Statistical Commission
Fifty-fourth session
28 February – 3 March 2023
Item 3(a) of the provisional agenda
Items for discussion and decision: Data and indicators for the 2030 Agenda for Sustainable Development

Background document
Available in English only

Examples of the implementation of the SDGs Geospatial Roadmap: disaggregating the SDGs by geographic location

Prepared by the Working Group on Geospatial Information of the Inter-agency and Expert Group on the Sustainable Development Goals Indicators

The **SDGs** Geospatial

ROADMAP

Examples of the implementation of the SDGs Geospatial Roadmap: Disaggregating the SDGs by Geographic Location

SUSTAINABLE DEVELOPMENT GOALS

SDGs GEOSPATIAL ROADMAP

Preface

The SDGs have multiple dimensions of data, inclusive of sex, age, income, migration, and disability status, as well as geographic location. These data dimensions enable analysis and identification of trends and patterns, making the SDGs actionable at more fine-grained levels. In transforming these data and statistics into information, policymakers can develop targeted policy action by generating knowledge and insights.

The disaggregation of SDG indicators by geographic location provides a mechanism to achieve a greater analytical potential of the data, turning them into a high-quality, consistent, integrable, accessible and timely tool for the generation of information that allows for more accurate and real-time decision-making. Disaggregation by geographic location, alone or in combination with other dimensions (sex, age, income, migration, disability status), allows for uncovering the existing hidden societal disparities, bringing to the fore of analysis vulnerable, precarious and marginalised segments of the population.

This paper discusses examples of how countries, and the SDG Custodians Agencies that support countries, are disaggregating the SDGs by geographic location in the context of the SDGs Geospatial Roadmap. It was collaboratively developed by the WGGI, following a broad process of qualitative consultation with NSOs and NGIAs representatives of both the IAEG-SDGs and WGGI. This paper was developed to highlight how countries have disaggregated SDGs by geographic location in the context of the SDGs Geospatial Roadmap. The SDGs Geospatial Roadmap aims to be an interactive living resource, which invites the statistical, data and geospatial information communities to contribute with new resources, services and examples of best practices. Further information on the Roadmap and the work of the WGGI is available here: https://ggim.un.org/UNGGIM-wg6/.

Please contact the co-Chairs of the WGGI for further information:

Mr. Kevin McCormack, Central Statistics Office, Ireland - Kevin.mcCormack@cso.ie

Ms Sandra Moreno, DANE, Colombia - slmorenom@dane.gov.co

This document is available as an interactive Storymap:

https://storymaps.arcgis.com/stories/85abafc5794a4a67ad78cb19f99a963f

Introduction: What is disaggregation by geographic location, and why is it important for the SDGs?

The 2022 Sustainable Development Goals (SDGs) Report[1] highlights how "the 2030 Agenda for Sustainable Development is in grave jeopardy due to multiple, cascading, and intersecting crises. COVID-19, climate change and conflict predominate. Each of them, and their complex interactions, impact all of the SDGs, creating spin-off crises in food and nutrition, health, education, the environment, and peace and security. To put the world on track to sustainability will require concerted action on a global scale". The vital need to 'put the world on track' has highlighted the fragility and limitations of existing national statistical systems underscoring the crucial role that innovations, such as geospatial information and its enabling technologies can play to help make up for the global the progress lost, placing us on a more sustainable path of global development.

Disaggregation by geographic location is the breakdown of data to smaller geographic areas. These geographic areas could be Administrative (i.e. from a national level to a local level) or could be grids, urban/rural areas or basins/sub-basins

The SDGs Geospatial Roadmap[2]

The role of geospatial information in the 2030 Agenda for Sustainable Development and its 17 SDGs is now well understood by various domains, including the United Nations Statistical Commission (UNSC) and the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM). For the SDGs, a milestone in this journey was the adoption of the SDGs Geospatial Roadmap by the UNSC, by its decision 53/101, made in March 2022. Through its three phases and their associated key actions, the Roadmap's recognises and accepts geospatial and location-based information as official data for the SDGs and their global indicators, providing actionable guidance on the use of geospatial information (inclusive of Earth observations and other forms of location-based data) for the production, measurement, monitoring, and dissemination of geospatially enabled SDG indicators. The SDGs Geospatial Roadmap provides simple and actionable guidance to the Inter-Agency and Expert Group on the SDGs, Member States and Custodian Agencies to bridge this gap and realise the innovation potential that using geospatial information and its associated technologies can bring to the SDGs.

"Geospatial information describes the physical location of geographic features and their relationship to other features and associated statistical information. Geospatial information is presented in many forms and mediums, including maps, EO, and aerial photography. It is a nation's 'digital currency' for evidence-based decision-making and a critical component of its national infrastructure and knowledge economy that provides a nation's blueprint of what happens where and the means to integrate a wide variety of government services and functions, inclusive of economic growth, national security, sustainable and equitable social development, environmental sustainability, and general national prosperity. A geospatially-enabled nation shares, integrates and uses a wide range of data to achieve social, economic and environmental benefits. This use and associated benefits extend across governments, businesses, and citizens, and from national to the city and small community levels."

Integrated Geospatial Information Framework - Part 1: Overarching Strategic Framework[3]

[1] The Sustainable Development Goals Report 2022: https://unstats.un.org/sdgs/report/2022/
[2] The SDGs Geospatial Roadmap: English: https://ggim.un.org/documents/SDGs-Geospatial-Roadmap.pdf | French: https://ggim.un.org/documents/La-feuille-de-route-geospatiale-des-ODDe.pdf | Spanish: https://ggim.un.org/documents/La-hoja-de-ruta-geoespacial-de-los-ODS.pdf
[3] The Integrated Geospatial Information Framework, Part 1: https://ggim.un.org/IGIF/part1.cshtml

The call for data disaggregation is a prevalent theme that runs across various international frameworks. In particular, the focus on vulnerable population groups and areas is underscored in the Global indicator framework[4] of the SDGs, through the 2030 Agenda's overarching principle of "leave no one behind". Indeed, at the very core of the monitoring framework, there is an overarching principle of data disaggregation stating that "SDG Indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics in accordance with the Fundamental Principles of Official Statistics[5]".

It is vital for the SDGs to be geocoded at the finest level of geography possible to enable their greatest potential to enable interoperability and comparability across geographic locations and data sources. Improving data disaggregation with the addition of geolocated data and statistics within the SDG indicator framework will positively impact the following:

1. Enable geographic disparities to be highlighted that may be hidden by higher-level geographies;
2. Ensure the harmonisation of information for measurement and monitoring;
3. Develop capacities and tools to enable a deeper understanding of SDG data and statistics;
4. Foster better decision-making at local levels of geography;
5. Promote routine and standardised comparative and trend analysis that cannot be observed in aggregated geographies; and,
6. Enable more detailed visualisation of SDGs at local levels.

The disaggregation of the SDGs by geographic location is essential for uncovering otherwise unknown or hidden disparities for vulnerable, precarious, and marginalised segments of the population – a crucial principle of the 2030 Agenda "leave no one behind". These include socially vulnerable groups at the sub-national level, as aggregate data often hides disparities not only within groups but also by spaces and places. Hence, the effectiveness of the SDG indicators depends not only on the statistical design of the data, but also on being geocoded by an adequate geography, such as an x- and y-coordinate. In adequately geocoding data, the ability to geographically analyse data becomes possible, including by identifying spatiotemporal patterns and understanding the interlinkages between different statistical dimensions. The analytical utility of the information carried in data depends largely on the effectiveness of its statistical design and geographical scope and scale. In this regard, geographic location is a vital integrator for every data dimension of the global indicator framework, as it is the basis for integrating all forms of data across dimensions and within geographies.

Therefore, this paper aims to support countries to disaggregate SDGs by geographic location, in the context of the SDGs Geospatial Roadmap. It does this by considering the importance of disaggregating the SDGs by geographic location, discussing what is needed to enable disaggregation by geography (emphasising the importance of the Global Statistical Geospatial Framework), identifying challenges and limitations, and highlighting specific examples and cases from countries and SDG custodian agencies.

What is needed to disaggregate by geographic location?

The socioeconomic fabric of a nation is consistently changing; the only constant is change. Thus, the effectiveness of the SDGs relies on up-to-date and fit-for-purpose data and statistics. As recommended by the UNSC, all statistical unit records should be geocoded, preferably with an x- and y-coordinate. Geocoding SDG data in this manner provides a basis for aggregation to various levels of geography. If geocoding with an x- and y-coordinate is impossible, then the statistical unit record should be geocoded to the smallest level of geography available.

