OVERCOMING DATA SCARCITY FOR ENERGY ACCESS PLANNING WITH OPEN DATA – THE EXAMPLE OF TANZANIA

C. Cader [a *], S. Pelz [a], A. Radu[b], P. Blechinger[a]

[a] Reiner Lemoine Institut gGmbH, Berlin, Germany –
(catherina.cader, setu.pelz, philipp.blechinger)@rl-institut.de
[b] alinradu86@gmail.com
*Corresponding author

Commission IV, WG IV/4

KEY WORDS: Open data, GIS for energy access planning, Mini-grids, Tanzania, SDGs

ABSTRACT:

The achievement of the United Nations Sustainable Development Goals (SDGs) is heavily influenced by access to data: Data is necessary to assess the current status quo as well as to measure progress and to find opportune contextualized solutions for development challenges. Specifically, the lack of energy access (as defined in SDG7) is an immense bottleneck for development in Tanzania and considering spatial planning plays a crucial role in locating the most appropriate electrification solution for each site; taking into account not only its inherent characteristics, such as local demand for electricity and economic activities, but also external factors such as the distance to existing energy transmission and distribution infrastructure.
Data in Tanzania is scarce, and this paper is an attempt to analyze the potential of open data sources to increase data availability to eventually provide improved foundation for decision making and investment flows for electrification planning.
Results show that data quality of the given sources is sufficient for providing a novel level of disaggregated spatial information which can serve as an additional information stream for all involved stakeholders.
From both perspectives, with national planning on the one hand and bottom-up initiatives on the other hand, it is important to understand the spatial aspects of any planning scheme to guarantee that a successful implementation phase will follow the planning stage.

1. INTRODUCTION

1.1 Problem definition

Improving access to modern energy services for all is not only a key challenge in itself, it is also strongly linked to other aspects of global development such as health, education and equality (McCollum et al., 2017). Alongside enabling factors such as regulatory support and market development, long-term improvement of electricity access requires realistic planning based on accurate data. In this paper we consider the importance of spatial data in energy access planning, where location specific, geo-referenced data has been proven crucial to understand the contextual challenges and opportunities (Mentis et al., 2016; Moksnes, Korkovelos, Mentis, & Howells, 2017, Zwoleff et al., 2009).
For improving energy access, generally two different approaches can be differentiated: an extension of the centralized, often nation-wide power grid infrastructure plus the installation of additional large-scale power plants to increase the electricity generation, opposed to decentralized energy systems, which can be composed of different technologies, such as renewable energy technologies (e.g. photovoltaics), battery storage and / or diesel generators, available in a large range of sizes. With regard to the size, systems may be small such as solar lanterns or solar home systems, or they can consists of different technology components and supply a larger demand for electricity, for example for a village. These systems are not interconnected to a superordinate system and can run in an island mode.
The employment of both options, central and decentralized systems, will eventually improve the energy access situation of a country or region; however, these variable solutions will result in different scenarios in terms of economic and ecological performance, as well as in their project implementation period and possible funding structures and related business models (Ahlborg & Hammar, 2014). The most important basis for identifying and evaluating decisions on the appropriateness of certain solutions for corresponding areas is a comprehensive data basis.
Specifically, we focus on the challenge of data scarcity, and methods to overcome this in Sub-Saharan Africa, using the case of Tanzania, where 67 % of the 56 million total population still lack access to electricity, with even lower access rates of 17 % in rural areas (IEA, 2017).

1.2 Novel approaches with open data

The base spatial data sets for energy access planning are settlement locations, population and status of electrification – to find out where energy access is still lacking, as well as existing and planned energy infrastructure, such as power plant locations and transmission and distribution grid infrastructure. On top of these base layers, other data sets can reveal more detailed information such as demographic and socioeconomic data and migration models which are key to estimating the potential demand and crucially, demand growth for electricity for a given settlement or region. Such information; which may be available for some countries, is often lacking for countries in the Global South due to general data scarcity, obsolete data or data with ambiguous geo-references (Tatem & Linard, 2011).

