, a close inspection was performed on all remaining tags. The OSM community has defined a series of tags that are relevant to road conditions, such as:

1. Key: smoothness, with eight possible values: excellent, good, intermediate, bad, very_bad, horrible, very_horrible, impassable.
2. Key: surface
3. Tag: tracktype, with a gradual variation from grade1 (solid) to grade5 (soft).

Due to the chosen study area of Tanzania, the typology and tags specified for OSM road data in Africa (https://wiki.openstreetmap.org/wiki/Highway_Tag_Africa) and in the East Africa Tagging Guidelines (https://wiki.openstreetmap.org/wiki/East_Africa_Tagging_Guidelines) were considered. Because of the lack of harmonization between OSM tags and official international methodologies for describing road conditions, OSM tags had to be mapped to the World Bank’s descriptions of road conditions. In the current use case, the definition for a “road in good condition” provided in the World Bank’s methodology was considered: “Paved road in excellent, good, or fair condition and unpaved road in excellent or good condition.” (World Bank, 2016). Table 1 presents the mapping for the Tabora region use case. All segments that did not meet the requirements for inclusion in the RAI calculation were removed (See Figure 3).

The next part of Step 5 was to clean the road network. A solid topology is not required for applying the World Bank’s methodology, however there is the possibility of distorted RAI values in the case of network inconsistencies. Another situation addressed in the automated mode is the existence of short road segments that are further away from any other connected road by at least 2 km. These short, disconnected segments would produce incorrect RAI values: even if a rural population has access to an all-season road segment in acceptable condition, it does not offer real accessibility, because it does not connect the population to

¹ Retrieved from <https://qgis.org/en/site/> on 3.02.2019

² EPSG.io is an open-source web service with a database of coordinates systems used in maps worldwide that allows discovery of coordinate reference systems utilized all over the world for creating maps and geodata and for identifying geo-position.

the larger network. This situation was addressed by imposing a condition on the buffer zones generated by roads shorter than 2 km, that it had to be connected to an urban area. The trade-off consists in the assumed risk of excluding from the RAI calculation parts of the road network that due to faulty digitization are composed only by shorter than 2 km segments.

RAI	OSM tags	
Paved road = excellent	highway=primary	smoothness=excellent
	highway=secondary	smoothness=excellent
	highway=tertiary	smoothness=excellent
	highway=unclassified	smoothness=excellent
	highway=service	smoothness=excellent
Paved road = good	highway=primary	smoothness=good/intermediate
	highway=secondary	smoothness=good/intermediate
	highway=tertiary	smoothness=good/intermediate
	highway=unclassified	smoothness=good/intermediate
	highway=service	smoothness=good/intermediate
Paved road = fair	highway=primary	smoothness=bad
	highway=secondary	smoothness=bad
	highway=tertiary	smoothness=bad
	highway=unclassified	smoothness=bad
	highway=service	smoothness=bad
Unpaved road = excellent	highway=track	tracktype=grade1
Unpaved road=good	highway=track	tracktype=grade2/grade3
not included	highway=*	smoothness=very_bad
	highway=*	smoothness=horrible
	highway=*	smoothness=very_horrible
	highway=*	smoothness=impassible
	highway=*	tracktype=grade4
	highway=*	tracktype=grade5

Table 1. Mapping from OSM tags to the World Bank's description of road conditions

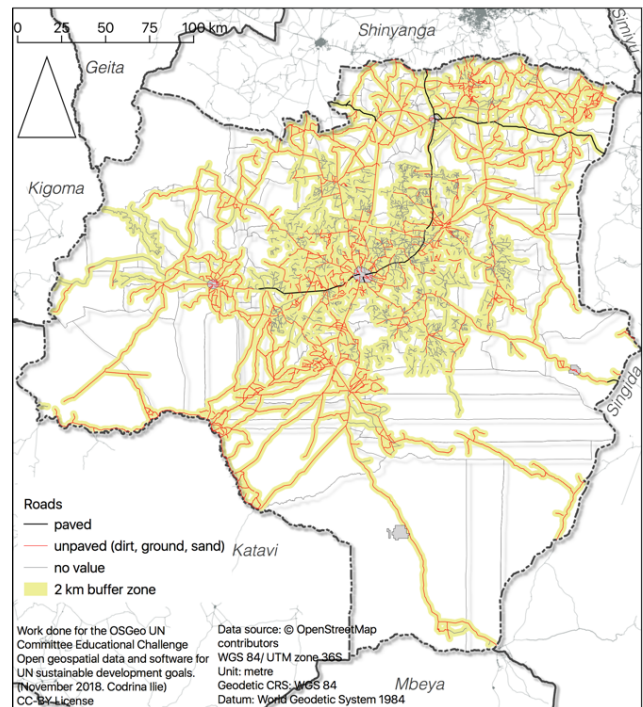


Figure 3. Roads in good condition in Tabora region of Tanzania

In the second stage, Step 6 represents the preparation of the final geospatial datasets: 1) the vector layer of the non-urban areas in the administrative units with population numbers; and 2) the vector layer with the 2 km buffer area around the all-season roads for each administrative unit, with population numbers. The RAI is obtained by dividing the rural population within the 2 km buffer area by the total rural population for each administrative unit.