[4] SDG Global indicator framework: https://undocs.org/A/RES/71/313
[5] The Fundamental Principles of Office Statistics GA Resolution 68/261 - https://undocs.org/A/RES/68/261

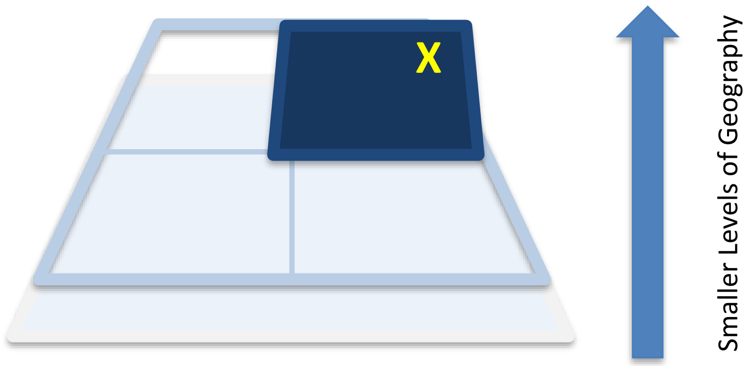


Figure 1: Geographies in context

Figure 1 provides a simple visual example of the importance of geocoding, where the 'x' is the "point" geography with x- and y-coordinates in a wider geographic area, which could be an administrative geography (such as administrative locality or a statistical enumeration unit) or a gridded geography. The SDGs Geospatial Roadmap is focused on guiding countries towards key actions that develop this capacity, and from the NSO perspective, a key framework is the Global Statistical Geospatial Framework[6] (GSGF). The GSGF and its Implementation Guide[7] offer actionable guidance, including ensuring the privacy and confidentiality of geospatially integrated statistical data. The below sections discuss the GSGF in the context of enabling the disaggregation of the SDGs by geographic location.

GSGF Principle 1: Use of fundamental geospatial infrastructure and geocoding and GSGF Principle 2: Geocoded unit record data in a data management environment

As recommended by the UNSC, all statistical data should be geocoded to the finest geographical scale possible, such as the x- and y-coordinate. While precise location geocoding is desirable, it is not always feasible with many types of social or industrial account data. Another important factor to consider is that the geographical level at which data are collected is not always the same level at which data are disseminated. A system of clearly defined spatial boundaries and location referencing is necessary to maintain linkages between different statistical geographies. Such a system of spatial referencing and hierarchy of locational units and their identifiers creates a fundamental geospatial infrastructure.

As with the generation of any statistical data, the calculation of the SDG indicators depends on the territorial/administrative geographic units or geographies defined in the fundamental geospatial infrastructure specific to each country. A hierarchical classification of geographical units allows for statistics to be disaggregated through the spatial reference codes assigned to primary observations. Principle 2 calls for each primary observation or statistical unit record of being geo-enabled or geocoded. The assignment of a unique geocode to each geography allows linking with other statistical data and geospatial information associated with the same geographic area. Using standard geocodes combined with a fundamental geospatial infrastructure allows for time- and new geography- robust and effective data linking and management.

[6] The Global Statistical Geospatial Framework: English: https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf Chinese: https://ggim.un.org/documents/GSGF_Chinese.pdf | French: https://ggim.un.org/documents/GSGF_French.pdf | Spanish: https://ggim.un.org/documents/GSGF-Post_Consultation_080719_Spanish_final_version.pdf

[7] The GSGF Implementation Guide: English only: https://unstats.un.org/UNSDWebsite/statcom/session_53/documents/BG-3x-EG-ISGI-GSGF-Implementation-Guide-E.pdf

GSGF Principle 3: Common geographies for the dissemination of statistics

Principle 3 discusses the importance of having common, standardised geographies to store, analyse and visualise statistical data. This common set of geographies ensures the consistent geospatial aggregation and dissemination of statistical data, irrespective of whether they have gridded, administrative, or statistical boundaries. These allow statistical data to be aggregated/disaggregated at different levels for their integration. An important aspect in defining the common geographies is the agreement on the maximum level of disaggregation, which should be determined with the consideration of, on the one hand, the technical and operational feasibility of achieving it; and, on the other, the established by law confidentiality of information rules, including the minimum count requirements for data release and dissemination.

GSGF Principle 4: Statistical and geospatial interoperability

Statistical and geospatial interoperability ensures that different types of information can be exchanged and integrated to inform analyses and decision-making. Improving interoperability means improving statistical and geospatial data and metadata standards and removing barriers (legal, institutional, and cultural) between the various stakeholders within national and international data ecosystems. The goal of statistical and geospatial information interoperability is to access, integrate, and link datasets across different systems and applications, enabling data users to have better information at different geographies. Producers of geospatial data operate using their data and metadata standards, which differ from the statistical data standards. Principle 4 calls for countries to recognise the need to integrate geospatial frameworks and data practices more explicitly into the statistical data production processes. This would improve geospatial and statistical data integration and interoperability between various data producers and consumers.

GSGF Principle 5: Accessible and usable geospatially enabled statistics

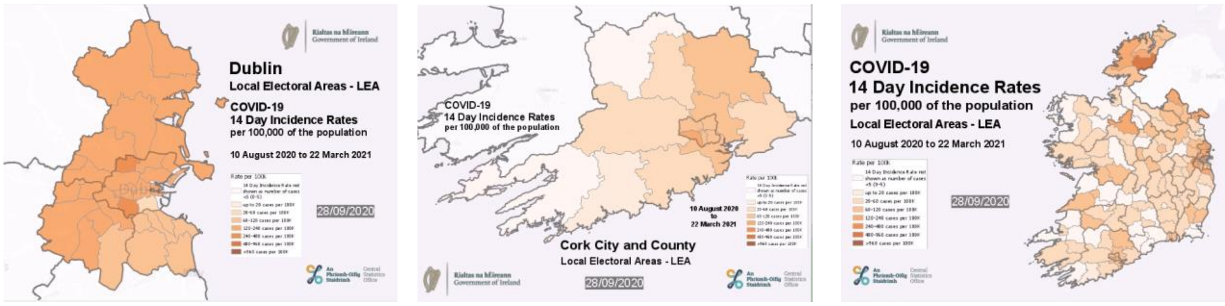
Principle 5 refers to data accessibility and its disclosure and dissemination. Geospatially enabled statistics must be accessible and usable following the prevailing international norms, standards, and good practices. Data users should be able to discover, access, view and manipulate information according to their interests and needs.

In the context of benefits for the SDGs, the commonly agreed definitions for the output geographies will determine the level of geographical detail for the calculation of each of the SDGs, aiming for a variety of agreed levels of geography, inclusive of second-level administrative boundaries, municipalities, cities or even at lower levels, if statistical disclosure guidelines are followed. Adopting a common geography standard for SDGs indicators estimation and dissemination requires collaboration between individual countries and stakeholders working across different data policy frameworks.

Some additional aspects must be considered in the design of geographical disaggregation of statistical variables, that include but not limited to the following:

- Define the levels at which data integrations will be made in advance;
- Avoid exclusions by design so that disaggregated geographies are inclusive of all areas of interest;
- Be clear about the purpose of geographic disaggregation; and,
  Ensure that updates to the underlying geographies can be made consistently and systemically.

Case: Monitoring the spread of COVID-19 infections, modelling future scenarios and informing policy decisions – The Irish perspective on COVID-19.



Dublin *Cork* *All Ireland*

Figure 2: A animated time series of the spread of COVID-19 infections, disaggregated to Local Electoral Areas in Ireland

In Ireland, the National Statistical Office and the National Mapping Agencyhave a long history of cooperation[8], traditionally centred on improving the quality and availability of geospatial data to meet the growing statistical needs, including for the Census' Geography. This relationship deepened in 2017 after Ireland joined the Federated Information System for the Sustainable Development Goals[9] (FIS4SDGs), an initiative led by the UN Statistics Division (UNSD) that aims to leverage state-of-the-art web technologies and services to improve the integration, accessibility and usability of official statistics, geospatial information, and other sources of data, including from outside the official statistical system, to support decision-makers at the local, national, regional and global levels in achieving the 2030 Agenda.

