UN-GGIM: Europe | Work Group on Data Integration | subgroup 2

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

---

INDICATOR 11.2.1 | Proportion of population that has convenient access to public transport by sex, age and persons with disabilities

[Tier II Indicator]

A | *Global metadata*

B | *National practices: Austria (NSI), France (NMCA), Ireland (NSI) Sweden (NSI) and Switzerland (NSI)*

C | *EU SDG indicator: Distribution of population by level of difficulty in accessing public transport (Eurostat)*

A. GLOBAL METADATA | UN SDG Metadata

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)

United Nations Human Settlements Programme (UN-Habitat)

At the Global level, all this data will be assembled and compiled for international consumption and comparison by the UN-Habitat and other partners. UN-Habitat and partners will explore several capacity building options to ensure that uniform standards for generation, reporting and analysing data for this indicator are applied by all countries and regions.

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)

The indicator should be disaggregated by age, sex and should also evaluate accessibility for persons with disabilities.

Based on available demographic data further potential disaggregation could be possible:

- Disaggregation by location (intra-urban)
- Disaggregation by income group
- Disaggregation by sex (female-headed household)
- Disaggregation by race (head of household)
- Disaggregation by ethnicity (head of household)
- Disaggregation by migratory status (head of household)
- Disaggregation by age (households, inhabitant)
- Disaggregation by mode of public transport

Quantifiable Derivatives:

- Proportion of urban area that has convenient access to public transport
- Proportion of population/urban area that has convenient access to public transport stop with universal accessibility for people with disabilities
- Proportion of population/urban area that has frequent access to public transport during peak hours
- Proportion of population/urban area that has frequent access to public transport during off-peak hours
- Proportion of urban central/suburban area that has convenient access to public transport

Frequency of dissemination: (Describe the time interval at which information is disseminated over a given time period)

The monitoring of the indicator can be repeated at an annual interval, allowing several reporting points until the year 2030. Monitoring at annual intervals will allow to determining whether the proportion of the population with convenient public transport is increasing significantly over time, as well as monitor what is the share of the global urban population living in cities where the convenient access to public transport is below the acceptable minimum. The proposed indicator has the potential to measure improvement within short term intervals.

Moreover, the disaggregated monitoring for this indicator will provide increasing attention on the access to transport especially among the vulnerable populations such as women, children, persons with disabilities and older persons.

Data release: Two year to five year windows will be applied, based on availability of new data.

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 situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

The actual and recommended data sources for this indicator are the following:

- Data on location of public transport stops in city: city administration or service providers, GIS data
- Dwelling units within 500m of public transport stops: Census, GIS data
- Number of residents per dwellings unit: Census/household survey
- Household surveys that collect information on the proportion of households that declare they have access to public means of transport within 0.5 km. These surveys can also collect information about the quality of the service.
- Urban typology: Due to its spatial nature, the use of the urban agglomeration is a precondition for the measurement and comparability of this indicator.

Geospatial data analysis and integration: (Describe spatial analysis methods, procedures and computations, including regarding data integration)

Identification of service areas: typically achieved using the buffering operation (using GIS) by constructing lines of equal proximity around each public transport stop. A common practice in public transport planning is to assume that people are served by public transport if they are within 0.5km (or 500m) of a public transport stop. Some studies measure the distance based on air, or Euclidean, distance, while others use network distance (that is, the walk distance computed using the street network) to reach a public transport feature.

Identification of the population served: once a service buffer is constructed, the next step is to overlay the buffer onto other polygons, such as Census tracts, for which socio-demographic data

(such as population figures, disabled persons, type of residence area, etc.) is available. These polygons are referred to as the analysis zones. Typically, a service buffer (denoted as i) intersects, either fully or partially, with more than one analysis zone j (j=1…..J). The population served by the public transport service in buffer i, Pi, is thus equal to the sum of the population in each of the intersecting areas, Pij. Pij is estimated based on the amount of interaction between service buffer i and analysis zone j. In estimating Pij it is assumed that the population is uniformly distributed within the analysis zones.

Integrating local temporal availability: the methodology described above covers public transport service solely based on spatial access to stops and does not address the temporal dimension associated with the availability of public transport. We note that temporal aspect of public transport availability is important because a service within walking distance is not necessarily considered as available if waiting times go beyond a certain threshold level that is required. This wait time for public transport is related to the frequency of the service as well as the threshold for tolerable waits for potential public transport users. We will leave out completely the temporal measurement for global comparison, but countries that can additionally capture this component are encouraged to collect and report this information as part of the disaggregation.

Finally, the population with access to public transport out of the entire city population will be computed as;

Percentage with access to Public transport = 100x (population with convenient access to Public transport) / (City Population)

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.

Population:

a. Resolution: Census tracts
b. Completeness: 100%
c. Temporal accuracy: Census or yearly
d. Socio-demographic data (sex and age, information on number of persons with disability including type of disability, …)

Public transport stops:

a. resolution: location by x, y coordinates
b. completeness: all stops within urban agglomerations
c. quality: stops should include information on
    i. accessibility to all special-needs customers, including those who are physically, visually, and/or hearing-impaired, as well as those with temporary disabilities, the elderly, children and other people in vulnerable situations.
    ii. frequency of service during peak travel times

iii. safety and comfort of station environment

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)

Geospatial data is in the heart and nature of this indicator. The location of the stations, the location of the population and the distance of 500m are all included in the description.

However the current reporting situation does not necessarily include data based on direct geospatial aspects, as the indicators reported can also have sample surveys as source.

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)

Indicator should reflect the urban dimension in the name as its methodological scope begins with the delimitation of urban agglomeration (e.g., Proportion of urban population with convenient access to public transport).

Global metadata focuses on population data availability based on Census tracts (enumeration areas). In the European context, an increasing number of countries have or are implementing point-based geocoding infrastructures – geocoded population data provides a more flexible and accurate way for calculation.

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)

This indicator is categorized under Tier II, meaning the indicator is conceptually clear and an established methodology exists but data on many countries is not yet available.

