

# GSGF Europe: Terms and Definitions

GEOSTAT 4

---

Title: GSGF Europe: Terms and Definitions

Author: GEOSTAT 4

Project: Eurostat ESSnet grant project GEOSTAT 4

Grant agreement number: 945503 - 2019-FI-GEOSTAT4

It is permitted to copy and reproduce the content in this report. When quoting, please state the source.

© GEOSTAT 4 and Eurostat 2022

---

## Content

|  |    |
|--|----|
| 1. Introduction .....                                  | 5  |
| 2. Terms and definitions .....                         | 5  |
| Accuracy .....   | 5  |
| Administrative geographies .....                       | 6  |
| Aggregated statistical- geospatial information.....    | 7  |
| Application Architecture .....                         | 7  |
| Big data .....   | 7  |
| Business Architecture .....                            | 7  |
| Business Function .....                                | 7  |
| Business Process.....                                  | 7  |
| Business Service.....                                  | 8  |
| Capability .....                                       | 8  |
| Clarity .....  | 8  |
| Coherence .....  | 8  |
| Comparability.....                                     | 8  |
| Completeness.....                                      | 9  |
| Consistency .....                                      | 9  |
| Coordinate.....  | 9  |
| Crowdsourcing.....                                     | 10 |
| Currency .....   | 10 |
| Data ecosystem.....                                    | 10 |
| Data management environment .....                      | 10 |
| Degree of urbanisation.....                            | 10 |
| Disaggregated statistical- geospatial information..... | 11 |
| Dissemination geographies .....                        | 11 |
| Enterprise architecture (EA).....                      | 12 |
| Enumeration geographies.....                           | 12 |
| Framework .....  | 12 |
| Functional geographies .....                           | 12 |
| Gazatteer .....  | 13 |
| Geocoding .....  | 14 |
| Geographical differencing .....                        | 15 |

---

|   |    |
|---|----|
| Geographical classifications .....                          | 15 |
| Geography .....   | 15 |
| Georeferencing .....  | 15 |
| Geosampling .....   | 16 |
| Geospatial core information .....                           | 16 |
| Geospatial services .....                                   | 17 |
| Geospatial data quality .....                               | 17 |
| Geospatial data uncertainty .....                           | 17 |
| Geospatial information .....                                | 18 |
| Geospatial object .....                                     | 18 |
| Geospatial reference dataset.....                           | 18 |
| Geospatial reference information .....                      | 19 |
| Geospatial statistics .....                                 | 19 |
| Grid .....  | 20 |
| Grid statistics .....                                       | 20 |
| Information Architecture (IA) .....                         | 21 |
| Integration of statistical and geospatial information ..... | 21 |
| Interoperability .....                                      | 21 |
| Infrastructure for geospatial information.....              | 22 |
| Lineage .....   | 22 |
| Linking .....   | 22 |
| Locality .....  | 23 |
| Location .....  | 23 |
| Location information.....                                   | 23 |
| Metadata .....  | 24 |
| Microdata .....   | 24 |
| National Geospatial Information Agency (NGIA) .....         | 24 |
| National Spatial Data Infrastructure (N)SDI.....            | 24 |
| Open data .....   | 25 |
| Open data licence .....                                     | 25 |
| Open standard.....  | 25 |
| Operating environment .....                                 | 25 |
| Population grid .....                                       | 25 |
| Precision .....   | 26 |

---

|                                       |    |
|---------------------------------------|----|
| Primary Sampling Units.....           | 26 |
| Punctuality.....                      | 26 |
| Quality Dimensions .....              | 26 |
| Reference Architecture.....           | 27 |
| Region .....                          | 27 |
| Regional statistics .....             | 27 |
| Relevance.....                        | 27 |
| Reuse .....                           | 27 |
| Scanner data.....                     | 28 |
| Service.....                          | 28 |
| Sharing .....                         | 28 |
| Spatial analysis .....                | 28 |
| Spatial extent.....                   | 28 |
| Spatial Resolution .....              | 29 |
| Spatial unit .....                    | 29 |
| Statistical area .....                | 29 |
| Statistical geography .....           | 30 |
| Statistical-geospatial framework..... | 30 |
| Statistical unit.....                 | 32 |
| Survey data .....                     | 32 |
| Temporal Resolution.....              | 32 |
| Thematic geospatial information ..... | 33 |
| Timeliness .....                      | 33 |
| Web scraping .....                    | 33 |
| 3. References.....                    | 33 |

---

## 1. Introduction

The document is an extension of the chapter Terms and definitions from GSGF Europe document and contains also other terminology defined during GEOSTAT 4 project that is important for statistical and geospatial communities working on statistical-geospatial data integration.

## 2. Terms and definitions

### **Accessibility**

Accessibility together with clarity refers to the conditions and modalities by which users can obtain, use and interpret data.

### **Accuracy**

Accuracy refers to the closeness of estimates to the unknown true value.

Explanatory text: to the measurement of an error on how far the dataset is from the “true” data or reality (real-world phenomena). To be clear, the term ‘error’ refers to how far a measurement is from the truth (e.g., a form of metadata that describe known uncertainties by data providers, such as measurement errors) and is closely related to quality (high quality is correlated with small errors). Error can also result when data fail to completely describe the event aimed to be analysed represented.