The outcome of Ireland's engagement with the FIS4SDG initiative was the development of Ireland's SDG data hub, an online collaboration platform for reporting on progress towards the goals and sharing information on related initiatives and is directly integrated into the national spatial data infrastructure platform, GeoHive[10]. Launched in 2017, this work has promoted the statistical potential of geospatial data by creating new national data sources and developing associated key partnerships across the public service. It has also influenced the work of the WGGI, which Ireland co-chairs with Columbia.

With the onset of the COVID-19 pandemic, there was an urgent requirement for national statistics to inform Ireland's response to the outbreak. Using the experience developed through Ireland's engagement with the FIS4SDG initiative and the WGGI, the CSO, OSi, the All-Island Research Observatory, the Health Protection Surveillance Centre, the Health Service Executive, the Health Intelligence Unit, and the Department of Health quickly collaborated to develop Ireland's COVID-19 Data Hub[11] which provided up to date data on Covid 19 in Ireland as well as several data visualisations such as geographical distribution of cases, as highlighted in Figure 2**Error! Reference source not found.** and animated in this paper's interactive Storymap.

At the global level, the World Health Organization (WHO) and the United Nations Department of Economic and Social Affairs (DESA) established a Technical Advisory Group (TAG) on COVID-19 Mortality Assessment in February 2021. This TAG serves as a broad scientific and strategic platform to facilitate the exchange of knowledge and application of methods on COVID-19 mortality to advise and support efforts to assist WHO and UN Member States to obtain accurate estimates of the numbers of deaths attributable to the direct and indirect impacts of the pandemic. Building on the experience gained through the FIS4SDGs and

[8] Central Statistics Office (CSO) and Ordnance Survey Ireland (OSi)
[9] The FIS4SDGs: https://unstats.un.org/unsd/statcom/50th-session/side-events/20190307-1L-Federated-Information-System-for-the-SDGs.pdf -
[10] Ireland's GeoHive: https://www.geohive.ie/
[11] Ireland's COVID-19 Geohive: https://covid19ireland-geohive.hub.arcgis.com/

Geohive and through leading the WGGI, Ireland has served as co-chair of this TAG, bringing the experience of the WGGI to the deliberations on global, regional, and country estimates.

Case: Disaggregating SDG 3 on good health and well-being – The Colombian perspective of community vulnerability and risk to COVID-19

In Colombia, the National Administrative Department of Statistics (DANE) Data Science Group of the National Planning Department spearheaded a collaboration with the Analytical Unit of the Institute for Health Technology Assessment (IETS) to develop a country-wide Vulnerability Index[12]. This Index disaggregates demographic data by health conditions, exposure to risk and other factors down to a census block level. Aligned with SDG 3, "Good Health and Well Being" (though impacts with many others), the Index helped inform public policy decision-making during the COVID-19 pandemic13, highlighting at-risk areas and clusters with higher potential for health complications in case of COVID-19 infection.

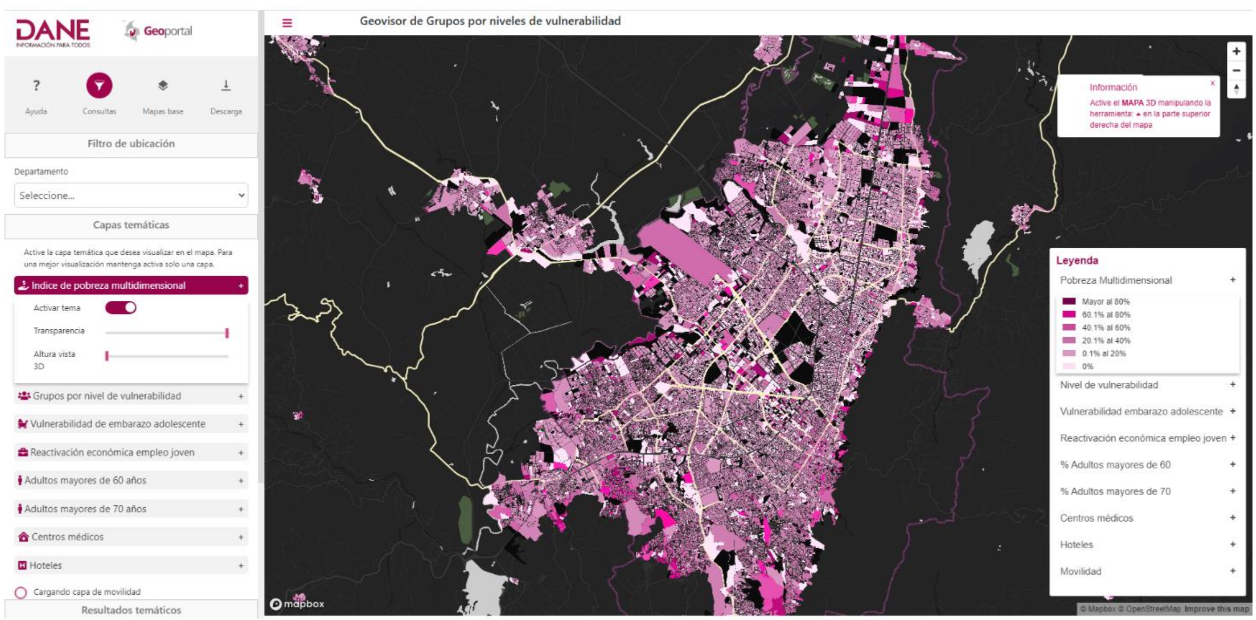


Figure 3: The COVID-19 Vulnerability Index for Colombia

The Index was developed using information from integrating data from the 2018 National Population and Housing Census with individual health service delivery records. According to demographic variables, comorbidities of the population and population density, each block of the municipal capitals is placed in one of five levels of vulnerability: low, medium-low, medium, medium-high, and high. Using a web-based

[12] Colombia's COVID-19 Vulnerability Index: https://geoportal.dane.gov.co/visor-vulnerabilidad/

[13] Particularly SDG Target 3.d "3.d Strengthen the capacity of all countries, particularly developing countries, in early warning, risk reduction and risk management for national and global health"

‘geoportal’ allows communication about differing geographies publicly, while informing more precise actions at smaller levels of geography.

Resources, opportunities, and challenges of using geographically disaggregated SDG indicator data

The UNSC stressed the importance of data disaggregation in the Global Indicator Framework, emphasising to the IAEG-SDGs to undertake efforts to develop the necessary statistical standards and tools while developing capacity at the country level for the disaggregation of the SDGs in their many dimensions. As a result, many relevant resources have been developed, such as the Practical Guidebook on Data Disaggregation for the SDGs[14] , which discusses in detail issues surrounding standards and methods for disaggregation. Yet, even with these resources, there are still limitations and challenges arising from the use of geographically disaggregated data.

The need for comparability: Developing and adopting standard definitions of disaggregation dimensions and categories

Defining such dimensions and their categories as sex (female/male) and age (child/adult/senior), though not without challenges, might be a more straightforward exercise, as common international standards for these categories have been defined and employed for many SDGs. The IAEG-SDGs has also developed a document entitled Overview of Standards for Data Disaggregation[15], which provides references to existing global and regional statistical standards for dimensions of disaggregation.

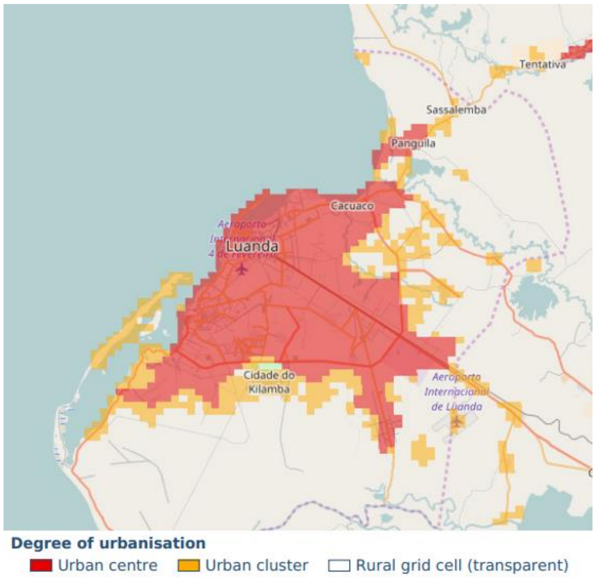


Figure 4: DEGURBA classification for Luanda, Angola (GHSL data) at 1 km² resolution

