Openness and Community Geospatial Science for Monitoring SDGs – An Example From Tanzania

Maria Antonia Brovelli

Politecnico di Milano, Italy

Maria Ilie Codrina

Technical University of Civil Engineering of Bucharest, Romania

Serena Coetzee

University of Pretoria, South Africa

This chapter focuses on two aspects that enable the monitoring of the United Nations (UN) sustainable development goals (SDGs): openness and community geospatial science.

20.1 Introduction

Openness typically refers to transparency, to free and unrestricted access to information, and to inclusive consensus-based decision-making. Community science is a branch of citizen science that involves a commitment from citizens, not only to collecting data, but also to designing and planning project activities in a more egalitarian (if not bottom-up) approach between (professional) scientists and citizen scientists. When the focus of the research is geospatial, we are dealing with community geospatial science. Examples of this approach are some of the projects related to OpenStreetMap (OSM), where citizens are significantly more than just “active sensors”, playing an instrumental role in the definition and shaping of the campaign [5].

Based on the two pillars of openness and community geospatial science, a relevant example for monitoring SDGs was developed for the OSGeo UN

Committee Educational Challenge [22]. In 2018, the Committee sent out a call for the development of educational material, comprising three challenges. The first two challenges were closely related to UN operations. 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. The hands-on training material was conceived in such a way to be available and replicable anywhere in the world. These characteristics are especially relevant in developing countries where data is often scarce and resources for buying software are limited. Winners were guided by mentors to ensure that the material met the requirements of the target audience. In this chapter, the material developed for the third challenge is described.

The training material refers to SDG 9 (Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation), and more specifically to the computation of indicator 9.1.1 (Proportion of the rural population who live within 2km of an all-season road) for a rural area of Tanzania. The methodology and technical tools (data and software) in this use case are presented in detail, so that anybody can replicate the method for other parts of the world. Since it is based on open data and open software, the method is cost effective and completely sustainable. Moreover, the possible lack of data can be overcome by actively involving people and communities in mapping their specific region of interest with freely available tools.

In the remainder of the chapter we present concepts and examples of open data and open software, followed by the concept of community geospatial science with reference to citizen science and volunteered geographic information (VGI). Next, we describe the use case from the training material and then conclude.

20.2 Open Data and Open Software

The principles of openness and transparency are widely advocated these days - as in open data, open software, open knowledge and open government - but what do they really mean?

The architects of the twenty-first century digital age proclaim that openness is their foundational value. The technological foundations that sustain this vision of openness are digital: the internet, mobile telephony and distributed systems. According to Russell (2014), ‘openness’ is a “marriage of technology and ideology and a fusion of technology, democracy, and entrepreneurial capitalism” [19]. The work described in this chapter exemplifies open principles: for the use case, open data is collected and analysed with open source software, and the training material is made available as an open educational resource.

Open source software has its origins in the early days of computing when programming problems were solved through scientific collaboration. Software was shared and each programmer added a new aspect to existing knowledge [4]. It evolved into a software development and licensing approach that ensures transparency through access to the source code and collaboration through a set of rights that protect the copyright to the source code. Through the free redistribution of the software and works derived from it, it is possible to create software products based on each other's work [18].

The interesting point, especially from the point of view of the poorest countries with limited resources for technologies, is that there is at least one mature sophisticated open source product for every geo-technology area and geospatial information need and application - from data collection in the field, crowdsourcing, desktop applications, spatial extensions to database management systems and software stacks. Together, they can be used to create sophisticated free and open Web and cloud-based systems [12, 21]. In developing countries, economic motivations rank high when choosing to use free and open source geospatial software [3].

OSGeo (Open Source Geospatial Foundation) is a not-for-profit organization aimed at fostering the global adoption of open geospatial technology. According to OSGeo, open source starts off with a license that provides royalty free (re)use of software. Next, open source guarantees access to the source code for audit and modification and the ability to redistribute the software with no additional costs¹. A wide range of open licenses are in use. Creative Commons licenses², a set of well-defined licenses that each describe a different permitted use of copyrighted materials by the public at large are widely used for content. The training material described in this chapter is made available under one of these Creative Commons licenses. For source code, OSI-approved licenses, such as the GNU General Public License (GPL), are frequently used³.

