Summary Report of the POPGRID workshop on

Gridded Population Data in the Context of the Sustainable Development Goals

Held on November 30, 2020

Introduction

Many of the Sustainable Development Goals' (SDGs) indicators have a direct connection to total population count, and gridded population data are valuable for policymakers to measure population growth, monitor change, and plan interventions. Traditionally, population counts are derived from census data. However, infrequent intervals of data collection, changes in reporting units' boundaries, inaccessibility to certain geographic locations, and other social constraints (e.g. language barriers between enumerators and indigenous communities) often leave many uncounted in the total population estimates. Over the past two decades, high-resolution satellite imagery, advanced Geographical Information Systems (GIS) and remote sensing capabilities have created the opportunity to produce gridded population datasets (based on different modeling approaches to disaggregate population data using spatial ancillary data and imagery) that can complement census data and fill in existing data gaps. For instance, gridded population data can easily be integrated with remote sensing data, aggregated to non-administrative, and alternative geographies. These benefits all work towards the goal of leaving no one behind.

Background and Objectives

New technologies, such as GIS, paired with new data sources and advanced methods, have provided an unprecedented opportunity to derive population estimates at fine granularity worldwide. However, to date, there has been a lack of effective and efficient information-sharing on the latest innovations and methods to support skills training in institutions, such as National Statistical Offices (NSOs).

The POPGRID Data Collaborative aims to bring together and expand the international community of data providers, users, and sponsors concerned with georeferenced data on population, human settlements, and infrastructure. The Collaborative works to improve data access, timeliness, consistency, and utility; support data use and interpretation; identify and address pressing user needs; reduce duplication and user confusion; and encourage innovation and cross-disciplinary use. It brings expertise and perspectives from diverse natural, social, health, and engineering science disciplines and sectors, and from government, academia, private industry, and nongovernmental organizations. POPGRID is managed by

the Center for International Earth Science Information Network (CIESIN) and the UN Sustainable Development Solutions Network's Thematic Research Network on Data and Statistics (TReNDS).

Funded by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ, a German development agency), SDSN TReNDS organized a workshop on November 30, 2020 for NSO representatives from DANE (Colombia) and DGEEC (Paraguay) to address population data challenges and raise awareness about the existence and nature of gridded population datasets produced by members of the POPGRID Data Collaborative. The workshop offered technical training to help NSO representatives in Colombia and Paraguay develop the skills needed to use gridded population datasets in different application areas.

Designed by Stefan Leyk, Professor at the University of Colorado Boulder and a member of the Collaborative, the virtual workshop provided training on the application of methods to use global gridded datasets on population for work related to SDGs. A total of 35 NSO representatives from the Data for Now Initiative[1] were invited to participate in the workshop. Professor Leyk instructed three 90-min sessions suitable for participants with different levels of experience in working with spatial data or in GIS software.

Summary of Session 1: Gridded Population Data and the POPGRID Data Collaborative

During this session, Professor Leyk discussed the motivation for and reasoning behind the creation of gridded population data as global data products; introduced the POPGRID Data Collaborative and its members; provided an overview of the global gridded population datasets available and important data properties; demonstrated the POPGRID Viewer; reviewed the methodological frameworks applied to produce the different data, and finished the session with a discussion on data uncertainty and limitations. The content was relevant to all participants, regardless of prior experience and training. Participants were also provided with a rich bibliography on the above topics.[2]

Summary of Session 2: Using Gridded Population Data in a GIS Environment

This session offered hands-on interactive training using three global gridded population datasets, WorldPop, GHS-POP, and Gridded Population of the World (GPW), in a GIS environment. Professor Leyk guided the participants step-by-step through an example to allow for participants to follow each step in real-time, and provided a second example as a demonstration that the participants could later practice with after the workshop. The two topical examples (see below) were selected based on the interests of the NSOs in SDG-related work.