Other reference points could be considered rather than just residential points, such as schools, work place, markets. This indicator would then measure the proportion of schools, work places and markets with convenient access (0.5 km) to public transport. Schools, work places and markets represent a different object type, not to be mixed with population. In principle, they could be measured too, but rather as "proportion of schools, work places and markets" with convenient access to public transport. This is a slightly different story.

The definition of stops could be a challenge as, typically, large stops have several entrances apart from each other and include all entrances as access points to public transport.

Information might not be easily available regarding population with disabilities, but another perspective would be to have information on stops accessible for people with disabilities, as anyone can be temporarily injured, push wheel chairs or baby strollers (e.g., proportion of population with access to public transport stops accessible for people with disabilities).

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)

B. NATIONAL PRACTICE | Austria (NSI)

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)

Statistics Austria.

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)

The indicator was only reported once based on the results of the EU-SILC (EU Survey on Income and Living Conditions) ad hoc module 2012 on housing conditions ilc_hcmp06, which was a sample household survey with a sample size of 6.000 households in Austria. The question suitable to answer SDG 11.2.1 was distribution of population by level of difficulty in accessing public transport.

There are no plans to repeat this questionnaire, so the frequency of dissemination remains "1".

Timeliness: (Length of time between data availability and the event or phenomenon they describe. Describe the average production time for each release of data)

The numbers were published about 6 months later describing the level of difficulties the households perceive to access public means of transport.

Data sources: (List the data sources and themes or variables in use, including conditions of access situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

EU-SILC sample survey, so the heads of households answered how they perceive the situation without giving absolute distances or obstacles.

Geospatial data analysis and integration: (Describe spatial analysis methods, procedures and computations, including regarding data integration)

There was no geospatial data used to answer this question.

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)

As the indicator was only published once and from a sample survey there is lots of room for improvement, both for the frequency and the computation including geospatial data. Whether the effort will be worthwhile and lead to a more accurate result remains open.

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)

The register based data sources available to Statistics Austria are similar to the situation described by Statistics Sweden. So point based population data is available for the production of statistics, hence can be used for spatial analysis. The data source for the population information is the population register, including characteristics such as e.g. age and sex and place of birth. However there is no accessible data source to find out the degree and type of disability of the residents.

As suggested in the metadata methodology the access to public transport is considered convenient when an officially recognized stop is accessible within a distance of 0.5 km from a reference point. The definition of a stop however can be challenging, considering that big stops have several entrances a few hundred meters apart from each other. The walking distance within the stop area from entrance to the platforms from which to enter the train can be more than 500m (as is the case in Vienna's main station). So in the analysis it might be a good idea to use all the entrances as access points to public means of traffic.

The metadata for indicator 11.2.1 describes "persons with disabilities" as all special-needs customers, including those who are physically, visually, and/or hearing-impaired, as well as those with temporary disabilities, the elderly, children and other people in vulnerable situations.

With the initiative of the European Commission from 2010 to create a barrier-free Europe (http://europa.eu/rapid/press-release_IP-10-1505_en.htm) a lot has been done in that respect.

Vienna has excellent open government data and some other Austrian cities also. So one can find, for example, the information whether the stations are accessible by elevators or escalators and where to find tactile paving as open data.

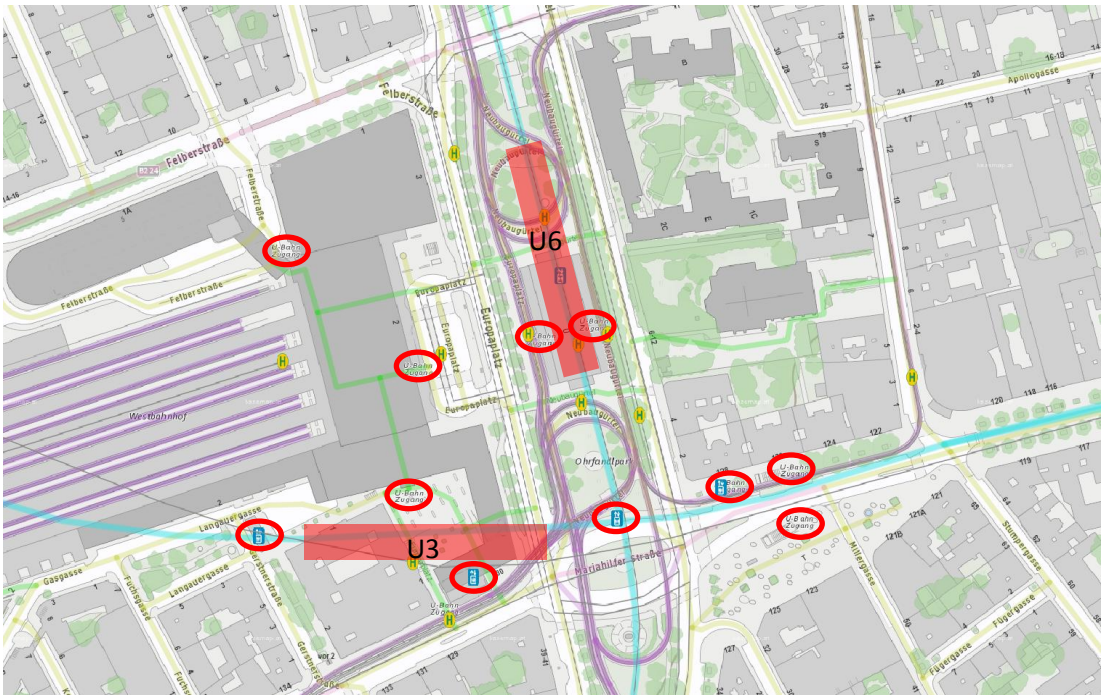
https://www.data.gv.at/katalog/dataset/79e7e34e-4d2f-46be-8a40-24a84b20b1ba

https://www.wien.gv.at/verkehr/oeffentlich/barrierefreiheit/

So even if the there is no data available about the disability of people, the analysis could still answer the question of the percentage of the total population that is serviced by stations accessible for people with disabilities. This could also be evaluated for population by age and sex and further population variables.