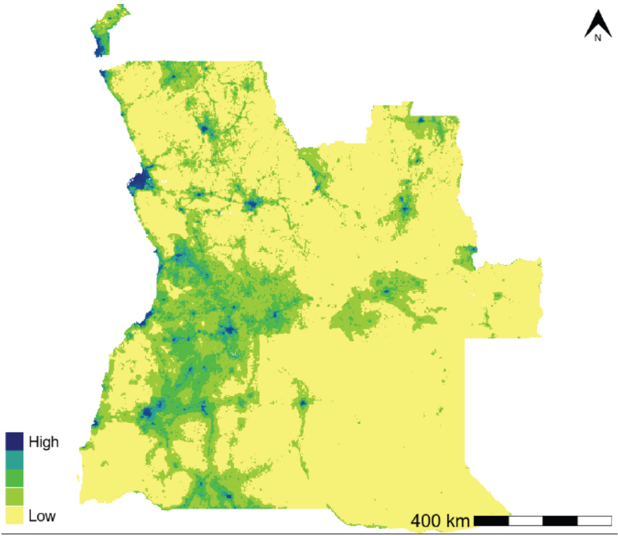


Figure 5: The coverage of Angola in WorldPop, at a 1 km² resolution

How data are disaggregated by geographic location differs between countries. Each country independently organises their geographies, with definitions varying greatly between what constitutes a

[14] Practical Guidebook on Data Disaggregation for the Sustainable Development Goals: https://www.adb.org/sites/default/files/publication/698116/guidebook-data-disaggregation-sdgs.pdf

[15] Overview of Standards for Data Disaggregation: https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/

city or a town, county or region, or rural and urban[16]. As a means of working toward a standard methodology between urban and rural areas, the UNSC, by its decision 51/112, adopted the Degree of Urbanization[17] (DEGURBA) as a methodology for the delineation of cities, towns, semi-dense areas, and rural areas for international and regional statistical comparison purposes, emphasising that the methodology is not intended to replace national definitions of urban and rural areas but to complement them.

However, population density classifications are not standardised at the global level. With the lack of a universal definition, the global comparability of geographically disaggregated population data becomes challenging. Here, global data sources such as WorldPop[18] offer global population grid products offer a means of bridging data gaps; these global sources of data can help provide global measurement and monitoring but also support the filling of gaps. Another global population dataset is the Global Human Settlement Layer[19] (GHSL), offering global population at defined resolutions.

Classifications, such as DEGURBA, directly depend on the chosen scale of the input data. The chosen resolution for DEGURBA is 1km. If the definition of DEGURBA classes changes to follow a more detailed scale of the input data, then the delineation between urban and rural areas will inevitably change. The SDGs indicators, particularly those directly utilising information on rural or urban dimensions for data disaggregation, such as SDG 9.1.1. “Proportion of the rural population who live within 2km of an all-season road” depends on the chosen scale/spatial resolution of the population data. Moving from raster grids to vector geography of administrative and statistical unit boundaries, the same is true. The scale at which the distinction between urban and rural areas will affect the resulting SDG indicator estimates. Statistics Canada, for instance, defines rural areas as “all territory lying outside population centres”, which, in turn, are defined as areas with “a population of at least 1,000 and a population density of 400 persons or more per square kilometre”[20]. The finest level of detail for retaining urban/rural distinction is the dissemination block. This definition and the scale at which the urban/rural disaggregation is possible is not necessarily comparable even with the neighbouring USA, where the definition of the urban area includes “at least 2,000 housing units or a population of at least 5,000”, with the smallest geographical unit being census block[21]. Choosing a common geography – a grid and common scale- 1km spatial resolution makes the GHSL and DEGURBA a solution suitable for international comparability of the SDG indicators reliant on urban/rural data disaggregation.

[16] Dijkstra, L., Hamilton, E., Lall, S., & Wahba, S. (2020). How do we define cities, towns, and rural areas? https://blogs.worldbank.org/sustainablecities/how-do-we-define-cities-towns-and-rural-areas
[17] The Degree of Urbanisation, a new global definition of cities, urban and rural areas.: https://ghsl.jrc.ec.europa.eu/degurba.php
[18] World Pop: https://www.worldpop.org/
[19] Global Human Settlement Layer: https://ghsl.jrc.ec.europa.eu/download.php?ds=pop --
[20] Statistics Canada. Dictionary, Census of Population, 2021. Rural area (RA): https://www12.statcan.gc.ca/census-recensement/2021/ref/dict/az/Definition-eng.cfm?ID=geo042
[21] US Census Bureau. Urban and Rural. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html

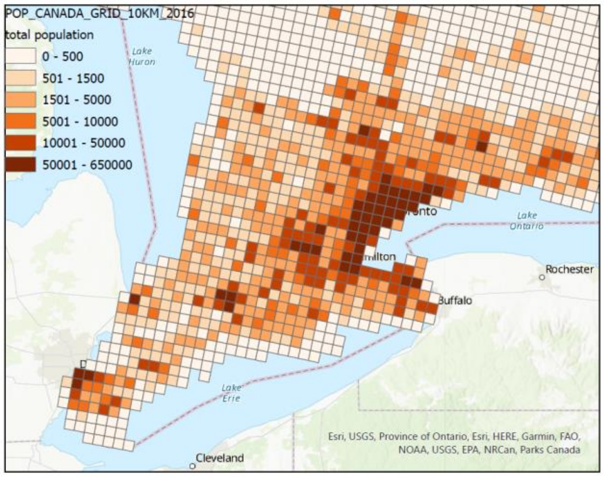


Figure 6: Canadian 10km² gridded population layer

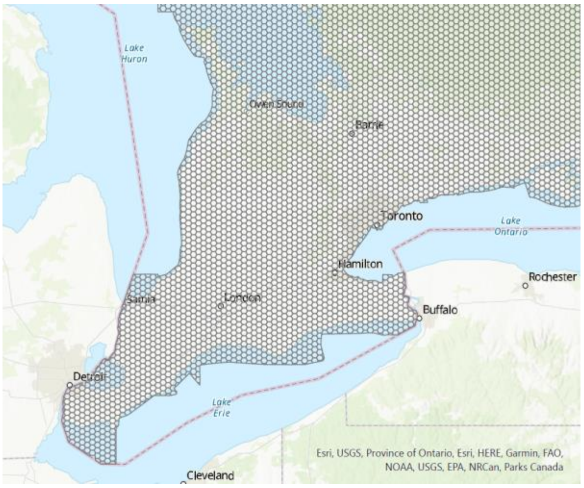


Figure 7: Hexagonal 25km² grid of Canada

One of the issues related to delineating urban and rural territories is the chosen spatial resolution. In the context of gridded geographies, the DEGURBA [22] classification has been led by the European Commission (EC). Canada, for instance, currently offers a square-based 10km² gridded population layer[23] and the Hexagonal Grid of Canada[24] with 25km² cell size, as illustrated in Figure 6 and Figure 7.

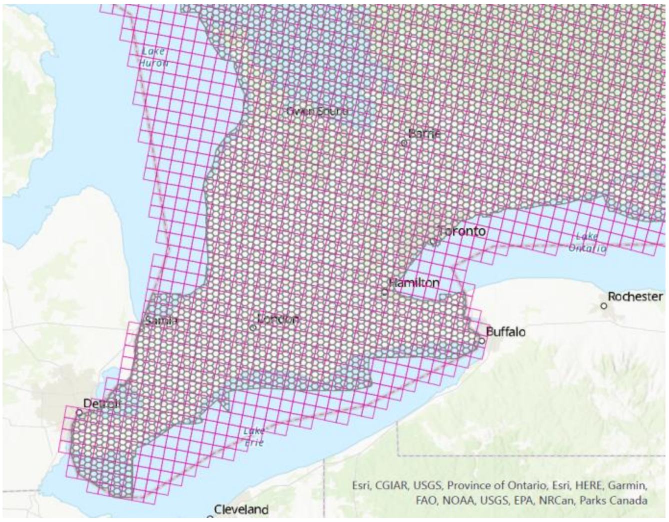


Figure 8: Overlaying differing geographies, intersecting square and hexagon grids

The challenge of geography is illuminated in the instance of Figure 8. In overlaying two different geographies, i.e. two grids of different base shapes limits. In this case, to compare between geographies, it will be necessary to aggregate both geographies to a 'common geography', from which a comparison can then be made.