More specifically, during this session, participants were introduced to the use of global gridded population data in typical applications relevant to sustainability aspects, and overcoming limitations due to spatially aggregated or irregularly collected census data. These are just two out of many possible examples that can be carried out in a GIS environment. Participants also learned how to compare analytical results based on different population grids, which provided important grounds for discussion on the topics of

---

[1] The Data For Now initiative aims to equip countries with the best knowledge, data sources, and technological solutions to produce more timely and high-quality data on the SDGs. Current country partners include: Colombia, Paraguay, Senegal, Rwanda, Ghana, Mongolia, Nepal, and Bangladesh.

[2] Please see the "The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use" paper for a comprehensive list of works provided in the bibliography.

uncertainty and fitness for use to ensure that the participants understood the importance of such data resources. The following two examples were covered:

1) *Interactive Exercise: How many people live in close proximity to inland surface water?*
Participants learned how to estimate the number of persons that live in close proximity to surface water (streams, water areas), which can have important implications for issues around water supply, transportation, as well as exposure to natural hazards. Additionally, knowing the number of people who live in close proximity to water can be critical for policymaking, planning, and mitigation purposes. The exercise demonstrated how population estimates can be extracted based on different population grids and using different distance parameters to define "proximate." This exercise used data from Paraguay.

2) *Demonstration: How many people face accessibility constraints to urban centers and critical services?*
In this exercise, Professor Leyk demonstrated how different population grids and novel gridded data on travel time can be used in combination to evaluate how many people have to travel more than a defined amount of time to reach the nearest urban center (e.g., places with more than 50,000 people). This type of analysis can be useful to better understand accessibility to public services, health care, and food supply that often depend on proximity to urban centers of a certain size and to identify population groups with limited accessibility because of immobility or long travel times. For this example, data from Colombia data were used.

For both exercises, pre-built geoprocessing models were provided to demonstrate the power of automation tools in GIS environments for making the work more efficient, and to improve productivity and sensitivity analysis when the user aims at computing different scenarios. For this session, the group used Quantum GIS (QGIS), an open GIS software product, developed by the QGIS Community.

Summary of Session 3: Understanding NSO Requirements and Technical Needs

The goal of this session was to obtain feedback from the NSO representatives on the training, the technology, and the use and applicability of gridded population data in their current and future work. For this purpose, we developed questions that systematically covered all important aspects on:
   A. Content and training;
   B. The different kinds of data needs;
   C. Further training needs;
   D. Plans for NSOs' work, SDG-related or other topics; and
   E. Opportunities to establish a communication and discussion platform between participants and POPGRID members

POPGRID members from CIESIN and the Joint Research Center of the European Commission joined this session and presented some ongoing data production efforts related to GHS-POP and GPW. This provided participants with the opportunity to ask questions directly related to the different data products and for POPGRID members to learn about NSOs' needs and requirements. The participants were also given ample

time to ask questions about the data, confidentiality issues, and the technology aspects related to such work.

Post-workshop and Debrief Activities

The questions discussed with the participants during session 3 were refined and revised after the workshop, and an online survey in English and Spanish has been shared with the workshop participants in order to receive more in-depth reflections and quantifiable responses to the different aspects of feedback.

We have received interest in this kind of workshop for other Data For Now country partners, as well as interest from technology and industry partners to support future training events.

Acknowledgments

We are grateful to Stefan Leyk for designing and instructing the workshop. We also express our gratitude to the translators, Jorge Castilla Viana and Juan Viyay Duggirala, for their hard work during the workshop sessions, as well as Maryam Rabiee, Bob Chen, Susana Adamo, and Sergio Freire for their support and contribution to the workshop.

Participating Organizations

• DANE, National Administrative Department of Statistics, Colombia

• DGEEC, General Directorate of Statistics, Surveys and Censuses, Paraguay

• Environmental Systems Research Institute (Esri)

• United Nations Statistical Division (UNSD)