# Working Group on Data Integration

## The integration of geospatial data with statistical data to grasp the territorial dimension in SDG indicators

UN-GGIM: EUROPE
UNITED NATIONS INITIATIVE ON
GLOBAL GEOSPATIAL
INFORMATION MANAGEMENT


**PIER-GIORGIO ZACCHEDDU** (Chair)

**FRANCISCO VALA** (Subgroup Leader)


UNECE Workshop on Data Integration:
Realising the Potential of Statistical and Geospatial Data
Belgrade, 21-23 May 2019

# UN-GGIM: Europe WG on Data Integration

## Work plan 2017-2019 – Tasks defined

Task 1

**Policy Outreach Paper** – Lead by Eurostat

- ✓ Promote the benefits of the integration of statistical and geospatial data aiming at responsible ministries but also relevant stakeholders
- ✓ Make use of recommendations and findings of WG reports already published

Task 2

**Select and analyse SDG indicators** – Lead by NSI PT - INE

- ✓ Meet the Sustainable Development goals 2030
- ✓ Analyse data integration aspects
- ✓ Reflect cross-cutting issues regarding the integration of geospatial and statistical data based on a Global, European and National perspective

At the international level, some of the most important and challenging issues are related to the **Sustainable Development Goals** and the **2030 Agenda**

The global indicator framework defined to monitor the progress towards SDG emphasizes the relevance of geographical disaggregation in order to cope with the motto of ...

*Leaving no one behind*

**Phase 1 | Organise** and define the work scope, main activities and expected inputs and outputs

1

31 May 2017 ✓

Scoping paper was drafted, reviewed and commented by WG members

**Phase 2 | Selection of indicators** for which integration between geospatial and statistical data is relevant for SDG

2

31 Dec. 2017 ✓

Identification of indicators based on WG GI and policy relevance for Europe

Stuctured comments on national practices, including identification of specific national indicators

Metadata analysis, including  tier III

**Phase 3 | Analysis of indicators** based on a structured and cross-level analysis of the selected indicators

3

[ 31 March 2018 ]
August 2018 ✓

Global metadata systematization and gap analysis, EU-SDG indicators gap analysis, national practices and specific national indicators

Link with GEOSTAT 3 – SGF ESS

**Phase 4 | Identification of best practices** to highlight the potential of geospatial approaches for producing SDG indicators

4

[ 16 Nov. 2018 ]
January 2019 ✓

Selection of best-practices linked with GEOSTAT 3

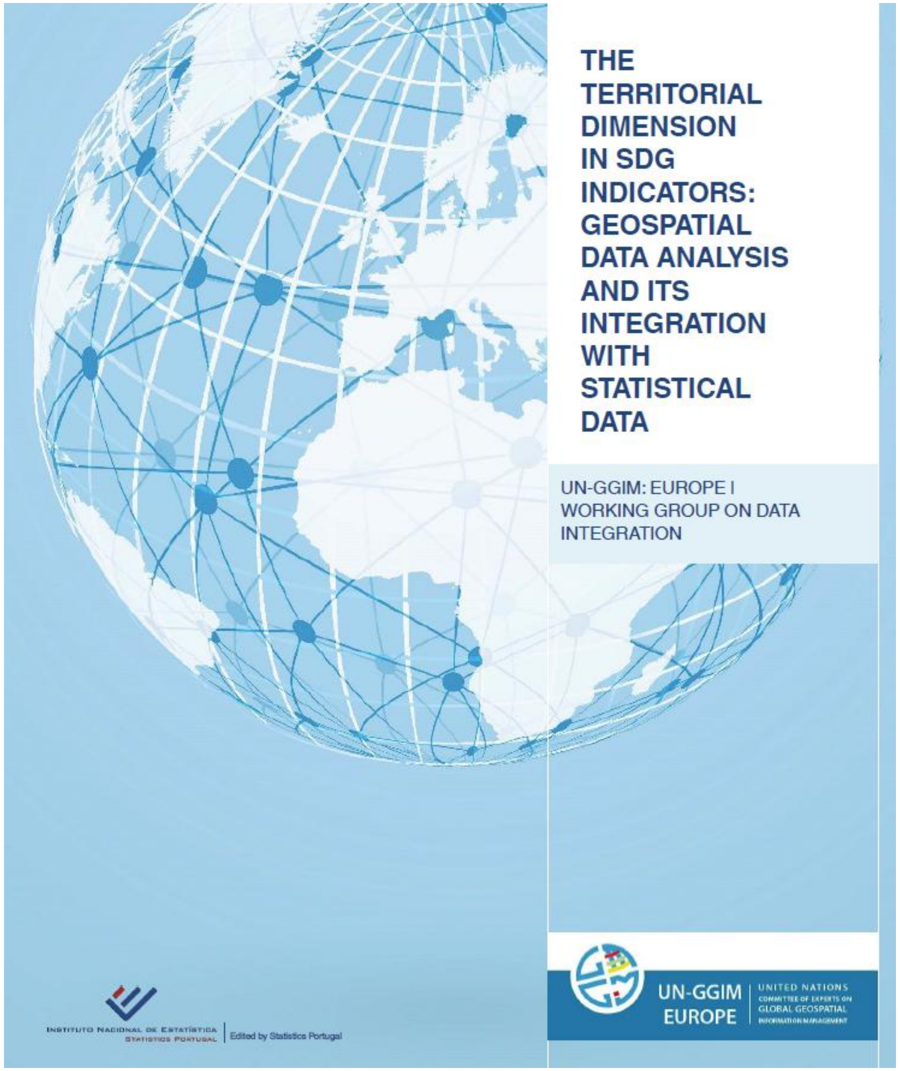
Propose a set of recommendations and final report

