# Meeting of the UN-GGIM Expert Group on the integration of statistical and geospatial information –Lisbon 24 May 2015

## Background document for session 4

## Proposal for a common statistical-geospatial terminology database

Date:                      12/5/2015

Version:                   0.1

Authors:                   PETRI

Revised by:

Approved by:

Public:

Reference Number:

## Document History

| Version | Date | Comment | | Modified Pages |
|---------|------|---------|---|----------------|
| 0.1 | 12/05/15 | Document created by PETRI | | All |
| | | | | |

# 1 PURPOSE OF THE DOCUMENT

This document responds to the request from the Expert Group on the integration of statistical and geospatial information to share knowledge of existing terminologies and practices and align the description of concepts in order to reach a common understanding among representatives of statistical and geospatial communities. As part of its program of work, the Expert Group decided to develop a web-based portal to share knowledge, including a glossary of statistical-geospatial concepts and terminology. The purpose of this document is to propose an initial set of definitions of key concepts. Once agreed these definitions should form the nucleus of a terminology database on a statistical-geospatial framework to be hosted by UNSD.

# 2 SOURCES

The terms and their definitions presented in this paper are mainly based on information in existing UN-GGIM reports, the Statistical-Spatial-Framework of the Australian Bureau of Statistics, INSPIRE, the European Statistical System and Wikipedia. A complete list of data sources can be found in Annex 1.

# 3 APPROACH

In order to agree on useful and widely accepted terms and concepts, the Expert group should establish an editorial board composed of experts from both communities (statistics and geospatial). This editorial board should first suggest additional terms to those presented in this paper. Starting from this list the editorial board should review each of them for their place in a hierarchy of concepts, thereby avoiding overlap, ambiguity and duplication with other terms. Moving along this structure from general to specific the editorial should review the proposed definition for each of the terms with to goal to reach consensus from both communities. Cross-references between terms should be established, too. At the end the list should be adopted by the Expert Group and then transferred into an on-line portal. The editorial board will remain in charge of revisions of existing terms and addition of new terms.

# 4 FRAMEWORKS AND CORE CONCEPTS

## 4.1 Statistical-geospatial framework

*(Synonym statistical spatial framework, location framework)*

A statistical-geospatial framework defines an information infrastructure composed of statistical and geospatial information that are connected and conceptually integrated. A statistical-geospatial framework has the goal to spatially enable all statistics along the full statistical production process. It connects spatial information that describes our physical man-made and natural environment, and statistics that describe their socio-economic and environmental attributes.

1) As such, in a basic form it may be understood as two separate frameworks, a geospatial framework and a statistical framework that are joint together technically, semantically and structurally at defined interfaces.

2) In a more integrated perspective, a statistical-geospatial framework may be understood as a system of objects that describe our environment and societies whereby our statistics describe their socio-economic-environmental attributes and temporal behaviour of these objects and the geospatial information describe their physical attributes and location in space.

The statistical-geospatial framework in its basic interpretation does not necessarily contain statistical or geospatial data. It could also be understood as a set of guidelines on how to integrate statistical and spatial data sources, generate new information through a number or processing steps, and improve the production process of information. From a user perspective, the framework should allow the user to discover, access, integrate, analyse and visualise statistical information for any area of interest.

To meet the above requirements the statistical-spatial framework supports the geocoding of statistics at unit record level using authoritative geospatial information. It also makes available a system of consistent geographies and grid systems for statistical dissemination. It also supports the technical integration of information in IT based production systems and exchange of information over the internet. This requires the existence of technical standards for data exchange, structural and semantical interoperability, dissemination, and metadata.

Moreover, it also needs to address organisational issues such as a data policy on scale, data maintenance and data quality and to regulate access conditions to data, dissemination rights of information, and data protection.

## 4.2   Geospatial reference framework

*(spatial reference framework for statistics)*

The geospatial reference framework contains a set of preferably authoritative geospatial information to be used to geocode directly or indirectly all public sector information at all levels of government, including data sources for statistical information. The components of the geospatial reference framework are not statistics though. The following categories of geospatial information form the geospatial reference framework (-> geospatial core information) can be defined.

Topographic data

- Detailed transport networks including public transport stops

- Hydrographic network

- Ortho-imagery and satellite imagery

- DEM

Geographies

- Administrative geographies (areas, regions)

- Statistical geographies (areas, regions)

- Address, building, dwelling register

- Land parcels (agriculture and estate)

- Cadastral maps

- Postal code geographies

## 4.3 GEOSPATIAL INFORMATION

*(synonym spatial information, geographical information, geospatial data and services)*

Generally, geospatial information is information that has traditionally been portrayed through maps or in association with maps. More technically geospatial information is defined as information with a direct or indirect reference to a specific location or geographical area on or near the surface of the Earth. Geospatial information generally relates to the natural and built environment, but also includes observations of people, and the social and economic outcomes of human activity. Geospatial information is stored in a geographic referencing system, usually a coordinate system of latitude, longitude and, increasingly, elevation.

