Photo by Jack Anstey on Unsplash

WHERE IS EVERYONE?

The Importance of Population Density Data: A Data Artefact Study of The Facebook Population Density Map

August 2021

THEGOVLAB

AUTHORS

ADITI RAMESH | RESEARCH FELLOW

Aditi Ramesh is a research fellow at GovLab. Prior to joining GovLab, Aditi spent two years working in the Indian education and technology policy space. As a researcher at the Aapti Institute, Aditi helped build The Data Economy Lab to drive an understanding and awareness of data stewardship. Here, she conceptualized participative, people-centric frameworks for data governance, with a focus on data sharing frameworks for migrant communities in urban India. Aditi was also a Fulbright fellow in India, where she studied youth civic engagement across government and private schools in Bangalore. Aditi holds a B.S. in Economics/ Mathematics from the University of Southern California.

STEFAAN VERHULST | CO-FOUNDER AND CHIEF RESEARCH AND DEVELOPMENT OFFICER

Stefaan G. Verhulst is Co-Founder and Chief Research and Development Officer of the Governance Laboratory @NYU (GovLab) where he is building an action-research foundation on how to transform governance using advances in science, data and technology. Verhulst's latest scholarship centers on how technology can improve people's lives and the creation of more effective and collaborative forms of governance. Specifically, he is interested in the perils and promise of collaborative technologies and how to harness the unprecedented volume of information to advance the public good.

ANDREW YOUNG | KNOWLEDGE DIRECTOR

Andrew Young is the Knowledge Director at The GovLab, where he leads research efforts focusing on the impact of technology on public institutions. Among the grant-funded projects he has directed are a global assessment of the impact of open government data; comparative benchmarking of government innovation efforts against those of other countries; a methodology for leveraging corporate data to benefit the public good; and crafting the experimental design for testing the adoption of technology innovations in federal agencies.

ANDREW J. ZAHURANEC | RESEARCH FELLOW

Andrew J. Zahuranec is Research Fellow at The GovLab, where he is responsible for studying how advances in science and technology can improve governance. In previous positions at the NATO Parliamentary Assembly and National Governors Association, he worked on issues as far-ranging as election security, the commercial space industry, and the opioid epidemic. He has a Master of Arts in Security Policy Studies from the George Washington University and a bachelor's degree in Political Science and Intelligence from Mercyhurst University.

Funding Disclosure Notice: The GovLab's Data Program is funded through a variety of diverse sources including Luminate, Microsoft, Facebook, Open Data Institute, Bertelsmann Foundation, Asia Foundation, Barilla Foundation, UNICEF, and Rockefeller Foundation.

We would like to thank the following individuals for their insights and contributions to this paper: Andres Gonzalez Garcia and Clara Pérez-Andújar Carretié (Universal Access Energy Lab); Andrew Schroeder (Direct Relief); Greg Yetman (CIESIN); Andreas Gros (Facebook) and Laura McGorman (Facebook).

I. INTRODUCTION

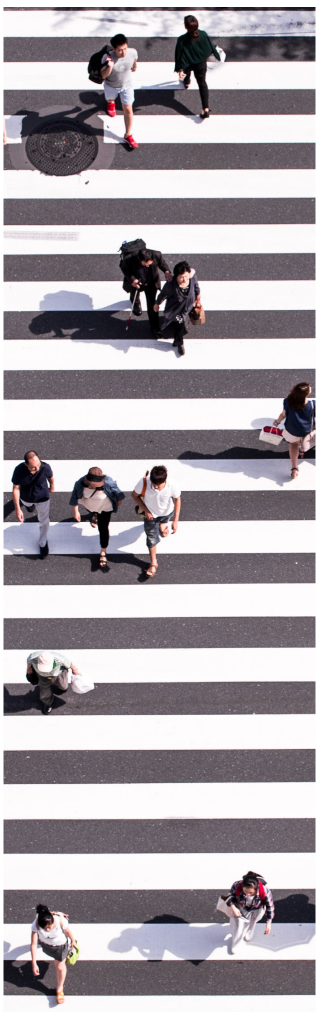


Photo by Ryoji Iwata on Unsplash

Data and information about population density offer significant opportunities to improve social, economic, and political outcomes. A variety of public and private stakeholders, including nonprofits, humanitarian actors, government agencies, and public research institutions, have used density indicators as inputs to support long-term infrastructure development projects, plan vaccination campaigns, and determine geographic areas most likely to be impacted by climate change. Population indicators are also critical to measure progress of UN Sustainable Developments Goals (SDGs)—more than one-third of these are defined with reference to population measurements.[1]

Some preliminary examples can help illustrate the potential of density indicators. In late 2020, a group of researchers used 2011 census data provided by the Government of India to understand the correlation between population density and mortality rate from Covid-19 in India: "After a detailed correlation and regression analysis of infection and mortality rates due to Covid-19 at the district level," the researchers found "moderate association between Covid-19 spread and population density."[2] mWater, a company that develops technologies to map and monitor water, sanitation, and health, embedded Facebook's High Resolution Settlement

[1] Weber, Eric M., et al. "Census-Independent Population Mapping in Northern Nigeria." *Remote Sensing of Environment*, Elsevier, 21 Oct. 2017, www.sciencedirect.com/science/article/pii/S0034425717304364.

[2] Bhadra, A., Mukherjee, A. & Sarkar, K. Impact of population density on Covid-19 infected and mortality rate in India. Model. Earth Syst. Environ. 7, 623–629 (2021). https://doi.org/10.1007/s40808-020-00984-7

Layer in their mapping tools to help NGOs providing safe drinking water track physical assets and coordinate with other organizations. [3] Another team of researchers found that Japanese cities with higher population densities facilitate greater energy–environment–economic (3E) performance. They found that "unprecedented increase in urban population" in Japan and around the world is "closely associated with the magnet effect and the economic achievement of cities."[4] These and other examples are discussed at greater length below.

In this paper, we explore new and traditional approaches to measuring population density, and ways in which density information has frequently been used by humanitarian, private-sector and government actors to advance a range of private and public goals. We explain how new innovations are leading to fresh ways of collecting data—and fresh forms of data—and how this may open up new avenues for using density information in a variety of contexts. Section III examines one particular example: Facebook's High-Resolution Population Density Maps (also referred to as HRSL, or high resolution settlement layer). This recent initiative, created in collaboration with a number of external organizations, shows not only the potential of mapping innovations but also the potential benefits of inter-sectoral partnerships and sharing. We examine three particular use cases of HRSL, and we follow with an assessment and some lessons learned. These lessons are applicable to HRSL in particular, but also more broadly. We conclude with some thoughts on avenues for future research.

---

[3] Nigeria Population Density. mWater Portal. https://portal.mwater.co/#/maps/4f1a33ddc1fb46acb55c59087c676b38

[4] Suzuki, S., Nijkamp, P. High urban population density as a facilitator of energy–environment–economy performance–development of an autoconfiguration target model in data envelopment analysis. Asia-Pac J Reg Sci 5, 261–287 (2021). https://doi.org/10.1007/s41685-020-00161-9

II. THE IMPORTANCE OF POPULATION DEMOGRAPHICS AND DENSITY MEASUREMENT

Photo by Luke Michael on Unsplash

In order to understand the history and importance of human population density data and measurement, we began by looking at the existing literature on the use and development of population density datasets and mapping technologies. Through desk research, we sought to understand the relevance of population demographics and density measurements for a number of stakeholders, such as researchers, government officials, and humanitarian organizations. We studied both traditional and emerging sources of population statistics, assessing and comparing their respective strengths and weaknesses. This analysis is presented below, as well as an assessment of key value propositions for population density data, which we describe using an existing taxonomy developed by The GovLab.[5]

A. The Concept of Population Density

As explained by the US Census Bureau, human population density "is typically expressed as the number of people per

[5] Verhulst, Stefaan, et al. "An Introduction to Data Collaboratives: Creating Public Value by Exchanging Data." *DataCollaboratives.org*, The Governance Lab, 2015, datacollaboratives.org/static/files/data-collaboratives-intro.pdf.

square mile of land area."[6] As *ScienceDirect* describes population density: "If conditions are not good for the species, the density will be low (organisms will have died or moved out of the sampled area), whereas if conditions are good the density will be high (organisms will have reproduced and/or immigrated into the area)."[7] Population densities are generally calculated for cities, countries, or continents, and are affected by a number of factors. Economic factors, like the availability of jobs and potential for economic mobility, have historically drawn large numbers of people to urban areas. Physical factors, such as climate, water supply, or soil quality, can also affect how and where people choose to settle. Cities located along a coastline, for example, are often more densely populated as they provide access to both a water source and, historically, opportunities for trading. Social and political factors, such as high crime rates or the lack of funding for public services, are also correlated to population growth rates and density.[8]

Under hypothetical conditions of unlimited resources, populations will grow at geometric or exponential rates. However, in the real world, it has been shown that population growth can be limited by the availability of resources, together with other cultural, environmental, and health factors. [9] These resources are sometimes called *density-dependent factors,* which cause a population's per capita growth rate to change (usually decrease as a population density increases). In addition to basic physical resources such as food and water, these factors can also include the presence of competition for resources, migration of individuals, and disease outbreaks. For instance, disease tends to spread more rapidly when human populations exist in closer proximity to one another.[10] Other factors, known as *density-independent factors,* affect populations irrespective of their densities. An example can be found in environmental stressors, such as inclement weather, fire, and floods, which impact the growth rate of populations regardless of how close people live to one another. Demographic factors and trends can also have significant impacts on population

[6] Cohen, Darryl. "Understanding Population Density." US Census Bureau, 2015. https://www.census.gov/newsroom/blogs/random-samplings/2015/03/understanding-population-density.html

[7] "Population Density." *ScienceDirect*, Insect Ecology (Fourth Edition), 2016, www.sciencedirect.com/topics/agricultural-and-biological-sciences/population-density.

[8] "What Factors Affect Population Density and Distribution?" *Internet Geography*, 31 Jan. 2021, www.internetgeography.net/igcse-geography/population-and-settlement-igcse-geography/what-factors-affect-population-density-and-distribution/.

[9] Bacaër Nicolas. "Verhulst and the Logistic Equation (1838)." *A Short History of Mathematical Population Dynamics*, by Bacaër Nicolas, Springer, 2011, pp. 35–39.

[10] Reid, Ari. "Examples of Density Dependent Factors." *Sciencing*, 2 Mar. 2019, sciencing.com/examples-density-dependent-factors-5928598.html.

density in different regions. Issues such as education and access to family planning in a region can affect fertility rates.[11] [12]

Population density data can help organizations better understand the causal relationships among these many variables and gain new insight into how and why people are distributed across regions.