The same challenge faces users of administrative boundary-based data. The smallest available administrative and census enumeration units are defined differently between countries. For instance, the smallest unit for which statistical information on sex and age is disseminated in the USA is called a **Census**

[22] The Economic Commission's Degree of Urbanisation https://ghsl.jrc.ec.europa.eu/degurba.php
[23] 10-km gridded population layer https://open.canada.ca/data/en/dataset/c6c48391-fd2f-4d8a-93c8-eb74f58a859b
[24] The Hexagonal Grid of Canada: https://open.canada.ca/data/en/dataset/4129e42c-bfa6-40f1-9b2a-19dc04136bb4

Block[25], while in Canada, it is called a **Dissemination Area**[26] – in other countries, Enumeration Areas and Small Area Geographies are also common terms. For Canada, a Dissemination Area includes an average population size (400-700 people), whereas in the USA Census Block specifically states that delineation is not based on population. Using disaggregated data requires acknowledgement that direct comparisons are not always possible even when using the finest geographical detail. This is particularly challenging for remote areas, where administrative or census unit boundaries can vary significantly in shape and coverage. As emphasised by the GSGF, stakeholders need to come together to develop common geographies that would enable consistency and comparability of integrated statistical and geospatial data[27]. Moreover, a significant challenge exists to ensure that geospatially enabled SDG indicators are disseminated in a manner that complies with prevailing national guidelines and international norms on statistical disclosure control.

Regardless of the geographic scale and their nomenclature, it is important to ensure that the resulting data are interoperable in agreed, open standards; GSGF Principle 4 on interoperability and Principle 5 on accessible and useable data.

Gaining greater insights by multidimensional disaggregation

Each indicator can have several dimensions of disaggregation. When indicator data are geocoded, it enables much greater insights to be achieved across dimensions, than it would if they were not. As the overarching framework for global development, the 2030 Agenda aligns with other global development frameworks; the Sendai Framework for Disaster Risk Reduction is one such framework. As the principal report on global disaster risk, The Global Assessment Report[28] 2022 (GAR) examines how "human action is creating greater and more dangerous risk, pushing the planet towards existential and ecosystem limits. Risk reduction needs to be at the core of action to accelerate climate change action and achieve the SDGs".

[25] What are Census Blocks https://www.census.gov/newsroom/blogs/random-samplings/2011/07/what-are-census-blocks.html
[26] Dissemination area https://www150.statcan.gc.ca/n1/pub/92-195-x/2011001/geo/da-ad/def-eng.htm
[27] Developing common geographies allow for consistency and comparability across countries - this is highly recommended as it relevant to alleviation issues surrounding Modifiable Areal Unit Problem (MAUP). MAUP is a statistical bias resulting from modifying areal units shape and geographical scale of data aggregation. The two effects of MAUP, zonal and scale effects of the spatial distribution will affect the observed patterns and analytical results. Whether utilizing cartographic grids or administrative, electoral or statistical unit boundaries, the challenge is to maintain the same desired level of spatial detail across the SDGs and other dimensions of disaggregation. The conclusions about patterns and correlations made at one spatial level of aggregation might not stand at another. Conclusions can be checked for robustness by changing to a different scale level if data availability permits. MAUP is inherent to all spatial data, i.e. GSGF Principle 3 on Gridded and Administrative geographies (GSGF, p.39), and might never be solved. A potential solution to MAUP is the choice of a grid (regular or hexagon) which could be used consistently across every country, yet this requires much further deliberation and review by the relevant groups on the integration of statistical and geospatial information and subsequent endorsement by the UNSC.
For further reading, see: Wong, D. (2008). The modifiable areal unit problem (MAUP). In: Fotheringham, A. S., & Rogerson, P. A. (Eds.) The SAGE handbook of spatial analysis, 105-125 and Manley D. (2014). Scale, Aggregation, and the Modifiable Areal Unit Problem. In: Fischer M., Nijkamp P. (Eds.) Handbook of Regional Science. Springer, Berlin, Heidelberg
[28] UNDRR Global Assessment Report 2022: https://www.undrr.org/gar2022-our-world-risk

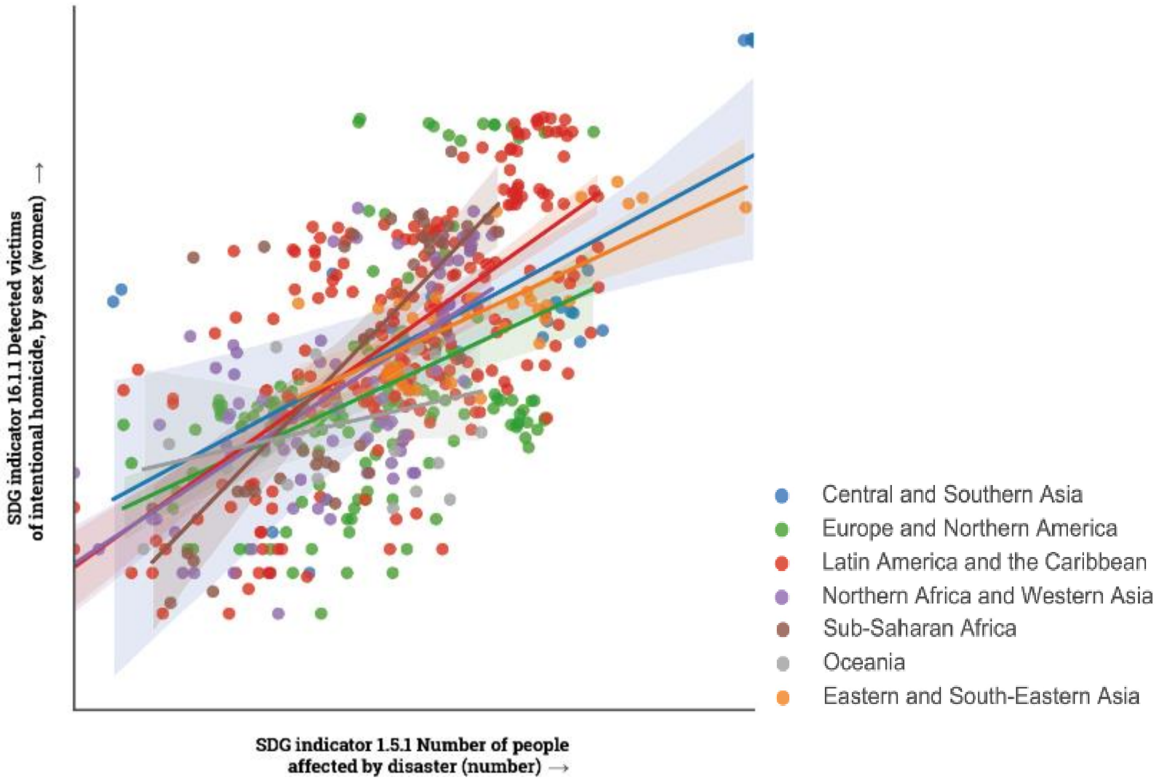


Figure 9: The relationship between the disasters and intentional homicides against women

Anchored by an in-depth consideration of the growing risks of climate change, the GAR 2022 also considers the broader societal impact of disasters by examining associations with other phenomena. For example, Figure 9 highlights that violence against women and girls increases in the aftermath of disasters, uncovering interesting insights that have not been previously seen. At the extreme end of the scale, this takes the form of intentional homicides[29]. By disaggregating by geographic location to the regional level, it is possible to understand which regions this is most prevalent, allowing for an increased focus on the need for coordinated action.

Even with the immense amount of data needed for the SDGs, there are many data gaps. It is a universal problem affecting developed and developing countries alike. Yet, to fill these data gaps, there is ongoing work to harness novel and existing datasets, beyond the 'traditional' data collected by national censuses. Data on sex and age are often maintained in civil registers, which might be available at a relatively small, granular geographical level, potentially at the x- and y-coordinate or address level. Other dimensions for disaggregation, such as migratory status, ethnicity and disability status, might not be available, but some countries collect these data through administrative records[30] or surveys[31].

[29] The repository and method used to calculate these insights is available on https://github.com/unstats/

[30] It is important to note that administrative records tend to omit non-registered or irregular migrants. Indeed, admin data are data generated by government operations and not with the primary purpose of producing official statistics. Hence, the coverage of target populations may be limited and estimates from admin records may be strongly biased if this under coverage is not considered and corrected.