## 4.4 Thematic geospatial information

Thematic geospatial is spatial data where the thematic attributes of geoobjects represent an elementary and intrinsic property of the object, in the absence of which these objects would not exist.

They can be used to directly create spatial statistics either stand-alone or in combination with geospatial reference data and other statistical data. Examples for this type of

- Land cover maps

- Land use maps

- Protected areas

- Other statistics collected on functional areas (geographies) (non-administrative or administrative);

## 4.5 Infrastructure for spatial information

Infrastructure for spatial information means metadata, spatial data sets and spatial data services; network services and technologies; agreements on sharing, access and use.

## 4.6 Geospatial core information

A set of essential geospatial data and services for geocoding other types of information. Examples are administrative boundaries, land-cover information, DEM, address points, ortho-photos and satellite images, transport and hydrographic networks, grid systems. In terms of services, geocoding and reverse geocoding, background map services, and routing could be considered as essential.

## 4.7 Geospatial reference information

Geospatial reference information is defined as information that is authoritative and to be used by all public stakeholders for a designated purpose. Reference information avoids conflicting results stemming from different data sources.

## 4.8 Spatial feature

*(Synonym spatial object, geospatial object)*

Geospatial object means an abstract representation of a real-world phenomenon related to a specific location or geographical area. Spatial features are represented with their geometry (point. line, polygon).

## 4.9 Location

Location is a general term to describe a place on or near the surface of the Earth. Location data is information that has any location component and is often used when referring to geospatial information.

## 5  DEFINITION OF DATA INTEGRATION PRACTICES

### 5.1 Integration of statistical and geospatial information

Integration of statistical and geospatial information describes the use of geospatial information for the production and dissemination of statistics. Integration of statistical and geospatial information can take place at any stage of the statistical production process, as described by the GSBPM. The integration includes geocoding of statistics, spatial analysis, and creating statistical maps. As part of the integration process the following steps may occur:

1) Geocode statistical information at unit-record level;

2) Processing and manipulation of statistical information using spatial analysis techniques with the purpose of selecting information or derive new information with a focus on their spatial characteristics, e.g. buffering around spatial features;

3) Supporting a more efficient and flexible statistical production process with geospatial information e.g. for surveying and sampling, field operation;

4) Combination of statistical end products with geospatial information in statistical maps;

5) Improving the quality of existing statistical products adopting spatial models, e.g. commuting information by calculating journey times based on detailed transport networks.

All statistical phenomena that can be associated to a location are in principle relevant for the integration of statistical and geospatial information. Location in this context means the location of the most individual observation at unit record level. In most cases the location will be a point with coordinates or a precise address. However, other spatial reference frameworks like lines or polygons are relevant as well representing e.g. road segments or areas with a certain land cover.

## 5.2   Linking

(*related term linked data*)

Linking defines a process of connecting structured data sources using a system of unique identifiers. Linking builds upon standard Web technologies such as HTTP, RDF and URIs . While integration describes the process of combining data from different thematic communities from a conceptual viewpoint, linking refers to technically connecting data in a machine-to-machine environment no matter what is the subject.

## 5.3   Geographic differencing

Geographic differencing is the process where the same data is obtained for two different but overlapping regions and the data from the smaller of these regions is subtracted from the data for the larger region. By utilising this method it is possible to obtain data for the area that is not common to both regions. Obtaining data for small areas using this method may result in a risk to privacy or confidentiality.

## 6   GEOCODING

## 6.1   Geocoding of statistics

*(synonym georeferencing of statistics)*

Geocoding of statistics describes the process of **geospatially enabling** statistical unit record data (i.e. data relating to individual persons, households, dwellings, businesses or buildings) by assigning a location descriptor (x, y, z coordinates, address, spatial id, spatial feature).

## 6.2   Geospatial enabling

*(synonym location enabling)*

Geospatial enabling describes the process of taking location information such as an address or administrative area code and linking this information to a geospatial feature.

The geocodes (location coordinates, address ids, or geographic areas codes), obtained from this process can be stored directly on the statistical unit record or linked in some way to the record. Unless geographical coordinates can be stored with the unit record, linking via key relationships is safer to avoid the changing geographies disrupt the time series.

# 7   SPATIAL STATISTICS

*(Synonym location-based statistics, geographically referenced statistics, location-enabled statistics, small area statistics)*

Location or extent are the main characteristics of spatial statistics. Spatial statistics is geocoded to small (in most cases below level 5) administrative or non-administrative geographies. Spatial statistics may also result from the integration of statistical and geospatial information during the statistical production process, although the product might be regional statistics. The cross-border perspective might be another important factor to define statistics as spatial.

The level of geography of spatial statistics should meet the perception of users in their area of interest ('What is the situation in my neighbourhood or in my areas of interest/ responsibility').