B. Population Density Use Cases

In recent years, a number of research studies have affirmed the causal relationships described above. For example, a group of researchers at the School of Architecture at Southeast University and the School of Landscape Architecture at Nanjing Forestry University used mobile signaling data to study the relationship between population density and the spatial distribution of urban public service facilities in downtown Shanghai. The results from this study demonstrated the value of mobile and land use data of service facilities to study spatiotemporal relationships.[13] Another study, produced by the Centre for Economic Performance, examines urban densities across six African countries and the relationship to household income differences across cities and neighborhoods, showing a negative correlation between household income and urban densities.[14]

Current literature also demonstrates the importance of new perspectives in population density mapping to deliver on its value as a tool for improved decision-making in the public interest.[15] As technology evolves, many are becoming aware of

---

[11] Pourreza, A., Sadeghi, A., Amini-Rarani, M. et al. (2021), "Contributing factors to the total fertility rate declining trend in the Middle East and North Africa: a systemic review". J Health Population Nutrition 40, 11. https://doi.org/10.1186/s41043-021-00239-w

[12] Lutz, W., Cuaresma, J.C. and Abbasi-Shavazi, M.J. (2010), "Demography, Education, and Democracy: Global Trends and the Case of Iran". Population and Development Review, 36: 253-281. https://doi.org/10.1111/j.1728-4457.2010.00329.x

[13] Shi Y., Yang J., Shen P. (2020), "Revealing the Correlation between Population Density and the Spatial Distribution of Urban PublicService Facilities with Mobile Phone Data" International Journal of Geo-Information

[14] Henserson J., Nigmatulina D., Kriticos S. (2018), Measuring urban economic density, Centre for economic performance https://cep.lse.ac.uk/pubs/download/dp1569.pdf

[15] Longwei Li & Dengsheng Lu (2016) Mapping population density distribution at multiple scales in Zhejiang Province using Landsat Thematic Mapper and census data, International Journal of Remote Sensing, 37:18, 4243-4260, DOI: 10.1080/01431161.2016.1212422; Eric M. Weber, Vincent Y. Seaman, Robert N. Stewart, Tomas J. Bird, Andrew J. Tatem, Jacob J. McKee, Budhendra L. Bhaduri, Jessica J. Moehl, Andrew E. Reith, Census-independent population mapping in northern Nigeria, Remote Sensing of Environment, Volume 204, 2018, Pages 786-798, https://doi.org/10.1016/j.rse.2017.09.024 ; Patel, N.N., Stevens, F.R., Huang, Z., Gaughan, A.E., Elyazar, I. and Tatem, A.J. (2017), Improving Large Area Population Mapping Using Geotweet Densities. Trans. in GIS, 21: 317-331. https://doi.org/10.1111/tgis.12214; Linard, C., Alegana, V.A., Noor, A.M. et al. A high resolution spatial population database of Somalia for disease risk mapping. Int J Health Geogr 9, 45 (2010). https://doi.org/10.1186/1476-072X-9-45.

the potential to create density indicators by combining traditional and non-traditional data sources. For example, a group of researchers at Vrije Universiteit Brussel combined field work data with high resolution remote sensing data to estimate the absolute number of inhabitants at the neighborhood level, an approach which was tested on Ngazidja Island (Union of the Comoros).[16] Another group of researchers at the German Aerospace Center and WorldPop note, "[i]n the production of gridded population maps, remotely sensed, human settlement datasets rank among the most important geographical factors to estimate population densities and distributions at regional and global scales.[17] As further illustrated in Section 2d, new and evolving geolocation innovations can also support the accurate mapping of population distribution. A comparative study conducted by a group of researchers from CIESIN, the University of Boulder, and other organizations, compares griddled population density datasets "to provide potential data users with the knowledge base needed to make informed decisions about the appropriateness of the data products available in relation to the target application and for critical analysis." [18]

As environmental and health factors continue to affect human populations, researchers are realizing the potential of new technologies, such as UAVs, to understand density-dependent factors in human populations.[19] For example, UAVs were used in the Solomon Islands to map a proposed re-settlement plan following Cyclone Ita, where over 50,000 people were affected or killed. Disaster responders relied on a "rotary-wing UAV (an 'Oktocopter') to assist with the damage assessment efforts," and specifically to "assess the extent of the flood damage in the most affected area along Mataniko River."[20]

[16] Mossoux S. et al (2018), Mapping population distribution from high resolution remotely sensed imagery in a data poor setting, Multidisciplinary Digital Publishing Institute

[17] Palacios-Lopez D., et al (2019), New Perspectives for Mapping Global Population Distribution Using World Settlement Footprint Products, Multidisciplinary Digital Publishing Institute (MDPI), doi:10.3390/su11216056

[18] Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistolesi, L., Levy, M., Tatem, A. J., and Pesaresi, M.: The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use, Earth Syst. Sci. Data, 11, 1385–1409, https://doi.org/10.5194/essd-11-1385-2019, 2019.

[19] Others are seeking to innovate in the collection, use, and governance of UAV data to advance geospatial mapping efforts. See for example: Verhulst, Stefaan, Aditi Ramesh, Andrew Young, Peter Rabley, and Christopher Keefe, "Establishing a Data Trust: From Concept to Reality." Place Blog, May 20, 2021. https://www.thisisplace.org/blog-1/introducingplace/establishing-a-data-trust

[20] "Humanitarian UAVs in the Solomon Islands." *iRevolutions.* 25 September 2014. https://irevolutions.org/2014/09/25/humanitarian-uavs-in-the-solomon-islands/

C. Strengths and Limitations of Traditional and New Population Density Statistic Sources

Various methods exist to determine population density, some long standing and others more recent. New methods of estimating population density, such as the use of satellite remote sensing technologies, are evolving to complement traditional methods, such as census surveys. The field of population density is also witnessing a growing number of Data Collaboratives, a new approach that extends the public-private partnership model, in which participants from different sectors exchange their data to create public value.[21] These efforts combine multiple population density statistic sources—satellite data, geospatial datasets, census data, and microcensus information collected through "bottom-up" field work—demonstrating the strong potential of merging new and traditional forms of acquiring density data to solve complex related challenges.

1. SURVEY DATA

Census and Registry / Administrative Data

Census data, collected through surveys "conducted on the full set of observation objects belonging to a given population by a government agency or counterpart,"[22] is the most traditional source of population density statistics. Census surveys are typically conducted on some regularly-occurring basis through a government statistical agency. The U.S. census, for example and as mandated by Article I, Section 2 of the Constitution, takes place every 10 years. In 2020, for the first time in history, US Census data was collected through online surveys.[23]

Census data usually includes more than just population counts; most census surveys will also gather other data points such as ethnicity breakdowns, income levels, and marital status. [24] In some cases, registry or administrative data is used to supplement or even replace census data. In Canada, for example, administrative records are "collected for the purpose of carrying out various non-statistical programs." [25]

[21] Verhulst, Stefaan, et al. "An Introduction to Data Collaboratives: Creating Public Value by Exchanging Data." *DataCollaboratives.org*, The Governance Lab, 2015, datacollaboratives.org/static/files/data-collaboratives-intro.pdf.

[22] "Census Survey." *UNESCO UIS*, uis.unesco.org/en/glossary-term/census-survey.

[23] "2020 Census." United States Census Bureau. https://www.census.gov/programs-surveys/decennial-census/decade/2020/2020-census-main.html

[24] "Census Data - Definition from KWHS." *KWHS*, 28 Feb. 2011, kwhs.wharton.upenn.edu/term/census-data/.

[25] "Use of administrative data." *Statistics Canada.* https://www150.statcan.gc.ca/n1/pub/12-539-x/2009001/administrative-administratives-eng.htm

Administrative records in Canada are being used to, for example, regulate cross-border population and trade flows, register vital statistics in line with legal requirements, collect taxes, and administer benefits programs. Administrative data has the potential to provide important insight into population density estimation modeling.

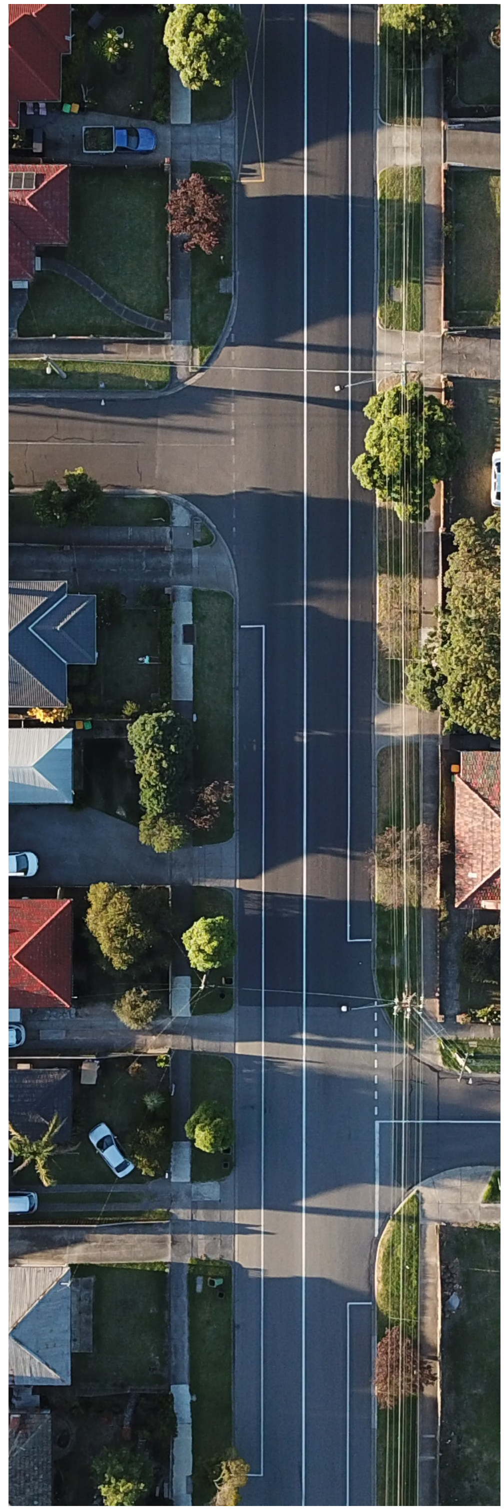


Photo by Tom Rumble on Unsplash

Benefits: Historically, census data has been used to inform decision-making and policy planning at the local, regional, or national level. Researchers and other development agencies also use census estimates for disaster preparedness or population growth modeling. Data collaboratives often combine census surveys with other, non-population information, resulting in new and innovative insights that increase the public good.[26] For example, Telefonica Research, a division of the major Latin American telecommunications company, combined census data with call detail records (CDRs), rainfall data, and satellite imagery to analyze activity in flood-affected areas in order to assess needs and allocate resources in the wake of a natural disaster.[27]

Risks and Challenges: Although they represent tried-and-tested methods, national census estimates do pose certain risks and limitations. They can be slowed by natural or public health disasters — at the time of this writing, for example, many countries are delaying their 2020 and/or 2021 census surveys on account of the COVID-19 pandemic. In resource-scarce settings,

[26] Verhulst, Stefaan G. "Data Collaboratives: Exchanging Data to Improve People's Lives." *Medium*, The Governance Lab, 25 Apr. 2015, sverhulst.medium.com/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a.