[31] However, dependent on the approach taken, survey data may omit the collection of information on certain population groups by design. In addition, the production of representative and reliable data at finer geographical levels requires a greater sample size, which might not be attainable due to the associated collection costs. Furthermore, multidimensional disaggregation, for example by disability status and by sex and a specific geographic area, poses additional constraints to maintain confidentiality.

Phase 1, Key Action 2 "Identify national data capacity and highlight potential data gaps"; and, Phase 2, Key Action 4 "Identify appropriate data, develop methods, and coordinate development support", are useful phases of the Roadmap to consider when seeking to fill data gaps.

Indirect estimation of disaggregated SDG indicators based on the integration of survey data with additional data sources

Sample surveys can be cost-effective means to collect detailed information at a relatively high frequency over time, have a long history in the production of official statistics, can be used for producing reliable estimates in various domains and can often provide some level of disaggregation. In this context, direct domain estimates of target parameters are statistics based solely on domain-specific sample data. One of the main requirements to achieve reliable disaggregated estimates by direct estimators is a sufficient domain sample size to yield adequate precision, or, in other words, a small estimated variance. When this circumstance is not verified, we are in the presence of so-called small areas, i.e. disaggregation domains for which too little or no sampling observations are available. These "small" disaggregation domains may be identified by geographical or any other partition of the target population. It should be noted that, in practical statistical applications, it is quite rare to have an overall sampling size that is large enough to guarantee enough observations for each dimension of disaggregation.

Therefore, indirect estimation techniques that integrate data from additional sources of information on the population of interest is often necessary. The range of possible approaches to produce indirect estimators is vast, and it goes from the implementation of design-based model-assisted approaches to model-based approaches such as Small Area Estimation (SAE) techniques. SAE model-based methods rely on explicit modelling to link the variable to be estimated with auxiliary variables retrieved by alternative data sources, such as censuses, administrative records, geospatial information systems, and other big data sources. Various empirical applications using SAE for SDG indicators have been developed, many of which are documented in the "SAE4SDG" Toolkit[32] produced by the Secretariat under the guidance of the Inter-Secretariat Working Group on Household Surveys (ISWGHS) and the IAEG-SDGs. There are several examples of how SAE is used for producing geospatially enabled SDG indicators, including:

- The UN Population Fund (UNFPA) on integrating survey and census microdata through SAE is enabling the production of SDG 5.6.1 highlighting information on family planning in Nepal[33];
- The World Bank, in mapping poverty through SAE, offers relevant insights for the production of granular sub-national estimates of SDG indicators under Target 1.1; and,
- The Food and Agricultural Organisation of the United Nations (FAO) produced case studies on SAE applied to the production of subnational estimates of indicators 2.3.1, 2.3.2, and 5.a.1[34], by integrating survey data with geospatial information – in the first case- and with census data.

Despite their increasing popularity, resorting to SAE should not be considered the solution to any data disaggregation problem, and there are various considerations that NSOs should make before engaging in the production of indirect estimates. Model-based approaches have stricter data requirements than direct estimation methods, and the quality of their results highly depends on the quality of the additional information used[35]. In this respect, the huge amount of digital and geospatial information produced by

[32] Small Area Estimation for the SDGs Toolkit https://unstats.un.org/wiki/display/SAE4SDG
[33] UNFPA (2020). Small Area Estimation. Better Data for More Effective Policies and Programmes https://www.unfpa.org/publications/small-area-estimation
[34] Using small area estimation for data disaggregation of SDG indicators https://www.fao.org/3/cb8998en/cb8998en.pdf
[35] More than a pretty picture: using poverty maps to design better policies and interventions https://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-6931-9

many tools and technologies nowadays offers alternative sources of auxiliary variables for SAE production. In addition, being based on models, after implementing SAE approaches, the underlying assumptions need to be carefully validated through adequate diagnostic techniques[36].

Phase 2, Key Actions 1: Identify relevant data and appropriate methodologies to develop SDG indicators; and, 4: Identify appropriate data, develop methods, and coordinate development support are useful phases of the Roadmap to consider when considering the indirect estimation of SDG indicators.

Data disaggregation through administrative records

Administrative data are “information collected primarily by governmental departments and other organisations, usually during service delivery or for registration or record keeping”[37]. Administrative data are often individual or organisation level data, supplemented with large scale geospatial data, such as address location that can be easily geocoded to x- and y-coordinates. In developed and developing countries, information gained from administrative sources can substantially contribute to improvingublic policies in food security, poverty, environmental health and economic development[38] and other domains.

Integrating administrative data with traditional data sources, such as surveys and censuses, can provide invaluable benefits to NSOs and national governments, as this offers the possibility of reducing costs related to data collection and respondents’ burden. Administrative datasets, such as tax files, can assist in estimating poverty-related indicators (SDG indicators 1.1.1, 1.2.1 and 1.2.2) and serve as an auxiliary data source for SAE, as they contain demographic and other socioeconomic characteristics. Other examples of administrative data include migration records supplemented with location information, such as the Canadian Longitudinal Immigration Database[39] and health care system data[40].

Supplementing traditional data sources with administrative data can strengthen existing data sets. Due to the resource limitations commonly encountered in developing countries, this is extremely valuable, because formal survey sampling is limited, and delays may prevent data from reaching policymakers promptly. On the other hand, administrative data can provide better geographic coverage, and are often collected more frequently and at a nominal cost. From a research perspective, administrative data can improve a variety of statistical methodologies, such as sampling frame construction and sample design; the use of administrative records to cover data gaps from surveys and censuses; forecasting; planning; and provision of small area estimates.

Producing and using geographically disaggregated SDG indicators

The production and use of geographically disaggregated SDG data can take several forms, leveraging several types of geospatial information including Earth observations and other forms of innovative data.

Disaggregating by geographic location using Earth observations

As a subset of geospatial information, Earth Observations (EO) is an all-encompassing term for data and information collected about our planet, whether atmospheric, oceanic or terrestrial. These EO includes space-based or remotely-sensed data, as well as ground-based or in situ data. EO data is borderless,

---

[36] The bias of small area estimates needs to be measured to assess estimates reliability. This is generally done by means of the mean square error (MSE), which provides a combined indicator of estimates precision (variance) and accuracy (bias).

[37] FAO (2016), Improving the Methodology for Using Administrative Data in an Agricultural Statistics System https://www.fao.org/3/ca6515en/ca6515en.pdf

[38] In this respect, the United Nations Statistics Division and the Global Partnership for Sustainable Development Data (GPSDD) have jointly convened the Administrative Data Collaborative, which is a multi-stakeholder collaborative of countries and regional international agencies, aiming to strengthen the capacity of countries to use administrative data sources for statistical purposes https://unstats.un.org/capacity-development/admin-data/

[39] Canadian Longitudinal Immigration Database https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5057

[40] Cadarette, SM & Wong, L. (2015). An introduction to health care administrative data. The Canadian journal of hospital pharmacy, 68(3), 232.

impartial and inclusive for all. They are an innovative data source for many SDG indicators and essential for describing environmental aspects of our planet. Designed for planetary-scale coverage, EO's characteristics make it an indispensable direct source of data for several SDG indicators and a supporting source of data for many others. Many EO data sources are freely available today, with consistent and timely global coverage. However, there are significant demands for consummate skills and resources, a gap to be bridged so that EO can fully realise its potential for the SDGs. There are many SDGs indicators where the use of EO can directly contribute to their production and several more where EO has a significant/supporting contribution[41], including the mountain green cover index (SDG 15.4.2) and its change detection at a local level, agricultural monitoring in support of food security (SDG 2.4.1), and ocean eutrophication (SDG 14.1.1).

A key consideration when using data from Earth observations is that the scale of disaggregation is dependent on both the spectral resolution and spatial resolution. Quality spectral resolution data allows for more accurate classification of various classes of built-up areas[42], leaving a 'higher quality' image. Spatial resolution is a measure of the smallest object that can be resolved by the sensor. For example, the WorldPop project produces 100m resolution gridded population estimates[43] disaggregating this information to a smaller level of geography (such as a census block) could decrease the validity of the statistic. Ultimately, EO data is being used to produce a variety of SDGs indicators; offering countries and their agencies a constantly improving source of data. For instance, the European Settlement Map now uses very high-resolution data to map human settlement distribution at 2m resolution. Night-time satellite imagery is another source for population density mapping used to delineate between rural and urban areas[44].

