BUILDING FOOTPRINT DATA

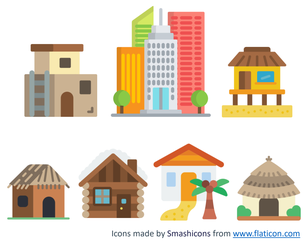
How comparable are existing data products for countries in Africa?

Heather R. Chamberlain, Edith Darin, Wole Ademola Adewole, Warren C. Jochem, Attila N. Lazar, Andrew J. Tatem

BUILDING FOOTPRINTS

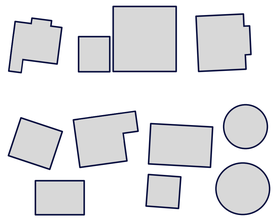
Building footprints consist of 2-dimensional outlines (vector polygons) of buildings. There are 3 main sources of building footprint data:

- **Authoritative datasets**, such as those from national mapping agencies
- **VGI initiatives**, such as OpenStreetMap
- **Automated feature extraction** from satellite or aerial imagery, often conducted by commercial companies



Icons made by Smashicons from www.flaticon.com

Mapped building footprints:



AN EVOLVING DATA LANDSCAPE

Over the past decade, growth and developments in computing power, machine-learning algorithms and the spatiotemporal resolution of satellite imagery, have led to rapid developments in automated feature extraction. Application of these methods to built infrastructure has led to new national and multi-country building footprint datasets, available as open geospatial data from multiple sources.

Timeline for publication of national or multi-country datasets:

- 2015
  - Ordnance Survey – OpenMap Local (GB)
  - Global growth of OpenStreetMap, particularly HOT-OSM
- 2018
  - Microsoft USA building footprints
  - Microsoft - Canada, Uganda, Tanzania building footprints
  - "Digitize Africa" building footprints for all of sub-Saharan Africa - Ecopia/Maxar
- 2021
  - Microsoft footprints – Nigeria, Kenya and S. American cities
  - Google v1
  - Microsoft "global" footprints
  - Google v2
  - Google v3
- 2024
  - Ongoing updates to Microsoft footprints (geographically-limited)

In settings with well-developed geospatial data systems, such datasets may complement existing authoritative sources, but in data-scarce settings, they may be the only source of data. With a growing number of building footprint datasets becoming available globally, data users need to understand the extent to which seemingly similar datasets are interchangeable and their potential advantages, disadvantages and limitations.

To address this knowledge gap, we review available building footprint datasets, focussing on countries in Africa, where there is considerable data scarcity and previously building footprint data has been very limited, with no national datasets from authoritative sources having been openly released.

Four building footprint data products were included in our comparative analysis:

- Ecopia "year 1" – Digitize Africa
- Google version 2 (v2)
- Microsoft (as of January 2023)
- OpenStreetMap (as of January 2023)

DATA APPLICATION

The development of new building footprint data products has increased data availability for urban areas, but also rural and other data-sparse settings, such as informal settlements. These developments have led to building footprint data being increasingly used in a wide range of contexts and applications, including in WorldPop's work to map and model population.

In WorldPop's work:

- As the basis for determining settled grid cells in "top-down" population modelling
- Calculating metrics related to count and area of buildings as geospatial covariates for population modelling
- Mapping and modelling settlement morphology and settlement types
- An essential input dataset for "bottom-up" population modelling
- Stratification of building types, e.g. residential, non-residential or mixed use
- Built settlement growth modelling to support projected population estimates

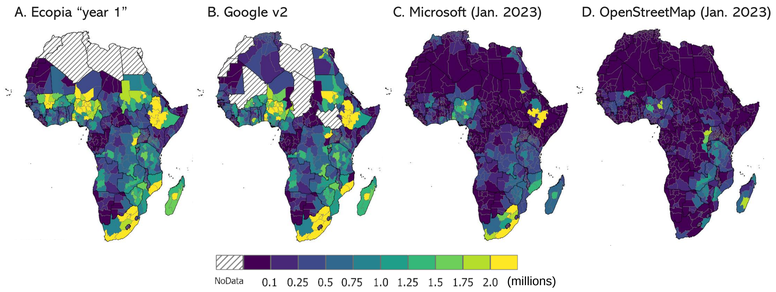
Other Examples:

- Mapping settlements – their locations, structure and morphology
- Assessment of damage (e.g. post-disaster)
- Survey and census logistics planning, implementation and coverage evaluation
- Estimation of vulnerability of infrastructure to current and future hazards