[27] Britto, Neil, et al. "The GovLab Selected Readings on Data Collaboratives." *The GovLab*, 10 June 2016, blog.thegovlab.org/post/the-govlab-selected-readings-on-data-collaboratives.

national registers are often incomplete or inaccurate due to a variety of factors, such as low survey response rates or the omission or undercounting of certain marginalized groups.[28] In economically or politically unstable countries, conflict and insecurity can also affect quality and accuracy of survey data collection. Furthermore, there have also been documented cases of corruption at the local or regional level to inflate population estimates, especially where counts are linked to resource allocation.[29]

Other Household Surveys and Opinion Polls

In the absence of reliable, up-to-date, and complete census data, researchers and practitioners have often turned to other survey methods. Household surveys are commonly collected by large development organizations such as UNICEF. [30] [31] Currently, there are three major household survey programmes in use: the Multiple Indicator Cluster Surveys (MICS), Demographic and Health Surveys (DHS), Living Standards Measurement Study (LSMS).[32] Other small-scale opinion polls and surveys are usually conducted by local or regional organizations. One such approach that has gathered pace in recent years, especially from research institutions, is bottom-up population estimation. Also known as "micro census surveys," this method identifies and rapidly surveys small, specific areas; the resulting information is then used to train spatial statistical models in combination with geospatial data to estimate population in unsampled areas with uncertainty quantified.[33][34] For example, grappling with a dearth of accurate demographic data to support government planning, development projects, and public health campaigns, researchers from WorldPop, Oak Ridge National Laboratory and the The Bill and Melinda Gates Foundation developed and tested a modeling framework to estimate population density using household survey data, together with national data, satellite imagery

---

[28] Ye, Yazoume, et al. "Health and Demographic Surveillance Systems: a Step towards Full Civil Registration and Vital Statistics System in Sub-Sahara Africa?" *BMC Public Health*, U.S. National Library of Medicine, 5 Sept. 2012, pubmed.ncbi.nlm.nih.gov/22950896/.

[29] Wardrop, N A, et al. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." *Proceedings of the National Academy of Sciences of the United States of America*, U.S. National Library of Medicine, 27 Feb. 2018, pubmed.ncbi.nlm.nih.gov/29555739/.

[30] *Introduction to UNICEF's Work on Statistics and Monitoring*. UNICEF, 14 July 2016, www.unicef.org/statistics/index_24287.html.

[31] Khan, M.B., Nausheen, S., Hussain, I. et al. Conducting household surveys on reproductive health in urban settings: lessons from Karachi, Pakistan. BMC Med Res Methodol 21, 38 (2021). https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01216-x

[32] "Key Facts on Household surveys." *Development initiatives.* 18 July 2017. https://devinit.org/resources/key-facts-on-household-surveys/.

[33] Wardrop, N. A., et al. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." *PNAS*, National Academy of Sciences, 3 Apr. 2018, www.pnas.org/content/115/14/3529.

[34] Leasure et al. "National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty." *PNAS,* 29 September 2020. https://doi.org/10.1073/pnas.1913050117

and digital maps. The researchers uses to the model to create an estimate of population sizes for every 100-m grid square in Nigeria using household survey data drawn from just 1,141 locations. [35] This method is often deployed in cases where there are limitations to census data—in Nigeria, the last national census had occurred in 2006 and did not include sufficient population counts and demographic information at the local level.

Benefits: Census estimates are often outdated and inaccurate, since country-level census surveys are frequently only administered once every 10 years or so. Household surveys can circumvent this problem by providing more current and accurate population estimates. As one research study estimating small-area population density in Sri Lanka suggests, household surveys are "cost effective in tracking local population density with greater frequency in the between-census years."[36]

Risks and Challenges: Household and micro census surveys are usually limited in scope because they are conducted at a much smaller scale than census surveys. As a result, they should be used alongside larger census initiatives or other geospatial data to create population density estimates. In addition, any smaller surveys should be developed with care, as population estimate data can have broader implications for society. In an article titled "Spatially disaggregated population estimates in the absence of national population and housing census data," Wardrop et al. note that "results [of surveys] affect all per capita rate estimates, shift political representation, and change claims to power or resources."[37]

2. GEOLOCATION INNOVATIONS

In recent years, a variety of new technological innovations have allowed for improvements to location and spatial data collection. Mobile devices such as smartphones or wearable technologies "interpret signals from their surroundings – including GPS satellites, cell towers, Wi-Fi networks, and Bluetooth – to generate a precise location measurement."[38] Geolocation innovations and research have also

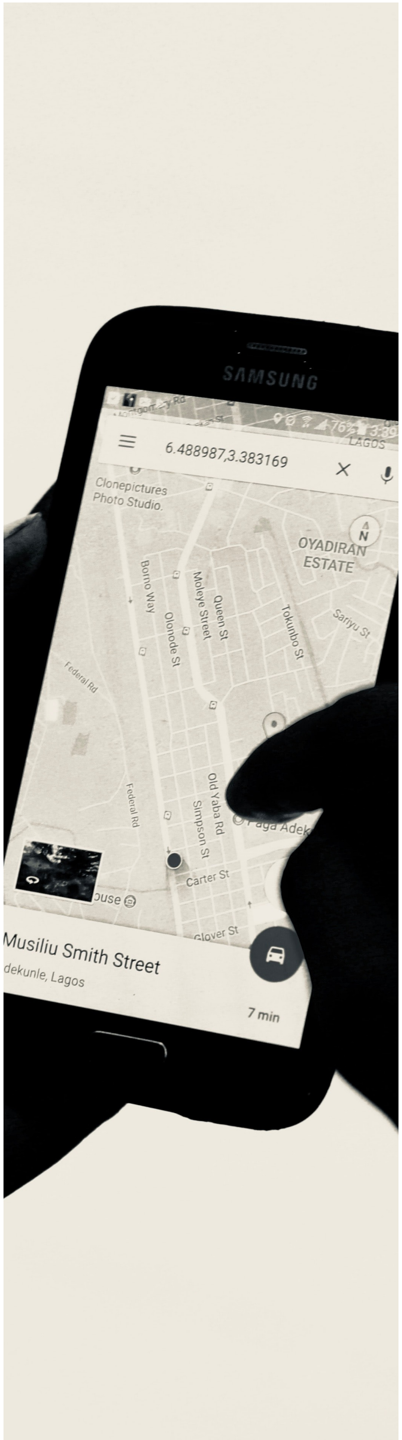
[35] Leasure et al. "National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty." *PNAS,* 29 September 2020. https://doi.org/10.1073/pnas.1913050117

[36] Engstrom, Ryan, et al. "Estimating Small-Area Population Density in Sri Lanka Using Surveys and Geo-Spatial Data." *PLOS ONE*, Public Library of Science, 5 Aug. 2020, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0237063.

[37] Wardrop, N. A., et al. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." *PNAS*, National Academy of Sciences, 3 Apr. 2018, www.pnas.org/content/115/14/3529.

[38] Gray, Stacey. "Understanding the 'World of Geolocation Data.'" *Future of Privacy Forum*, 22 May 2020, fpf.org/blog/understanding-the-world-of-geolocation-data/.

Photo by Lagos Techie on Unsplash



complemented the development of artificial intelligence and machine learning systems, and remote sensing and drone imagery is expanding rapidly to provide new consumer-level insights to several private companies.[39] However, with the rise of new technologies, there are also several important ethical and social implications to consider, such as the lack of user awareness on data collection and reuse processes. The use of drones and drone imagery is highly regulated by many countries, especially in regions where there are human settlements.

More broadly, new and increasingly affordable technological innovations, such as mobile phone and tablet devices, have made it easier for field workers to collect survey data and blend old and new forms of data collection. The rise of cheap GPS sensors and digital data collection, combined with a growing awareness of these innovations, have given international researchers, humanitarian organizations, and commercial companies the ability to better harness the power of geographical information.[40] A team of researchers, from the School of Electrical and Data Engineering at the University of Technology Sydney, experimented with mapping crowd densities using WiFi access points at train station platforms. The final results were compared with distributions acquired from CCTV cameras and exhibited strong accuracy.[41]

Benefits: New data sources from both the private and public sector have great potential to complement and enhance official statistics such as census data. Many companies are forging cross-sector collaboration to provide actors working in the public interest with

---

[39] "The Past, Present, and Future of Geospatial Data Use." *Trajectory Magazine*, 1 Feb. 2018, trajectorymagazine.com/past-present-future-geospatial-data-use/.

[40] Yetman, Greg. GovLab <> CIESIN - Population Density Data. Zoom, February 5, 2021.

[41] Farzad Tofigh, Guoqiang Mao, Justin Lipman, Mehran Abolhasan (2018), "Crowd Density Mapping Based on Wi-Fi Measurements on Train Platforms", School of Electrical and Data Engineering, University of Technology Sydney, Australia https://www.guoqiangmao.com/ewExternalFiles/Tofigh18Crowd.pdf

functional access to certain datasets. This in turn also can benefit companies seeking to demonstrate corporate responsibility, generate revenue, and bolster their reputation. [42]

Risks and Challenges: Accurate, real-time location data that can inform policy needs to be balanced with citizen privacy and other individual rights. Without adequate anonymization and aggregation at the data collection stage, geospatial data can be used to unjustly surveil members of the public. Additionally, while public-private partnerships can bolster data for good programs, they can also create risks of governments being "locked into" licenses with proprietary software and service providers, potentially hindering interoperability and creating a single point of failure. [43] It is also important to diagnose any biases or blindspots that could be baked into new geolocation innovations and perpetuated by them. Deep learning models, for example, tend to overestimate sparsely populated patches, or underestimate densely populated areas.[44]

3. SATELLITE IMAGERY

Satellite imagery is also a critical tool to help map population densities. Imaging satellites generate data that "can estimate population distributions at fine spatial scales across entire countries,"[45] by using sensors to collect radiation from the earth's surface that is then translated into imagery.[46] It is not just the raw imagery itself that is important – it's processing these images to extract information about buildings, roads, neighbourhood types, land covers etc. that can be an important piece of population density estimation models. In recent years, a number of research studies have shown the potential of satellite imagery to provide sufficient population information in areas where census information may be unreliable. In a study that is still pending peer review, a group of researchers from Stanford University and the World Bank used satellite imagery from rural images in India and population labels from the 2011 SECC census, to achieve a model that exhibits greater success than

[42] Klein, Thilo, and Stefaan Verhulst. "Access to New Data Sources for Statistics: Business Models and Incentives for the Corporate Sector." *SSRN*, 22 Mar. 2018, papers.ssrn.com/sol3/papers.cfm?abstract_id=3141446.

[43] Johnson, Peter A., et al. "The Cost(s) of Geospatial Open Data." *Wiley Online Library*, John Wiley & Sons, Ltd, 23 June 2017, onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12283.

[44] X. Huang, D. Zhu, F. Zhang, T. Liu, X. Li and L. Zou, "Sensing Population Distribution from Satellite Imagery Via Deep Learning:Model Selection, Neighboring Effects, and Systematic Biases," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5137-5151, 2021, doi: 10.1109/JSTARS.2021.3076630.

[45] Wardrop, N. A., et al. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." *PNAS*, National Academy of Sciences, 3 Apr. 2018, www.pnas.org/content/115/14/3529.

[46] " How Does Satellite Imaging Work?" *Global Forest Link*, globalforestlink.com/tutorials/how-does-satellite-imaging-work/.