# *The territorial dimension in SDG indicators: geospatial data analysis and its integration with statistical data*

## AIM

To address the territorial dimension of the Sustainable Development Goals indicators by focusing on the contribution of geospatial data analysis and its integration with statistical data based on a global, European and national perspective

The work took into consideration, at the global level, the activities of the IAEG-SDG WG GI, and also the background and experiences of European and national initiatives addressing the SDGs from a geospatial perspective



# The territorial dimension in SDG indicators: geospatial data analysis and its integration with statistical data

# Background for selecting the SDG indicators

- ✓ **IAEG-SDG WG GI Short list of indicators** directly or indirectly benefiting from geospatial information
- ✓ **GEO list of indicators** that can directly or indirectly be supported by earth observations
- ✓ **Eurostat analysis on the spatial dimension** in SDG indicators – present in all the 17 SDG, but especially in goals 6, 11 and 15
- ✓ **EU Urban Audit contributions to UN SDG Agenda** by mainly focusing on the scope of goal 11
- ✓ **EU SDG Indicator set to monitor EU policies** in the perspective of 2030 Agenda
- ✓ **National indicators** defined within national SDG monitoring strategies

# Method for selecting the SDG indicators

✓ Structured comments for 26 indicators using a template focusing on:

1. current reporting situation with metadata analysis
2. gap analysis on the global suggested methodology and data integration
3. identification of corresponding EU SDG and specific national indicators

## 4 SDG INDICATORS WERE SELECTED

## Based on:

- ✓ number of contributions and inputs
- ✓ maturity of indicators computation
- ✓ possibility of calculation at European and national levels
- ✓ policy relevance in the European context

### Template for indicator analysis

| |
|---|
| **1. CURRENT REPORTING SITUATION** |
| **Responsibility:** (Identify the agency responsible for the indicator and the situation regarding the ESS and NSS projects (including dissemination) and /or INSPIRE conformance) |
| **Indicator disaggregation:** (List the indicator disaggregation by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts to support the monitoring of the implementation of the SDGs) |
| **Frequency of dissemination:** (Describe the time interval at which information is disseminated over a given time period) |
| **Timeliness:** (Length of time between data availability and the event or phenomenon they describe. Describe the average production time for each release of data) |
| **Data sources:** (List the data sources and themes or variables in use, including conditions of access, timeliness and frequency of dissemination, situation regarding the ESS and NSS projects (including dissemination) and /or INSPIRE conformance) |
| **Geospatial data analysis and integration:** (Describe spatial analysis methods, procedures and computations, including regarding data integration) |
| **Data quality requirements:** (List in general terms the requirements for the sources and themes in use with relevant parameters: Resolution, completeness, logical consistency, positional accuracy, temporal accuracy etc. List if certain international standards are being followed, including classifications/nomenclatures. Data quality should allow computing results to the needed level of resolution and disaggregation). Please take into account the EURO-SDMX Metadata Structure (ESMS) 2.0 |
| **Current use of geospatial data for the indicator:** (Describe the current use of geospatial data, as suggested by the existing metadata – the 'as-is' situation) |
| **2. SUGGESTED METHODOLOGY** |
| **GAP analysis:** (Describe what changes in use of applied methods are needed to go from the suggested/current procedure for monitoring the indicator, to a future procedure which better fulfils the reporting requirements - going from the 'as-is' situation in the present metadata proposal to a 'to-be' situation) |
| **3. SUGGESTED GEOSPATIAL DATA INTEGRATION** |
| **GAP analysis:** (Describe what changes in use of data needed to go from the suggested/current procedure for monitoring the indicator, to a future procedure which better fulfils the reporting requirements - going from the 'as-is' situation in the present metadata proposal to a 'to-be' situation) |
| **List required geospatial data:** (Develop a list from the GAP analysis, which lists the geospatial data sources and themes which are required to support the to-be situation, including INSPIRE conformance) |
| **Data quality requirements:** (List in general terms the requirements for the suggested sources and themes with relevant parameters: Resolution, completeness, logical consistency, positional accuracy, temporal accuracy etc. List if certain international standards should be followed including classifications/nomenclatures. Data quality should allow computing results to the needed level of resolution and disaggregation). Please take into account the EURO-SDMX Metadata Structure (ESMS) 2.0 |
| **Data availability:** (List the data availability for the suggested sources and themes or variables: 1) Geographically: national/ regional/global (as well as comparability across countries), 2) Source: Accessible through services or download, 3) Commercial/ legally: license conditions - are data free or are there restriction on use; 4) Timeliness; 5) Frequency of dissemination) |
| **Data collection:** (Describe how the geospatial data for the indicator can be collected/made available, and issues to overcome – are there many sources to collect from, do they need to be integrated and normalized etc.) |
| **Geospatial data analysis and integration:** (Describe which analysis, procedures and computations are needed to provide the results needed to support the reporting requirements - 'to-be' situation) |

# Selected SDG indicators

## 11.2.1

*tier II indicator*

Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities

**Indicator coordinator**: Austria (NSI)

**Contributors**: Austria (NSI), France (NMCA), Ireland (NSI), Sweden (NSI), Switzerland (NSI)

## 11.3.1

*tier II indicator*

Ratio of land consumption rate to population growth rate

**Indicator coordinator**: Portugal (NSI)

**Contributors**: Finland (NMCA), Ireland (NSI), Italy (e-GEOS), Portugal (NSI and NMCA)

## 11.7.1

*tier III indicator (currently tier II)*

Average share of the built-up area of cities that is open space for public use for all, by sex, age and persons with disabilities

