Data Release Statement

Automatic national census pre-Enumeration Areas for Zimbabwe (2021)

version 1.0

WorldPop, University of Southampton

05/12/2024

This document outlines the data sets and methods used to generate automatic national census pre-Enumeration Areas for Zimbabwe in 2021. Assumptions, limitations, and use constraints are provided.

Enumeration Areas (EAs) are the operational geographic units for collecting and disseminating census data and are often used as a national sampling frame for various types of surveys. In certain low income or conflict-affected countries, EA demarcations are incomplete, outdated, or missing. Even for stable and middle or high-income countries, creating and updating EAs is a challenging yet essential task in preparing for a national census. Commonly, EAs are created by manually digitising small geographic units on high-resolution satellite imagery or physically walking the boundaries of units, both of which are time, cost, and labour intensive. We have developed a user-friendly tool that could be employed to generate draft EA boundaries automatically. The tool is based on high-resolution gridded population and settlement datasets, GPS household locations, building footprints and uses publicly available natural, man-made and administrative boundaries. Initial outputs were produced in multiple countries around the world including Burkina Faso, Somalia, Paraguay, Togo, Zimbabwe, Niger and Guinea. The results indicate that the EAs are in line with international standards including boundaries that are following ground features and easily identifiable, no overlaps, the boundaries are nested within administrative boundaries, and they are compact and free of pockets and disjoints.

Tool applications:

1- Generate national census preEAs in countries where the dataset does not exist
2- Update national digitised census EA in countries where the dataset is outdated
3- Generate national population sampling frame to serve health and socio-economic survey.

Suggested Citation

Qader S H, Kuepie K, and Andrew J. Tatem (2024). Automatic national census pre-Enumeration Areas for Zimbabwe in 2021, version 1.0. WorldPop, University of Southampton. Doi: 10.5258/SOTON/WP00797

Data Use Constraints

data provided; and shall not be liable for incidental, consequential, or special damages arising out of the use of any data that they offer.

Contacts and Data Queries

WorldPop appreciates feedback regarding this data set, such as suggestions, discovery of errors, difficulties in using the data, and format preferences.

Please contact: WorldPop Email: release@worldpop.org for more information.

1. Introduction

These data were produced by the WorldPop Research Group at the University of Southampton. This work was part of the Geo-referenced Infrastructure and Demographic Data for Development (GRID3) programme, supported with funding from the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (OPP1182425). Programme partners included the United Nations Population Fund (UNFPA), the Center for International Earth Science Information Network (CIESIN) within the Earth Institute at Columbia University, and the Flowminder Foundation.

For the creation of automatic pre-EAs in BFA, the pre-EA tool was employed. This tool has been developed in a form of QGIS plugin and it has been implemented in QGIS platform. Currently, the tool is not publicly available, but it is our aim to make it publicly available once all the development and accompanying user guide are finalised.

2. Input Data

- National Commune administrative boundaries obtained from HDX (HDX 2021).
- Digitized road types across ZWE extracted from Open Street Map (2020) (OSM, 2015).
- Rasterized building footprints obtained from WorldPop (Dooley et al. 2021)

3. Methods Overview

WorldPop at the University of Southampton in close collaboration with UNFPA and GeoData have developed an automatic designation tool for delineating census pre-Enumeration Areas (EAs) and national population sampling frames in the absence of recent census data. This tool was employed to generated national census pre-EAs and building blocks for entire ZWE.

The workflow can be summarized in two main steps:

i. Data preparation, cleaning and pre-processing

This step includes the collation of accessible datasets and GIS processing of these to check and correct geometry (as needed) and thus to create appropriate inputs to the pre-EA tool. These data preparation and processing includes:

- Clip all data lines to the official ZWE country boundary
- Re-project the input data to WGS 1984 UTM Zone 35S
- Generate single line road features in place of matched pairs of divided road lanes (to solve issues of the double or triple parallel highway or motor way lines) using Merge Divided Roads tool in ArcGIS
- In the case of Zimbabwe, uncrossable features such as main roads and rivers were tested in the creation of preEAs. This is because, Zimbabwe has around 2000 commune boundaries where the EAs must be nested in. These administrative boundaries have been digitised manually but they are not entirely matching the ground features on high resolution satellite imagery. As a result, the lines are not matching the existing digitised features that are publicly available (for example Open Street Map). For an experiment, we have created the uncrossable boundaries, and they were set in the tool, but it has created various complex shapes and numerous silver polygons. However, more compact shape and consistent outlines were produced where the uncrossable boundaries was not set in the tool.
- Due to lack of digitised boundaries in rural areas and to create adjacent boundaries around settlements, Euclidean Allocation technique was

employed. Euclidean Allocation computes, for each cell, the nearest source (settled area) based on Euclidean distance. But here we are not interested in distance calculation, Euclidean Allocation was created to define and delineate proximal regions around individual data polygons by using polygon boundaries. The rasterized building footprints was converted to polygons, and it was the input data to the Euclidean allocation technique.