LandScan, a popular global population distribution dataset. [47]

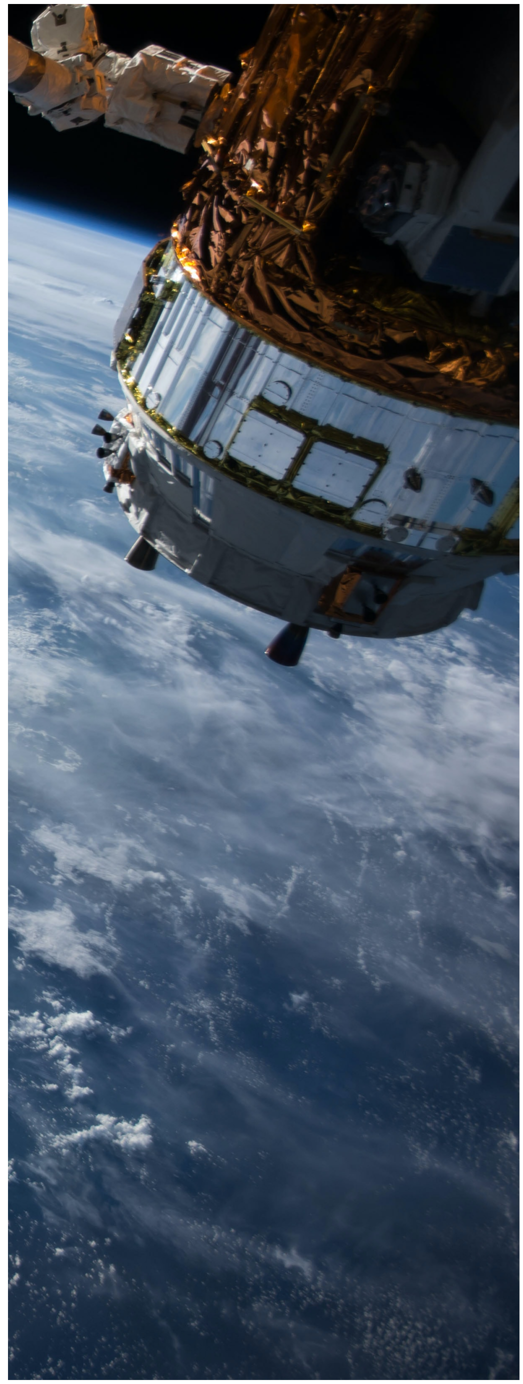


Photo by NASA on Unsplash

Benefits: Satellite imagery can be an important tool to rapidly estimate the density of small, rural populations in information-scarce settings, where census data is typically limited. In 2017, for example, the Bill and Melinda Gates Foundation created a population map to chart human settlements across Nigeria. This map used a detailed analysis of buildings in satellite imagery, along with over 2,000 on-the-ground surveys. In an article in the *Scientific American,* researchers that worked on this project expressed hope that this data would support infrastructure development, natural resource allocation, and much more.[48] Satellite imagery is also an unobtrusive method for data collection, and can provide insights more rapidly, or in real-time. When combined with ground truth data, satellite derived variables can provide more robust results. [49]

Risks and Challenges: Processing satellite imagery requires immense computing capabilities. As a result, such a task is usually outsourced to private technology companies who then sell the data to third-party developers. In addition, these images are only as good as their resolution, and require validations other datasets to ensure accuracy. [50] Publicly available, free satellite images (e.g. Sentinel and

[47] Hu, Wenjie et al. "Mapping Missing Population in Rural India: A Deep Learning Approach with Satellite Imagery." *ArXiv.* https://arxiv.org/pdf/1905.02196.pdf

[48] Tellefson, Jeff. "Satellite Images Reveal Gaps in Global Population Data." *Scientific American.* 11 May 2017, https://www.scientificamerican.com/article/satellite-images-reveal-gaps-in-global-population-data/

[49] "Spotting elephants from space: a satellite revolution." University of Oxford. 18 Dec. 2020, https://www.ox.ac.uk/news/2020-12-18-spotting-elephants-space-satellite-revolution

[50] Laituri, Melinda. Satellite imagery is revolutionizing the world. But should we always trust what we see? *The Conversation.* https://theconversation.com/satellite-imagery-is-revolutionizing-the-world-but-should-we-always-trust-what-we-see-95201

Copernicus data and Images) are typically very low resolution, and researchers still need access to ground truth data to improve image resolution and provide useful fodder for population density modeling and estimation. Satellite imagery also needs to be analyzed and used with caution, especially in conflict zones or areas that face political instability, where the mis-use data or faulty handling of data could harm certain communities. There are also several ethical issues that need to be tackled as image resolutions increase, and as newer sources of data are used to improve image resolutions.

4. NEW CROWDSOURCING INNOVATIONS

Population density indicators can also be generated through crowdsourcing techniques. While crowdsourced data is frequently a component of population estimation models, they are not considered population density data on their own. As a recent article states, "crowdsourcing geospatial data refers to generating a map using informal social networks and web 2.0 technology."[51] This data relies on a large number of digital users to voluntarily (or for minimal compensation) contribute relevant information, creating an open source dataset that fluctuates over time. Crowdsourced geographic information, for example, was used together with other datastreams on population density and flows in the aftermath of the 2010 Haiti earthquake to better target relief efforts in data-scarce regions. [52]

Benefits: In contrast to strictly controlled census and geospatial datasets, crowdsourced data and tools usually require minimal oversight.[53] The flexible nature of new crowdsourcing methodologies allows for real-time updating of population density data, permitting more rapid information generation in the case of an emergency event. While crowdsourced data alone may not always grant a complete picture of the on-the-ground events, there is clear potential to combine expert and citizen knowledge to create population density statistics that can help increase the public good.

Risks and Challenges: Because crowdsourced datasets largely rely on voluntary efforts, contributors' "decisions, opinions, and preferences could be significantly

[51] Heipke, Christian. "Crowdsourcing Geospatial Data." *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, 12 Aug. 2010, www.sciencedirect.com/science/article/abs/pii/S0924271610000602.

[52] Zook, Matthew, et al. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake." *World Medical & Health Policy* , 2010, www.un-spider.org/sites/default/files/Volunteered%20Geographic%20Information.pdf.

[53] Rice, Matthew T, et al. *Crowdsourced Geospatial Data: A Report on the Emerging Phenomena of Crowdsourced and User-Generated Geospatial Data*. Nov. 2012, apps.dtic.mil/docs/citations/ADA576607.

represented and/or influence their contribution."[54] These datasets may contain biases and pose limitations with regard to reliability and accuracy of datasets—though research suggests that crowdsourced data can be as accurate as expert-generated data for remote sensing.[55] In addition, the fact that crowdsourced oversights typically occur with minimal oversight also allows for greater possibility of errors, inaccuracies, or subjectivities in data. However, it is important to note that this would depend on the type of data being sought through the crowd sourcing endeavor and the method being used to collect or deliver the data to the platform.


Photo by NASA on Unsplash

D. Toward a taxonomy of applications of population density statistics

As demonstrated above, population density statistics collected from various sources—both traditional and emerging--have the potential to help map a changing world. Below, we discuss four key value propositions for using population density to achieve public goals.

1. *Situational awareness and response:* Increased access to population density data might enable stakeholders across sectors to better understand how to advance public good efforts, including fighting disease outbreaks and contagion risk, supporting aid and relief efforts, and enhancing the distribution of information. We are already seeing signs of such advances in the wake of the COVID-19 pandemic, where several research institutions and organizations have sought to understand how urban population compositions affect disease spread. For example, the World Bank

[54] Ibid.

[55] See, Linda et al. "Comparing the Quality of Crowdsourced Data Contributed by Experts and Non-Experts." *PLOS ONE.* 13 July 2013. https://doi.org/10.1371/journal.pone.0069958.

developed a methodology using Facebook's high resolution population density maps, which we describe further below, to help city leaders around the world target relief and resources by better identifying areas with highest exposure and contagion risk. [56]

2. *Cause and effect analysis:* Population density data can help policymakers and others arrive at a better understanding of the causal relationships among citizen movements, residence patterns, public health, and economic and social well-being. One research study in the Netherlands studied the impacts of population density and crowding on health and social adaptation, using population density data to study cities, neighborhoods, and other geographical areas to understand the relationship between density and pathology. [57]

3. *Prediction:* New predictive capabilities enabled by access to enhanced and more granular population density datasets can allow stakeholders to better assess future risks, needs, and opportunities of those populations. This data can help policymakers create digital tools to support infrastructure planning, design, and development, or support researchers in efforts to develop models for predicting population movements and their impact on a wide variety of phenomena. Consider, for example, the previously mentioned mWater, which helps other development organizations and nonprofits provide safe, accessible drinking water to populations around the globe. Among its innovations, the company used embedded Facebook population density maps to create a digital tool that allows organizations to plan and coordinate the development of pumps, wells, and other water treatment facilities.

4. *Impact Assessment:* Better access to population data can help determine whether, and which, institutional interventions positively affect specific population clusters. It can also help governments evaluate the impact of policies and facilitate product- or service- enhancing iteration. In 2018, for instance, the OECD used population density indicators to assess national policies regarding urban sprawl, and proposed recommendations to encourage greener and more sustainable cities.[58]

---

[56] Bhardwaj, Gaurav, et al. *Cities, Crowding, and the Coronavirus: Predicting Contagion Risk Hotspots*. World Bank, 21 Apr. 2020, openknowledge.worldbank.org/bitstream/handle/10986/33648/Cities-Crowding-and-the-Coronavirus-Predicting-Contagion-Risk-Hotspots.pdf?sequence=1&isAllowed=y.

[57] Choldin, Harvey M. "Urban Density and Pathology." Annual Review of Sociology, vol. 4, 1978, pp. 91–113. JSTOR, www.jstor.org/stable/2945966. Accessed 27 Aug. 2021.

[58] *Rethinking Urban Sprawl:Moving Towards Sustainable Cities*. OECD, 2018, www.oecd.org/environment/tools-evaluation/Policy-Highlights-Rethinking-Urban-Sprawl.pdf.

III. DATA ARTEFACT: THE FACEBOOK POPULATION DENSITY MAP

Photo by Tim Mossholder on Unsplash

The above discussion outlined the benefits and risks of different methods for accessing and using population level data in the public interest. In this section, we focus on one particular application that has been used to address a range of public challenges: Facebook's High-Resolution Population Density Maps, often referred to as the High Resolution Settlement Layer (HRSL). HRSL has been deployed by research institutions, non-profits, and international development agencies to support a range of projects, from disaster planning to vaccination campaigns. The density maps, which estimate population levels within 30-meter grids in nearly every country in the world,[59] *are not based on data collected through Facebook's social media or messaging platforms*, but instead use a combination of machine vision AI, satellite imagery (aerial data), and census information compiled by Facebook's AI and data science team in collaboration with Columbia University.[60] The resulting density maps are publicly available through the Humanitarian Data Exchange (HDX) data store[61] and the Registry of Open

[59] "Using HDX to Access High Resolution Population Density Maps and Demographic Estimates." *Facebook Data for Good*, 18 Oct. 2019, dataforgood.fb.com/docs/high-resolution-population-density-maps-demographic-estimates-documentation/.