Spatial statistics is used to answer questions from a spatial perspective e.g. What is close? How many within a distance? How many per surface area?

## 7.1   Aggregated statistical information

Aggregated statistical information is aggregated from geocoded unit record level data into the dissemination geography, as opposed to disaggregated statistical information that is created using a spatial distribution model and larger statistical geographies as source data.

## 7.2   Regional statistics

*(Synonym sub-national statistics)*

Regional statistics are statistics that is geocoded to administrative and functional geographies down to administrative level 4.

## 7.3   Grid statistics

Grid statistics are spatial statistics geocoded to rectangular grid cells. Each grid cell as the same size and carries a unique code. Ideally the code carries also geocoding information, e.g. of the lower left corner of the grid cell.

# 8   GEOGRAPHIES

## 8.1   Geographical classifications

*(Synonym sub-national typologies, regional typologies, territorial typologies)*

Geographical classifications are a method to group geographies according objective criteria. Examples are classifications based on population density, functional aspects (labour market areas), or geography (mountain areas, maritime areas). Often geographical classifications are based on statistical or administrative geographies and are used to compare statistics between different areas with the same characteristics (e.g. urban with other urban areas).

## 8.2    Degree of Urbanisation

The Degree of urbanisation (DEGURBA) is a classification of municipalities based on population densities and urban clusters. Based on the share of local population living in urban clusters and in urban centres, it classifies municipalities into three types of area: *thinly populated area* (rural area); *intermediate density area* (towns and suburbs/small urban area), and *densely populated area* (cities/large urban area).

## 8.3    Spatial Unit

A spatial unit is an artificial demarcation with digital boundaries of an area that can be mapped. Spatial units can aggregate to cover large areas.**Locality**

Locality is a term used by different people to mean different things and assumptions should not be made about what the term means in any given usage. An increasingly important official use of the term is in connection with the census. A locality in this sense is a contiguous built-up area use for settlement reaching a minimum population threshold.

## 8.5    Region

Refers to a geographic area or boundary of any type: for example Suburb, Local Government Area, Local Administrative Unit, Statistical Areas at various sub-national levels.

## 8.6    Statistical area

Spatial unit used for the dissemination or collection of statistics.

## 8.7    Dissemination geography

*(Synonym output system, output areas)*

System of often hierarchically nested geographies with the goal to be either particularly suitable for analysis (functional geographies, grids) or to

## 8.8    Statistical geography

A statistical geography provides the extra dimension of location to statistics. A statistical geography effectively divides the area of interest, on which the statistics are collected or disseminated, into spatial categories, called statistical areas that allow the user to see not just how the data varies but also where it varies. An effective statistical geography is one which supports many uses and enables comparisons over time. Statistical geographies are often hierarchically nested areas to collect or disseminate statistics. The construction of statistical geographies may be functional but also population or socio-economic driven. If based on population they are designed around a consistent number of people or households within each hierarchical level. Statistical geographies may coincide with administrative, postal or other geographies. Examples are the European NUTS geography. Statistical geographies may also coincide with dissemination geographies.

## 8.9    Functional geography

Functional geographies are defined by characteristics other than their surface area or administrative level. Examples are geographical characteristics such as mountain areas, social characteristics such as less-favoured areas, areas in need for development, areas by type of economic activity etc.

## 8.10    Administrative geography

Administrative geographies are the spatial representation of the administrative division of a county. The largest administrative subdivision of a country is called the "first-level administrative level".

## 8.11    Enumeration geography

Enumeration geography is the division of a country into areas for census purposes. They represent the smallest area for which in most countries population information is available. However, in certain countries enumeration areas are further subdivided into blocks e.g. bounded by physical features such as streets or rivers.

## 9    SPATIAL ANALYSIS

*(synonym location analytics)*

The process of examining the locations, attributes, and relationships of spatial features in spatial information through overlay, distances, spatial selection, intersection, aggregation and other analytical techniques in order to address a question or gain useful knowledge. Spatial analysis extracts or creates new information from geospatial information.

## 10 ANNEX 1 – INFORMATION SOURCES

(1) Statistical Spatial Framework of the Australian Bureau of Statistics http://www.nss.gov.au/nss/home.nsf/pages/Statistical%20Spatial%20Framework%20Homepage

(2) Report of a task force of the European Statistical System on the integration of statistical and geospatial information

(3) UN-GGIM: Europe working group B on the integration of statistical and geospatial information (in draft, upcoming)

(4) INSPIRE directive http://inspire.ec.europa.eu/

(5) Expert Group report http://ggim.un.org/docs/meetings/UNSG_EG/ESA_STAT_AC.279_L4_Report%20of%20the%20EG%20Meeting_rev.pdf

(6) Global Forum report http://ggim.un.org/docs/meetings/GGIM4/GGIM4%20Report_en.pdf

(7) UN Data revolution report http://www.undatarevolution.org/

(8) EFGS reports www.efgs.info

(9) Statistics explained article on grids http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_grids