ii. Running the pre-EA tool

After pre-processing the input datasets, the tool's parameters need tuning in order to produce the pre-EA boundaries in line with the desired country's criteria. The run of the pre-EA tool consists of two steps: creating building blocks from the inputs and then merging them to pre-EAs based on the specified criteria. These two steps are briefly described here:

A. Building Blocks: The aim of splitting process is to subdivide the area of interest into regions (building block polygons) that are as small as possible so that the subsequent merging process has enough flexibility to combine them into optimal pre-EAs. These building block polygons need to follow ground features so the final pre-EA boundaries can be clearly identified by the field teams, hence it is important to use good quality input datasets. Such inputs are natural and man-made features such as roads, railways and waterways from existing digitised sources and administrative boundaries. The estimated number of structure (building) for each small unit is then computed from high resolution gridded building count (Dooley et al. 2021). It is the most useful if the size of the grid cells is similar to size of the split units, hence we use a custom-built, robust, high-resolution dataset. This step will be done automatically in the tool

B. Merging: After computing the estimated number of buildings and area for all the small units generated in the building block step, the tool

merged the units until they reach a user specified building and/or area threshold. By setting weights, the user can set the importance of the total building, total area or shape of the final pre-EA polygons. This step will be done automatically in the tool

Tool Input:

- Road and railway datasets
- Waterway dataset
- Uncrossable features
- Euclidean Allocation polygons
- Administrative boundary
- Rasterized building footprints count (100x100m)

Figure 1 shows the specific setting that were set in the pre-EA tool to generate automatic census pre-EAs for Zimbabwe (old preEA version).