Besides software, open data or knowledge is based on the principle that some information should be shared and available to anyone, without any restrictions to rights of access or use. According to Open Knowledge International, open data and content can be freely used, modified, and shared by anyone for any purpose [17]. Generally, transparency and collaboration are well aligned with the principles that democratic governments stand for and with the principles embodied in the Charter of the United Nations. Organizations, such as Open Knowledge International, promote the use of open data and knowledge, e.g. to support citizens in taking action on social problems. Monitoring SDGs with open data makes it possible for citizens to track the status of SDGs in their countries or cities. It empowers them with an understanding of the challenges at hand so that they can work on addressing these, for example, by supporting or lobbying for appropriate initiatives or by holding governments and other actors to account. OpenStreetMap, started in

¹www.osgeo.org

²<https://creativecommons.org/licenses/>

³<https://opensource.org/licenses>

the UK in 2004, is the most widely known example of global open geospatial data. The success of its open characteristics is further described in the next section.

Through open education, barriers to education are removed by making educational resources freely available for anyone to study and use or by eliminating admission requirements. Through GeoForAll, OSGeo promotes open education among its members based on the belief that knowledge is a public good and that open principles in education provide opportunities for everyone. Teachers and educators provide open access to their educational resources for teaching related to geo-technologies and geospatial data. Access to education is a challenge in the poorest countries, confirmed by the UN SDG 4 on Education. Open and freely available educational resources contribute to addressing this challenge.

The UN OSGeo Committee works towards identifying and developing open source geospatial software and services that meet the requirements of UN operations, taking full advantage of the expertise of mission partners, including partner nations, technology developed by contributing countries, international organizations, academia, NGOs, and the private sector ⁴. The 2018 Challenges were aimed at supporting these goals.

20.3 Community Geospatial Science

Public participation in scientific achievements has a long history but the last few decades have seen more attention and an impressive increase in the number of people involved. Citizen science, i.e. scientific research conducted, in whole or in part, by amateur (or non-professional) scientists, the term used for denoting such an approach, is a diverse practice, encompassing various forms, depths, and aims of collaboration between scientists and citizen researchers and a broad range of scientific disciplines [8].

Different classifications of citizen science projects exist based on the degree of influence and the extent of the contributions by citizens. Haklay *et al.* (2018) distinguish three kinds of citizen science projects [9]:

1. Long-running citizen science, the traditional projects, similar to those run in the past [11, 1].
2. Citizen cyberscience, strictly connected with the use of technologies [6], and which can be subclassified into:
 - (a) volunteer computing, where citizens offer the unused computing resources of their computers;

⁴<https://wiki.osgeo.org/wiki/UnitedNations.Committee>

- (b) volunteer thinking, where citizens offer their cognitive abilities for performing tasks difficult for machines; and
 - (c) passive sensing, where citizens use the sensors integrated into mobile computing devices to carry out automatic sensing tasks.
3. Community science, involving a more significant commitment from citizens, also in designing and planning the project activities in a more egalitarian (if not bottom-up) approach between scientists and citizen scientists [10, 13, 2]. Community science is further classified into:
- (a) participatory sensing, where citizens use the sensors integrated into mobile computing devices to carry out sensing tasks;
 - (b) Do-It-Yourself (DIY) science, which implies that participants create scientific tools and methodologies for carrying out their research; and
 - (c) civic science, which is “explicitly linked to community goals and questions the state of things” [9].

Because of the bottom-up approach, community science is the most interesting when it gets to activities and programs that are closely related to the life and well-being of people. If a geospatial dimension is involved, i.e. location plays a fundamental role in interpreting the phenomena under study, we can refer to this as community *geospatial* science.

The best example of community geospatial science is OpenStreetMap (OSM) ⁵. Many people consider it to be an object (i.e. a map or its modern version, a geodatabase): “a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license” [5]. It is also commonly referred to as a geo-platform or project where as many as 5 million users contribute, edit, download and assess the data that is shared. OpenStreetMap is most of all a community of communities [20], in the sense that the OpenStreetMap community is diverse and incorporates the motivations of many different groups, depending on how they approach their volunteer activity.

Examples of communities are the community “dedicated to humanitarian action and community development through open mapping”, the Humanitarian OpenStreetMap Team (HOT) ⁶; the community that is “putting the world's vulnerable people on the map”, Missing Maps ⁷; the community that works to close the gender gap in OSM, GeoChicas⁸; the community of university students, YouthMappers ⁹; the community that helps end female genital mutilation and aids community development in rural Tanzania, Crowd2Map¹⁰.