[60] *Methodology: High Resolution Population Density Maps + Demographic Estimates.* Facebook Data for Good, 23 Feb. 2021, dataforgood.fb.com/docs/methodology-high-resolution-population-density-maps-demographic-estimates/.

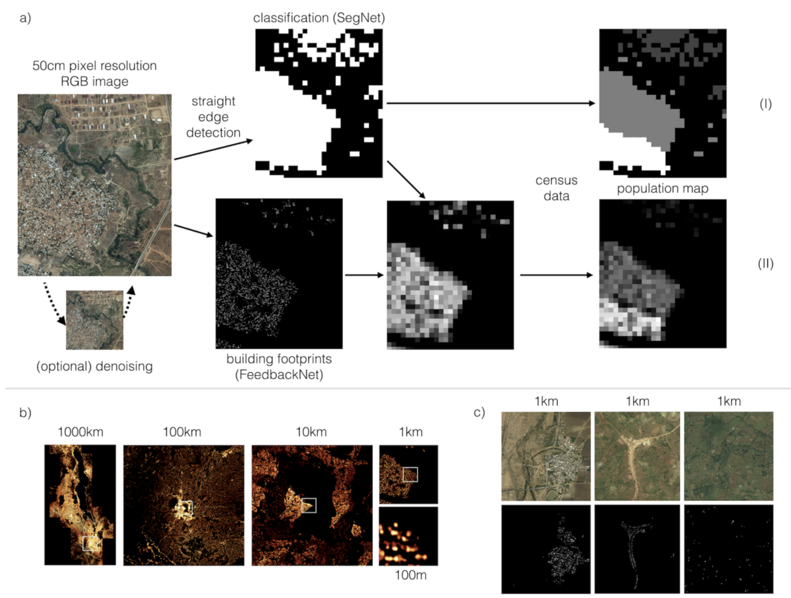
[61] "What We Do." *The Centre for Humanitarian Data*, centre.humdata.org/what-we-do/.

Data on Amazon Web Services (AWS). [62]

In order to better understand the evolution and usefulness of HRSL, The GovLab conducted interviews with stakeholders who were important to the creation and development of the technology, Greg Yetman, Associate Director for the Geospatial Applications Division at CIESIN, and Andreas Gros, Data Scientist at Facebook. We examined use cases of the data artefact, which were identified and curated through a desk research, document review, and consultations with experts in the field and personnel at Facebook supporting the use of the HRSL. A number of these case studies are integrated into the analysis below, and the following section contains a more in-depth consideration of three particular case studies.

A. Origin of HRSL

The inception of HRSL dates back to 2013, when Facebook launched a project known as the Connectivity Lab, which sought to bring fast, reliable internet to those lacking it around the world.[63] As part of this project, Facebook turned to population density datasets to understand how and where people were distributed, and areas with especially inadequate internet connectivity given population density.



Source: Facebook Research

The team behind this project quickly found that information on large expanses of the earth's population was lacking. To address this gap, Facebook set out to create its own population density indicators. The team approached DigitalGlobe (now known as Maxar Technologies), an American vendor of space imagery and geospatial content,

[62] "High Resolution Population Density Maps + Demographic Estimates by CIESIN and Facebook." *Registry of Open Data on AWS*, registry.opendata.aws/dataforgood-fb-hrsl/.

[63] https://connectivity.fb.com/

and the two companies settled on an arrangement in which Facebook purchased a large dataset of satellite imagery with a 6-month access period, as well as indefinite access to the derived dataset.[64] Facebook initially bought data for 23 countries.

The first step in creating HRSL was to determine where people were living. At the start of the project, the team began prototyping a computer vision system to identify homes, commercial establishments and other buildings. To identify these buildings, Facebook used tools known as image classifiers that had already been built for users with visual impairments and other security purposes. Eventually, after navigating several initial stumbling blocks, the team trained a new machine learning system to improve accuracy.

Facebook also collaborated with partners who had experience with population density maps, including OpenStreetMap (OSM), "a free editable map of the world that is being built by volunteers and released with an open-content license,"[65] and the Humanitarian OpenStreetMap Team (HOT). To train its machine learning algorithms, Facebook used data from OSM, allowing its researchers to add over 100 million buildings to the training set.

To build its maps, Facebook's spatial computing team (a group of trained AI researchers and data scientists) created a pattern-recognizing algorithm through neural networks that could locate structures in satellite imagery. Facebook then initiated a partnership with the Center for International Earth Science Observation at Columbia University (CIESIN), as they had already had experience working with population density indicators through their Gridded Population of the World Project (GPW). As a next step to create HRSL, CIESIN overlays "general population estimates based on publicly available census data and other population statistics."[66] Lastly, CIESIN validated these estimates against other publicly available information, such as household surveys from USAID.[67]

CIESIN handled all data licensing and negotiation processes for the census data of the countries it worked with. Developing these relationships was a continuously evolving

[64] *Derived data* is an element of data derived from other data elements using a mathematical, logical, or other type of transformation.

[65] "Mapping the World to Help Aid Workers, with Weakly, Semi-Supervised Learning." *Facebook AI*, Facebook Data for Good, 9 Apr. 2019, ai.facebook.com/blog/mapping-the-world-to-help-aid-workers-with-weakly-semi-supervised-learning.

[66] Ibid.

[67] "Population Density Maps." *Facebook Data for Good*, 22 Jan. 2021, dataforgood.fb.com/tools/population-density-maps/.

effort, made possible through a large network of national agencies, international research institutes, and colleagues. It is worth noting that, despite the many innovations in this project, census surveys continued to form the backbone of population estimation, and were critical to help identify density-dependent factors for population estimation.[68]

B. Uses of HRSL

Initially, the HRSL was primarily used internally at Facebook. Since the user base for Facebook's social media and messaging platforms is not representative of the population at large, the company considers HRSL when planning changes to its infrastructure or policies; HRSL allows Facebook to get a better idea of which populations their changes may affect, and thus make efforts to de-bias many of its products. For example, HRSL has been used extensively by the Facebook connectivity team to determine where certain infrastructure investments would be most impactful for increasing internet connectivity, and it is also leveraged by Facebook's location services.[69]

According to Andreas Gros and other stakeholders from the Facebook Data for Good team, Facebook's intention has always been to make its population density data available for public access, and the application has found a number of external uses too. The initial data code was posted on the CIESIN website, and over time, HRSL transitioned to be hosted on the Humanitarian Data Exchange (HDX), a platform that hosts a collection of datasets about crises around the world, as well as through Amazon Web Services (AWS). The challenge with publishing on HDX, as well as AWS, however, is that HRSL doesn't update in real-time, and sometimes data does not update quickly enough to capture real-time changes in settlements or short-term population growth. This is because Facebook first publishes a basic population density dataset that is mainly based on high-resolution satellite images and census data, both of which do not update globally every year. [70]

One of the early external applications for HRSL was in the field of disaster response. NetHope (a consortium of nearly 60 global nonprofit organizations seeking to

---

68 Yetman *supra* note 35.

69 Gros, Andreas. GovLab <> Facebook - Population Density Data. Zoom, January 25, 2021.

70 "Guide to Publicly Available Demographic Data." *University of Virginia.* https://demographics.coopercenter.org/guide-to-publicly-available-demographic-data

improve IT connectivity in developing countries and areas affected by disaster[71]) and other affiliated organizations were some of the first users of the dataset. Subsequently, practitioners have deployed a range of use cases for HRSL, from impact assessments of flood risk to infectious disease modeling. The World Bank, for example, uses Facebook's population density maps for infrastructure development projects, to understand how and where micro-grid insulation plants should be built.[72] Surveyors can also use Facebook's data to mitigate bias in the survey process by understanding how and where exactly buildings are located to more effectively randomize the survey process. HRSL has also been used to understand and measure the progress of the UN's Sustainable Development Goals.[73] In the following section, we explore a range of additional applications and uses of HRSL.

C. Comparison with Other Population Density Maps

Below, we provide an overview and assessment of four other noteworthy developments in the population density dataset and mapping ecosystem: the Gridded Population of the World Project, WorldPop, GHSL, and LandScan.[74]

1. **The Gridded Population of the World Project (GPW)**, an initiative by the CIESIN and now a core partner of HRSL, collects census data from various sources (commercial and governmental) in every country in the world and then integrates this data in space and time to reach population estimates on administrative data.[75] Created in 1994, GPW was the first globally produced population data set. GPW is solely based on census data. [76] Several providers of population products use the GPW dataset and model it for their own needs. The GRUMP project was also later created, building off of GPW, "to construct a common geo-referenced framework of urban and rural areas by combining census data with satellite data."[77]

[71] *NetHope*, nethope.org/.

[72] Kilic et al. "A first look at Facebook's high resolution population maps." *World Bank Blogs.* 18 November 2016. https://blogs.worldbank.org/opendata/first-look-facebook-s-high-resolution-population-maps

[73] Palacios-Lopez, Daniela et al. "High-Resolution Gridded Population Datasets: Exploring the Capabilities of the World Settlement Footprint 2019 Imperviousness Layer for the African Continent." *Remote Sensing.* 17 Mar. 2021, file:///Users/GovLab/Downloads/remotesensing-13-01142-v2.pdf

[74] Gros *supra* note 59.

[75] Yetman *supra* note 35.

[76] Yetman *supra* note 35.

[77] "GRUMP v1." Socioeconomic Data and Applications Center. https://sedac.ciesin.columbia.edu/data/collection/grump-v1/about-us

2. **WorldPop** provides open population distribution datasets for Central and South America, Asia, and Latin America, building upon the work of GPW to incorporate covariance layers – in which the original dataset is linked with ancillary information, such as climate patterns or environmental factors.[78] [79] Recently, WorldPop has also recently employed bottom-up population density estimation methods, to "fill gaps in a census where full enumeration is not possible due to poor access, or financial limitations." [80] In 2017, WorldPop, Flowminder, and the United Nations Population Fund, following approval from the government of Afghanistan, worked with Afghanistan National Statistical Office to provide bottom-up population estimation modeling "integrate satellite-based mapping of all residential compounds in the country with other geospatial datasets and recent small area population estimation in a spatial statistical modeling framework."[81] As a result of this effort, new population estimates were produced for 100 meter grid cells across the country, at a national, provincial, and district level.[82]

3. **GHSL,** or the Global Human Settlement Layer, is a spatial dataset created by the EU Joint Research Centre (JRC). Population estimates from GPW are disaggregated "from census or administrative units to grid cells."[83] GHSL consistently measures changes over time in population estimates.

4. **LandScan,** by Oak Ridge National Laboratory, is a global population distribution dataset that represents "an ambient population (average over 24 hours) distribution", relying less on census inputs than the other datasets.[84] LandScan's core feature is that it attempts to estimate both daytime and nighttime population density data – daytime populations can look very different from those at night in some commercial or residential areas, such as airports or urban downtown centers. While this variation is important to capture, the dataset has also suffered from a general sense of low accuracy and spatial resolution.[85]

[78] "Geospatial Covariates." *Spatial Data Repository* , USAID, https://spatialdata.dhsprogram.com/covariates/.

[79] Gros *supra* note 59.

[80] "WorldPop gridded population estimate datasets and tools." *WorldPop,* https://www.worldpop.org/methods/populations.