Figure 10 and Figure 11 highlight the global (FAO) and national (Japanese[45]) examples of how EO data can be used to produce an indicator, specifically SDG indicator 15.4.2 – the Mountain Green Cover Index. In light of the many data gaps in how countries have reported against the global indicator framework, the WGGI recommends and encourages the use of global datasets in circumstances where national capacities are not as developed as those at the global level.

[41] The WGGI shortlist 'results of the analysis of the Global Indicator Framework with a "geographic location" lens' details some of these indicators in detail. https://ggim.un.org/meetings/2017-4th_Mtg_IAEG-SDG-NY/documents/WG's_Initial_Shortlist-Table_A_B.pdf

[42] Bozheva, A. M., Petrov, A. N., & Sugumaran, R. (2005). The effect of spatial resolution of remotely sensed data in dasymetric mapping of residential areas. *GIScience & Remote Sensing*, *42*(2), 113-130. DOI: 10.2747/1548-1603.42.2.113

[43] Top-down estimation modelling: Constrained vs Unconstrained - WorldPop: https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained

[44] Spinosa, A. Wider urban zones: use of topology and nighttime satellite images for delimiting urban areas. *Rev Reg Res* 42, 141–159 (2022). https://doi.org/10.1007/s10037-022-00169-y

[45] Japan's National Experience of Producing SDG 15.4.2: https://storymaps.arcgis.com/stories/d93fb8faa2e84f2fad508ff8859abc93

Figure 10: The FAO Land Cover (Re)Classification Dashboard showing Green (Classes such as Forest, Wetland, Agriculture and Grassland) and Orange (i.e. Settlements, Bareland, and Otherland)

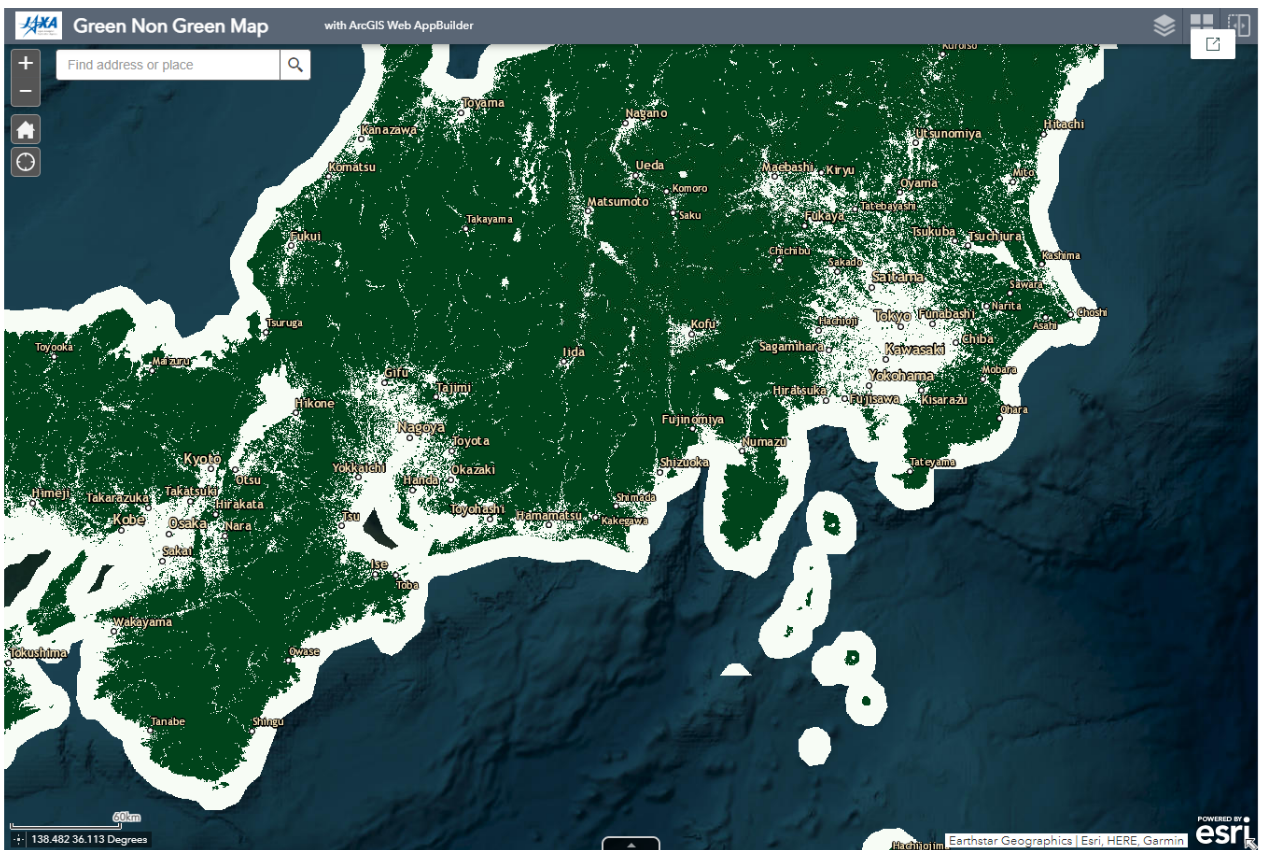


Figure 11: A Green/Non-Green Map of a segment of the Honshu island of Japan, produced by JAXA

SDGs GEOSPATIAL ROADMAP

Further progress can be achieved through leveraging existing resources, such as the 'Global and Complementary Geospatial Data for the SDGs'[46] and 'Specification of land cover datasets for SDG indicator monitoring'[47] reports developed by the WGGI. These reports identify and recommend agreed minimum validation criteria or common parameters that SDG Custodian Agencies could use to validate the effectiveness of EO through its metadata.

While a significant number of indicators can only be produced by geospatial information alone, almost all indicators would benefit from their use in production, measurement, or monitoring. All indicators produced should be geospatially enabled to allow for disaggregation by geographic location at sub-national levels, where possible. In turn, consistent production will allow for progressive measurement and monitoring at these levels of geography.

Integrating several forms of data

SDG 9.1.1 is a key indicator that estimates the proportion of the rural population with adequate access to the transport system. It is defined as the proportion of the rural population living within 2 km of an all-season road. Two kilometres represents a 20–25-minute walk (subject to the topography)[48]. An all-season road is a road that is drivable on all year but may be temporarily unavailable during inclement weather. This indicator relies on three major items of geospatial data: population, road network location and the "all-season" status of those roads. Figure 12 demonstrates how these data sources can be integrated to calculate the proportion of the rural population that lives within 2 km of an all-season road for Kemin, a district in Chuy Region, Kyrgyzstan.

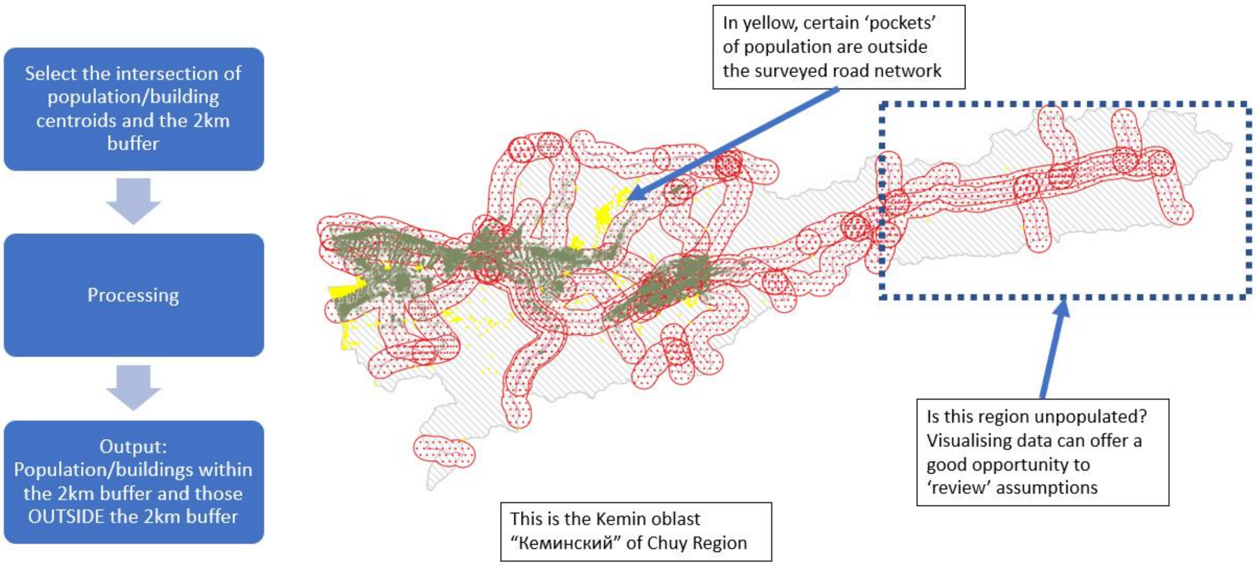


Figure 12: SDG 9.1.1 Proportion of the rural population who live within 2km of an all-season road for Kemin, Kyrgyzstan