On the other hand, accuracy provides a measurement of error in which an accurate dataset is close to a represented phenomenon with only small errors. Briefly, accuracy means closeness of agreement between a test result or measurement result and the true value (Source: ISO) and is the opposite of error. Thus, the term of error addresses the discrepancies between the encoded spatial, temporal and thematic attributes values and “actual” values for a given entity in which accuracy aims to measure it. In this regard, accuracy measures spatial, temporal and thematic errors, preferably recognising the interdependence of space, time and attributes. Inaccuracy is usually mistaken for the concept of ‘imprecision’ in which this last one arises from limitations imposed by data granularity (or data resolution) under which the observation is made, therefore the observation’s contexts of the phenomena generally lead to imprecise data. The most common components of error in geospatial and geospatial-temporal data include inaccuracy, imprecision and incompleteness.

**Map Accuracy:** Map accuracy refers to the level of accuracy of the (digital) cartographic representations resulting from different levels of generalisation and intrinsically linked to the scale of representation.

---

**Spatial Accuracy:** Spatial accuracy or positional accuracy refers to the proximity between the position of an object in the database and its true position in the real world.

Explanatory text: It could address absolute or relative accuracy. Absolute accuracy refers to the accuracy of data elements concerning a coordinate scheme. Relative accuracy refers to the accuracy of map features relative to one another. Positional accuracy is relatively easy to measure, especially in vector point-based data in which metrics are well-defined for point spatial entities, normally Euclidean distance between the encoded location and the location as defined in the specification, and summary statistics for a set of points. Spatial accuracy can be subdivided into horizontal (2D) and vertical (3D) accuracy elements and encompasses assessment methods that are based on comparison to source, comparison to a standard or higher accuracy, and deductive estimates or internal evidence.

**Temporal Accuracy:** Temporal accuracy refers to the agreement between encoded and 'true' temporal coordinates. The assessment of temporal accuracy depends on the ability to measure time objectively using a standard temporal coordinate system.

**Thematic Accuracy:** Thematic accuracy or attribute accuracy refers to the proximity between attributes values and their "true" values.

Explanatory text: The most common situation on thematic accuracy addresses categorical attributes through misclassifications (e.g., in map classification generating an unreliable spatial pattern). The metrics of thematic accuracy vary with the measurement scale, including quantitative attributes metrics and classification accuracy assessment in remote sensing addressing various types of misclassifications and developing error matrixes.

## **Administrative geographies**

Administrative geographies are the spatial representation of the administrative division of a country.

Explanatory text: The largest administrative subdivision of a country is called the "first-level administrative level". Local administrative units (LAUs) represent a system for dividing up a territory for the purpose of developing statistics at a local level. These units are usually low-level administrative divisions within a country, ranked below a province, region, or state.

Reference/related themes/topics: Geographical Classifications

---

## Aggregated statistical- geospatial information

Aggregated statistical-geospatial information is aggregated from geocoded unit record level data into the dissemination geography, as opposed to disaggregated statistical geospatial information.

Reference themes: Spatial statistics

## Application Architecture

Application Architecture is a description of the major logical grouping of capabilities that manage the data objects necessary to process the data and support the business - it details the structure of components, their inter-relationships, and the principles and guidelines governing their design and evolution over time. Source: CSPA Architecture Patterns 2018

## Big data

Big data is an all-encompassing term for any collection of data sets so large or complex that it becomes difficult to process them using traditional data-processing applications.

## Business Architecture

Business Architecture covers all the activities undertaken by a statistical organization, including those undertaken to conceptualize, design, build and maintain information and application assets used in the production of statistical outputs. Business Architecture drives the Information, Application and Technology architectures for a statistical organization. Source: Statistical Network BA definition

## Business Function

Business Function is something an enterprise does, or needs to do, in order to achieve its objectives. Source: GSIM

## Business Process

Business Process is a set of process steps to perform one or more Business Functions to deliver a Statistical Program. Source: GSIM



---

## Business Service

Business Service is the means of accessing a Business Function. It will perform one or more Business Processes. *Source: GSIM*

## Capability

Capability means something that enables us to carry out an activity: An ability that an organisation, person or system possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organisation, people, processes and technology to achieve (Source: The Open Group Architecture Framework TOGAF).

Explanatory text: Capabilities provide the unit/organization with the ability to undertake a specific activity. A capability is only achieved through the integration of all relevant capability elements, e.g., methods, processes, standards and frameworks, IT systems and people skills.

## Clarity

Clarity together with accessibility refers to the conditions and modalities by which users can obtain, use and interpret data. Clarity with accessibility forms one single quality dimension: "accessibility and clarity" that states that European statistics should be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

## Coherence

Coherence refers to the adequacy of the data to be reliably combined in different ways and for various uses.

## Comparability

Comparability refers to the measurement of the impact of differences in applied statistical concepts, measurement tools and procedures where statistics are compared between geographical areas, sectoral domains or over time.

---

## Completeness

Completeness refers to the presence or absence of data features, their attributes and relationships, including errors of omission and commission. The level of completeness is related to the degree of representation of all objects in a certain universe, including spatial objects, time references and attributes.

**Data Completeness:** Data completeness refers to the most common and well-known concept within completeness' quality dimension related to completeness which is the measurable error of omission observed between the dataset and the universe defined.

**Thematic Completeness:** Thematic completeness or attribute completeness refers to the degree to which all relevant attributes of a feature have been encoded.

## Consistency

Consistency refers to the degree of adherence to logical rules of data structure, attribution and relationships as well as the absence of apparent contradictions in a dataset.