In fact anyone can be temporarily disabled e.g. injured or when pushing a push chair during parenthood, so this interpretation of the indicator would be very meaningful.

Map of Wien Westbahnhof station and underground stations U3 and U6 with its entrances:



 	Various entrances to underground system / Elevators to reach platforms. Often >= 300m from/to the actual platform!
	Underground platforms
	Station points – in case of undergrounds these are of course not on/linked with the road network
	Trams and train lines
	Underground lines
	Pedestrian ways
	various road classes
Source: Background Map Roads and traffic lines Entrances, elevators, station points	Basemap https://www.basemap.at/ GIP http://www.gip.gv.at/ OGD Austria https://www.data.gv.at/ https://www.data.gv.at/katalog/dataset/f1f6f15d-2faa-4b62-b78b-80599dd1c66e

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)

B. NATIONAL PRACTICE | France (NMCA)

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)

INSEE (the National Institute of Statistics and Economic Studies) in articulation with IGN (the National Geographic Institute)

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 situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

a) Spatial analysis to delimit the built-up area of the urban agglomeration:

The metadata recommends: “Delimit the built-up area of the urban agglomeration and calculate the total area (square kilometres). Area of delimitation should be aligned with Census enumeration areas to match with demographic data.”

INSEE publishes population/Census data on Census enumeration districts in urban areas, called IRIS (aggregated units for statistical information).

There has been a project between IGN and INSEE to improve the geometry of these statistical units, namely better accuracy and better consistency with transport data.

b) Inventory of the public transport stops in the city:

IGN has data on railway stations and data on other public transport stops may likely be provided by local authorities. Main challenge would be to gather the information available from various providers at national level.

c) Computation of the service area:

The methodology recommends use of the network distance.

IGN has data on road network: this data is at 10K, accuracy around 1m and continuously updated so completeness is ensured. The motorways (that can't be used by pedestrians) may be excluded using the attribute formOfWay.

d) Assessment of the population within the service areas:

Two potential hypotheses could be applied:

1. The first one is based on the data that is available for users (i.e. that respects the privacy restrictions). Typically, it is the method that would likely be used by local authorities if they had to compute this indicator

2. The second one is based on the assumption of point-based statistics as recommended by the GEOSTAT-2 project.

Hypothesis 1

Population data is available only on Census enumeration districts (called IRIS in France).

The transfer of population information from the statistical units (IRIS) to the service areas may be done, using the Buildings data of IGN and the process described in INSPIRE data specifications (B1.2 – computation of population at night): the principle is to repartition of the population according the pro rata of the "liveable surface" of residential buildings within the service area compared with the liveable surface" of residential buildings in the IRIS area:

- Extract residential buildings
- Compute their "liveable area": geometry area x number of floors above ground
- Compute the number of inhabitants of each residential building: (population of IRIS / sum of liveable areas of all residential buildings within IRIS) x liveable area of the given building
- Sum up the number of inhabitants for the residential buildings within the service area

The IGN data on Buildings is at scale 10K, accuracy around 1 m, continuously updated but with some difficulties (completeness is not totally ensured). Until now, the attribute "number of floors" is not present but it may be derived from the "height above ground". The main

issue is about the attribute "currentUse" that does not enable users to extract residential buildings in a reliable way: in practice, the "industrial", "commercial" or "agriculture" buildings may be excluded (if their use may be identified from aerial images) but it is not possible to make distinction between "office" buildings and residential ones. IGN is struggling to get "number of floors above ground" and to improve the quality of "currentUse" attribute in order to have reliable information on residential buildings. This might be done with partnership with Cadastre.

At European level, attributes "numberOfFloors" and "currentUse" are in the INSPIRE data model and are considered as first priority information by UN-GGIM: Europe WG on Core Data.

Hypothesis 2

Population data is available as point-based statistics. In this case, the methodology just consists to sum up the number of inhabitants of each "point" located in the service area. Due to privacy issues, this method might be applied only by the NSI.

e) Identification of mass transit stops

This step may be necessary to assess the "accessibility related to transport planning". "Mass transit stops" are reference points other than "homes": they include public services (schools, hospitals, …), working places, commercial centres or areas.

IGN has data on POI (Point of Interest) including public services, such as schools and hospitals. Globally, data has good quality (scale 10K), but no indication on the capacity of the public service.

There is a Business Register, called SIRENE. Some selection might be done, e.g. regarding the "size" of enterprises and the selected enterprises might be geocoded, using the national address geographic database (BAN).

Geospatial data analysis and integration: (Describe spatial analysis methods, procedures and computations, including regarding data integration)

Computation of the service area:

The intersection of the "network buffers" around public transport stops and the urban extent is pure GIS function and does not require additional data.

Assessment of the population within the service areas

Hypothesis 1:

Population data is available only on Census enumeration districts (called IRIS in France).

The transfer of population information from the statistical Units (IRIS) to the service areas may be done, using the Buildings data of IGN and the process described in INSPIRE data specifications (B1.2 –

computation of population at night): the principle is to repartition of the population according the pro rata of the "liveable surface" of residential buildings within the service area compared with the liveable surface" of residential buildings in the IRIS area.

- Extract residential buildings
- Compute their "liveable area" : geometry area x number of floors above ground
- Compute the number of inhabitants of each residential building: (population of IRIS / sum of liveable areas of all residential BU within IRIS) x liveable area of the given BU
- Sum up the number of inhabitants for the residential buildings within the service area

Hypothesis 2:

Population data is available as point-based statistics. In this case, the methodology just consists to sum up the number of inhabitants of each "point" located in the service area. Due to privacy issues, this method might be applied only by the NSI.

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.

IGN has data on road network. This data is at 10K, accuracy around 1 m and continuously updated so completeness is ensured. The motorways (that can't be used by pedestrians) may be excluded using the attribute formOfWay.