⁵<https://www.openstreetmap.org>

⁶<https://www.hotosm.org/>

⁷<https://www.missingmaps.org/>

⁸<https://wiki.openstreetmap.org/wiki/GeoChicas>

⁹<https://www.youthmappers.org/>

¹⁰<https://crowd2map.wordpress.com/>

Starting from the community level, OSM has been able to attract the attention of institutional entities, ranging from small local ones (such as the local civil protection agencies) to national mapping agencies and international organisations, like the World Bank and the UN (see for example the Open Cities Africa project¹¹) [7]. Universities are often involved in community geospatial science, even if their role is that of co-creators of solutions together with the other involved actors.

Apart from the social ecosystem that has been established in this global community geospatial science experience, the technological ecosystem that has been established is worth mentioning. One of the main reasons for the success of OSM is that the technology behind the project allows everybody to contribute, independently of their level of expertise. The tools and systems, developed by different actors in the social ecosystem of OSM (volunteers; small and medium companies; universities; local, national or international agencies) are generally characterized by being free and open source, i.e. they can be passed on for further development by other people in the community; and by the different applications very often accessible simply through the personal account on the OSM platform.

Apart from facilitating contributions by individuals, the OSM ecosystem is designed to elicit and simplify collaboration. One fundamental tool (the Tasking Manager) for instance allows the subdivision of large areas to be mapped into a set of smaller ones, each of a size that can be mapped by an individual. This facilitates collaboration among mappers and avoids problems of overlap and confusion. Moreover, this tool allows the validation of the mapped data, so that a quality assessment of the mapped area is possible.

Citizen science and community geospatial science represent a new step in the history of science and these examples, like OSM and its communities, are relevant cases of what can be done within the new paradigm of collaboration and openness.

20.4 The Use Case and Training Material

The UN 2030 Agenda defines the challenges to be overcome in order to achieve prosperity for all in a sustainable manner for the entire planet up to the year 2030 [14]. These challenges are embodied in the 17 SDGs that are broadly interdependent and address all aspects, from poverty to peace and justice, from environmental protection to human health, from food security to gender equality. To track progress towards achieving the 2030 Agenda, it is essential to understand the attainment level for each of the 17 SDGs. Progress can be

¹¹<https://opencitiesproject.org/>

quantified by measuring, gathering data and calculating the indicators that define each goal using a consolidated methodology [15].

As the 2030 Agenda fulfilment has a worldwide scope, the monitoring and reporting activities must be done globally. Yet, the quantity, quality, precision and consistency of the necessary data vary significantly across the globe and so do the resources to gather it. However, in the last two decades, we have witnessed a significant increase in the availability of open data, from open public data to citizen science data, and in the case of geoscience, satellite imagery and related products ¹². In this context and within the framework of the third OSGeo UN Committee Educational Challenge related to open geospatial data and software for UN SDGs, we developed a replicable use case to demonstrate that open geospatial data, which very often are contributed by citizens and communities, are available globally; that free and open source solutions for geospatial have sufficiently developed to conduct a global geospatial analysis at small and intermediate scales; and that these data and software can be used to monitor a geospatial SDG indicator.

The selection of the indicator has been done following these guidelines:

1. to have a spatial dimension;
2. to avoid indicators that are already addressed through an advanced initiative, such as the GEO Wetlands Initiative¹³, WHO Interactive Air Pollution Maps¹⁴, GEO AquaWatch¹⁵ or ESA CoastColour¹⁶.

Indicator 9.1.1, Proportion of the rural population who live within 2 km of an all-season road (C0901010), was selected, which supports the target of developing quality, reliable, sustainable and resilient infrastructure, including regional and transborder infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all. In December 2018, the UN published an updated but still in progress work on the metadata for indicator 9.1.1. that includes clear definitions for the terminology, methodology, data sources and their availability [16]. A significant limitation is highlighted: “*The Indicator relies substantially on data collected by road agencies and national statistics offices for their operational work. As such, its update is dependent on the frequency of update of the road condition surveys and national census.*” ([16], subchapter Methodology, paragraph Comments and limitations). This is exactly the type of limitation we aim to address with our use case. In comparison to data collected by a multitude of national agencies, globally produced datasets have the advantage of worldwide coverage and a coherent and consistent technical and logical structure. Yet, both kinds of dataset can have a similar scope. Even though the open dataset is the result

¹²https://wiki.osgeo.org/wiki/Training_Material_for_UN_Open_GIS_OpenData

¹³Retrieved from <https://www.earthobservations.org/activity.php?id=122> on 3.02.2019

¹⁴Retrieved from <http://maps.who.int/airpollution/> on 3.02.2019

¹⁵Retrieved from <https://www.geoaquadwatch.org/> on 3.02.2019

¹⁶Retrieved from <http://www.coastcolour.org/> on 3.02.2019

of a collaborative community effort, its structure and semantics are described and allow a clear understanding of it. Thus, reliable cleaning of the data can be done, if necessary.