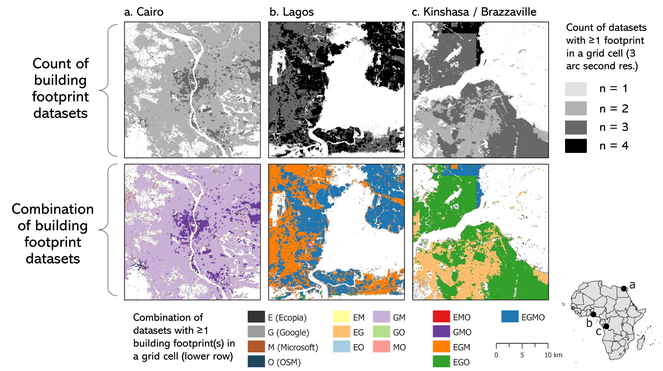
HOW COMPARABLE ARE DATASETS?

Subnational mapping of building footprint counts per administrative level 1 (AL1) unit shows considerable heterogeneity between the four different data products.

COUNTS OF BUILDING FOOTPRINTS PER AL1 UNIT



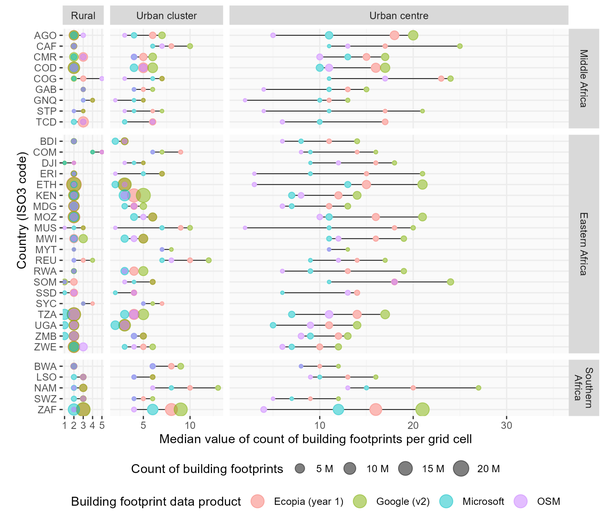
Even within major cities, considerable differences in spatial coverage and dataset completeness are observed. Creating a binary raster version of each building footprint dataset and comparing these between data products shows big variations in coverage, shown here for Cairo, Lagos and Kinshasa.



WHAT IS THE IMPACT?

Aside from differences driven by dataset coverage and completeness, we also see considerable variations in count and area of building footprints at a local level, between data products. These observed differences have potentially large impacts when building footprint data are used in analyses to support research and decision making.

Calculating the median count of building footprints per grid cell (3 arc second resolution) shows such differences are found across countries, and rural/urban settings – shown here for 3 of the UN Africa Regions, with rural/urban stratification based on GHS-SMOD classes:



All building footprint datasets conceptually are trying to represent the same thing, but our analysis has shown that for countries in Africa, existing building footprint data products differ considerably in their representation of buildings – the datasets are not interchangeable.

Data users therefore need to assess the data for suitability in terms of the context, time period, spatial scale and use case that is of interest. Data producers can also take steps to enhance useability of their products, including providing (i) information on both spatial and temporal coverage of input imagery, and (ii) enhanced data documentation with details of data processing, footprint geometry and context-specific rates of omission/commission.

FULL PAPER & OUTPUT DATASETS

Chamberlain, Heather R., Edith Darin, Wole Ademola Adewole, Warren C. Jochem, Attila N. Lazar, and Andrew J. Tatem. "Building footprint data for countries in Africa: to what extent are existing data products comparable?." *Computers, Environment and Urban Systems*, **110** (2024): 102104. https://doi.org/10.1016/j.compenvurbsys.2024.102104

Chamberlain, Heather R., Warren C. Jochem, Attila N. Lazar, and Andrew J. Tatem. "Gridded datasets of building count and area metrics for the UN Africa Region, version 1.0" [DATASET]. *WorldPop, University of Southampton* (2024). https://doi.org/10.5258/SOTON/WP00776

ACKNOWLEDGEMENTS

This work was part of the GRID3 project with funding from the Bill & Melinda Gates Foundation (BMGF) and the United Kingdom's Foreign, Commonwealth & Development Office (FCDO) [grant numbers: INV-009579 (BMGF) and INV-045694 (BMGF and FCDO)].

Poster prepared by Heather Chamberlain (h.chamberlain@soton.ac.uk)