Explanatory text: It refers to the fidelity of the relationships encoded in the dataset. It is also appropriately addressed as the dataset internal validity measurement as pieces of data and information that are consistent with each other or with what we already know. In the spatial domain, consistency is usually addressed to assess the conformance of certain topological rules - topological consistency -, including topological inconsistencies and cleaning as a prerequisite for GIS processing, and also identifying redundancies in spatial attributes. Consistency in the temporal domain is related to inconsistencies in the timeline and temporal overlaps, in which the underlying idea is that only one event can occur at one time at a given location and therefore, an inconsistency occurs if a different entity appears at the same location on two maps of the same time reference.

## Coordinate

Type of direct geographical position of an object on Earth, often expressed as latitude (y) and longitude (x) in which the intersection of these two lines or axis determines the geographical point of a place. The lines in the coordinate system can also create networks (grids)

Explanatory text: In GIS environment, spatial coordinates represent corresponding locations on Earth relative to other locations in a digital manner, usually through a vector with point format and attributes identifying the geographic coordinates (two fields with longitude - x, and latitude - y; or three fields with the previous presented and elevation - z). In addition, temporal coordinates could be added to geospatial data by referring, for instance, the date of the event. eastings and northings.

---

## **Crowdsourcing**

Crowdsourcing is “the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers” (Source: Merriam-Webster).

## **Currency**

(synonym – timing)

Currency or timing refers to the degree to which a dataset is up-to-date, which could be used as an application-specific measure of temporal accuracy.

## **Data ecosystem**

In this context, a data ecosystem refers to a network of organisations involved in the delivery of data products and/or services through cooperation. A data ecosystem combines data from numerous providers and builds value through the usage of processed data e.g., the statistical-geospatial (data) ecosystem.

## **Data management environment**

A data management environment holistically encompasses the tools, storage and environment for acquiring, validating, storing, protecting and processing required data to ensure the accessibility, reliability and timeliness of the data for its users. (Source: GSGF global)

## **Degree of urbanisation**

The Degree of urbanisation (DEGURBA) is a classification of grid cells and local administrative units (LAUs) based on population number and density as a combination of geographical contiguity and population density, measured by minimum population thresholds applied to 1 km<sup>2</sup> population grid cells.

---

Explanatory text: The basis for the Degree of urbanisation classification is data for equal-size 1km<sup>2</sup> population grid cells avoiding distortions related to multi-sized unit, enabling to detect the presence of individual rural areas, towns and suburbs, or cities and providing more accurate data for the three categories when aggregated to produce regional and national data.

First, the grid cells are classified into three clusters groups according to sum of population, population density and neighbouring cells classification: I) rural grid cells – all grid cells outside the urban clusters/centres; ii) urban clusters (or moderate-density clusters – a cluster of contiguous grid cells that share a common border including grid cells that only touch diagonally at corners with a population density of at least 300 inhabitants per km<sup>2</sup> and a minimum population of at least 5 000 inhabitants; and iii) urban centres (or high-density clusters) - a cluster of non-diagonal contiguous grid cells, excluding those cells with only touching corners, with a population density of at least 1 5000 inhabitants per km<sup>2</sup> and collectively at least 50 000 inhabitants after gap-filling.

Secondly, once all grid cells have been classified and urban centres, urban clusters and rural grid cells identified, these results will be overlayed into local administrative units (LAUs), based on the share of local population living in urban clusters and in urban centres. It classifies LAUs 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). “Urban areas” refers to an aggregation composed of information covering cities as well as towns and suburbs (in other words, densely populated areas and intermediate density areas).

Reference themes: Geographical Classifications

## **Disaggregated statistical- geospatial information**

Disaggregated statistical- geospatial information is created using a spatial distribution model and larger statistical geographies as source data.

Reference themes: Spatial statistics

## **Dissemination geographies**

(Synonyms - output system, output areas)

System of often hierarchically nested geographies with the goal to be either particularly suitable for analysis (functional geographies, grids) or e.g., to present statistics for population and housing census.

Reference themes: Geographical Classifications

---

## **Enterprise architecture (EA)**

Enterprise Architecture is about understanding all of the different elements involved in making up an enterprise and how those elements interrelate. It is an approach to enabling the vision and strategy of an organisation, by providing a clear, cohesive, and achievable picture of what's required to get there. (Source: Statistical Network BA definition)

## **Enumeration geographies**

Enumeration geographies are the spatial representations of the division of a country into areas for census purposes. They represent the smallest area for which in most countries population data and information is available. However, in certain countries enumeration areas are further subdivided into blocks e.g., bounded by physical features such as streets or a river. Enumeration geographies are the spatial support when conducting enumeration procedures through rules of enumeration in order to help users better understand the condition of the data and testing the geographical coverage of data collection.

Reference themes: Geographical Classifications

## **Framework**

A framework is a structure for content or process that can be used as a tool to structure thinking, ensuring consistency and completeness (TOGAF).