**Indicator coordinator**: Sweden (NSI)

**Contributors**: Ireland (NSI), Sweden (NSI and NMCA), Switzerland (NSI)

## 15.1.1

*tier I indicator*

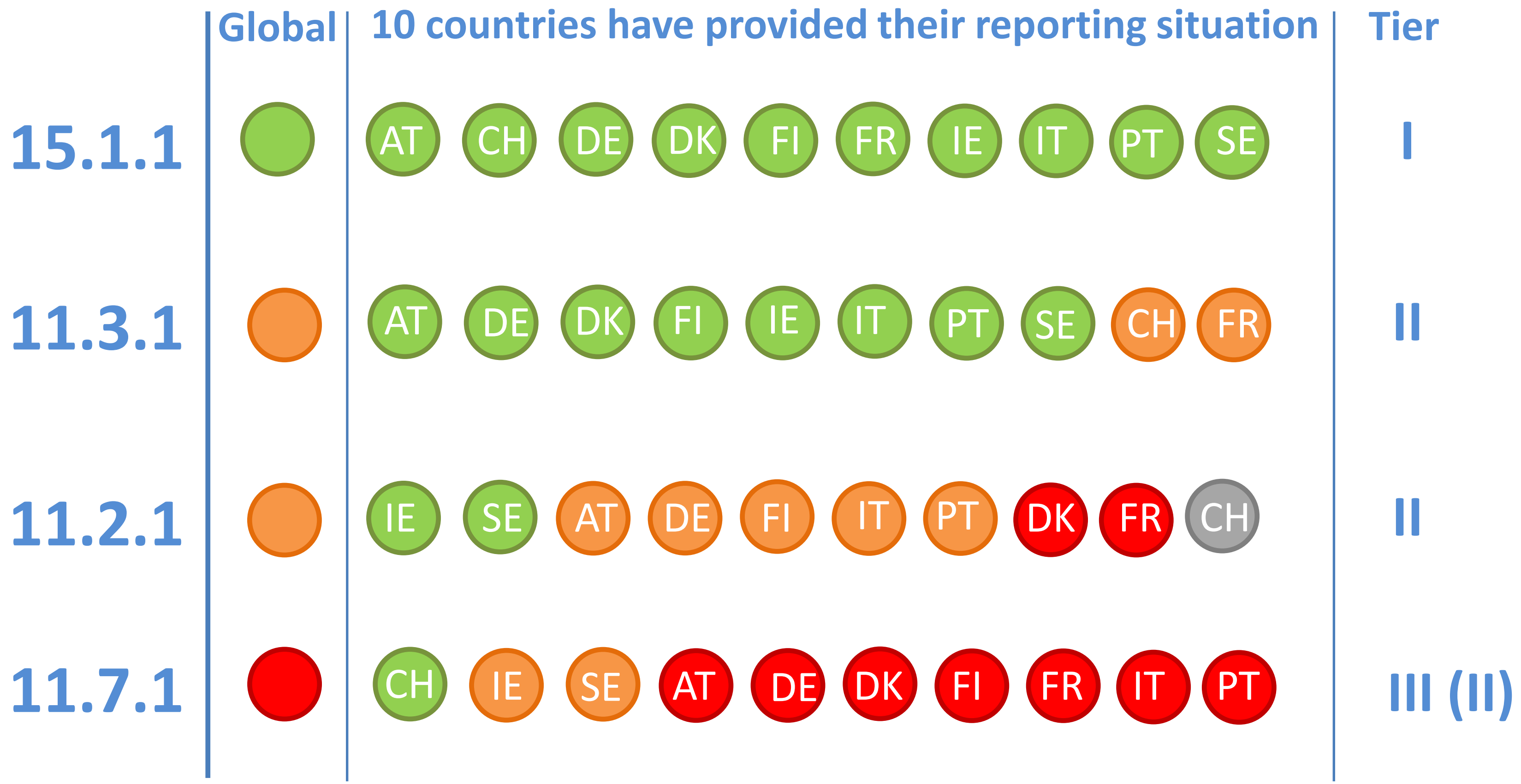
Forest area as a proportion of total land area

**Indicator coordinator**: Italy (e-GEOS)

**Contributors**: Austria (NMCA), Finland (NMCA), France (NMCA), Germany (NMCA), Italy (e-GEOS), Spain (NMCA)

# National report assessment

Inputs from WG members on their current official reporting situation



# Measuring accessibility using spatial modelling and analysis

## 11.2.1 *Proportion of population that has convenient access to public transport*

11 SUSTAINABLE CITIES AND COMMUNITIES

Measuring accessibility has a strong spatial character, since it is intrinsically associated to the physical distance to a place

The 11.2.1 tier II indicator showed that geospatial data and modelling is at the core of this indicator

| | |
|---|---|
| Concepts | The Global Human Settlement Layer (GHSL) definitions could be used at the global level and, at the European level, a common definition of urban areas should be considered based on available common territorial typologies, such as the Degree of Urbanization (DEGURBA).<br><br>The global metadata would benefit from a clear definition of stops and of terms such as comfortable environment and frequent service. |
| Data sources | Point-based population data provides a more flexible and easier way to calculate the indicator as no disaggregation procedure is needed.<br><br>Public transport data on stops can entail different levels of detail on available routes, number of services and timetable information whether the indicator is being computed at the global, European or national level.<br><br>A more general approach on the selection of public transport stops can be followed for global and European comparability and at the national level a more refined and differentiating approach can be applied when detailed data on public transport is available. |
| Computation and algorithm | Network distances, such as the shortest path, can provide more accurate results than Euclidian distances, but it requires quality data on the road network, including pedestrian walks to account for walking access.<br><br>The proposed global computation takes into account the places of residence as reference points, but other reference points could be considered, such as schools, workplaces or markets. |
| Challenges regarding the use of geospatial data | ▪ Availability of comparable point-based data on public transport data, especially including information on routes, frequencies and timetables.<br><br>▪ Availability of road network data for computation of distances, including for walking distances.<br><br>▪ Complying with the proposed global disaggregation by disability. |