[81] "Mapping Afghanistan's Population." *Case Studies*, WorldPop, www.worldpop.org/case_studies/mapping_afg_pop.

[82] *Population Density Mapping in Afghanistan Using Remote Sensing Technology*. UNFPA, paris21.org/sites/default/files/2017-09/UNPFA.pdf.

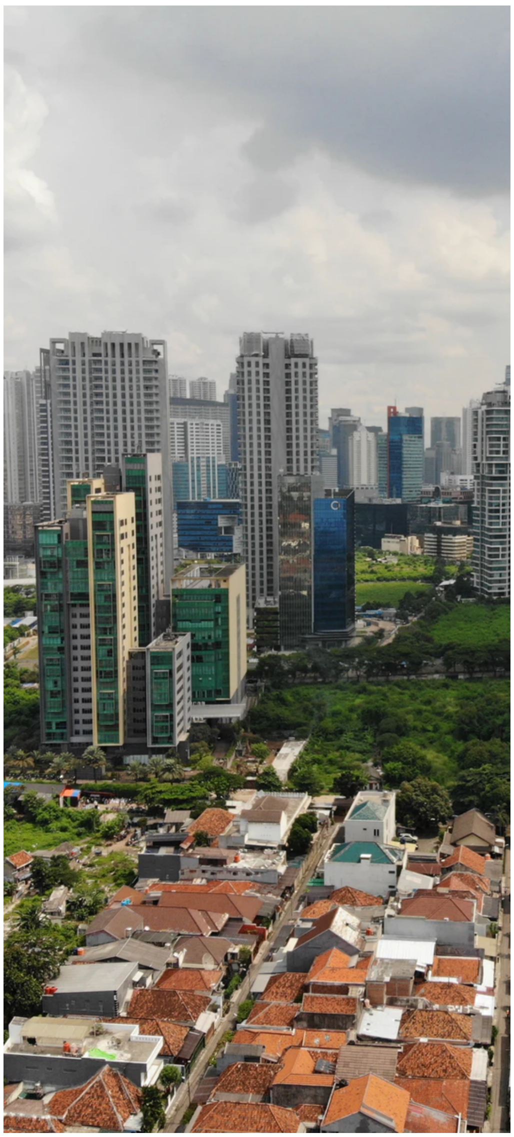
[83] "GHSL - Global Human Settlement Layer." *GHS-POP R2015A*, European Commission, 6 July 2016, ghsl.jrc.ec.europa.eu/ghs_pop.php.

[84] *LandScan Global 2019 Release*. Oak Ridge National Laboratory, 21 Oct. 2020, landscan.ornl.gov/.

[85] Yetman *supra* note 35.

IV. USE CASES

Photo by Afif Kusuma on Unsplash



The preceding discussion outlined some ways in which Facebook's HRSL project has been deployed by humanitarian, civil society, and government organizations around the world. In this section, we consider three use cases in greater depth: an effort to combat Covid-19 in Indonesia, one to reduce the risk of a cholera outbreak in Mozambique, and another to support national electrification plans in low- and middle-income countries. Table 1 below contains brief descriptions and analysis of additional use cases. In July 2021, NetHope and Facebook Data for Good released a report detailing the experiences of individuals and organizations who had used Facebook's Data for Good dataset, noting that "Facebook high resolution population density maps are being used by a wide variety of non-profit organizations around the world, helping organizations become more data driven." [86]

As two of these initiatives consist of efforts to combat disease outbreaks, they contain obvious relevance to responses to the COVID-19 pandemic. In the following sections, we consider lessons learned and avenues for future research.

[86] *Facebook-NetHope High Resolution Population Density Maps Feedback Report.* Facebook and NetHope, 9 July 2021, https://app.box.com/s/0azcdr19vjhl4rlp04dl6wplzjorg3xg

A. Pulse Lab Jakarta's efforts to identify areas with high transmission risk and transmission potential for the spread of Covid-19 in West Java, Indonesia

Background

In September 2020, Indonesia registered the second highest transmission rate of Covid-19 cases in Southeast Asia. Pulse Lab Jakarta (PLJ),[87] a joint data innovation facility by the United Nations and Government of Indonesia, used several population data sources and advanced analytics to develop indicators to help predict the potential spread of Covid-19, and to prevent the resulting economic fallout.[88]

Project Overview

For this project, PLJ launched an initiative with Jabar Digital Service (JDS), UN agencies, and their main government counterpart, the Ministry of National Development Planning (Bappenas), to develop metrics that could assess how to safely lift large-scale social restrictions across Indonesia.[89] The project aimed sought to support the provincial government's loosening of Pembatasan Sosial Berskala Besar (PSBB), or large scale social restrictions, and restore economic and social activity.[90]

The project used two major sources of data. The first was Village Potential Census Data (PODES), a government census of villages last conducted in 2018. The second major source of data was derived from Facebook's population density maps. As PLJ puts it, "to have a better understanding of the transmission potential of each area, there is a need for data that has information on transmission factors, preferably at the smallest granular level to support the localised intervention mechanism."[91] PLJ merged these data sources to identify two relevant metrics: i) transmission potential index— the "baseline measure representing the possible capacity of coronavirus transmission in each village before hitting the first case"—and ii) transmission risk, "an ex-post measure, where villages without cases basically have zero risks

[87] While we were unable to secure an interview with a stakeholder adjacent to an initiative conducted by Pulse Lab Jakarta, we spoke with Andrew Schroeder, VP of Research and Analysis for Direct Relief, to better understand his perspective as a user of the HRSL.

[88] *Pulse Lab Jakarta.* https://www.unglobalpulse.org/lab/jakarta/

[89] Pulse Lab Jakarta. "Alternative Use of Traditional Data in Times of COVID-19." *Medium*, Pulse Lab Jakarta, 9 Mar. 2021, medium.com/pulse-lab-jakarta/alternative-use-of-traditional-data-in-times-of-covid-19-558914f23f7f.

[90] Ibid.

[91] Ibid.

regardless of their baseline score and villages with cases have positive risks associated with the number of cases and the baseline prospect."[92]

Impact and Lessons Learned
The insights developed from both metrics were used to develop policy interventions for Covid-19 response across Indonesia. As PLJ noted, "[w]hen faced with crucial decisions on the where, when, and how in relation to restarting economic activities or reopening schools for example, the insights we co-developed can help in part to inform policy makers' decision making."

One of the chief lessons learned stemmed from limitations of traditional government data sets. As PLJ also reflected, the "static spatial and temporal measures" they created from PODES "need to be coupled with more up to date dynamic data on a host of important factors, such as mobility and case incidence data," in order to effectively inform decision-making.[93]

However, the team behind the initiative also noted the importance of finding new ways to leverage traditional statistics, as they were able to do in subsequent iterations of the project. For all its shortcomings, census data also offered several advantages. First, census datasets contain publicly available, reliable information—though often not the most granular data. Second, census data is received through the government, which is seen as a more official and reliable source of information in many settings. Finally, census data is already structured in a way that is accessible and easy to use. In a final analysis, all of these advantages point to the benefits of combining traditional data sources with new ones, in the process creating new data structures and suggesting new, more innovative avenues for analysis and insight.

B. Using Facebook's Population Density Maps for Cholera Vaccination of 252,000 children in Mozambique

Background
In 2019, just six weeks after the devastation that followed Cyclone Idai, Northern Mozambique was struck again by Cyclone Kenneth – the strongest tropical cyclone ever to make landfall in the region. Idai sparked a massive Cholera outbreak. While a

[92] Ibid.
[93] Ibid.

large vaccination campaign was already underway, there was widespread concern that Kenneth could exacerbate the ongoing epidemic.

Project Overview

Prior to Kenneth's landfall, Dr. Caroline Buckee, Dr. Ayesha Mahmud and Rebecca Kahn from Harvard University School of Public Health led a team of researchers in collaboration with Direct Relief, Nethope Crisis Informatics, Facebook Data for Good and Jen Chan from Northwestern University School of Medicine to identify high risk areas and develop a model-based estimation of cholera spread. The goal of the modeling effort was to help prioritize where, given limited time and resources, vaccines should be distributed in Mozambique.

The team created a cholera outbreak risk model for several districts, based on four measurements: flooding risk index (derived from existing data about Cyclone Idai and projected flooding from Cyclone Kenneth); previous annual cholera incidence; sensitivity to previous outbreaks arising from the El Nino-Southern Oscillation cycle (irregularities to winds and sea surface temperatures in the eastern Pacific Ocean); and a model that simulates movement patterns of infected travelers.[94] The model used HRSL as an input, which provided population estimates for the areas affected. The team created a web-based tool to help examine these risk factors more closely. [95]

The model predicts how affected populations would behave in real time. As Direct Relief notes, "[i]n the case of Cyclone Idai, density maps can show the percentage change in population, measuring how a certain fraction of the

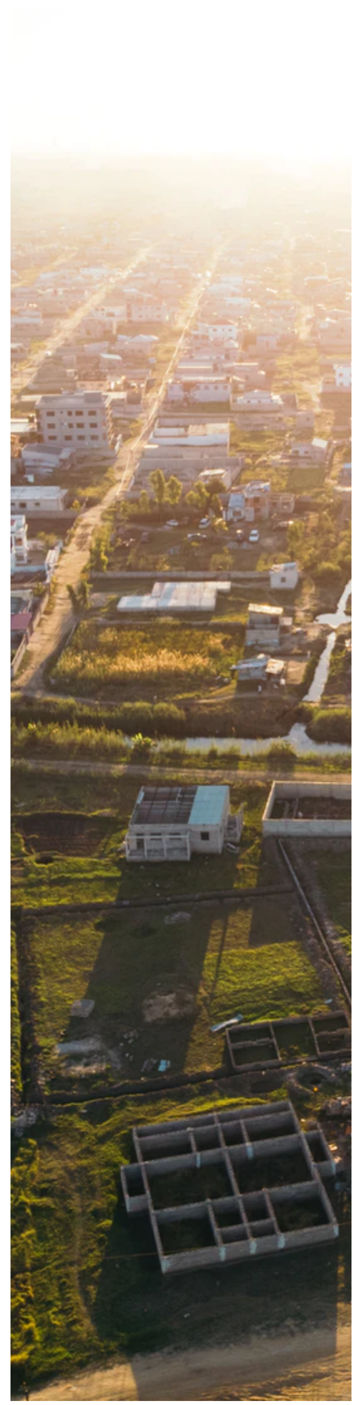


Photo by Edvin Johansson on Unsplash

[94] Kahn, Rebecca, et al. "Rapid Forecasting of Cholera Risk in Mozambique: Translational Challenges and Opportunities." *Prehospital and Disaster Medicine*, U.S. National Library of Medicine, Oct. 2019, www.ncbi.nlm.nih.gov/pubmed/31477186.

[95] Schroeder, Andrew, et al. "Implementing Real-Time Disaster Risk Models: Collaborations across Government, Non-Profits, Academia and Private Sector in Mozambique and the Bahamas." Understanding Risk. 2021.

population is moving.”[96]

Impact and Lessons Learned

The result of this modeling helped the humanitarian community assess and mitigate the risk of an outbreak. Though its impact has not been independently verified, this modeling effort laid the groundwork for similar work in northern Mozambique, which lay in the inland path of Cyclone Kenneth. [97] Those who participated in this effort also posted the data, code, and methodology for this modeling effort on GitHub, for other interested parties and researchers.