## **Functional geographies**

(Synonyms - hybrid geographies)

---

Functional geographies (or hybrid geographies) are defined by characteristics other than their surface area or administrative level presenting themselves as relevant geographies for the collection and dissemination of statistics and attending suitable analysis, spatial planning and policy design purposes through their territorial flexibility feature. In this sense, functional geographies are non-traditional geographies different from the conventional territorial systems. That can be geographical characteristics such as transboundary areas, socio-economic characteristics such as less-favoured areas, areas in need for development, labour market areas, areas by type of economic activity, built-up areas of cities, etc. Functional geographies are increasingly prevalent within the statistical production process and NSIs activities considering the instability of administrative geographies for time-series analysis, the inconsistency of grid population for comparison of population, and ultimately the streamlining of processes with increasingly vague and ambiguously defined geographical entities. In this regard, NSIs are increasingly looking towards other geospatial alternatives for publishing official statistics using a more functional and hybrid approach without neglecting the need to prevent data disclosure by differencing on a wide range of geographical bases.

Reference themes: Geographical Classifications

## **Gazatteer**

Gazetter is a nomenclature service which works as a georeferenced directory of geographic names or name of places, allowing the digital maps associated with a place to be located in the Earth's surface. It is also the oldest method of converting georeferences arising from the need to easily identify and describe locations.

Explanatory text: Gazatteer may be a toponomy database of place' names at national level provided by the NMCA OR other governments agencies. It is also a term commonly given to the index in an Atlas that relates place names to geographic coordinates (latitude and longitude), and to relevant pages in the Atlas where information about a certain place can be found and consulted.

The gazetteer is a useful locator service and usually it works only in one direction as conversion between georeferences – from place name to latitude and longitude. In addition, from a cartographic point of view, a gazetteer is a directory of elements of a type or types of entities containing information relating to their geographic position in which each entity beside including its geographic coordinates it may also include purely descriptive information, using other geographical identifiers such as postal codes or cadastral identifiers.

Gazetteers may be provided by several companies allowing the user to send a place name and to receive in return the coordinates that match the place name requested, or even a place name layer as a more usable format of gazetteer to the user perspective.

---

Gazetteer services can also be implemented following various standards such as the Gazetteer profile of Web Feature Service (WFS) as a gazetteer service is an application profile of the WFS specification providing a web service that allows users to obtain location from a georeferenced vocabulary of name of known geographical entities (rivers, mountains, cities, etc.). Concerning other broader geospatial standards, gazetteers are encompassed in the Open Geospatial Consortium's ISO 19112:2019 (Geographic Information – Spatial referencing by geographic identifiers). This document defines the conceptual schema for spatial references based on geographic identifiers specifying a conceptual scheme for a gazetteer and therefore, enabling gazetteers to be constructed in a consistent and harmonised manner.

## Geocoding

Geocoding is the process of transforming a description of location or unreferenced location information (such as address, a name of a place or other location information) to a location's measurable position on the earth's surface. In other words, geocoding is a way to make data geospatially enable to indicate where the data are in space by linking unreferenced location information associated with a feature through a unique identifier (e.g., housing unit or business) to a set of coordinates within a coordinate system (also referred to as a spatial reference system). These resulting coordinates are the geocode.

Explanatory text: Geocode refers to the spatial representation of descriptive location information such as an address string. Geocodes are, preferably, fine scale geospatially referenced objects that are stored as a geometry data type, such as: location coordinates (i.e. x, y, z coordinates) and/or small area geographies (e.g., mesh blocks, block faces or similar small building block geographies). Larger geographical units, such as enumeration geographies, can be used as geocodes where finer scale geospatial units are not available.

The outputs of a geocoding process are geographical features with attributes, which can be mapped in a GIS environment and location-enabled, and then used for location-based modelling and spatial analysis processes.

In statistics, geocoding is defined as the process of geospatially enabling statistical unit records or other non-spatial data (such as address lists or housing unit records) by creating x- and y- (and potentially z-) coordinates and linking them to each statistical record.

Once geocoding is performed on individual statistical unit records, they (or the associated data) can be aggregated into larger geographical units (e.g., states, provinces, or municipalities) for the purpose of geostatistical analysis. The records are ready for further applications such as methodologies to ensure confidentiality and avoid data disclosure.

The condition for good quality of geocoding is proper physical address, property or building identifiers, or other location descriptions, in order to assign accurate coordinates and/or a small geographical area to each statistical unit record.

---

Reference themes: Geocoding and georeferencing

## **Geographical differencing**

Geographical 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.

Reference themes: Data integration practices

## **Geographical classifications**

(synonyms - sub-national typologies, regional typologies, territorial typologies)

Geographical classifications are a method to group geographies according to 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).

## **Geography**

Geography is a division of space into smaller units, typically areas.

Explanatory text: Geography also have played a long tradition and strategic role as a fundamental component within the statistical production process By using geogrphical classification for processing and disseminating statistical data and as a supportive field to achieve greater geographic granurality of statistical data and ensure their comparability through consistent and common geographies.

## **Georeferencing**

(synonym - geospatial referencing)



---

Georeferencing is generally defined as the process of associating statistical unit records (or other non-spatial data) to a specific location in space for use in geospatial analysis and consists of a set of broad processes that include geocoding. Georeferencing is the process of referencing data against a known geospatial coordinate system, by matching to known points of reference in the coordinate system e.g. image rectification to survey points or addresses linked to parcel centroids, so that the data can be viewed, processed, queried and analysed with other geographical data. Georeferencing can be defined shortly as “relating data with where the fact happens” (GSBPM document).

Reference themes: Geocoding and georeferencing

## **Geosampling**

Sampling that takes measures of spatial autocorrelation, spatial variation principles and spatial relationships and properties of data into consideration and is also supported by a geospatial-based sampling design and frame.