For the present use case, we selected Tabora Region, one of the 31st administrative regions in the central-western part of Tanzania. However, the same use case is replicable for any other region in the world, provided the data are available. This is possible because we based the computation only on open global datasets. In [Table 20.1](#), we show the geospatial information required to calculate the Rural Access Index (RAI) for a region, together with the open global datasets in our use case.

The most challenging geospatial information to obtain is the road network and the road condition. At the moment, the only global dataset that could provide such information is OpenStreetMap. The amount and quality of OpenStreetMap data for various regions around the world can vary significantly. As our region of interest is located in Africa, specific OpenStreetMap developments for Africa must be taken into consideration, namely the Highway Tag Africa Typology of Road Network in African countries¹⁷, a roads classification designed for the context of African countries, and the East Africa Tagging Guidelines¹⁸, which provide guidance for tagging roads in Kenya, Tanzania, Uganda, Rwanda, Burundi, and South Sudan. Even though the road classification is specific to Africa, the clear and consistent definition of each element¹⁹ and tag²⁰ in OpenStreetMap makes this use case reproducible in any other part of the world. With this in mind, apart from describing the processing steps for the Tabora Region RAI calculation in detail, we present the conceptual workflow for calculating the RAI in [Figure 20.1](#).

The workflow consists of three main stages: (1) preparation of the geospatial data; (2) calculation of the RAI; and (3) presentation of the results. The first stage is the most time consuming as preparing the data implies a thorough analysis, ranging from the structure (i.e. format) to aspects related to consistency, precision, scale etc., and subsequent cleaning of the dataset. The outcome of the first stage is a dataset with only the information necessary for the analysis, any redundant information is removed.

The use case and the training material are based on a well-known free and open source GIS package, QGIS²¹. For the RAI calculation, a topologically correct road network is desirable, but not necessary. Because OpenStreetMap is an open collaborative mapping project, there may be inconsistencies in the data. After close analysis of the Tabora region road network, a series of in-

¹⁷Retrieved from https://wiki.openstreetmap.org/wiki/Highway_Tag_Africa on 3.02.2019

¹⁸Retrieved from https://wiki.openstreetmap.org/wiki/East_Africa_Tagging_Guidelines on 3.02.2019

¹⁹An element is the basic component of the OpenStreetMap conceptual data model of the physical world

²⁰A tag describes the element to which is attached and it is defined by a key and a value that are conventions agreed upon by the OSM community and openly published on the OSM wiki.

²¹Retrieved from <https://qgis.org/en/site/> on 3.02.2019

TABLE 20.1
Geospatial information and corresponding open datasets used in the RAI calculation for the Tabora Region. Data from different sources: 1) *gadm.org*, 2) *worldpop.org.uk*, 3) *sedac.ciesin.columbia.edu*, and 4) *openstreetmap.org*

Theme	Geospatial Information	Dataset Used	RAI Layer	Producer / Collector of Data
Administrative Units	Administrative Units	Database of Global Administrative Areas ¹	Administrative Units	University of California, Berkeley, Museum of Vertebrate Zoology, and the International Rice Research Institute (Global Administrative Areas 2009)
World Population	Estimates of numbers of people per grid square	WorldPop ²	Population Numbers	GeoData Institute, University of Southampton
World Population	Polygon representation of urban areas with city or agglomeration name and time series	Global Rural-Urban Mapping project (GRUMP), v1 ³	Urban Geometries	Socioeconomic Data and Applications Center (SEDAC)
General Geospatial Data	OpenStreetMap is built by a community of mappers that contribute and maintain data about roads, trails, cafés, railway stations, and much more, all over the world	OpenStreetMap ⁴	Road Network, Road Condition	OpenStreetMap contributors

FIGURE 20.1
Conceptual workflow for calculating the RAI

consistencies were identified and corrected to eliminate as much as possible any artificial results in the RAI. Even though the use case was developed for a particular region, the types of inconsistencies addressed include the most common situations, thus making the study replicable. The types of inconsistencies and solutions for addressing each one of them are presented in detail in the training material.

The geospatial layer of the RAI is obtained by dividing the rural population within the 2km buffer area around the all-season roads in a specific administrative unit by the total rural population for that administrative unit. The last step of the workflow provides suggestions for presenting the results of the RAI analysis (Figure 2) so that the indicator can be used as the basis for future management and development of the region in question. The complete training material, for those who want to repeat the calculation of the RAI for Tabora or do something similar in other regions of the world is available at the website²².