## Box 3 I

Measuring access to public transport in European cities (DG REGIO)

DG REGIO calculation using geospatial information, population distribution and public transport stops timetable took into account:

- Walking distances (service areas around each stop) using the street network (accounting for obstacles such as rivers, steep slopes, highways and railroads). It was assumed that people would be willing to walk five minutes (417m) to a bus/tram stop or 10 minutes (833m) to train/metro.

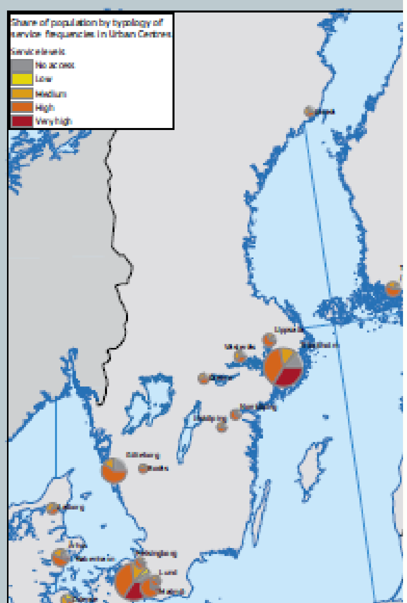
- Time table data on departures on a normal weekday (6:00 – 20:00) by considering five groups based on access and departure frequency [see Figure 8]: i) no access: people cannot easily walk to a public transport stop, it takes more than 5 minutes to reach a bus or tram stop and more than 10 minutes to reach a metro or train station; ii) low access: people can easily walk to a public transport stop with less than four departures an hour;

iii) medi
stop wit
can eas
departu
train sta
v) very
with mo
station v

- Each
by the
availabl
assume
the mos
of the g
case of
average

### Figure 8 I

Access to public transport in urban centres in [



Source: Poelman, H. and Dijkstra, L. (2015). *Measuring access to public transp*

## Box 4 I

Sweden national complementary indicator on Housing in proximity of public transports

As a national complementary indicator, the Swedish National Board of Housing, Building and Planning proposed to measure the number of dwellings and new dwellings developed in proximity of public transport stops. The rationale for this indicator is to follow up the sustainability of urban planning; assuming that housing close to public transports will require less need for cars. The indicator will be updated annually in order to follow the trend of new housing in proximity of public transports.

The steps for calculating the indicator are the following:

**Step 1** public transport data, select only those stops that match the desired frequency of departures.

image showing all public transport stops (yellow dots) and transport stops considered "frequently trafficked" (yellow dots surrounded by a bigger white dot)

**Step 2** create service areas around each public stop: Buffers with varying sizes (400, 1 000 and 2 000 meters).

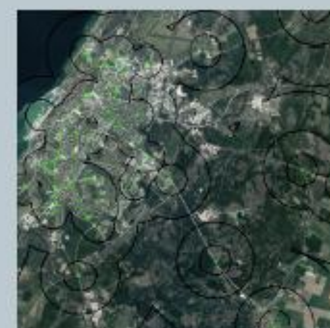


image showing frequently trafficked public transport stops with service areas

**Step 3** conduct a point-in-polygon operation to find out which dwellings are within the range of the service areas (both in total and dwellings in buildings completed during the reference year of interest).

image showing buildings with registered dwellings together with service areas of public transport stops

**Step 4** conduct a point-in-polygon operation also on population data geocoded to the level of address locations to find out how many people live within the range of the service areas.

**Step 5** use the total figure for dwellings and population by county and municipality to calculate a share.

**Step 6** publish the information in the Statistical database from which the National Board of Housing, Building and Planning can retrieve data either by means of searches or by means of machine-readable data served through an API.

The results show that, on national level, 78% of all dwellings are located within 400 m from a "frequently trafficked" public transport stop and 90% of the dwellings were located within 1 000 m.

Among the new dwellings (completed throughout the year of 2015) some 83% were located within 400 m from a frequently trafficked public transport stop.

Source: Lantmäteriet, Trafiklab and Statistics Sweden.

# Deriving new metrics integrating land cover and population data

## 11.3.1 *Ratio of land consumption rate to population growth rate*

11 SUSTAINABLE CITIES AND COMMUNITIES

The geospatial analysis combining land cover and population data provides the possibility of deriving new metrics that are relevant to grasp important dimensions on human settlement planning and management