Explanatory text: This geospatial-based sampling framework is usually divided into the spatial or geometrical characteristics of each individual observation and/or the spatial coverage of the sample. The geosampling framework should encompass three important components: I) sampling scheme; ii) sampling density; and iii) sample size. The sampling scheme refers to the spatial pattern of the sample observations - including square grid as a systematic sampling scheme, but also simple random sampling and stratified random sampling as other various sampling schemes used for spatial statistics. The selection of a sampling scheme depends on the situation and purpose of an application. The sampling density refers to the number of observations per unit area. The sample size is the total number of observations. These three fundamental components of any geosampling framework define its spatial coverage which is the set of distance and direction vectors between the observations of the sample. In addition, the sampling scheme and sample size influence the assessment results in a statistically valid manner.

## **Geospatial core information**

A set of essential geospatial data. Examples are administrative boundaries, cadastral parcels, land-cover information, digital elevation model (DEM), address points, outlines of buildings, ortho-photos and satellite images, transport and hydrographic networks, grid systems.

Reference themes: Core Concepts

---

## Geospatial services

Services for geocoding and reverse geocoding of data, background map services, and routing could be considered as essential.

Reference themes: Core Concepts

## Geospatial data quality

Geospatial data quality is defined based on the assumption that there is geographical truth to compare with a dataset, and assessed through data quality components that describe and measure uncertainty in geospatial data.

Explanatory text: The core principle is that the closer a geospatial dataset is to the truth, the higher its quality. Geospatial data quality also implies characteristics of both error and accuracy regarding different dimensions of geospatial data, including space, time and attributes. The concern for geospatial data quality issues is clearly expressed in the development of data transfer and metadata standards and systems as quality parameters, at national and international level. Quality standards are important to set how accurate, precise and complete data could really be while trying to capture real-world phenomena and setting the acceptable difference between what is allocated in the dataset and what actually exists. Geospatial data quality is even more relevant in the absence of an authoritative data source in which uncertainty questions become even more central and a requirement.

## Geospatial data uncertainty

Geospatial data uncertainty refers to a subfield of Geographic Information Science (GIScience) and is used as an umbrella term to encompass data quality, accuracy, error, vagueness, fuzziness, and imprecision. In this regard, it is a term closely related to data quality.

Explanatory text: Over the past three decades, the geospatial community has addressed the issue of geospatial data uncertainty in varied and meaningful ways, from semantics, modelling to visualisation. Uncertainty has a pervasiveness feature since it is introduced and propagated at every step of the geographical knowledge's production and spatial analysis process – from conceptualisation and generalisation to measurement and analysis. Geospatial data uncertainty may be studied as spatial uncertainty that is also intertwined with other related issues, such as scale as a fundamental characteristic of any geospatial data. This is an extremely relevant issue since all geospatial data are scale and context-dependent and therefore, scale – cartographic scale, process scale and spatial extent – has impacts on spatial analysis, representation, results and interpretations and affects geospatial data accuracy, error and associated uncertainty. In the end, the treatment of uncertainty in geospatial data is a prerequisite for further processes related to data, such as aggregation, spatial analysis, representation and visualisation.

---

## Geospatial information

(synonyms - spatial information, geographical information, also geospatial data)

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 may include observations of people, and the social and economic outcomes from human activity, especially concerning their spatial distribution and spatial relationships. Geospatial information is stored in a geographical referencing system, usually a coordinate system of longitude, latitude and, increasingly, elevation.

Reference themes: Core Concepts

## Geospatial object

(synonyms - spatial object, spatial feature)

When a statistical unit is linked to location, it becomes a geospatial object which has a direct link to geospatial terminology and geospatial elements. Geospatial object means an abstract representation of a real-world phenomenon related to a specific location or geographical area. Geospatial objects are represented with their geometry (point, line, polygon).

Reference themes: Core Concepts

## Geospatial reference dataset

(synonyms - spatial reference dataset for statistics)

The geospatial reference dataset contain 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.

Explanatory text: The components of the geospatial reference dataset are not statistics though. The following categories of geospatial information form the geospatial reference dataset:

Topographic data

- Detailed transport networks including public transport stops
- Hydrographic network

---

## Geographies

- Administrative geographies (areas, regions)
- Statistical geographies (areas, regions)
- Addresses, buildings, dwellings
- Land parcels (agriculture and estate)
- Cadastral parcels
- Postal code geographies

Reference themes: Core Concepts

## 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.

Reference themes: Core Concepts

## Geospatial statistics

(Synonyms - geospatially enabled statistics, location-based statistics, geographically referenced statistics, location-enabled statistics, small area statistics)

Location or geospatial-extent are the main characteristics of geospatial statistics. Spatial statistics are 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 of functional perspective might be another important factor to define statistics as geospatial.

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

Geospatial 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? Geospatial statistics is also used to measure and monitor spatial SDG indicators.

Reference themes: Spatial statistics

---

## Grid

Grid is a network composed of two or more sets of curves in which the members of each set intersect the members of the other sets in an algorithmic way. (Source: [https://isotc211.geolexica.org/concepts/ ISO 19123:2005](https://isotc211.geolexica.org/concepts/ISO_19123:2005))

### Grid statistics

Grid statistics (or gridded statistics) are spatial statistics aggregated to equal-sized grid cells in which each grid cell carries a unique code. Ideally the code carries also coordinates of the lower left corner of the grid cell. The grid cell is also referred to as the spatial support, a concept in Geospatial statistics referring to the area over which a variable is measured or predicted.