The initial model developed for this effort was for the Beira area. However, by the time the model was fully tested, it was no longer operationally relevant as the cyclone had already passed through. In addition, the model considered data at a district level, but the campaigns took place one level below this, which limited planning capabilities for a massive vaccination campaign.

Nonetheless, despite the limited utility of the model itself, the data it relied on was ultimately used in the Ministry of Health’s planning and response efforts. In particular, it informed a large-scale vaccination campaign strategy for over 250,000 children in northern Mozambique. This unintended benefit shows the advantages of taking an agile, flexible approach to how population density datasets are used, and suggests the value of preparedness, pre-existing data partnerships, and streamlined data collection and aggregation practices.[98] [99]

In addition, as Andrew Schroeder noted, the effort also had a significant educational impact on the Ministry of Health. It demonstrated the value of predictive modeling and building early partnerships to have data on hand in the event of an emerging threat. This further illustrates the importance of adaptability in how initiatives respond to needs and conditions on the ground, and shows the vital role played by Minimum Viable Models combined with clean, integrated population density information. “Keeping these kinds of cumulative learnings going is crucial,” Schroeder said, “because it's not going to be one event that singularly changes everyone’s mind.”[100]

---

[96] Schroeder, Andrew. “Cyclone Idai: Analyzing Population Movements in Mozambique with Facebook Disaster Maps.” DirectRelief. March 14, 20129. https://www.directrelief.org/2019/03/cyclone-idai-analyzing-population-movement-in-mozambique-with-facebook-disaster-maps/.

[97] Ibid.

[98] Ibid.

[99] Schroeder, Andrew GovLab <> Andrew Schroeder - Direct Relief in Mozambique. Zoom, April 20, 2021.

[100] Schroeder *supra* note 84.

In an interview, Schroeder and his colleagues at Direct Relief also raised some benefits, challenges, and lessons learned from using Facebook data during this initiative. For example, they pointed out that Facebook's population density data set provided neighborhood-level calculations, allowing for more targeted modeling and responses. In addition, they noted that Facebook's platform data was unfortunately too spotty for Mozambique in order to be useful, given the low penetration of smartphone users. As a result, the Facebook population density map for Mozambique —which is not derived from Facebook's user base—was seen as the most beneficial tool for the setting. Uncertainties remain, however, given the significant length of time since the last census, which informed the Mozambique's population density map.

C. Universal Access Energy Lab and Waya Energy's use of HRSL for Rural Electrification Planning

Background

According to the Universal Access Energy Lab (UAEL), of the 840 million people worldwide who lack access to electricity, 570 million live in sub-Saharan Africa.[101] This represents over half the region's population, and projections predict that 9 out of 10 people who lack access to electricity in 2030 will be in sub-Saharan Africa. [102] The UAEL developed the Reference Electrification Model (REM) to help plan detailed electricity networks in developing countries, and connect unelectrified people to the existing infrastructure.

Project Overview

REM uses various data streams to determine the optimal "electrification modes (e.g., grid-connected, microgrid, or isolated system)," estimate costs of electrification for each unit, and "produce preliminary engineering designs of recommended systems."[103] As a part of this effort, UAEL uses Facebook's HRSL to determine the location of residential customers and expected growth through 2024. Waya Energy, a spin-off of UAEL, partners with governments, utilities, agencies, and private-sector organizations to implement this model. Government planners and entrepreneurs can use REM to make better decisions about how to plan and implement electrification

[101] *Universal Energy Access Laboratory - About.* MIT & IIT-Comillas, 2020, universalaccess.mit.edu/#/main.
[102] *Universal Energy Access Laboratory - About.* MIT & IIT-Comillas, 2020, universalaccess.mit.edu/#/main.
[103] Ibid.

projects across the country. Since 2018, for example, Waya Energy has been supporting the development of the National Electrification Plan for Rwanda.

Impact and Lessons Learned
To date, REM and Waya Energy's planning software has been used to design off-grid systems in India, Uganda, Peru, and Indonesia.[104] In Rwanda, Waya Energy and its partners developed the *Design of the National Electrification Plan in Rwanda*, which helped build an implementation plan for national electrification, and aims to connect three million customers. This plan quadrupled energy access in Rwanda in the past seven years, and will continue to support the government's goal of universal access through 2024.[105]

Andres Garcia, Co-Founder of Waya Energy, highlighted some of the challenges and opportunities in using HRSL to support national electrification. For some countries, such as Colombia and Ecuador, the team is unable to rely exclusively on HRSL due to its lack of accuracy in rain forest areas in the Amazon. Due to the lack of precision, UAEL has to re-check and validate the population data with ground-truth data or satellite imagery. He hopes that future iterations of HRSL will allow them to expand their work.[106] As it stands, the HRSL provides population density data with household-level granularity, which is crucial for the REM model. [107]

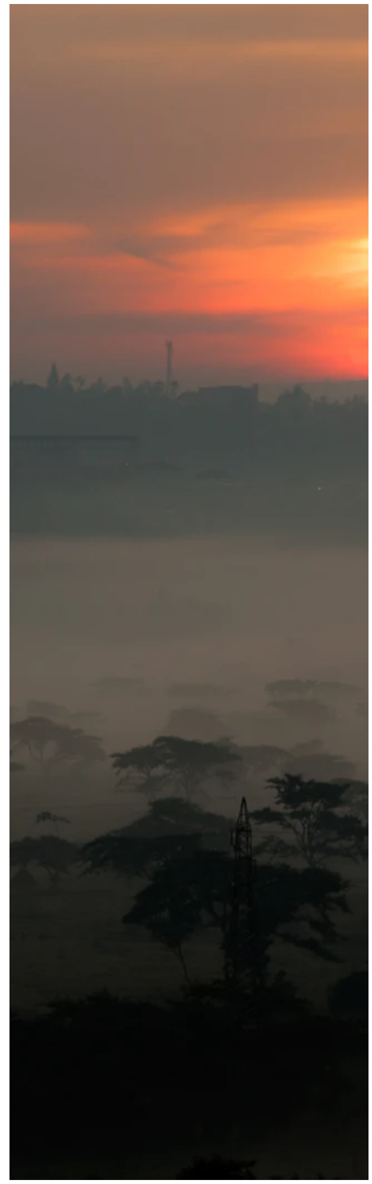


Photo by maxime niyomwungeri on Unsplash

[104] "Our Projects." *Waya Energy*, 2021, waya-energy.com/projects.

[105] "Electricity Access in Rwanda Quadrupled in the Last 7 Years as More Households Get Connected." *Rwanda Energy Group*, 2021, www.reg.rw/media-center/news-details/news/electricity-access-in-rwanda-quadrupled-in-the-last-7-years-as-more-households-get-connected/.

[106] Garcia, Andres Gonzalez and Clara Pérez-Andújar Carretié. Use of HRSL in national electrification plans by the Universal Access Energy Lab. Zoom. June 2, 2021.

[107] Garcia *supra note* 88.

Table 1: Example Use Cases for HRSL

Project Title	Country	Collaborative Type	Description
Cause and Effect Analysis			
Identifying areas with high transmission risk and transmission potential for the spread of Covid-19 in West Java, Indonesia	Indonesia	Data Pooling	Pulse Lab Jakarta (PLJ) undertook efforts to help mitigate the spread of the pandemic and prevent economic fallout. In a project initiated with Jabar Digital Service (JDS), UN agencies, and their main government counterpart, the Ministry of National Development Planning (Bappenas), to develop an understanding of how to safely lift *Pembatasan Sosial Berskala Besar (PSBB)*, or large scale social restrictions, across Indonesia. The project is aimed at supporting the West Java province government's decision to restore economic and social activity as PSBB loosens.
Situational Awareness			
Map of health infrastructure gaps in Ethiopia for COVID response	Ethiopia	Research and Analysis Partnership	Recognizing the scarcity of data-driven responses to Covid-19 in developing countries, Village Data (VIDA), with support from the European Space Agency, "built a data-based decision making tool that uses earth observation, on-ground data and artificial intelligence to improve how infrastructure, policy and investments can be planned." VIDA released a paper focused on one such use of the tool. The paper examines questions on the accessibility of hospitals and health centers, test centers' access to grid power, and likely changes to hospital capacity given certain trends.
Understanding the need for safe drinking water	Global	Intelligence Generation	mWater, a technology company for water, sanitation and health, embedded HSRL in their app to help NGOs providing safe drinking water track their physical assets and coordinate with other organizations.

Prediction			
Cities, crowding, and coronavirus: Predicting contagion risk spots	Global	Research and Analysis Partnership	The World Bank developed a methodology, using high resolution population density maps, to help city leaders around the world identify areas with highest exposure and contagion risk to appropriately direct supplies and resources. The report outlines a methodology based on the practical inability to keep people apart (population density) and conditions in which people have little option but to cluster.
Planning waste-water networks using open source GIS Datasets	Saudi Arabia, India	Intelligence Generation	NetCreate used Facebook's population density maps to support the planning and design of wastewater networks, creating a digital tool "to assign the route of least resistance from each property to the lowest point in the catchment along defined roads."
Using HSRL to support vaccine distribution in DRC and Mozambique	DRC, Mozambique	Intelligence Generation	The Gates Foundation uses population density maps to guide health workers' vaccine distribution, procure the appropriate number of vaccines, and very the vaccination rates of children under five in the Democratic Republic of Congo and Mozambique.
Impact Assessment			
Rapid Mapping of Areas Marked for Demolition in Nairobi	Nairobi	Intelligence Generation	Cadasta Foundation and Pamoja Trusts created a public story map to highlight the number of people that would be displaced by municipal construction in Nairobi and advocate for government protection of their interests. HSRL was used to understand how many people live in these areas.

V. ASSESSMENT: BENEFITS, CHALLENGES, AND RISKS

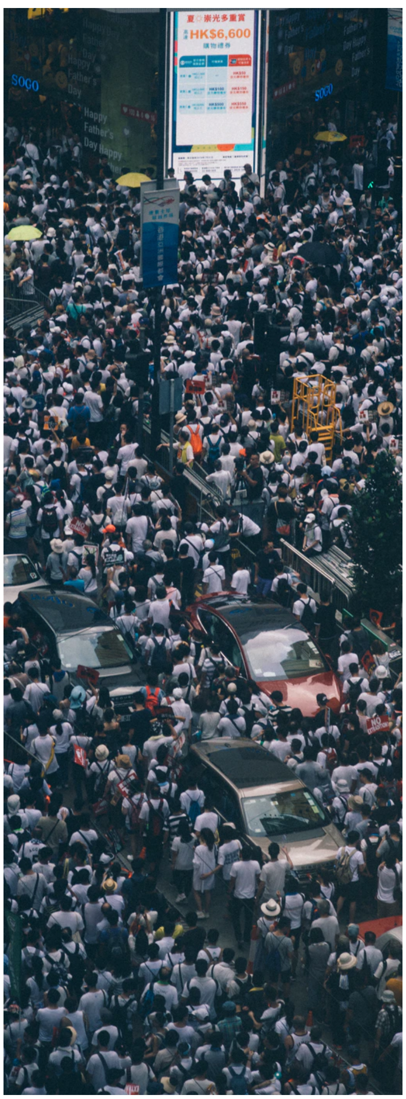


Photo by Joseph Chan on Unsplash

Based on the prior analysis and supplementary interviews with stakeholders such as Andreas Gros, Greg Yetman, Andrew Schroeder, Andres Gonzalaz Garcia, and Clara Pérez-Andújar Carretié, we can identify several benefits, challenges, and risks associated with HRSL. While not comprehensive, these elements are important for data practitioners to be mindful of before they use HRSL.