The 11.3.1 tier II indicator was a very straightforward example on this type of data combination

| | |
|---|---|
| Concepts | Global metadata should be more precise on the definition and use of urban area/city and built-up area as they refer to two different concepts. Urban area/city operationalization would benefit from an international statistical definition to deal with issues of comparability. The EU definitions as presented in the TERCET regulation can be taken as a reference for a worldwide concept following the discussions in UN Statistical Division for a global definition of cities as proposed by the European Commission, OECD and partners.<br><br>The concept of built-up area should be used as a metric to capture artificial land and the expansion of land consumption over time and it is, therefore, a clear distinct concept from the one relating to urban area/city. |
| Data sources | GHSL should be considered as a ready to use product and/or European Space Agency Land Cover products, but special attention should also be given to stimulate European remote sensing derived products initiatives worldwide, such as Copernicus Imperviousness High Resolution Layer and CORINE Land Cover.<br><br>National data sources can provide more detailed and high quality geospatial data for greater territorial disaggregation of the indicator. |
| Computation and algorithm | The algorithm on Land Use Efficiency (LUE) as proposed by the Joint Research Centre (JRC) should be considered as it deals with issues of zero population growth and periodicity of the available information by recommending a normalization of the results for a 10 year reference.<br><br>JRC developed a toolbox in open code format which provides a good way to increase indicator harmonization and comparability. |
| Challenges regarding the use of geospatial data | ▪ Accessibility, periodicity, timeliness and methodological stability of the data sources in order to measure progress over time.<br><br>▪ Identification of precise land cover components or categories to derive built-up areas based on different geospatial data products, including at the national level.<br><br>▪ Spatial resolution of input data and its impact on the quality of the territorial disaggregation of statistical outputs. |

## Box 6 I

ESS Territorial typologies relevant for urban delimitation - the grid-based typology, DEGURBA and FUA - and the Global DEGURBA approach

At the European level, the Regulation (EU) 2017/2391 of the European Parliament and of the Council of 12 December 2017, defines the territorial typologies (TERCET) to be used and published by the Commission (Eurostat), including typologies composed of territorial units at the levels of NUTS, LAU and grid cells. The typologies to identify urban at grid and LAU levels are particularly relevant for Goal 11 monitoring but also to structure other indicators according to urban and rural segmentation:

**The grid-based typology (1 km²)** defines [Figure 10]:

- 'Urban centres' or 'High density clusters': Contiguous (without diagonals) 1 km² grid cells within the 'urban cluster' with a density of at least 1 500 inhabitants/km² and a minimum of 50 000 inhabitants in the cluster (gaps in the cluster are filled).
- 'Urban clusters' or 'Moderate density clusters': Contiguous (including diagonals) 1 km² grid cells with a density of at least 300 inhabitants per km², and a minimum of 5 000 inhabitants in the cluster.
- 'Rural grid cells' or 'Low density grid cells': 1 km² grid cells with density below 300 inhabitants/km² and other cells outside urban clusters.

**The Degree of urbanization**
level [Figure 11]:

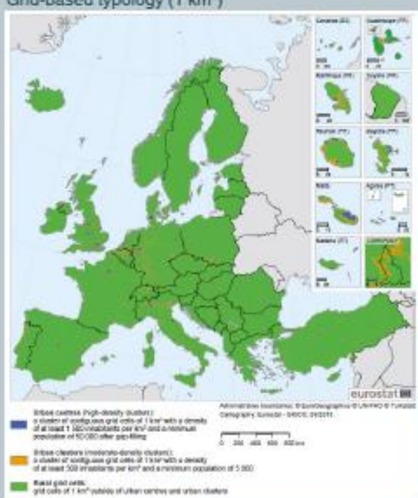
- 'Cities' or 'Densely populated
where at least 50% of the pop
- 'Towns and suburbs' or 'Interm
territorial units where less tha
'rural grid cells' and less than
- 'Urban areas': 'Cities' and 'T
- 'Rural areas' or 'Thinly popul
units where at least 50% of the

**The Functional urban areas**

- 'Cities' plus their 'Commutin
territorial units from which a
population commute to the
included and exclaves are ex

**Figure 10 I**

Grid-based typology (1 km²)



Source: Eurostat, JRC, EFGS, REGIO-GIS (Eurostat - Statistics Explained – Territorial Typologies)

**Figure 11 I**

Degree of urbanization for LA

Source: Eurostat, JRC and
Directorate-General for Regio
Statistics Explained – Territorial T

These typologies have been discussed in the UN Expert Group Meeting on Statistical Methodol
Rural Areas and in particular a **refined DEGURBA definition** intended to be applicable worldwide
Division and SDG monitoring. This Global DEGURBA details the three classes into six allowir
'Cities', 'Towns', 'Suburbs', 'Villages', 'Dispersed rural areas' and 'Mostly uninhabited areas' at bot

## Box 7 I

Calculation of the indicator based on the Land Use and Land Cover Map (COS): example for Portugal

In December 2018, Statistics Portugal published a new set of Land Use and Land Cover Statistics based on the Land Use and Land Cover Map (COS) produced by the Directorate-General for Territory (the Portuguese NMCA) that includes the calculation at municipality level of the global SDG 11.3.1 indicator based on the Land Use Efficiency (LUE) formula as proposed by the JRC [Figure 12] and of the corresponding EU SDG indicator defined by Eurostat to monitor Goal 11 at the EU level [Figure 13].

For both indicators, the megaclass 1 of COS nomenclature "artificial land", excluding the class 133 corresponding to "areas under construction" was used and the area of artificial land occupied in each municipality was extracted based on a common territorial delimitation of municipalities as defined by the Official Administrative Map of Portugal. The LUE indicator was calculated using data from COS 2010 and COS 2015 and data from annual resident population estimates for the reference years of 2010 and 2015. The use of the JRC formula allowed to deal with those situations with zero growth and, thus, provided consistent results for the different municipalities.

The result for Portugal's mainland for the period 2010-2015 was -10% and only 15 municipalities scored positive LUE values, i.e., an increment of population faster than the increase of artificial land [Figure 12]. The indicator on artificial land *per capita* was calculated using data from COS 2015 and population data was also derived from the annual resident population estimates for the reference year of 2015. The results show that 60 municipalities, mainly located in the metropolitan areas of Lisboa and Porto, recorded less area of artificial land *per capita* than the one registered for Mainland Portugal [Figure 13].