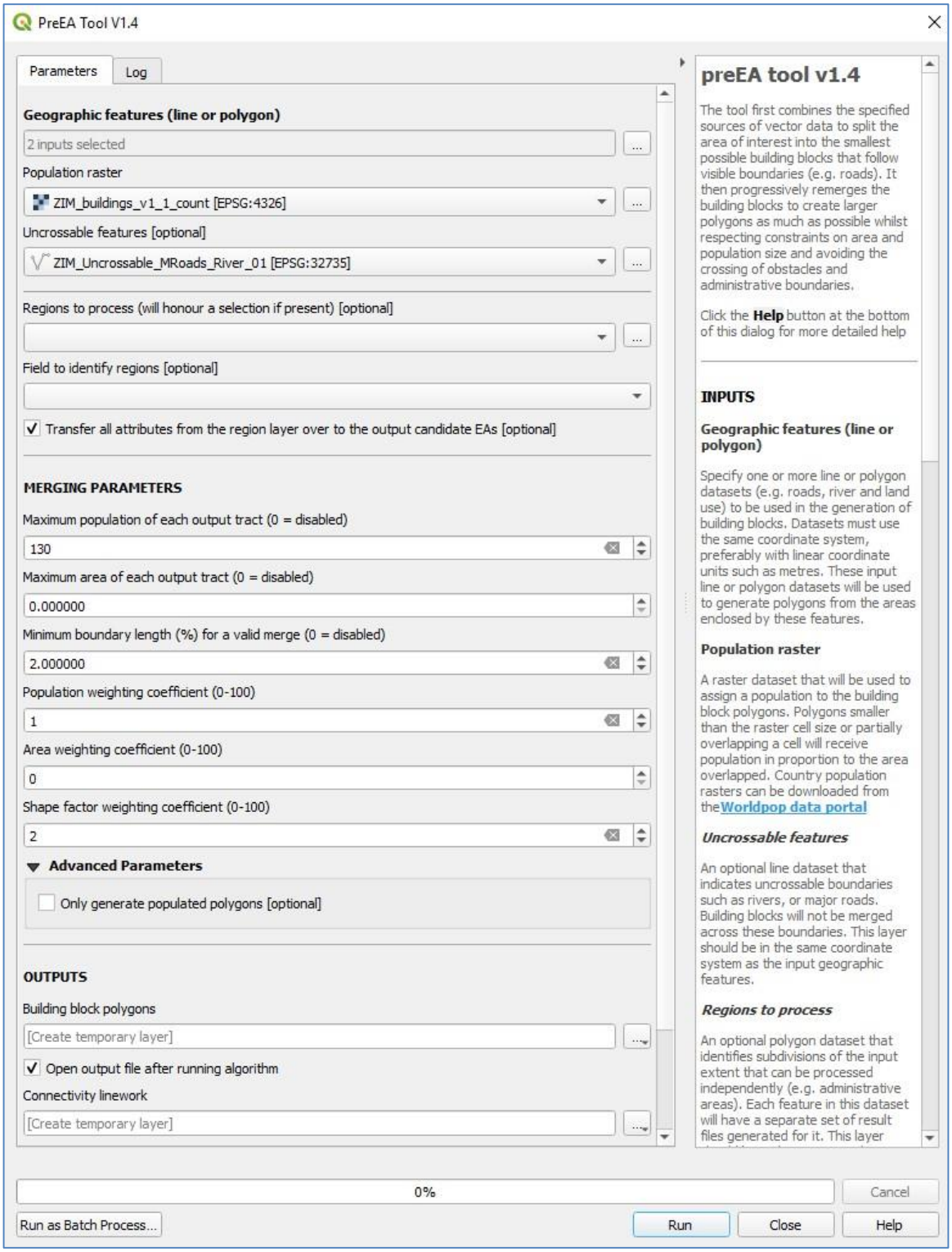

Figure 1. The pre-EA tool interface with the Zimbabwe application settings (The tool has been developed further)

4. Release Content and File Descriptions

- **ZWE_BuildingBlocks_v1.0.shp**

  These are the polygons generated from the input data, with population assigned but prior to any merging taking place. The attribute table for the building block contains the estimated total building (structure) number (preEA_Bldn), the surface area (m^2) and the administrative name. The projection is the projected coordinate system WGS84 (WGS_1984_UTM_Zone_35S).

- **ZWE_PreEAs_v1.0.shp**

  The polygons generated following the completion of the merging process. The attribute table for the pre-EAs contains the country code (ZWE), the estimated total building (structure) number (preEA_Bldn), the surface area (m^2) and the administrative name. The projection is the projected coordinate system WGS84 (WGS_1984_UTM_Zone_35S).

Field names in the attribute tables:

1- preEA_EAID: Random ID for the preEAs
2- preEA_Bldn: Total estimated number of buildings (structures) in the given preEA
3- preEA_Area: Estimated total geographic area in m^2 for the given preEA

Important Note:

If the current tool's settings and defined constraints that are used to create this version of the datasets are not suitable for your intended application, please get in touch with us and we can generate new outputs based on your desired criteria. The preEA tool has been developed further after this work and it can consider new parameters to ensure a better output.

6. Assumptions and Limitations

1- The 'settled' grid cells included in all rasterized building footprints is defined by where the building footprints datasets contain polygons. We have not masked out any areas, nor do we perform any quality checks of the building footprints data prior to processing (Dooley et al. 2021)

2- The distribution of the buildings might not represent current building landscape in the country because satellite imageries from different years were used to extract building footprints (e.g. due to cloud coverage).

3- In terms of the input vector datasets used in producing the census pre-EAs, OSM were used as the primary data sources for roads, railways and waterways. The lack of spatial coverage of these datasets has limited the flexibility of the creation of optimal pre-EAs. As a result, some pre-EAs were left with buildings larger than maximum building constraints as there was not enough spatial data to split them further.

4- Many pre-EAs can be found with low building numbers, including zero, as the tool was obliged to halt merging because either the uncrossable boundary blocked the merging or the geographic areas has reached the maximum threshold (9 km^2) that ensures that the EA would not become massive and unmanageable for the enumerator. The pre-EAs with zero building can, of course, be merged with their neighbouring pre-EAs if needed. User can request a new version based on its desired criteria.

5- We recommend that the pre-EA outputs should be carefully reviewed in the lab and manually edited as needed prior to census cartography perhaps by using the split polygons that are also provided. Then whilst in the field (i.e. cartography), the pre-EA boundaries should be validated and finalised.

7. Acknowledgments

Funding for the development and dissemination of this data set was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (OPP1182425).

Data preparation, cleaning, pre-processing, and the preEA tool application were led by Sarchil Qader. Additional support with final output review and country engagement were provided by Mathias Kuepie (UNFPA). The whole WorldPop group and GRID3 partners are acknowledged for overall project support.

8. References

OpenStreetMap contributors. (2015) Planet dump [Data file from 2019]. Retrieved from https://planet.openstreetmap.org

Dooley, C. A., Leasure, D.R., Boo, G. and Tatem, A.J. 2021. Gridded maps of building patterns throughout sub-Saharan Africa, version 2.0. University of Southampton: Southampton, UK. Source of building footprints “Ecopia Vector Maps Powered by Maxar Satellite Imagery”© 2020/2021. doi:10.5258/SOTON/WP00712.

HDX (2021). Zimbabwe - Subnational Administrative Boundaries. https://data.humdata.org/dataset/zimbabwe-administrative-levels-0-3-boundaries