One option to tackle these challenges is the recent global development of open source geo-spatial data sets, for example data sets derived from freely available satellite imagery such as

Landsat or Sentinel, or the development of crowdsourced databases, e.g. OpenStreetMap (OSM). In this paper, we conduct a comparative analysis of deriving base data for energy access planning by using the high resolution settlement layer (HRSL) and OSM in combination with energy, health and water infrastructure data and administrative data extracts for the case of Tanzania. By using open source software for geo-data analysis and processing in combination with freely available datasets, simple replicability is given to conduct this analysis for other countries or regions of interest and to create transparency about the derived data sets by enabling replication and validation.

2. METHODS

2.1 Input data collection

One key dataset for the analysis is the High Resolution Settlement Layer[1] (HRSL). This raster dataset derives human settlements from high resolution satellite imagery and assigns population values scaled by national census data of the respective country. The HRSL contains population information with a resolution of approximately 30 m per pixel at the equator (1 arc-second).
For the administrative delimitation and reference of respective settlements data openly provided by the National Bureau of Statistics of Tanzania was used. Vector files with the respective village polygons are used. They include information on the number of people in each; however, the information is only available for the polygon without a higher resolution of the data on how the population is distributed within each of the polygons. Some villages may have an overall denser structure population structure, while others may show a circular shape in contrast to others which are stretched out long formed in the shape of a valley or alongside a road.
Openly available OSM data is currently available online for some regions in Tanzania[2]. As a crowd-sourced data platform, OSM data sets are open and additional data or more detailed information is added by a large user group globally.

2.2 GIS processing

In order to obtain data on village level instead of pixel level, we have resampled the HRSL data to a lower resolution, by aggregating pixels into more uniformly populated places. The resampling method was done with the QGIS warp function (Bilinear interpolation and a factor f=2). The resulting file was then vectorized with the purpose of obtaining polygons with population information.
Since the obtained vector dataset still contained a very large number of entries (some of them irrelevant to our research due to wrong classification or insignificant size), smaller structures such as isolated farms (which may be likely to be electrified via solar home systems) were eliminated. As a further step towards understanding the population data, the obtained vector file was buffered and used to calculate zonal statistics (in combination with a vector file, “TZ Villages” which included the official administrative polygons and census data). As an output, we obtained population data statistics based on each village polygon.
The grid data shapefile contains information from four sources: AICD, REA, World Bank and OpenStreetMap. However, the AICD and World Bank data is not accurate, therefore it was excluded. Moreover, the 600 kV lines are mostly mislabeled and corrected respectively.
In order to calculate the distance from each population cluster to the grid, we have applied QGIS algorithms: V.distance was used to calculate the actual distance from the centroids of each population cluster to the gridlines.

2.3 Webmap development

The processed datasets resulting from the applied methodology are published within an online webmap[3]. For the development of this two software packages were used:

- Leaflet[4] - an open-source JavaScript library for mobile-friendly interactive maps. Licensed under BSD 2-Clause. Copyright (c) 2010-2017, Vladimir Agafonkin
- tippecanoe[5] - to build vector tilesets from large collections of GeoJSON features. Licensed under BSD 2-Clause. Copyright (c) 2014, Mapbox Inc.

The webmap allows an interactive exploration of the spatial data without the requirement of specific GIS knowledge: different zoom levels allow on the one hand gaining insights on a national level, while detailed information on sub-regions can also be assessed in a higher degree of accuracy. In addition, different information layer can be turned on- and off in response to specific interest of the respective viewer.
Also, dynamic filtering allows to define own threshold criteria for certain numerical updates.

3. RESULTS

The results of the analysis showcase how openly available data can improve the available database for energy access planning across geospatial areas and regions, alongside the example of Tanzania.

3.1 Data availability

The HRSL data set has a global coverage and a very high resolution, leading to a comparably data quality across space. Due to the crowd-sourced character of OSM data on the other hand, data quality and available data quantity is much harder to assess, since it might differ extensively from one region compared to another regions.

3.2 Results of village clustering

A comparison of the total aggregated village population figures of Tanzania from official national sources compared to the total population extracted from the spatially disaggregated HRSL layer shows a positive match: The official population is defined at 55.6 million, while HRSL data accumulates to 53.3 million. The clustering process covered ca. 80% of Tanzania’s population (41 million people), while about 20 % of the population is identified to live in less dense settlements in a highly dispersed sourrounding.