The dissemination of this new set of Land Use and Land Cover Statistics comprised several challenges on complying with the standard statistical methodological document made available describing all the methodological procedures, concepts and classifications associated with a statistical operation, as this constitutes the first statistical operation disseminated by Statistics Portugal based on a geospatial data source and on its integration with statistical data.

**Figure 12 I**

LUE by municipality 2010-2015

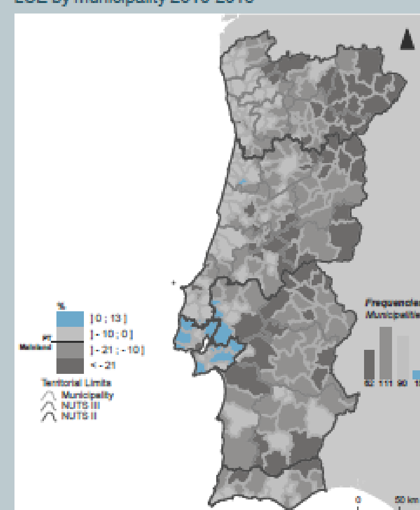
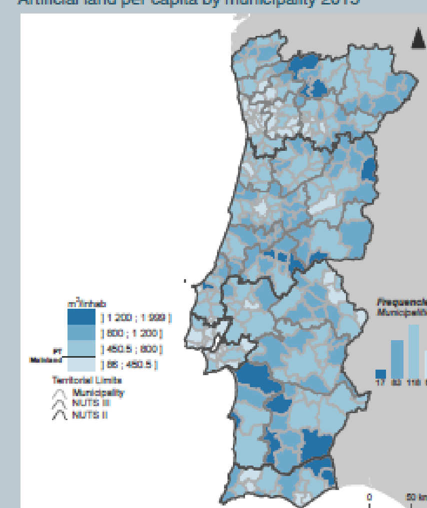


**Figure 13 I**

Artificial land per capita by municipality 2015



Source: Statistics Portugal/Directorate-General for Territorial Development, Land use and Land Cover Statistics.

# Addressing challenging indicators based on land use and cadastral data

## 11.7.1 *Average share of built-up area of cities that is open space for public use*

The SDG monitoring framework includes a number of more challenging indicators due to the lack of data availability and existing established methodology

The 11.7.1 former tier III (now tier II) indicator showed that land use and cadastral data, obtained using different geospatial based products, can play a significant contribution for its operationalization

| | |
|---|---|
| Concepts | The concepts of urban area/city and built-up area should be defined in line with other indicators of Goal 11. In this vein, the concept of built-up area should be measured as artificial land and available common and harmonised territorial typologies (e.g. EU TERCET) should be used to capture the urban dimension.<br><br>The global metadata should benefit from a more detailed description on the different dimensions that are supposed to be captured by the concept of open space for public use, especially, in order to achieve comparability across countries at the global level. |
| Data sources | GHSL provides global coverage to measure built-up areas and other remote sensing derived products, such as Copernicus Imperviousness High Resolution Layers and CORINE Land Cover provide ready to use comparable data at the European level.<br><br>National data sources might have more detailed information and are able to better address the conceptual definitions of the indicator, namely the differentiation between public and private open space. |
| Computation and algorithm | Cadastral data can provide better data coverage for a more detailed territorial disaggregation and a more consistent and stable classification for measurements over time.<br><br>The global metadata should include a reference on the substitutability of field data/ non-geospatial inventory data for geospatial information as a valuable approach to increase comparability across countries. |
| Challenges regarding the use of geospatial data | ▪ Definition and availability of data sources in order to measure the concept of public open space ensuring international data comparability.<br><br>▪ Availability of data sources with ownership (public vs. private).<br><br>▪ Combination of different geospatial data sources in order to grasp the necessary definitions to calculate the indicator: urban delimitation, built-up areas and open space for public use. |

## Box 10 I

Measuring built-up areas that are open space for public use based on the Irish PRIME2 database

For the national context of Ireland, the identification of built-up areas that are an open space for public use can be derived from the PRIME2 geospatial database maintained by the Irish NMCA - Ordnance Survey Ireland (OSi). The central premise behind PRIME2 is to have topologically consistent polygons that cover the surface of Ireland. These polygons are grouped into five broad categories: way, water, vegetation, artificial and exposed (non-vegetative ground such as sand and mud).

For the built-up areas, artificial and way objects can be considered to be in scope. Artificial objects represent man-made ground cover such as concrete, tarmacadam, gravel, sloping masonry, rail bed, among others including gardens. Way objects in the PRIME2 database also represent all drivable and walkable roa to sidewalks and lane w classified as 'artificial' an

PRIME2 contains over 5 (building), line (road) and p form and function object physical form (e.g. buil hospital, church etc.). Th types recorded in PRIME selected for open public beach, and cemetery. Fig as open public space in th

**Figure 16 I**

Built-up areas in Dublin city centre

**Figure 17 I**

Objects classified as ope

The following categories of objects can be selected in as public space: car park, roundabout, sidewalk, traff allocated to streets in the centre of Dublin.

**Figure 18 I**

Objects classified as la

## Box 11 I

Sweden national complementary indicator on Access to urban public green areas

For the context of the national SDG indicator framework and monitoring, Statistics Sweden proposed a complementary national indicator measuring people's access to public green areas in urban areas.

This indicator is based on existing official statistics. To calculate this statistics, a method has been developed, based on a combined use of Earth Observation (EO) data and geocoded register data and other geospatial information (geocoded population data, cadastral parcels, buildings, road networks etc).