Examples of grid statistics, especially concerning population densities exist from the late 18<sup>th</sup> century when this approach was adopted in statistical atlases, however their implementation over very large areas in a harmonised way is relatively recent due to the introduction of new technologies and increased computing power.

Grid statistics are an alternative way to publish data, opposite to the traditional hierarchical system of administrative units ranging from local administrative units (LAUs) through regions and countries to supranational aggregates using geographical classifications, commonly used for reporting official statistics. Despite the fact that this system provides high-quality data from small areas and meets specific requirements from territorial authorities at each level of the hierarchy, it is not so suitable for modelling and representing some socioeconomic and environmental events which are often independent of administrative boundaries.

Grid cells all have the same size, allowing for easy comparison, and are stable over time, ensuring consistency in time-series data and temporal combined data. Grid cells also integrate easily with other types of data. Grid systems can be defined to form areas that match a specific purpose or study, which make this system fit-to-purpose for users and grid statistics are produced based on a grid system. Depending on the spatial resolution (size of the grid cell), grid statistics ensure a more accurate representation of reality, and that is evident in real spatial distribution of the population.

Reference themes: Spatial statistics

---

## Information Architecture (IA)

Information Architecture (IA) classifies the information and knowledge assets gathered, produced and used within the Business Architecture. It also describes the information standards and frameworks that underpin the statistical information. IA facilitates discoverability and accessibility, leading to greater reuse and sharing. Source: Statistical Network BA definition

## Integration of statistical and geospatial information

Integration of statistical and geospatial information is a process of using geospatial information for the production and dissemination of statistics. Integration 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:

- Geocode statistical information at unit-record level;
- Processing of statistical information using spatial analysis techniques with the purpose of selecting information or deriving new information with a focus on their spatial characteristics, e.g., buffering around spatial features;
- Supporting a more efficient and flexible statistical production process with geospatial information e.g., for surveying and sampling, field operation;
- Combination of statistical products with geospatial information in statistical maps;
- 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 with 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.

Reference themes: Data integration practices

## Interoperability

Interoperability is a capability to communicate, execute programs or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units. (Source: ISO/IEC 2382:2009, 2121317)

---

Explanatory text: Interoperability refers to ability of information systems to operate in conjunction with each other encompassing communication and semantic protocols, hardware, software, application and data compatibility layers. (Source: Glossary – ANVIL: A Network Virtual Interoperability Laboratory). It is also the ability of two or more systems to exchange geospatial information and to make mutual use of the information that has been exchanged through the application of open standards (Source: Bishr, 1997). In the geospatial community, the term of interoperability includes the technical, semantic and organisational dimensions, concerning the applicability of interoperable statistical data application between technical operators and legal actors, ensuring data privacy, transparency and reuse aspects and working towards an explicit common standard and good practices.

Reference themes: Core concepts

## Infrastructure for geospatial information

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

Reference themes: Core concepts

## Lineage

Lineage refers to source materials, methods of derivation and data transformations applied to a dataset. Lineage aims to be precise enough to identify the sources of individual features or records in the dataset, for instance, if a dataset was derived from different sources, lineage information is useful to track the source and document the feature evolution.

## Linking

(related term - linked data)

Linking defines a process of connecting structured heterogeneous and semantically integrated data sources using a system of unique identifiers and/or ontology network to model information of diverse datasets. Linking builds upon standard Web technologies such as [HTTP](#), W3C's [RDF \(Data Cube Vocabulary\)](#) and [URIs](#).

---

Exploratory text: 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 the subject is. Meanwhile, linking also entails various issues related to datasets heterogeneity such as multithemed, multiresolution, multitemporal, multilingual and multiformat datasets. Linked data facilitates the reuse and connection of multiple and diverse statistical and geospatial data sources, overcoming current integration burdens and problems of data silos associated with these information domains, and offering a great opportunity for data integration. In this regard, the proliferation of initiatives related to linked data, especially oriented to Data Cloud and exposing, sharing, and integration data on the Web should be promoted towards a collective understanding of geospatially enabled statistics to enrich data analysis, connecting global issues to local ones.

Reference themes: Data integration practices

## Locality

Locality is a term used to define different things (eg. place, spot, district) and assumptions should not be made about what the term means in any given usage. An increasingly important use of the term in official statistics is in connection with the census. For census purposes, a locality is defined as a distinct population cluster (also designated as inhabited place, populated centre, settlement and so forth) in which the inhabitants live in neighbouring sets of living quarters and that has a name or a locally recognised status. (Source: OECD glossary of statistical terms)

Reference themes: Geographical Classifications

## Location

Location is a general term to describe a place on or near the surface of the Earth.

Reference themes: Core concepts

## Location information

Location information can include addresses, property or building identifiers, as well as other location descriptions, such as enumeration geographies and other standardised (e.g., what3words reference) and non-standardised (e.g., village names) textual descriptions of a location.

Reference themes: Geocoding and georeferencing



---

## Metadata

Metadata are data about data, that define and describe and provides information about other data.

## Microdata

Microdata are data on an individual object obtained e.g., from sample surveys, censuses and administrative systems.