[1] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL© 2016 DigitalGlobe. https://ciesin.columbia.edu/data/hrsl/#data. Accessed 10th January 2018.
[2] www.openstreetmap.org. Accessed 20th April 2018.
[3] https://catcad.github.io/maptz/ Accessed 20th April 2018.
[4] http://leafletjs.com/. Accessed 20th April 2018.
[5] https://github.com/mapbox/tippecanoe. Accessed 20th April 2018

The resulting dataset (which includes a column with the distance to grid measurements) was then joined with the original cluster file, based on unique IDs. The on and off grid data was then split in two different files, for a better overview, by selecting the clusters based on its connectivity status. The "on-grid" shapefile contains ca. 7,000 grid-connected clusters. The "off-grid" shapefile contains 29,000 entries. Population in grid-connected clusters sums up to about 28 million people, while population in clusters without grid connection is ca. 13 million. Here it needs to be considered that grid-connected does not necessarily imply electricity access, since grids might be temporarily or
permanently undersupplied. A significant majority of all clusters has a population below 1,000 (Tab. 1).

	10 km grid distance		**20 km grid distance**	
Cluster population	[num]	[people]	[num]	[people]
0-999	8,095	1,397,744	3,688	577,414
1,000-1,999	468	657,916	190	269,281
2,000-4,999	285	854,640	118	376,989
5,000-9,999	83	559,855	48	328,023
10,000-19,999	28	360,715	15	191,337
>20,000	13	1,212,391	9	1,107,418
Total	**8,972**	**5,043,260**	**4,068**	**2,850,463**

Table 1. Number and population of population clusters in 10 and 20 kilometre distance to the grid; categorized according to the cluster population size.

From the cluster shapefile, some entries were deleted, since they were either not within the borders of Tanzania or were not covering built-up areas. Another encountered issue was the fact that island cities and villages were considered "off grid" and showed abnormal distances to the grid, although they are actually connected. This was causing errors in the ranking system so they were removed. Moreover, the sites with less than 100 inhabitants were also removed from the dataset (some of them were inaccurately classified as villages or out of scope for mini grid development).

3.3 Validation with official figures

By comparing the deviation of the calculated population layer within the administrative village polygons with the official population values, the high accuracy of the HRSL data becomes apparent: The largest share of cluster are within a threshold of 500 people higher or lower than the provided figures.

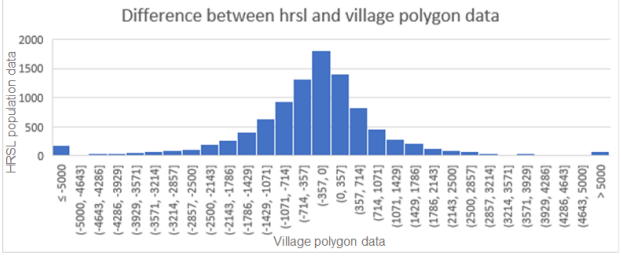

Figure 1. Difference of HRSL population data and village polygon data with official figures for each village. Some villages lack the official figures.

The largest variance occurs for sites where no population is given in the official statistics, which lead to deviations of < 5,000 people (Fig. 1).

The overlay of the OSM data with the official village polygons shows that in some cases, village structures have evolved across the borders of administrative villages (Fig. 2).

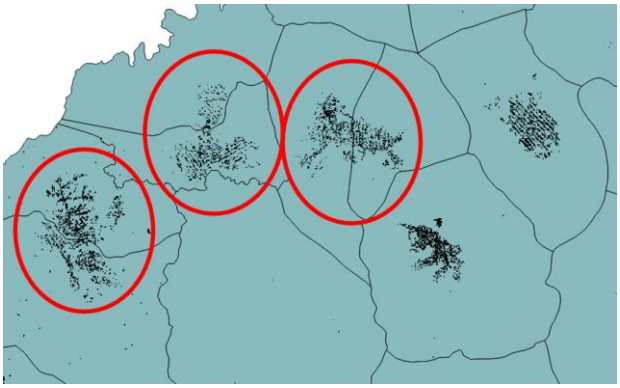

Figure 2. The red circles highlight cases where cross-border village structures are identified in contrast to detected villages solely within their administrative spatial extent.

This specific characteristic needs to be considered for energy access planning, since the option to supply agglutinated neighboring village structures in one instance to decrease costs.