The IGN data on Buildings is at scale 10K, accuracy around 1 m, continuously updated but with some difficulties (completeness is not totally ensured).

IGN has data on POI including public services, such as schools and hospitals. . Globally, data has good quality (scale 10K), but no indication on the capacity of the public service.

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)

Assessment of the population within the service areas

The IGN data on Buildings is at scale 10K, accuracy around 1 m, continuously updated but with some difficulties (completeness is not totally ensured). Until now, the attribute "number of floors" is not present but it may be derived from the "height above ground". The main issue is about the attribute "currentUse" that does not enable users to extract residential buildings in a reliable way: in practice, the "industrial", "commercial" or "agriculture" buildings may be excluded (if their use may be identified from aerial images) but it is not possible to make distinction between "office" buildings and residential ones. IGN is struggling to get "number of floors above ground" and to improve the quality of "currentUse" attribute in order to have reliable information on residential buildings. This might be done with partnership with Cadastre.

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)

B. NATIONAL PRACTICE | Ireland (NSI)

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)

Central Statistics Office Ireland, National Transport Authority.

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)

It is possible to disaggregate the indicator by the following variables from the Census file:

- By sex
- By age
- By mode of transport
- By ethnicity

However, publication of the above would have to respect office policy on statistical disclosure control. Certain variables may only be published if grouped into broad categories.

As with other countries the available data sources allow for these breakdowns:

- Proportion of urban area that has convenient access to public transport
- Proportion of urban central/suburban area that has convenient access to public transport

Frequency of dissemination: (Describe the time interval at which information is disseminated over a given time period)

Given that the Census is the source of the demographic data for this indicator, the data can be disseminated at five year intervals.

Timeliness: (Length of time between data availability and the event or phenomenon they describe. Describe the average production time for each release of data)

As described the demographic data is available every five years. Data containing additions, deletions and modifications from the database of the National Mapping Agency (called PRIME2) are available every three months with the full country being surveyed over a three year period. The data in relation to public transport stops are the responsibility of the National Transport Authority, and are available on a regular basis.

Data sources: (List the data sources and themes or variables in use, including conditions of access situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

- National data on officially recognized public transport stops including coordinates and traffic frequency for each stop is published as open data on the data.gov.ie website by the Irish National Transport Authority

- Public transport networks and stations, along with the road network are also available in the Ordnance Survey Ireland databases. All mapping products are free to public service bodies under the National mapping agreement

- Resident population and population characteristics can be obtained from the most recent Census which has been geocoded to address point level. During the fieldwork around 1.5% of dwellings enumerated were not listed in the national address database, and these were retrospectively geocoded according to the location stated on the enumerator maps.

- The delineation of urban settlements is conducted following the processing of Census results, i.e., every five years. For the calculation of an urban area, the definition applied by the Census is used. Since the 2011 Census a settlement is defined as being a cluster of 50 or more occupied dwellings, with a maximum distance between any dwelling and the building closest to it of 100 metres, and where there was evidence of an urban centre (shop, school etc.). The criteria for extending an existing settlement depends on whether there are newly created estates or dwellings within 100 metres of the original settlement boundary.

  At national level an urban area is a settlement in excess of 1 500 persons. Five of the urban settlements are considered to be cities .i.e. populations exceeding 50 000 (Dublin, Cork, Galway, Limerick and Waterford).

Geospatial data analysis and integration: (Describe spatial analysis methods, procedures and computations, including regarding data integration)

Datasets for all national public transport operators are gathered and maintained by the National Transport Authority on the data.gov.ie website. These datasets include coordinates along with extensive information about routes, trips and traffic frequency for each stop. Data are provided through an API under open data license in GTFS format (Google General Transit Feed Spec) which can be imported into a GIS. Information relating to identifying stops with “frequent service” during peak or off-peak travel times can be done by using the time-table information connected to each stop. The public transport stops would be combined with the settlements boundary to determine which of these stops are located in urban areas.

Using the Euclidian distance buffering operation (0.5 km) on the public transport stops can produce areas which meet the proximity requirements of the indicator. Given that the Census dataset is point

based, the created buffer areas can be placed over the address points enumerated in the Census to calculate the required population characteristics.

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.

Population (Census dataset)
Completeness: 100%
Temporal accuracy: five-yearly
Positional accuracy: 1 metre for 98% of dataset, 10 metres for remaining 2%

Public transport stops

Resolution: x, y coordinates
Completeness: 100%
Temporal accuracy: Annual
Positional accuracy: Unknown

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)

As already described the geospatial data are used for calculation of Euclidian distance, buffer creation and point in polygon analysis.

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)

Ideally emphasis should be placed on calculating the 'shortest distance' route as opposed to Euclidian distance. The meaning of terms such as "safe and comfortable environment" and "frequent service" in the indicator metadata would need to be clarified further in order to refine the methodology.

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)

B. NATIONAL PRACTICE | Sweden (NSI)

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)

The global indicator does not exist yet within the current statistical framework of Sweden, accordingly no agency has yet been assigned responsibility for the indicator. Most likely responsibility will reside with Transport Analysis, an agency responsible for transport policy including official statistics on transport. Possibly also with support from Statistics Sweden as the calculation of the indicator involves micro data (geocoded population data) with privacy restrictions.

As a national complementary indicator for Goal 11, (11.2.1 Proportion of population that has convenient access to public transport), The Swedish National Board of Housing, Building and Planning proposed to measure "Number of dwellings and new dwellings developed in proximity of public transport stops".

The rationale on the national complementary 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 has already been adopted for the monitoring of the national environmental objectives (15. A Good Built Environment). Statistics Sweden agreed to collaborate with the National Board of Housing, Building and Planning, providing calculations of the indicator for the monitoring framework.

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)

Global indicator:

According to the proposed approach and data sources described in the global metadata, the following requested groups of disaggregation would be possible meet:

- By location (intra-urban)
- By income groups
- By sex
- By age
- By mode of transport

Disaggregation by race and ethnicity is not possible as these variables may not be registered according to Swedish law.