## National Geospatial Information Agency (NGIA)

National body whose task is mainly to ensure availability and access to authoritative geospatial information, mostly government geospatial data to accomplish efficient and effective implementation and management of geospatial information. In general, the development, production and dissemination of geospatial information promoted by these reference national authorities aim to support the territorial development, policy implementation and evidence-based decision-making in an open, transparent and informed manner. Furthermore, NGIAs should ensure available geospatial data for policy-making on any geographical level responding to the need for more detailed data with a higher geographical resolution to help design policy responses with a territorial aspect.

## National Spatial Data Infrastructure (N)SDI

(also Spatial Data Infrastructure (SDI), Geospatial infrastructure)

A (National) Spatial Data Infrastructure is the technology, policies, standards, good practices, and human resources necessary to acquire, process, store, distribute and improve utilisation of geospatial data. A successful (N)SDI implementation addresses the following considerations:

- Maintenance of data and systems;
- Redundancies should be built into the dissemination solution to prevent a single point of failure;
- Final review and pre-processing before release (data disclosure and confidentiality) to prevent disclosure issues; and,
- Generalisation and thinning of spatial data should be implemented to ensure that the data meets the minimum level of quality and are useable at defined scaled supporting both large and small-scale needs. This can impact both cartographic and data storage issues.

---

## Open data

Data to be freely accessed, used, modified, integrated and shared for any purpose.

## Open data licence

Open data licence is an agreement to allow users to publish, provide, and use data freely as open data.

## Open standard

Open Standards are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus driven process. Open Standards facilitate interoperability and data exchange among different products or services and are intended for widespread adoption. (Source: [ITU](#))

## Operating environment

Here, an operating environment is a sum or a collection of all internal and external factors, which have a direct or indirect bearing on the functioning of organisations, such as economic, social, political, legal and technical factors.

## Population grid

A population grid is composed of a set of equally-sized cells overlaying a particular territory, for which information is collected relating to the number of inhabitants. In the European context, 1 km<sup>2</sup> grid cells are used representing a good compromise between analytical capacity and data protection concerns. These grids are a powerful tool to describe the spatial distribution of population (or other socioeconomic phenomena that are independent of administrative boundaries) which may be used to analyse the interrelationships between human activities and the environment.

Population grids are stable over time, unlike administrative boundaries or other political or legal-based geographies, and may be used for spatial aggregations to various territories of interest. There are three methodological solutions foreseen for establishing the total number of inhabitants living in each of these 1 km<sup>2</sup> grid cells: i) aggregation method; ii) disaggregation method; and iii) hybrid method.

---

## Precision

Precision refers to the degree of measurement of an error and to the amount of detail that can be discerned, not having the same connotation and technical definition as 'Accuracy' but inversely related to this concept.

Explanatory text: It also refers to the level of detail in the reporting of a measurement. It is also known as granularity or resolution within the geospatial community, including GIS and related fields, as a conceptual alternative to avoid misinterpretations with the statistical concept of precision as an observational variance. Data precision or data resolution is limited because no measurement system is infinitely precise in which most datasets are intentionally and inevitably generalised, including processes of data elimination, merging, reduction in detail, smoothing and aggregation to construct a simplified scenario of real-world phenomena with a small fraction of the attributes and their relationships. Spatial resolution and temporal resolution are well-known and widely used concepts concerning precision.

## Primary Sampling Units

Primary Sampling Units (SPUs) are a sub-division of the population often based on geographical criteria, in which firstly the PSUs are selected and then the individuals in those units. This approach allows the collection to be spatially concentrated reducing costs when the survey is conducted face-to-face. It is also appropriate for neighborhood studies.

## Punctuality

Punctuality refers to the possible time lag existing between the actual delivery date of data and the target date when it should have been delivered, for instance, with reference to dates announced in some official release calendar or previously agreed among partners (Source: OECD Glossary of Statistical Terms)

## Quality Dimensions

Quality Dimensions refers to relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity, and are used to describe and measure uncertainties in the data.

---

## Reference Architecture

A Reference Architecture is a generic architecture that provides guidelines and options for making decisions in the development of more specific architectures and the implementation of solutions. A reference architecture can be at any point of the architecture continuum. (The Open Group)

## Region

Regions refers to a geographical area or boundary of any type: for example Suburb, Local Government Area, Local Administrative Unit, Statistical Areas at various sub-national levels, The term of region also has two distinct meanings in a statistical context: first, refers to a geographical area at a subnational level based on NUTS; and second, a region may also refer to a supranational level, as in a region of the world (for example, Latin America, South-East Asia or Europe).

Reference themes: Geographical Classifications

## Regional statistics

(Synonym sub-national statistics)

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

Reference themes: Spatial statistics

## Relevance

Relevance refers to the degree to which data and information meet current and potential needs of the users.

## Reuse

Reuse means common use of a single implementation of a service, with only one organization acting as the service provider (the one who runs the service). (Source: CSPA)

---

## Scanner data

Scanner data are the data recorded by the retailers when consumers make purchases. They include, for each article sold in a store on a given day, the quantity and the sales price of articles sold. (Source: [Insee](#))

## Service

Service is a representation of a real-world business activity with a specified outcome. It is self-contained and can be reused by a number of business processes (either within or across statistical organizations). A Statistical Service will perform one or more tasks in the statistical process. Statistical Services will be at different levels of granularity. (Source: CSPA)

## Sharing

Sharing means exchanging concepts, designs or software, where each user of a service creates and operates its own implementation of that service. (Source: CSPA)

## Spatial analysis

(synonym - location analytics)

Spatial analysis is 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 that can be used for example for monitoring SDGs.