3.4 Resulting webmap

The resulting layers, showing administrative; population and electricity access level information, together with the grid infrastructure, were then plotted in a webmap overlaid on OpenStreetMap, with filtering options based on the presence of social infrastructure, population and the distance to the grid (Fig. 3).

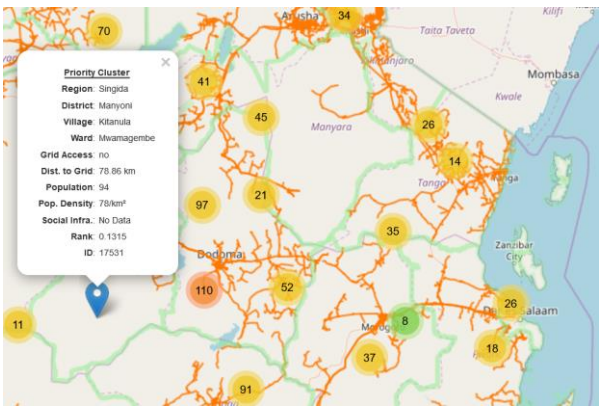

Figure 3. Screenshot of the interactive web-map, zoomed into eastern Tanzania and showcasing all village attributes of one village cluster in the pop-up.

4. DISCUSSION

The various datasets are characterized by different advantages and disadvantages – while the main limitation for the crowd-sourced OSM data is a lack of universal coverage and quality differences for different areas, the HRSL data only considers settlements and population structure without including other attributes useful for energy access planning. The analysis for the case of Tanzania illustrates that HRSL and OSM data can be combined with existing energy infrastructure data to improve base data for energy access planning for cases with limited or no data availability from other data sources such as the respective ministries.

This case study provides insight into the challenge of combining open data sets and ambiguously licensed or restricted ministerial data, such as detailed geo-spatial information on planned electrification projects such as grid extension or decentral mini-grid development. A positive effect from open licenses of datasets, is that the results may be published as an open source web map, which can offer relevant information to different stakeholders, such as governments or private sector companies. As an example, we show how the combined dataset enables the identification of priority settlements for mini-grid electrification which can be useful for both private sector developers, and rural energy access planners in establishing a project pipeline and zoning for future grid expansion. The results obtained from the analysis based HRSL and OSM data are validated against data sets available for Tanzania through other, country-specific open data portals.

5. ACKNOWLEDGEMENTS

This work results partly from a project financed by the International Finance Corporation. The authors gratefully acknowledge the support from the International Finance Corporation. Any conclusions or recommendations expressed in this paper do not necessarily reflect the views of International Finance Corporation.

6. REFERENCES

Ahlborg, Helene & Linus Hammar (2014). Drivers and barriers to rural electrification in Tanzania and Mozambique – grid-extension, off-grid, and renewable energy technologies. Renewable Energy 61. World Renewable Energy Congress – Sweden, 8–13 May, 2011, Linköping, Sweden, pp. 117–124.

IEA (2017). World Energy Outlook – Energy access database.

McCollum, D., Gomez Echeverri, L., Busch, S., Pachauri, S., Parkinson, S., Rogelj, J., Stevance, A.-S. (2017). Connecting the Sustainable Development Goals by their energy inter-linkages (IIASA Working Paper No. WP-17-006). Laxenburg, Austria: IIASA.

Mentis, D., Andersson, M., Howells, M., Rogner, H., Siyal, S., Broad, O., Bazilian, M. (2016). The benefits of geospatial planning in energy access - A case study on Ethiopia. Applied Geography, 72, 1–13. https://doi.org/10.1016/j.apgeog.2016.04.009

Moksnes, N., Korkovelos, A., Mentis, D., & Howells, M. (2017). Electrification pathways for Kenya – linking spatial electrification analysis and medium to long term energy planning. Environmental Research Letters. https://doi.org/10.1088/1748-9326/aa7e18

Tatem, A., & Linard, C. (2011). Population mapping of poor countries. Nature, 474, 36.

United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. URL: https://sustainabledevelopment.un.org/content/documents/21252030%16620Agenda % 20for % 20Sustainable % 20Development % 20web . pdf (visited on January 5, 2018).

Zvoleff, Alex, Ayse Selin Kocaman, Woonghee Tim Huh, & Vijay Modi (2009). The impact of geography on energy infrastructure costs. Energy Policy 37 (10), pp. 4066–4078.