According to the proposed approach and data sources described in this document, the following quantifiable derivatives will be possible:

- Proportion of urban area that has convenient access to public transport
- Proportion of population/urban area that has frequent access to public transport during peak hours

- Proportion of population/urban area that has frequent access to public transport during off-peak hours.
- Proportion of urban central/suburban area that has convenient access to public transport

National complementary indicator:

- Number of dwellings in total within 400m, 1 000m and 2 000m from frequently trafficked public transport stops.
- Number of new dwelling (completed during last year) within 400m, 1 000m and 2 000m from frequently trafficked public transport stops
- Population, total and by sex, within 400m, 1 000m and 2 000m from frequently trafficked public transport stops

All variables are also reported as shares (%) of the total for national level and by county and municipality.

Frequency of dissemination: (Describe the time interval at which information is disseminated over a given time period)

Global indicator:

This is a theoretical question as the indicator has not yet been launched. However, with regards to the approach and data sources described in this document, annual dissemination will be possible I needed. Data sources to be used are well structured and continuously updated. Hence, processing time is estimated to be fairly limited.

Timeliness: (Length of time between data availability and the event or phenomenon they describe. Describe the average production time for each release of data)

Global indicator:

This is a theoretical question as the indicator has not yet been launched. However, with regards to the approach and data sources described in this document, annual dissemination will be possible I needed. Data sources to be used are well structured and continuously updated. Hence, processing time is estimated to be fairly limited.

Data sources: (List the data sources and themes or variables in use, including conditions of access situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

Global indicator:

- National data on officially recognized public transport stops including coordinates and traffic frequency for each stop. Data is available under open data license in GTFS format (Google General Transit Feed Spec) and provided jointly by the public transport service providers (www.trafiklab.se).

- Population data geocoded to address point location. In Sweden, population data is retrieved from administrative data from the central population registry, which can be geo-enabled by use of geocoded authoritative address and/or building data (INSPIRE conformant) from the NMCA.
- Geographic delimitation of urban areas (localities) following national methodology, conducted by Statistics Sweden.

National complementary indicator:

- National data on officially recognised public transport stops including coordinates and traffic frequency for each stop. Data is available under open data license in GTFS format (Google General Transit Feed Spec) and provided jointly by the public transport service providers (www.trafiklab.se). Public transport stops are considered to be "frequently trafficked" if there are at least one departure per hour during working days between 06:00 and 20:00. All modes of transport are regarded.
- The geocoded national dwelling register. The national dwelling register is authoritative data maintained by the NMCA as part of the Real Property Register. The dwelling register is part of the NSDI (National Spatial Data Infrastructure) but the use is restricted by confidentiality reasons. One of the main purposes of the register is production of statistics.
- Population data geocoded to address point location. In Sweden, population data is retrieved from administrative data from the central population registry (Tax administration), which can be geo-enabled by use of geocoded authoritative address and/or building data (INSPIRE conformant) from the NMCA.

Geospatial data analysis and integration: (Describe spatial analysis methods, procedures and computations, including regarding data integration)

Global indicator:

a) Spatial analysis to delimit the built-up area of the urban agglomerations:

This step is in principle already completed as Statistics Sweden recurrently (since 1960) delimits the geographical extent of urban areas ("localities") as part of the production of urban official statistics. A locality consists of a group of buildings normally not more than 200 metres apart, and must fulfil a minimum criterion of having at least 200 inhabitants. Thus, localities include the largest cities as well as small areas with just over 200 inhabitants. Delimitation is conducted as an automatic geospatial process involving high quality authoritative geospatial data from the NSDI in combination with point-based population data geocoded to the level of address location. The result is a national polygon dataset representing the urban extent of each locality (some 2 000 in Sweden). Data is available under open data license agreements.

b) Inventory of the public transport stops in the city:

Fortunately, a complete national dataset (covering the whole country and all modes of transport) on officially recognized public transport stops is gathered and maintained by a consortium representing all public transport service providers in the country (www.trafiklab.se). The dataset includes

coordinates along with extensive information about routes, trips and traffic frequency for each stop. Data is provided through an API (Application-Programming-Interface) under open data license in GTFS format (Google General Transit Feed Spec).

c) Computation of service areas

Service areas can be easily computed through Euclidian distance buffering operation (0.5 km) on the basis of the public transport stops. As proposed in the metadata description, there might also be relevant to identify stops with "frequent service" during peak travel times. This can be done by using the time-table information connected to each stop. However, no definition is provided in the metadata description. A good point of departure for further discussions on criteria for "frequent services" can be found in Poleman and Dijkstra (2015) paper on *Measuring access to public transport in European cities*.

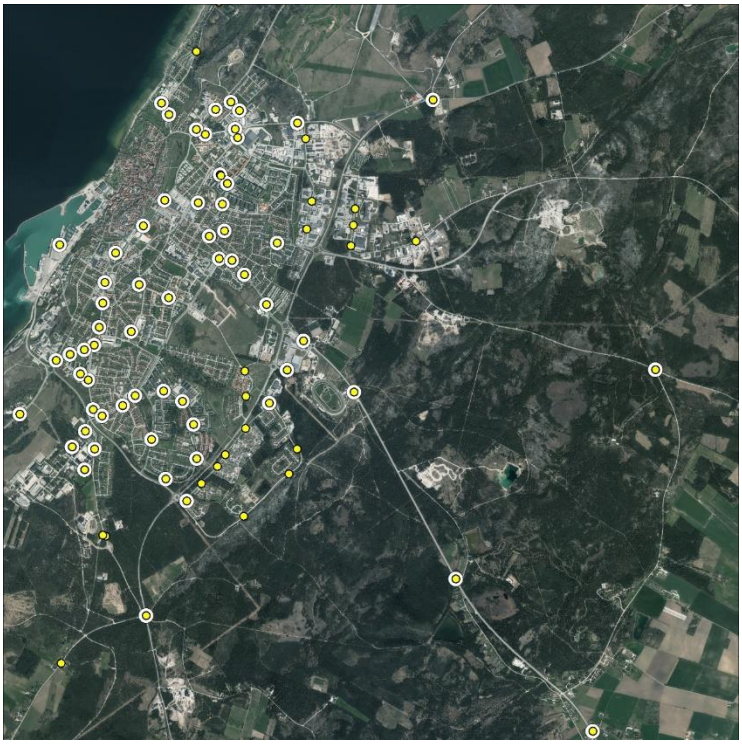
d) Assessment of the population within service areas:

When the service areas have been created for the entire county they can be overlaid with the urban delimitation data (localities) to define those service areas located within urban agglomerations. Once this is done, it is a quite simple operation to assess the population within these service areas. Point-in-polygon statements involving urban service areas (polygons) and high resolution point-based population data (points) will accurately return the number of people permanently settled within 0.5 km from public transport stops. This figure can be related to the total number of people living in the urban agglomeration (which is also calculated from point-based population data.

National complementary indicator:

The computation of the national complementary indicator on housing proximity of public transport is based on the following steps:

Step 1 is to process the public transport data in order to select only those stops that matches the desired frequency of departures. As data comes in GTFS format, the specification is open and interpretable. An automated routine has been set up to filter out the relevant stops. The image below shows all public transport stops (yellow dots) and transport stops considered "frequently trafficked" (yellow dots surrounded by a bigger white dot).

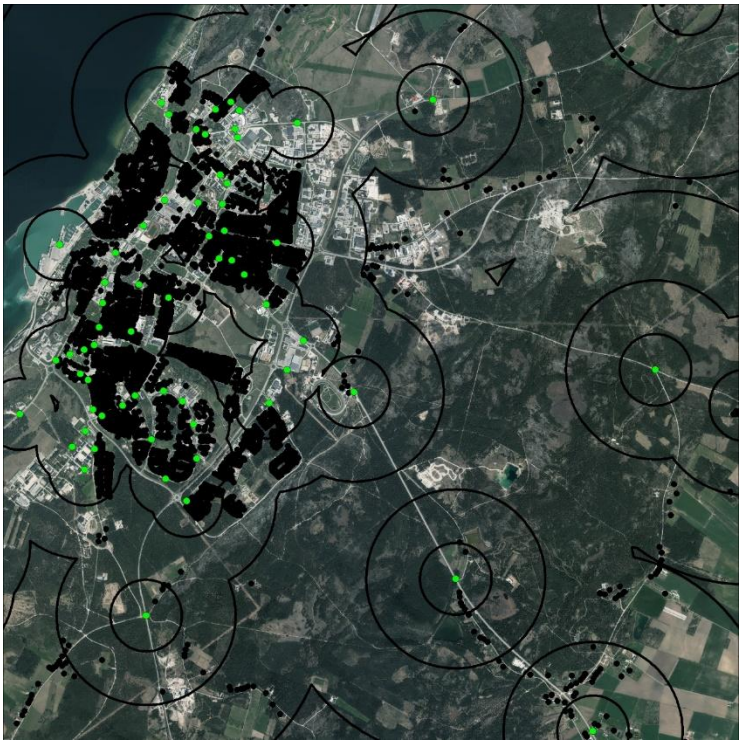


Source: © Lantmäteriet, Trafiklab and Statistics Sweden.

Step 2 is to create services areas around each public transport stop (buffers) with varying sizes (400, 1 000 and 2 000 meters). The image below shows frequently trafficked public transport stops with service areas.

Source: © Lantmäteriet, Trafiklab and Statistics Sweden.

Step 3 is to prepare the dwelling register (link dwelling data to coordinates of buildings) and to 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). Dwellings are considered "new" if the building year of the building in which the dwelling is located is from the previous year. The image below shows buildings with registered dwellings together with service areas of public transport stops.



Source: © Lantmäteriet, Trafiklab and Statistics Sweden.

Step 4 is to 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 is to calculate the total figure for dwellings and population by county and municipality to be able to calculate a share.

Step 6 is to 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, on a national level, that 78% of all dwellings are located within 400 meters from a "frequently trafficked" public transport stop and 90% of the dwellings were located within 1 000 meters. Among the new dwellings (completed throughout the year 2015) some 83% were located within 400 meters from a frequently trafficked public transport stop. The indicator will be updated annually in order to follow the trend of new housing in proximity of public transport.

Though data on officially recognised public transport stops available under open data license in GTFS format (Google General Transit Feed Spec) is typically not authoritative data, it can provide very useful data in many countries. The GTFS is an open data model that can be used in a rather flexible way. E.g. exceptions from regular time tables can apply using several different approaches. Hence, it is very important to analyse and interpret the data in order to find out how create the selection rules for departure frequency.

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)

Global indicator:

The proposed method entails a few problems with regards to the full scope of the indicator. For example, accessibility for disabled persons cannot be assessed using geocoded population data retrieved from administrative sources, as this information is not registered. The proposed approach also entails constraints as to determine if a stop is in a "safe and comfortable environment". There is currently no information to support such considerations.

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)

B. NATIONAL PRACTICE | Switzerland (NSI)

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)

Swiss Federal Office for Spatial Development ARE and three other federal offices (thematic responsibility); Swiss Federal Statistical Office FSO (responsible for the data sources used).

Since Switzerland already has a highly developed and dense network of public transport, this indicator does not seem to be a serious challenge at the national level. Generally, the main issue is to maintain and consolidate the present level of public transport service, with only marginal increases in density, frequency and capacity depending on increasing population and mobility demands.

Therefore, Switzerland emphasizes on the access to public transport for persons with disabilities where improvements are required and political targets set. Consequently, Switzerland has defined at the national level the indicator "Autonomous utilization of public transport by persons with disabilities". Its target is to measure the percentage of seriously handicapped people between 15 and 64 years of age living in private households which can use public transport autonomously (without aid by third persons) and without difficulties.