20.5 Conclusion

The Tabora region use case for calculating an SDG geospatial indicator proves that through the exclusive use of open global datasets, some of which contributed by citizens, and free and open source software, complex geospatial analyses can be conducted to better understand, manage and protect our environment. The use case was deliberately developed using only global datasets so that it can be replicated for any other region in the world.

²²https://wiki.osgeo.org/wiki/Training_Material_for_UN_Open_GIS_OpenData

Undoubtedly, the issue of data quality remains when considering an open collaborative environment such as OpenStreetMap. However, this kind of initiative can and should work as a driving force towards improving the open datasets, either by defining new significant attributes - referred to as tags, in OSM – or by cleaning and maintaining the attributes in the dataset. Working with citizens and communities like those surrounding OpenStreetMap, following a community geospatial science paradigm, “will ensure that the challenges are addressed for all populations in different locations, leaving no one behind” (Rajabifard, 2019).

Bibliography

- [1] Rick Bonney, Heidi Ballard, Rebecca Jordan, Ellen McCallie, Tina Phillips, Jennifer Shirk, and Candie C Wilderman. Public participation in scientific research: Defining the field and assessing its potential for informal science education. a caise inquiry group report. *Online Submission*, 2009.
- [2] Jessica Breen, Shannon Dosemagen, Jeffrey Warren, and Mathew Lipincott. Mapping grassroots: Geodata and the structure of community-led open environmental science. *ACME: An International E-Journal for Critical Geographies*, 14(3), 2015.
- [3] Junyoung Choi, Myung-Hwa Hwang, Hyungtae Kim, and Jaeseong Ahn. What drives developing countries to select free open source software for national spatial data infrastructure? *Spatial Information Research*, 24(5):545–553, 2016.
- [4] Arnulf Christl. Free software and open source business models. In *Open source approaches in spatial data handling*, pages 21–48. Springer, 2008.
- [5] OpenStreetMap Wiki contributors. About openstreetmap, 2017.
- [6] François Grey. Viewpoint: The age of citizen cyberscience. *Cern Courier*, 29, 2009.
- [7] The World Bank Group. Measuring rural access: using new technologies. Technical report, Washington, D.C 20433, 2016.
- [8] Trisha Gura. Citizen science: amateur experts. *Nature*, 496(7444):259–261, 2013.
- [9] Mordechai Muki Haklay, Suvodeep Mazumdar, and Jessica Wardlaw. Citizen science for observing and understanding the earth. In *Earth Observation Open Science and Innovation*, pages 69–88. Springer, Cham, 2018.

- [10] Paul Jepson and Richard J Ladle. Nature apps: Waiting for the revolution. *Ambio*, 44(8):827–832, 2015.
- [11] Hiromi Kobori, Janis L Dickinson, Izumi Washitani, Ryo Sakurai, Tatsuya Amano, Naoya Komatsu, Wataru Kitamura, Shinichi Takagawa, Kazuo Koyama, Takao Ogawara, et al. Citizen science: a new approach to advance ecology, education, and conservation. *Ecological research*, 31(1):1–19, 2016.
- [12] Rafael Moreno-Sanchez. Free and open source software for geospatial applications (foss4g): A mature alternative in the geospatial technologies arena. *Transactions in GIS*, 16(2):81–88, 2012.
- [13] S Nascimento, Â Guimarães Pereira, and A Ghezzi. From citizen science to do it yourself science; 2014.
- [14] United Nations. Transforming our world: the 2030 agenda for sustainable development. *United Nations General Assembly. Resolution 70/1*, 2015.
- [15] United Nations. Work of the statistical commission pertaining to the 2030 agenda for sustainable development. *United Nations A/RES/71/313*, 2017.
- [16] United Nations. 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. *United Nations, Department of Economic and Social Affairs, Statistics Division*, 2018.
- [17] Open Knowledge Foundation (OKF). The open definition, 2018.
- [18] OSD. The open source definition, 2018.
- [19] Andrew L Russell. *Open standards and the digital age*. Cambridge University Press, 2014.
- [20] Patricia Sols. Building mappers not just maps: challenges and opportunities from youthmappers on scaling up the crowd in crowd-sourced open mapping for development, April 2017. AAG Mapathon; AAG Annual Meeting.
- [21] Stefan Steiniger and Andrew JS Hunter. The 2012 free and open source gis software map—a guide to facilitate research, development, and adoption. *Computers, environment and urban systems*, 39:136–150, 2013.
- [22] OSGeo Website. Osgo un committee challenge, 2018.