A.Benefits

- **Filling information gaps:** Interviewees spoke at length about how HRSL provides local policymakers and researchers, typically in information and resource-scare settings, with the information they need to understand population densities and patterns. Insights from this information, they argued, can help decision-makers make short-term policy decisions, such as whether to institute lockdowns amid the pandemic, or allow them to arrange for long-term interventions, such as better waste-water networks. HRSL can offer opportunities for various data users, seeking to conduct impact assessment, prediction, situational awareness, cause and effect analysis to achieve public goals.

- **Granular and accurate population information:** HRSL hosts highly detailed and granular public-facing population density datasets. Interviews indicated that this detail makes HRSL a resource for decision-makers who need rapid, neighborhood-level insights for response planning to, for example, distribute aid or supplies efficiently.

- **Iterating on the technology:** While several gaps in HRSL still exist, the Data for Good team continues to iterate on the technology. Gros explained, for example, that Facebook has ambitions to develop building type classifiers in the existing algorithm. This would help more accurately assign densities to buildings—a residential building, for example, will hold more people than a commercial one.[108]

B. Challenges and Risks

- **Reliance on underlying census data:** Although the CIESIN team creates modeled estimates for population in each country based on current growth rates, they are still estimates based on underlying census data, which is imperfect and outdated in many countries. CIESIN's underlying population estimates may be particularly outdated in parts of the world where a census has not taken place in many years, leading to the overall population numbers to over- or underestimate the total number of people living in any given area. However, there is potential to use bottom-up data collection methods in collaboration with national governments to supplement the lack of reliable or available census data.

- **Poor imagery in some places:** HRSL relies on a cloudless mosaic of satellite imagery around the world, which can be as recent as the past year, but can also be very old depending on the cloud cover when Maxar's satellites pass over particular geographies. This can lead to settlement detection in the algorithm that is out of date, as well as poor accuracy in particular geographies based on

[108] Gros *supra* note 59.

the topography. Facebook is working to improve their accuracy by incorporating radar data, but this is an ongoing process. [109]

- **Risks and sensitivities for vulnerable populations:** In his interview, Gros described the challenge with publishing HRSL in conflict zones or politically unstable areas, and stressed the need to mitigate negative political and social implications that may come from opening population data for public use.[110] Yetman noted the importance of openly publishing standards for data collection and use, and to anonymize datasets to avoid exploitation of data for nefarious purposes.

- **The need for better visibility and reporting:** Gros also noted that Facebook has limited visibility into how the data is being used outside of voluntary reporting from partners.[111] An improved understanding of how the data is being used and where opportunities for improvement and iteration lie, Facebook would be better-positioned in developing future iterations of the product and tailoring its offerings to the specific needs of its partners or users. Additional data collection from users, especially passive data collection, could create new risks, however, and would necessitate an appropriate governance strategy.

- **Low data capacity:** Lastly, Schroeder describes that users of HRSL sometimes face capacity constraints. Because the data is made freely available on HDX or AWS with little additional guidance, users need to be equipped with technical knowledge and software to functionally use the dataset. As a result, often only organizations with technical expertise and capacity are able to fully utilize the potential of HRSL. [112] As such, there is a need to build the capacity of humanitarian actors, by increasing access to training workshops or appropriate software, to fully take advantage of the benefits of HRSL.

[109] Gros, Andreas et al. "How maps builty with Facebook AI can help Covid-19 vaccine delivery." *Facebook AI.* 15 Apr. 2021, https://ai.facebook.com/blog/how-maps-built-with-facebook-ai-can-help-with-covid-19-vaccine-delivery/.

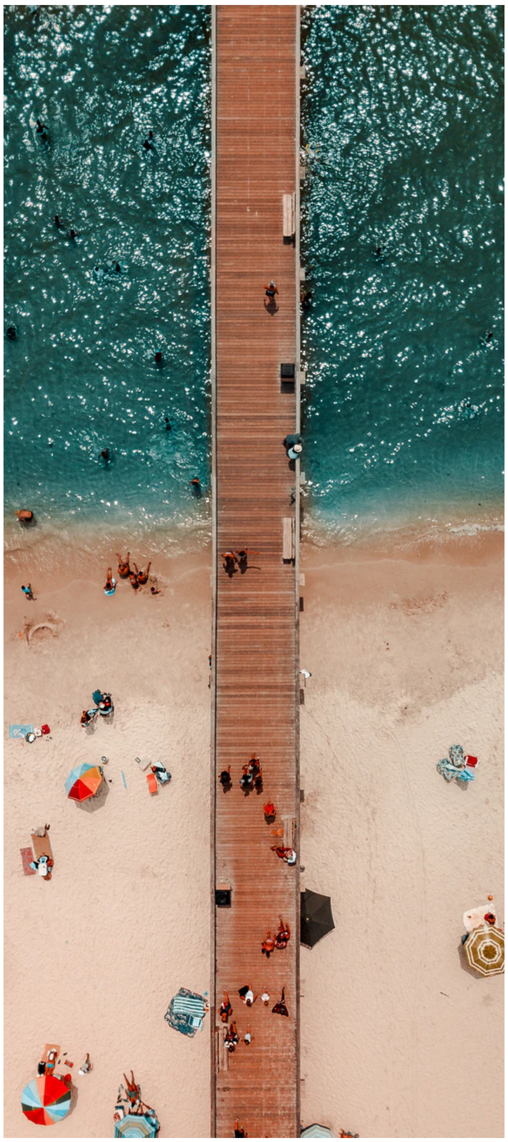
[110] Ibid.

[111] Gros *supra* note 59.

[112] Schroeder *supra* note 84.

VI. FUTURE RESEARCH AVENUES

Photo by Alex Perez on Unsplash



Along with new technological innovations and advances to population density data products, population indicators are increasingly being used to map and understand a changing world. Governments use population projections to gauge future demand for food, water, and other energy services. Population projections and modeling can help policymakers and decision makers plan for future crises.[113]

As the above discussion makes clear, however, many areas remain unexplored, and there are several fields in which population density data may become increasingly important in the coming years. In this final section, we examine some areas that contain particular potential, and that we believe should be the focus of future research.

1. Long-term Infrastructure Development: Geospatial data can be helpful to government and humanitarian actors in strategizing and executing development or planning projects. Combining geospatial data with augmented reality or virtual reality could help create an immersive context to re-create hazardous environments such as oil fields or nuclear plants,

[113] *Understanding and Using Population Projections*. PRB, 3 Dec. 2001, www.prb.org/understandingandusingpopulationprojections/.

or visualize large-scale architecture projects. For example, virtual structural analysis programme (VSAP) technology helps users virtually examine the architecture of buildings.[114]

Greg Yetman discussed how population density data, including those shown through Facebook's HRSL, is useful in providing city and state actors with hyper-detailed estimates that contribute to long-term resiliency and planning projects. The New York City fire department, for example, was working with Facebook to understand how the population is dispersed across the city to inform their own planning in advance of crisis events. [115] In low-income communities, geospatial imagery can also help inform critical needs. In the article, *The Past, Present, and Future of Geospatial Data Use* Behm et. al note "geospatial technology is being used to detect and geo-locate waste tire piles in Africa, which are a significant breeding ground for disease-carrying mosquitoes."[116]

2. **Disaster Preparedness:** With the rise of cell phone data, digital checking data, and location data global actors have used population indicators to respond to disasters such as fires, floods, hurricanes, cyclones, and earthquakes. In the aftermath of Hurricane Maria in Puerto Rico in 2017, 97% of the island's infrastructure was destroyed. To deploy help, CrowdRescueHQ Puerto Rico Map created an open platform, harnessing the power of developers and cartographers, to navigate post-hurricane recovery.[117] However, there still remain several untapped areas for population density data for humanitarian response and preparedness. For instance, detailed geospatial mapping could be used by first responders to reduce response times and increase responder safety. It is also important to note that some use cases of remote sensing data can also lead to unethical outcomes and denial of humanitarian aid.

3. **Simulation and Analysis:** Population density datasets and modeling efforts can support simulation exercises to help humanitarian actors make better decisions and react appropriately in the event of an actual emergency. Population data can also contribute to high-level population analysis. Governments, for example, can use social vulnerability analyses to understand the needs and conditions of various

[114] Ibid.

[115] Yetman *supra* note 35.

[116] "The Past, Present, and Future of Geospatial Data Use." *Trajectory Magazine*, 1 Feb. 2018, trajectorymagazine.com/past-present-future-geospatial-data-use/.

[117] *ArcGIS Web Application*. www.arcgis.com/apps/webappviewer/index.html?id=24da7c00f59a4876989c4e7b02e0a3ab&extent=-7550145.754%2C1989803.0956%2C-7256627.5654%2C2138702.4267%2C102100.

populations and drive policies with a better understanding of demographic variances.[118]

4. **Last Mile Service Delivery:** Population density data can also be used to access last-mile populations, helping in efforts to inform timely decision-making about rural populations and deliver critical services such as electricity, internet, or health. For example, the National Primary Health Care Development Agency used high-resolution satellite imagery, machine learning, and advanced data science to map and plan vaccination campaigns across northern Nigeria.[119]

5. **Climate Change Response and Agriculture Forecasting:** Changes to our natural environment are heavily impacted by population density. In the coming decades, density indicators can support both knowledge and planning for agricultural forecasting and climate change resiliency projects, such as coastal zone retreat planning. Some data collaborative initiatives are already doing this. CIESIN, in a project with City University and funded by the World Bank, are predicting population movements due to climate change impacts through 2030.[120]

6. **Urban Planning:** Finally, the availability of open source geospatial datasets can accelerate urban planning efforts. Increased use of autonomous driving vehicles and their IoT sensors can lead to other applications of geospatial data to enhance safety and monitor real time trends in traffic, weather, road conditions as well as urban planning.

Overall, there is great potential to discover and harness the power of geo-spatial data to augment and enrich a wide variety of applications in order to increase the public good. As computational capabilities continue to grow, and tools like the HRSL continue to improve, we can deliver rapid insights in information-scare settings to bridge resource asymmetries. Early indications point to the potential of data collaboration in the space to further social, economic, and environmental goals, though these efforts will require fit-for-purpose governance strategies and frameworks to ensure their effectiveness and legitimacy.

---

[118] Schroeder *supra* note 84.

[119] "Outside the box: how Nigeria won the fight against polio." GRID3. 31 July 2020. https://grid3.org/news/outside-the-box-how-nigeria-won-the-fight-against-polio

[120] Yetman *supra* note 35.

THEGOVLAB