In the third and final stage, the RAI is presented as a map. See Figure 4.

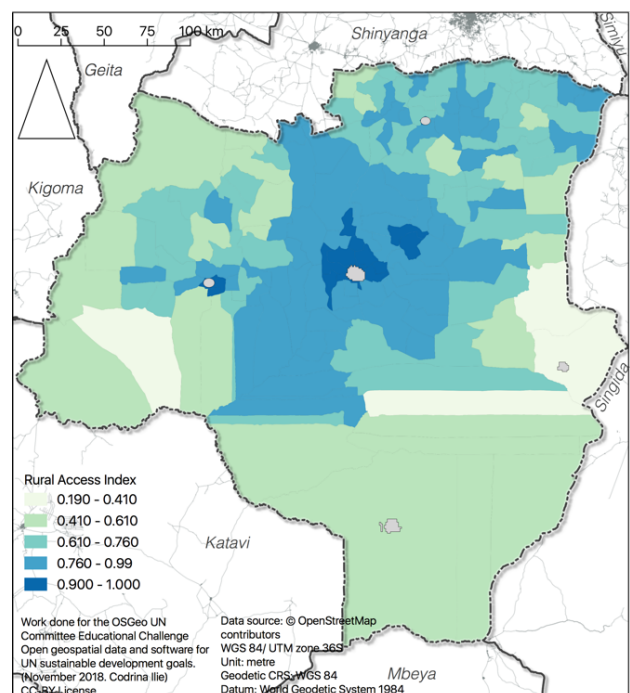
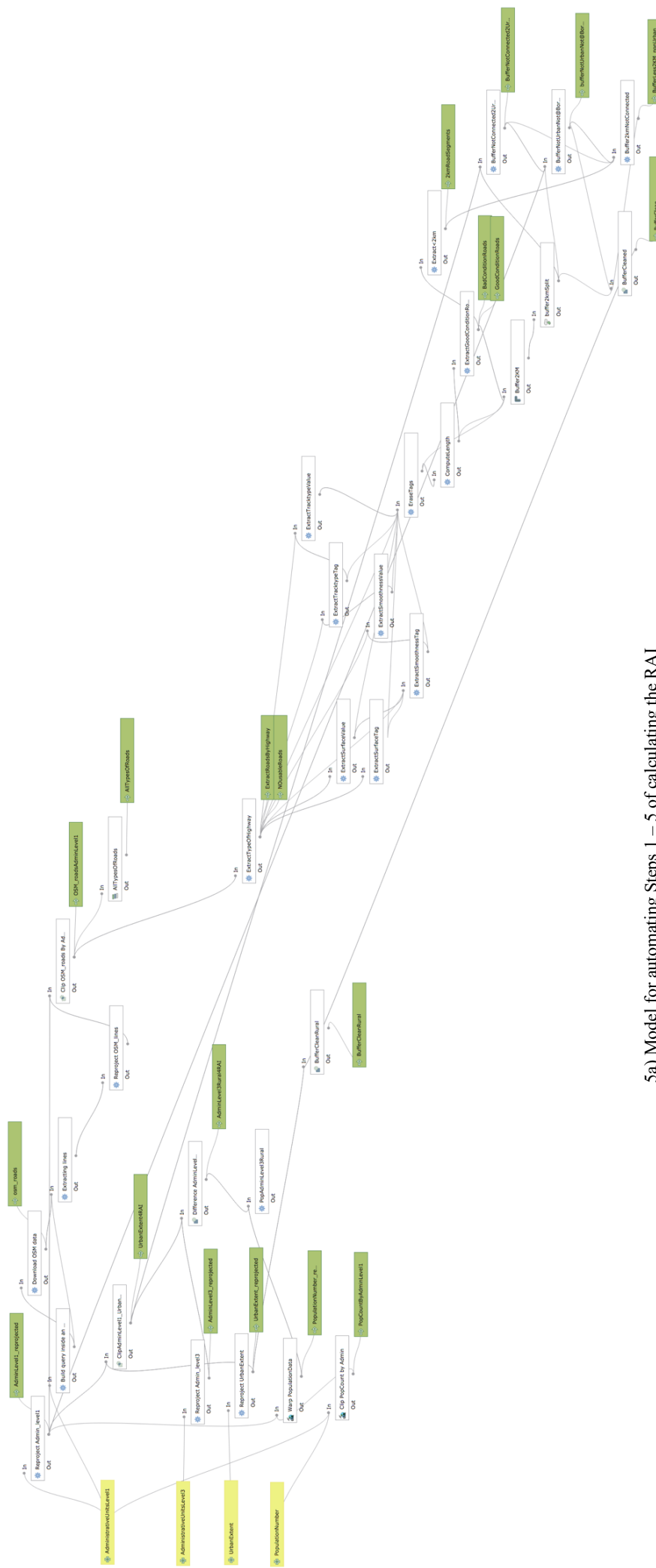


