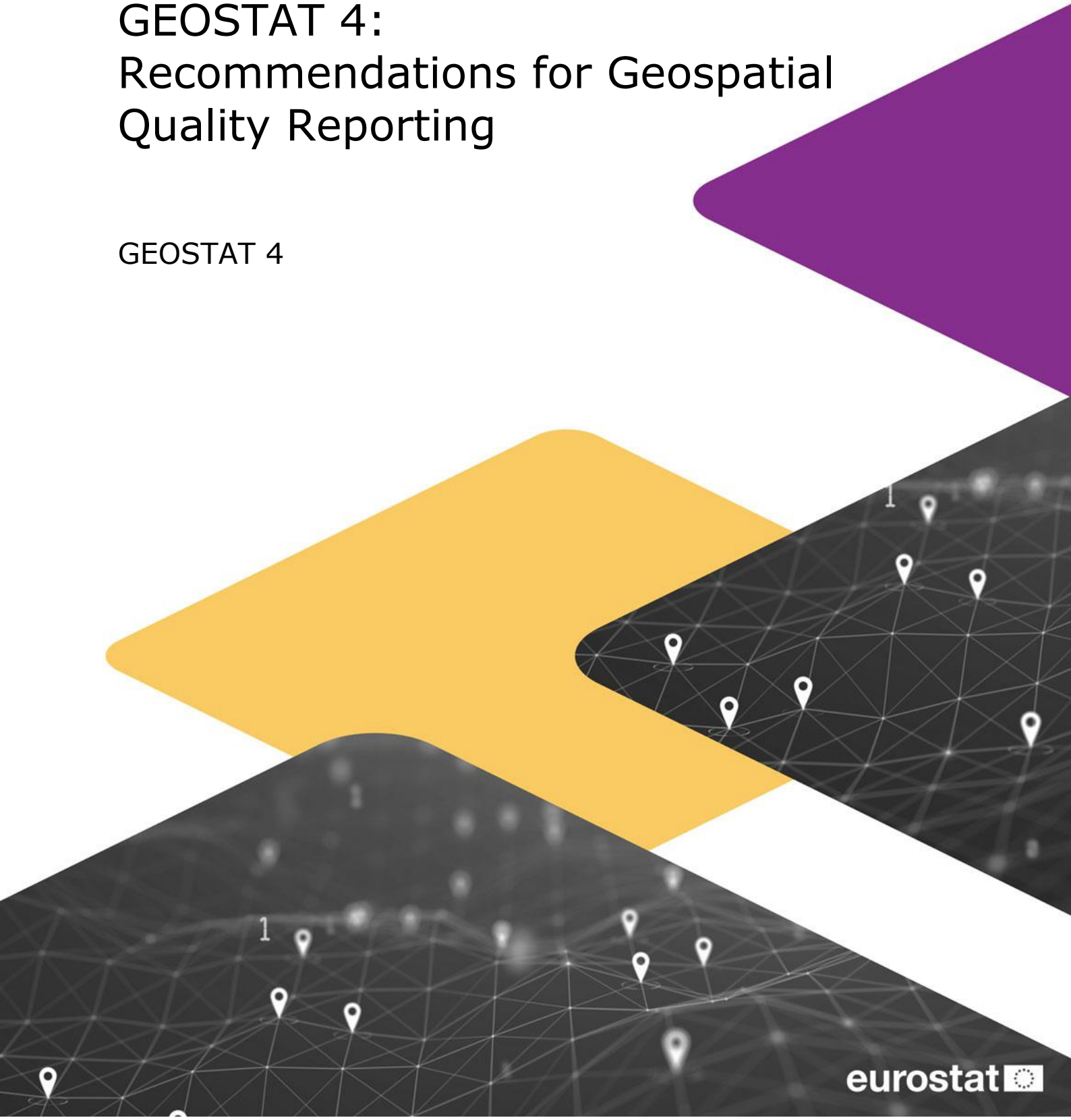
GeoStat

GEOSTAT 4: Recommendations for Geospatial Quality Reporting

GEOSTAT 4



eurostat

Title: GEOSTAT 4: Recommendations for Geospatial Quality Reporting

Author: GEOSTAT 4

Project: Eurostat ESSnet grant project GEOSTAT 4

Grant agreement number: 945503 - 2019-FI-GEOSTAT4

1. Introduction

The integration of statistical and geospatial information can contribute to all dimensions of output quality and must consider all phases within the statistical process. Its significant impact on relevance and comparability, for example, and its importance on sampling, collection and dissemination activities, within the scope of statistical production, is clear.

Assessing the quality of inputs and outputs of the geospatial and statistical data used across the GSGF is fundamental to its implementation, as well as to determine possible improvements to address specific quality concerns within the statistical business process. Likewise, considering and evaluating the quality of geospatial data and processes specifically is vital for understanding its impact on the quality of geospatially enabled statistical products.