[46] Global and Complementary Geospatial Data for the SDGs report: http://ggim.un.org/documents/Report_Global_and_Complementary_Geospatial_Data_for_SDGs.pdf

[47] Specification of land cover datasets for SDG indicator monitoring https://ggim.un.org/documents/Paper_Land_cover_datasets_for_SDGs.pdf

[48] Measuring rural access for SDG 9.1.1 https://onlinelibrary.wiley.com/doi/full/10.1111/tgis.12721#tgis12721-bib-0008

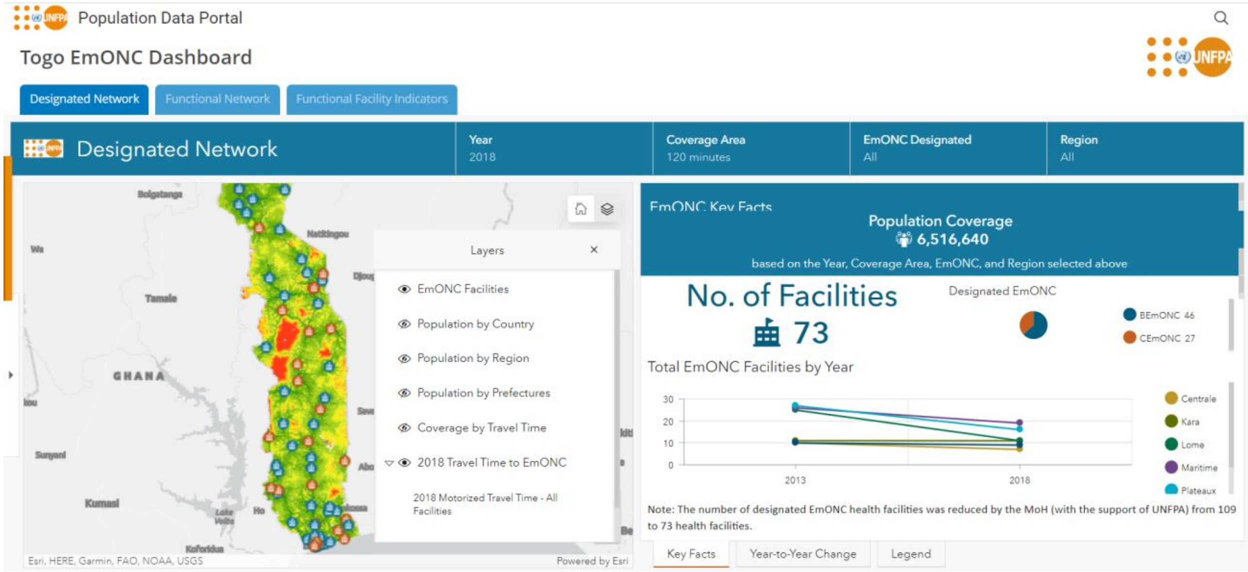


Figure 13: The Togo Emergency Obstetric and New-born Care Dashboard

Another example is highlighted in Figure 13, a dashboard that supports the production and use of indicators relevant to Togo's Emergency Obstetric and New-born Care[49], developed by UNFPA. The dashboard helps identify the access to health services (including sexual and reproductive health services), disaggregated to local levels, enabling policy insights into which villages have more people located out of service areas and other key questions relating to SDG 3 on Good Health and Well-being and SDG 5 on Gender equality.

[49] Togo Emergency Obstetric and Newborn Care dashboard https://pdp.unfpa.org/apps/06c598cd6aed43bca2c8b1c24204b7d6/explore

Summary and future areas of work

For each of the examples discussed in this paper, the SDGs Geospatial Roadmap provides countries with a foundation enabling them to better harness geospatial information and to develop their capacities for the production, measurement, monitoring, reporting and dissemination of geospatially-enabled SDG indicators.

The disaggregation of indicators by geographic location and integrating/analysing them with data disaggregated by income, sex, age, and other statistical dimensions will help countries make better decisions informed by data. Due to the interconnected and interrelated nature of the SDGs, statistical disaggregation alone is not enough, hence why geospatial information is of utter importance for achieving the overarching aims of the SDGs and the 2030 Agenda.

Several areas of advancement exist to enable countries to realise the transformational potential of geospatial information for the SDGs. These areas could include: examining common geographies for both countries and custodian agencies to report and disseminate SDG indicators; updating recommendations on the use of globally available datasets; or developing novel methodologies aimed at filling the data gaps that current exist in the global indicator framework. This is the future that the WGGI now turns to investigate under the guidance and purview of the IAEG-SDGs.

This document is available as an interactive Storymap here:

https://storymaps.arcgis.com/stories/85abafc5794a4a67ad78cb19f99a963f

The SDGs Geospatial Roadmap is available here:

English: https://ggim.un.org/documents/SDGs-Geospatial-Roadmap.pdf
French: https://ggim.un.org/documents/La-feuille-de-route-geospatiale-des-ODDe.pdf
Spanish: https://ggim.un.org/documents/La-hoja-de-ruta-geoespacial-de-los-ODS.pdf

Storymap: https://storymaps.arcgis.com/stories/226e3f606f7940e1b5738e5bcab0cef3

Examples from Burundi, Kyrgystan, and Rwanda

https://storymaps.arcgis.com/stories/482140f9d56647c794469db6da2d07bc

Further information and other resources from the IAEG-SDGs WGGI

https://ggim.un.org/UNGGIM-wg6/

Notes

The designations used and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. The term "country" as used in this publication also refers, as appropriate, to territories or areas. The designations "developed regions" and "developing regions" are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process. The boundaries and names shown and the designations used on the maps on this site do not imply official endorsement or acceptance by the United Nations.

United Nations Statistical Commission

The United Nations Statistical Commission (UNSC), established in 1947, is the highest body of the global statistical system. It brings together the Chief Statisticians from member states from around the world. It is the highest decision-making body for international statistical activities especially the setting of statistical standards, the development of concepts and methods and their implementation at the national and international level. UNSC is a subsidiary body of the UN Economic and Social Council (ECOSOC).

United Nations Committee of Experts on Global Geospatial Information Management

The United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) is the apex intergovernmental body to discuss, enhance and coordinate global geospatial information management activities by involving Member States at the highest level, to work with Governments to make joint decisions and set directions on the use of geospatial information within national and global policy frameworks, and to develop effective strategies to build geospatial capacity in developing countries. UN-GGIM is also a subsidiary body of ECOSOC.

The IAEG-SDGs Working Group on Geospatial Information (WGGI)

In September 2015, Member States adopted the 2030 Agenda for Sustainable Development and tasked the United Nations Statistical Commission to develop the global indicator framework. The overarching principle of the 2030 Agenda for Sustainable Development is that no one should be left behind. At its 46th Session in March 2015, UNSC established the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs), composed of Member States and including regional and international agencies as observers. The IAEG-SDGs was tasked to develop a global indicator framework for the 17 goals and 169 targets of the 2030 Agenda, and to support its implementation. The IAEG-SDGs, in its report to the UNSC (in March 2016) noted that the integration of statistical data and geospatial information will be key for the production of a number of indicators. As a means to address these issues, and to address specific areas relevant to the production of SDGs indicator, the IAEG-SDGs created the Working Group on Geospatial Information (WGGI) at its third meeting in Mexico City. The WGGI is composed of representatives from the IAEG-SDGs, SDG Custodian Agencies, and experts from the wider geospatial and Earth Observations communities, including from UN-GGIM. Together, the WGGI diligently works to provide expertise and advice to the IAEG-SDGs, custodian agencies and the broader statistical community as to how geospatial data, Earth Observations and other new data sources can reliably and consistently contribute to the production and dissemination of the indicators.

[Page Intentionally Left Blank]

The **SDGs** Geospatial

ROADMAP

SUSTAINABLE DEVELOPMENT GOALS