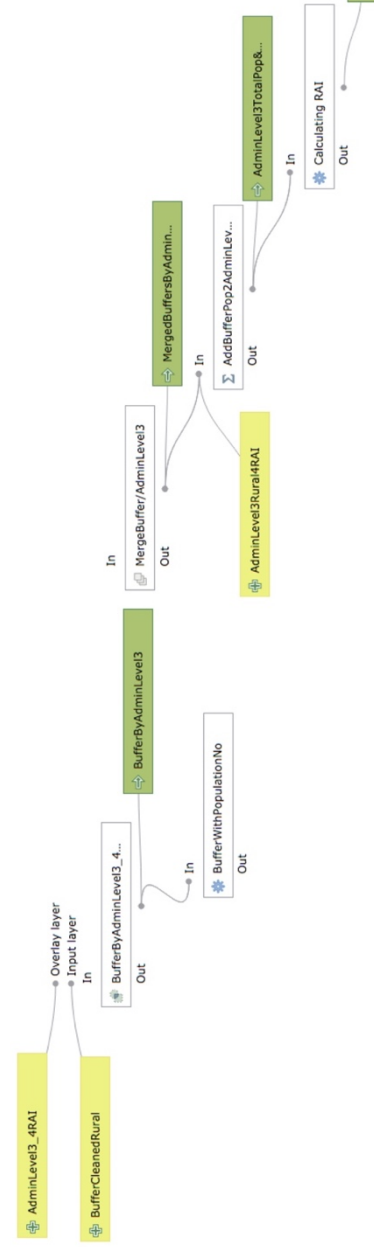
Figure 4. RAI for the Tabora region in Tanzania

3.2 Model to automate the SDG 9.1.1 calculation

The entire workflow can be studied and improved as necessary by analysing the model built to automate the calculation of SDG 9.1.1. (see Figure 5).



5a) Model for automating Steps 1 – 5 of calculating the RAI



5b) Model for automating Step 6 of calculating the RAI

Figure 5. Workflow for calculating SDG 9.1.1 in the training material

4. CONCLUSION

The training material for the third challenge of the UN OSGeo Committee Educational Challenge provides a step-by-step guide on how data is collected, updated and analysed in order to calculate the state of SDG indicator 9.1.1. Because open data and open software are used, calculating the indicator is cost effective and completely sustainable. Any gaps in the data can be filled by actively engaging local communities and/or citizens from anywhere in the world. The training material is published with a Creative Commons Attribution-ShareAlike (CC-SA) license at https://wiki.osgeo.org/wiki/Training_Material_for_UN_Open_GIS_OpenData.

Anyone anywhere in the world can follow the steps in the training material to calculate SDG indicator 9.1.1. for their specific part of the world. It is therefore a very powerful contributor towards building capacity in measuring progress on SDG 9.

ACKNOWLEDGEMENTS

The authors would like to thank OSGeo for sponsoring the prize money for the third challenge of the 2018 OSGeo UN Committee Educational Challenge.

REFERENCES

Center for International Earth Science Information Network - CIESIN - Columbia University, CUNY Institute for Demographic Research - CIDR, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical – CIAT, 2017. *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 01*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4Z31WKF>.

Center for International Earth Science Information Network - CIESIN - Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia, 2013. *Global Roads Open Access Data Set, Version 1 (gROADSv1)*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4VD6WCT>.

Eom, K. S., Arias, R., Brovelli, M. A., Criloux, G., Kang, H. K., Li, K. J., 2017. United Nations Open GIS initiative: The first year of activities, *GEAM. GEOINGEGNERIA AMBIENTALE E MINERARIA*, 151(2), 5-8.

UNSD (United Nations Statistics Division), 2018. *SDG Indicators Metadata repository, Metadata for Indicator: 9.1.1. Proportion of the rural population who live within 2 km of an all-season road*. Last updated: 20 December 2018. <https://unstats.un.org/sdgs/metadata/files/Metadata-09-01-01.pdf> (1 April 2019).

OSGeo, 2018. *OSGeo UN Committee Educational Challenge*. OSGeo news item. <https://www.osgeo.org/foundation-news/osgeo-un-committee-educational-challenge/> (1 April 2019).

OSGeo, 2019. *United Nations Committee*. OSGeo wiki page.

https://wiki.osgeo.org/wiki/UnitedNations_Committee (1 April 2019).

World Bank (2016). *Measuring rural access: using new technologies (English)*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/367391472117815229/Measuring-rural-access-using-new-technologies> (1 April 2019).

WorldPop (www.worldpop.org – School of Geography and Environmental Science, University of Southampton), 2013. United Republic of Tanzania 100m population. Alpha version 2010 and 2015 estimates of numbers of people per grid square, with national totals adjusted to match UN population division estimates (<http://esa.un.org/wpp/>) and remaining unadjusted. DOI: 10.5258/SOTON/WP00288.