The indicator is based on the results of the Swiss Health Survey, it represents therefore a partly subjective self-estimation of the persons questioned.

In addition, Switzerland will publish another indicator relevant for target 11.2, which is the "Average distance to next public transport stop". Switzerland calculates this indicator mainly to supply information to target 9.1 concerning the quality of infrastructures.

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)

For the main indicator observed, the "Autonomous utilization of public transport by persons with disabilities", a spatial disaggregation is neither possible nor essential. Since the Swiss Health Survey is based on telephone interviews of a sample of 21 000 persons representing the entire population of 8 million, the share of respondents with disabilities is already quite low. Any further disaggregation, according to geographic or socio-economic criteria, is therefore considered inappropriate.

However, the second indicator "Average distance to next public transport stop" can be easily calculated for regions, cantons (NUTS3 units), agglomerations and even communes (LAU2 units). Theoretically, it could also be differentiated according to certain socio-economic characteristics of the population (persons or households), such as sex, age, and others.

Frequency of dissemination: (Describe the time interval at which information is disseminated over a given time period)

Since the Swiss Health Survey is repeated every five years, it is envisaged to update the indicator "Autonomous utilization of public transport by persons with disabilities" at the same five year interval. While the data sources for the second indicator "Average distance to next public transport stop" would allow an annual update, the calculation of the distances between each residential building and each stop of public transport is time-consuming and expensive. Switzerland envisages, therefore, updating this indicator every two to three years only.

Timeliness: (Length of time between data availability and the event or phenomenon they describe. Describe the average production time for each release of data)

Typically, these indicators could be published approximately two years after the reference dates of the surveys which are used as data source.

Data sources: (List the data sources and themes or variables in use, including conditions of access situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

Indicator "Autonomous utilization of public transport by persons with disabilities":

- Swiss Health Survey, survey of 21 000 computer-assisted telephone interviews supplemented by a detailed questionnaire which may be filled out online or on paper. It is updated every five years and not geocoded in detail. The origin of the respondents is identified geographically up to the cantonal level (NUTS3).

Indicator "Average distance to next public transport stop":

- Since 2010, the Swiss Census is annually updated based on the exploitation of the registers of inhabitants of cantons and communes which is complemented by sample surveys. The Population and Households Statistics (STATPOP) provides key figures on the size, structure (age, sex, civil status, nationality, etc.), development and spatial distribution of the population as well as information on the number and size of households. It is fully geocoded up to the level of a single building and its results can therefore be spatially disaggregated at will.
- National road network as basis for the distance calculations. This network is part of the large-scale topographical landscape model of Switzerland produced und maintained by the Swiss Federal Office for Topography swisstopo (i.e. Switzerland's National Mapping Agency). It is completely revised and actualized at a six-year periodicity, i.e. every year; one sixth of the national territory of Switzerland is updated.
- List of the public transports stops from the Federal Office of Transport (FOT). A selection of public transport stops that are in operation and used for passenger service was made out of this list. A further analysis of cadences was made using data from Federal Office for Spatial Development ARE. All these data are available in the Swiss geographical portal https://map.geo.admin.ch.

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.

While the results of the Swiss Health Survey used for the indicator "Autonomous utilization of public transport by persons with disabilities" represent the personal, subjective opinions of the questioned individuals, the data sources for the indicator "Average distance to next public transport stop" are considered highly reliable and accurate.

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)

For the indicator "Average distance to next public transport stop" two of the most important geo-spatial data sources available in Switzerland are directly used as base for the calculations (see section on Data sources). These calculations are performed using a GIS infrastructure and GIS functionalities.

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)

The present example shows to which extent it is possible to use existing data and data collection programs to fulfil the reporting requirements as they are understood in Switzerland. Due to limited resources all possibilities are used in order to define variables and indicator calculations which allow to be implemented without requiring expensive adjustments of the present or establishment of new surveys or measurement programs.

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)

Described in previous sections.

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)

- Population and economic data geocoded to address
- National road network
- List of public transport stops (Federal Office of Transport)
- Other data layers from the Swiss geographical portal (https://map.geo.admin.ch)

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.

Described in previous sections.

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)

Described in previous sections.

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.)

Described in previous sections.

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)

Described in previous sections.

C. EU SDG Distribution of population by level of difficulty in accessing public transport | Eurostat

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)

Eurostat.

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)

Per Member State by degree of urbanization.

Frequency of dissemination: (Describe the time interval at which information is disseminated over a given time period)

Ad hoc module to the EU-SILC in 2012 and it is not clear if it will be repeated.

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 situation, resolution, positional accuracy, frequency and timeliness regarding the ESS and NSS projects and /or INSPIRE conformance).

EU Survey on Income and Living Condition (EU SILC).

Geospatial data analysis and integration: (Describe spatial analysis methods, procedures and computations, including regarding data integration)

Not applicable as no spatial analysis involved.

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.

EU SILC data on the distribution of population by level of difficulty in accessing public transport by income quintile and degree of urbanization reveal which population groups are most affected by limited accessibility in EU countries.

Unit of measure is percentage of population total, in cities, in towns and suburbs and in rural areas facing:

i. very low
ii. low
iii. high
iv. very high difficulties in accessing public transport.

As such it is based on peoples' perception not on spatial analysis. See http://ec.europa.eu/eurostat/web/sdi/sustainable-cities-and-communities for more information on this indicator.

Resolution is by degree of urbanization per Member State. Selected users require information per individual city or NUTS 3 area. An important discussion is therefore if this indicator should be available for individual cities and for all NUTS3 areas e.g. also rural regions which tend to have the most severe shortcomings in terms of public transport.

General comments: (Please provide any additional information or comments)

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)

The DG REGIO working paper (2015) on *Measuring access to public transport in European cities* (Hugo Poelman and Lewis Dijkstra) method is based on actual walking distance along road networks, instead of Euclidean distances. It uses the stop as access point to public transport. The population served is based on gridded population with grid cell sizes 1km or smaller or house blocks. It considered the EU Urban Audit city definition of urban centres (high-density cluster) for 29 EU cities.