The data compiled for the green area statistics can also be used as a foundation to estimate the global indicator on the totality of urban public space. Some of the categories of public open space are already calculated [Figure 19] and some can be retrieved on the basis of the data compiled.

In brief, green spaces are mapped using EO data. Public access is determined using cadastral parcels in combination with ownership data. Large enough adjacent green spaces (>= 0.5 ha) accessible for the public are qualified for 'green areas'.

Proximity to population is calculated using Euclidian distance between population data linked to physical address location and green areas. The share of urban population with access to green areas within 200 meters from the place of permanent residence is calculated.

**Figure 19 I**

Public green areas in Malmö by size


## Box 8 I

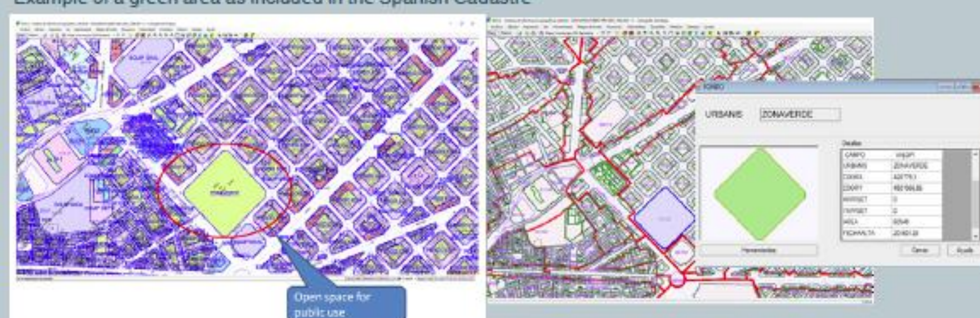
Identifying open public space with cadastral information: the case of Spain

In the case of Spain, the cadastre includes in its database the urban qualification of each real estate, including green areas. The Spanish cadastre provides information on the status, area (m²) and coordinates [Figure 14], and information can be analysed at different territorial levels (e.g. by municipality, building blocks).

In contrast to fieldwork, the use of such data will result in less precise classification of types of public open space, but on the other hand it will have a better coverage and possibly also a more consistent and objective classification.

**Figure 14 I**

Example of a green area as included in the Spanish Cadastre




Source: Dirección General del Catastro (http://www.catastro.meh.es).

# Increasing the scope of indicators disaggregation with earth observation data

## 15.1.1 *Forest area as a proportion of total land area*

Earth observation data can provide an objective and consistent view of the earth for different periods in time, at different scales and ensuring a coherent basis for comparability between different countries

The 15.1.1 tier I indicator was a very good example of how earth observation data can increase the scope of territorial disaggregation

| | |
|---|---|
| Concepts | Common technical specifications for targets to be managed by the indicator, in particular forest and inland waters should be better defined.<br><br>A shared forest definition according to FAO standards that could be captured by geospatial layers would provide a way to increase data comparability at the global level. |
| Data sources | Global (ESA Land cover) and European (e.g. Copernicus High Resolution Layers, CORINE Land Cover) geospatial layers, based on remote sensing techniques, could be used for this indicator allowing a more detailed territorial segmentation of the indicator.<br><br>At the European level, geospatial data layers are generated on a regular basis and the situation is the same for satellite data, since EU Sentinel satellites are planned to remain in orbit for several years. |
| Computation and algorithm | Vector or raster format geospatial layers should be used in order to allow proper data aggregation/disaggregation.<br><br>Different levels of technical algorithm specifications can be considered for a single country depending on its objectives, but common, harmonized specifications and its availability in open formats provide a way to increase data comparability across countries. |
| Challenges regarding the use of geospatial data | ▪ Stability of the geospatial data sources and of its methodology regarding processing workflows with a high degree of automation.<br><br>▪ Clear definition of who is responsible for the generation of the output product quality at the global level.<br><br>▪ Capacity building in using earth observation data and derived products to produce statistical indicators. |

**Box 12 I**

Calculation of the indicator based on ESA Land cover data: the case of Italy

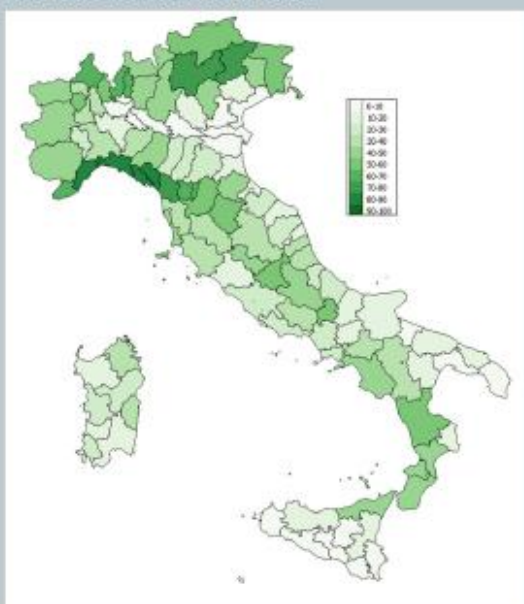
A simulation of the indicator computation has been carried out by e-GEOS using ESA Land Cover and is based on the following workflow:

- 2012 ESA Land Cover data and Italy's NUTS 1, NUTS 2, NUTS 3 and commune administrative borders provided by ISTAT (Italian NSI)
- extraction of the Italy area subset of the ESA Land Cover
- computation of the forest area over Italy by selecting the following classes:

  ESA Land Cover classes: 40 (Mosaic natural vegetation (tree, shrub, herbaceous cover)/cropland), 50 (Tree cover, broadleaved, evergreen, closed to open), 60 (Tree cover, broadleaved, deciduous, closed to open), 70 (Tree cover, needle-leaved, evergreen, closed to open), 80 (Tree cover, needle-leaved, deciduous, closed to open),

**Figure 20 I**

Forest area as a proportion of total land area based on 2012 ESA Land Cover data, NUTS 3 level



90 (T
shrul
cove

- compu
  by cor
- compu
  admin
- calcula
  x (fore
  area)

At nationa
based on
compared
FRA 2015
obtained f
NUTS 3 [F

**Figure 21**

Forest are
ESA Land

**Box 13 I**

Calculation of the indicator based on Copernicus High Resolution Layers (HRL): the case of Italy

A simulation of the indicator computation has been carried out by e-GEOS using Copernicus HRL and based on the following workflow:

- 2012 Copernicus HRL Forest Type geospatial layer and Copernicus HRL Water Bodies geospatial layer
- administrative borders geospatial layer provided by ISTAT
- computation of the total country area by administrative borders geospatial layer
- computation of the total land area by subtracting water bodies geospatial layer from the total country area
- computation of indicator based on the ratio HRL Forest Type / Total land area

At national level, the indicator for Italy computed for 2012 based on Copernicus HRL was 36.9%, also a close values to those reported in FRA2010 (36.7%) and FRA2015 (37.7%).The following figures show the results obtained by using only Copernicus geospatial layers for the case of Italy considering a disaggregation at NUTS 3 [Figure 22] and commune level [Figure 23].

**Figure 22 I**

Forest area as a proportion of total land area based on 2012 Copernicus HRL, NUTS 3 level

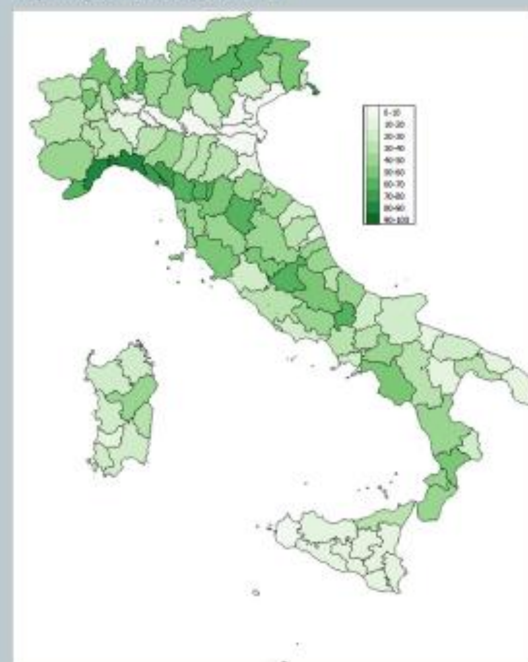
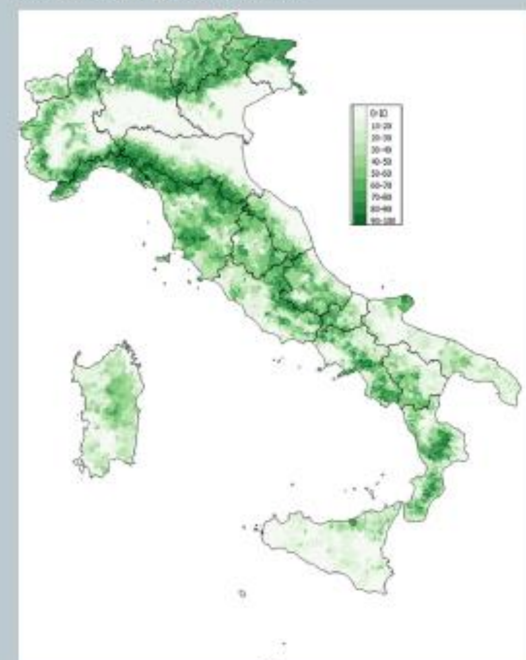


**Figure 23 I**

Forest area as a proportion of total land area based on 2012 Copernicus HRL, Commune level



Source: Copernicus High Resolution Layers.

# 12 RECOMMENDATIONS

*towards a more effective geospatial data integration to address SDG statistical indicators*

1| Harmonize relevant geospatial data themes

2| Implement Cadastral and Land Cover data as key national authoritative data

3| Use geospatial layers generated from Earth Observation data

4| Create capacity building initiatives for National Statistical Institutes to take full advantage of Earth Observation based data

---

5| Define and implement National Spatial Data Infrastructures having in mind the requirements for statistical production

6| Implement consistent and stable sub-national spatial units

7| Develop and use population grids and other grid-based statistics

8| Adopt harmonised and comparable concepts, definitions and classifications and build consensus among Geospatial Agencies and National Statistical Institutes

---

9| Ensure availability and accessibility of processing workflows, including open formats of programming codes

10| Develop initiatives that promote availability, accessibility and usability of geospatial data

11| Increase the collaboration with researchers and data providers

12| Increase cooperation between National Statistical Institutes and Geospatial Agencies

# Working Group on Data Integration

## The integration of geospatial data with statistical data to grasp the territorial dimension in SDG indicators

UN-GGIM: EUROPE

UNITED NATIONS INITIATIVE ON
GLOBAL GEOSPATIAL
INFORMATION MANAGEMENT


**PIER-GIORGIO ZACCHEDDU** (Chair)

**FRANCISCO VALA** (Subgroup Leader)


UNECE Workshop on Data Integration:
Realising the Potential of Statistical and Geospatial Data
Belgrade, 21-23 May 2019