## Spatial extent

Spatial extent refers to the geographical boundaries of a study area and influences the level of detail that can be represented, influencing the scale at which a geographical phenomenon can be observed.

---

## Spatial Resolution

Spatial resolution refers to the minimum data unit size or smallest feature that can be distinguished in which in geographical terms is addressed as minimum mapping unit size depending on map scale.

Explanatory text: This concept is also well-developed in the field of remote sensing referring to the pixel size (ground area fixed) within a digital environment. The concept of resolution is also closely related to scale. Spatial resolution or spatial granularity are relevant quality components and data features since imprecision in geospatial data arise from them according to the level of detail the observations are made, leading to limitations imposed by computational representations, processing and further interpretations.

## 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 larger areas. Demarcations of spatial units can follow physical demarcations as national border lines to support the creation and management of basic administrative units.

Reference themes: Geographical Classifications

## Statistical area

Spatial units used for statistical purposes, including data collection and dissemination of statistics, ultimately to support the statistical production systems, related processes and activities.

Reference themes: Geographical Classifications

---

## 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.

Reference themes: Geographical Classifications

## Statistical-geospatial framework

(synonyms - 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 entire 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.

Explanatory text: A statistical-geospatial framework in a basic form it may be understood as two separate frameworks, a geospatial framework and a statistical framework that are joined together technically, semantically and structurally at defined interfaces.

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 and models on how to integrate statistical and spatial data sources, generate new information through a number of 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-geospatial 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, data protection and confidentiality.

Examples of frameworks:

### Global Statistical Geospatial Framework

The Global Statistical Geospatial Framework (GSGF) provides a common method for geospatially enabling statistical and administrative data, as well as ensuring that this data can be integrated with geospatial information.

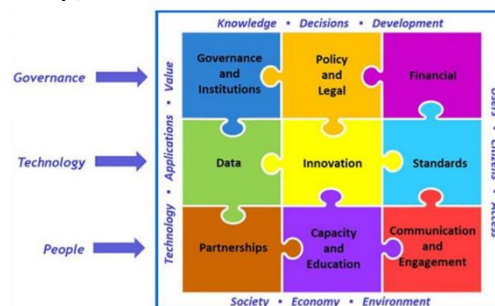


The 5 principles of the GSGF set a foundation to put in place common language, processes, standards and methods across both statistical and geospatial communities.

### Integrated Geospatial Information Framework

The Integrated Geospatial Information Framework (IGIF) provides a basis and guide for developing, integrating, strengthening and maximizing geospatial information management and related resources in all countries. It will assist countries in bridging the geospatial digital divide, securing socio-economic prosperity, and to leave no one behind.

The IGIF comprises three parts as separate, but connected, documents: Part 1 is an Overarching Strategic Framework; Part 2 is an Implementation Guide; and Part 3 is a Country-level Action Plan. The three parts comprise a comprehensive Integrated Geospatial Information Framework that serve a country's needs in addressing economic, social and environmental factors; which depend on location information in a continually changing world. The Implementation Guide communicates to the user what is needed to establish, implement, strengthen, improve, and/or maintain a national geospatial information management system and capability.





---

The IGIF focuses on location information that is integrated with any other meaningful data to solve societal and environmental problems, acts as a catalyst for economic growth and opportunity, and to understand and take benefit from a nation's development priorities and the Sustainable Development Goals.

Reference themes: Frameworks

## Statistical unit

A statistical unit describes the object of interest being studied/analysed. This can for instance include persons, households, businesses, buildings or parcels/units of land.

Explanatory text: GSIM describes unit in a Business Process as follows: <https://statswiki.unece.org/display/clickablegsim/Unit>.

Reference themes: Geographical Classifications

## Survey data

Survey data is defined as output data from a survey used as a research method for collecting data from a pre-defined group of respondents to gain information and insights into various topics of interest.

Explanatory text: In general, survey data could be obtained and request manually and electronically to citizens and organisations by direct interview (face-to-face), by filling in the questionnaires (with or without the support of the interviewer) and by telephone or Internet. Survey data also comes from Primary Data Sources, such as Census and Social Surveys, and unlike Secondary Data Sources (e.g., administrative data) and Big Data sources, are designed to be used in statistical production, including concepts, definitions and classification that are state and know and having a well-established body of knowledge, mostly related to statistical validation. Survey data are usually structured, the interest variables are directly available and do not need "heavy" pre-processing to be used in statistical production.

## Temporal Resolution

Temporal resolution refers to the minimum duration of an event that is discernible and it is related to the duration of the recording interval and the rate of change in the event which will affect data variability. In general terms, a shorter event requires higher temporal resolution to achieve higher accuracy in which spatial and thematic resolutions add new complexities due to interactions changes.

---

## Thematic geospatial information

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

Explanatory text: 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);

Reference themes: Core concepts

## Timeliness

Timeliness refers to the period between the availability of the information and the event or phenomenon it describes.

Explanatory text: It is also related to the term “Currency” on assessing the level of currentness of the dataset and further information disseminated, including by the users’ perspective, in order to have a timelier view of social, economic and environmental facts

## Web scraping

Web scraping is a technique to extract data from web sites. Typically, automated processes are implemented but web scraping can also be done manually.

## 3. References

The terms and definitions in the document are based on revised terms from Terminology from efgs.info. Other reference materials and information sources which were used for a revision, update and new definitions of terms are added to appropriate terms.