The main change in method would be to base it on objective criteria instead of perception of people and in particular on the service frequency based on time table information.

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)

Completely different approach, instead of a survey, it would be based on geospatial information, population distribution and time table information, so it is a principle methodological change and not a gradual change.

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)

- Population distribution at grid or block level best with cell size 100m;
- Detailed road network (scale 10 000) with footpaths;
- Transport stops locations;
- Time table information for all transport networks of a country and all public transport modes.

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.

Time table and transport stop information would need to be available for all transport networks that is currently the biggest obstacle for implementing this method.

Population grid and transport network information is available, e.g. GEOSTAT 2011 population grid and OSM (Open Street Map) or commercial road networks.

The data should be updated every three years, more frequent data is not likely to reveal any significant trends. Data should be fairly accurate, e.g. scale 1:10 000.

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)

Population grid and transport network information is available. DG REGIO study used building blocks corresponding to polygons of the Copernicus Urban Atlas layer (2006). In areas with no data population was estimated to 100m x 100m grid (downscaling from the EU 2006 population grid).

TomTom or alternatively OSM provide transport stops and road network including footpath information. DG REGIO study used TomTom MultiNet

Transport stops and time table information is not available for all transport networks and for all MS, which is an obstacle for calculating the indicator at EU level. DG REGIO study used national data sources on public transport were used. E.g.:

- Sweden: integrated data: http://www.trafiklab.se/api/gtfssverige

- Finland: integrated data: http://developer.reittiopas.fi
- Germany (Berlin): integrated urban transport data: http://www.gtfs-data-exchange.com/agencies

Google also has transport timetables available but with coverage gaps

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.)

TomTom is of commercial nature. OSM is free and open. Transport timetables are available from Google, but coverage is extremely patchy.

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)

DG REGIO paper on *Measuring access to public transport in European cities* calculation using geospatial data and public time table information took into account.

1) How many people could easily walk to a public transport stop? For bus and trams, it was assumed that people would be willing to walk five minutes (417 metres) to a bus or a tram stop. For a train or a metro, it was assumed people would be willing to walk 10 minutes (833 metres) as they generally offer a higher speed. The walking distance was calculated using a street network. This means that it takes into account the density of the street network and obstacles such as rivers, steep slopes, highways or railroads, which cannot easily be crossed on foot;

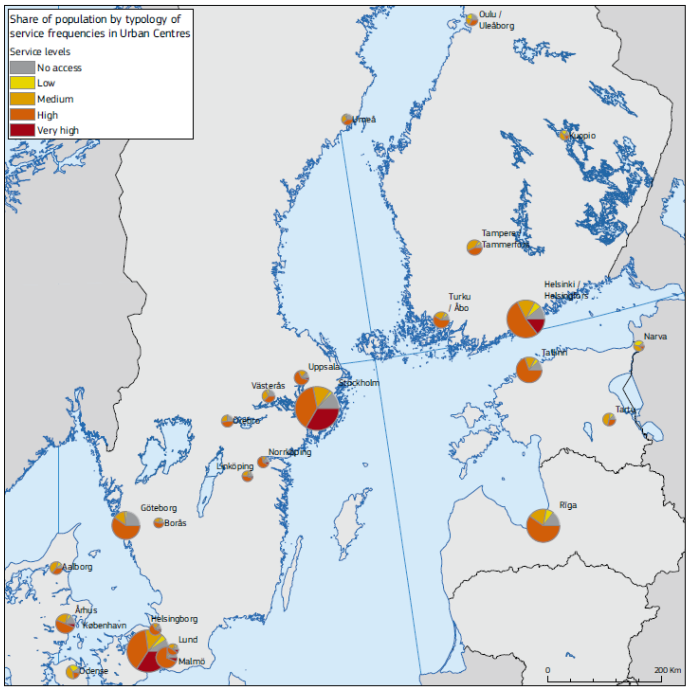
2) The number of departures on a normal weekday between 6:00 and 20:00. The average per hour to create frequency classes was calculated. Stops that were less than 50 metres apart were grouped, which means that in most cases departures in both directions on the same route were taken into account. So for example, a bus stop with only one route with six departures an hour would have three departures in one direction and three in the other;

3) Five groups based on access and departure frequency were created:
   - No access: people cannot easily walk to a public transport stop, in other words 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.
   - Low access: people can easily walk to a public transport stop with less than four departures an hour.
   - Medium access: people can easily walk to a public transport stop with between 4 and ten departures an hour.
   - High access: people can easily walk to a bus or tram stop with more than 10 departures an hour OR people can easily walk to a metro or train station with more than 10 departures an hour (but not both).

- Very high access: people can easily walk to a bus or tram stop with more than 10 departures an hour and a metro or train station with more than 10 departures an hour. Very high access is only possible in cities with a metro and/or a train network and depends heavily on the extent of this network.

4) Accessibility areas ('service areas') around each of the stops were created. They were defined as a five-minute walk (at 5 km/h) to bus and tram stops, and as a ten-minute walk to stops of high-speed modes (metro and train). The service areas were created using all streets of the road network accessible to pedestrians, instead of Euclidean distances, and took into account the existence of barriers (e.g. motorways, railways, water bodies).
Each of the service area polygons is characterised by the sum of the hourly average number of departures available at the stop around which it is created. The service areas tend to partly overlap each other, especially in an urban environment. In these overlapping areas, people have the choice between two or more stops nearby, where the departure frequency can be different. The study assumed that the stop with the most frequent departures is the most probable choice. For this reason, the service areas within each of the groups of transport modes were intersected, and to each of the overlapping areas the maximum value of the hourly average number of departures was attributed. Mapping this result showed the best available level of service (within each of the groups).

Map 3: Access to public transport in urban centres in Denmark, Sweden, Finland and Estonia



Source: Poelman, H. & Dijkstra, L. (2015). *Measuring access to public transport in European cities*. Regional Working Paper, DG REGIO.