Describing and reporting the quality-related components of geospatial data and processes provides the necessary information to evaluate compliance with potential requirements and needs – from both the perspective of producers and users. Furthermore, complete and comprehensive descriptions of the dataset quality will encourage the sharing and use of data and contribute to the assurance of the reliability, objectivity and transparency of the processes and outputs as some of the principles and guiding values in European statistics[1].

For reporting purposes, it is important to guarantee comparability between assessment methods and criteria, through specific quality standards and guidelines, while keeping in mind the difference between the perspectives of producers and users regarding data and process quality.

[1] The Regulation (EC) No 223/2009 as amended by the Regulation (EC) 2015/759 on European Statistics establishes the principles and institutions guiding the development, production and dissemination of European Statistics in which follows the principles expressed in the Article 338 of the Treaty on the functioning of the European Union (UE). Those principles – the principles of professional independence, impartiality, objectivity, reliability, statistical confidentiality, transparency and cost effectiveness – are enumerated in the regulation as well as the guiding values in European statistics and are further elaborated in the European Statistical Code of Practice (CoP).
The amendment of the Regulation (EC) No 223/2009 recognised some weaknesses regarding to the statistical quality-management framework aiming to address those weaknesses towards a robust quality management for European Statistics, in particular by making public Eurostat's assessment of the quality of national contributions to European Statistics as well as drafting and publishing quality assessment frameworks applicable to confidential data based on transparency principles.

This document aims to establish guidelines for the correct and standard assessment and reporting of quality aspects related to geospatial data and its integration within the production of official statistics. Recommendations are proposed on what could be included in a quality report and which indicators are suitable for this purpose.

In chapter 3 a list of proposals for quality indicators is presented, as well as their relationship to GSGF: Europe Principles, GSBPM phases, and quality dimensions in the context of the ESS and ISO 19157 (quality of geospatial dataset). In chapter 5, a list of recommendations for geospatial data and processing quality reporting is described, within the scope of existing Single Integrated Metadata Structure (SIMS) concepts and sub-concepts. Both of these exercises should provide enough information and insight to set up a comprehensive quality reporting system, in terms of geospatial statistics, including geospatial input data, processing and product perspectives.

2. Quality reports in the European Statistical System (ESS)

Quality reports have a long tradition in the ESS. The basic intention is to provide so called metadata related to a comprehensive quality framework about a certain statistical product. When referring to frameworks in the ESS, we rely on the following elements:

The European Statistics Code of Practice (CoP)

The CoP sets the framework of quality and standards for developing, producing and disseminating European statistics, along the lines of the institutional environment, statistical processes and statistical output. It sets out the 16 key principles and a set of 84 indicators of best practice to evaluate their implementation by all the European and national institutions dealing with statistics (NSIs and other national authorities). Since its last version (2017), the CoP includes the Quality Declaration of the ESS, establishing the commitment of the ESS as a whole to the production and dissemination of independent high quality statistical information at European, national and regional levels.

The ESS Quality Assurance Framework (QAF)

The QAF complements and breaks further down the Code of Practice (CoP), identifying possible methods, tools and good practices that can provide guidance and evidence for the implementation of the CoP, at process and institutional levels.

For the purpose of providing statistics by the ESS, quality criteria are formulated in Regulation (EC) No 223/2009[2] of the European Parliament and of the Council on European Statistics amended by Regulation 2015/759 and duly aligned with general quality management principles from ISO 9001:2015 and the EFQM Excellence Model. In Article 12 of the regulation, Quality in European Statistics is defined by the following criteria:

(a) 'relevance', which refers to the degree to which statistics meet current and potential needs of the users;

(b) 'accuracy', which refers to the closeness of estimates to the unknown true values;

(c) 'timeliness', which refers to the period between the availability of the information and the event or phenomenon it describes;

(d) 'punctuality', which refers to the delay between the date of the release of the data and the target date (the date by which the data should have been delivered);

(e) 'accessibility' and 'clarity', which refer to the conditions and modalities by which users can obtain, use and interpret data;

(f) 'comparability', which 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;

(g) 'coherence', which refers to the adequacy of the data to be reliably combined in different ways and for various uses.

The implementation of the ESS Quality Framework is further underpinned by statistical regulations and facilitated by quality related standards, guidelines, methods and tools, providing particular focus on quality assessment and reporting activities.

Quality reporting in the ESS

Quality reporting in the ESS is based on certain concepts developed in consistency with the quality definition provided above. The basic structure is the so called Single Integrated Metadata Structure (SIMS).

[2] This Regulation provides the birth of the European Statistical System (ESS)

There are two types of reports distinguished according to purpose. The first one, a user-oriented quality report intends to inform users about the quality of a certain statistical product. The second one, a so-called producer-oriented quality report is intended as an internal tool in order to constitute quality information of certain processes which are observed over time. It is worth mentioning that the distinction need not to be seen as a mandatory non-deliverable in the sense that a producer-oriented quality report cannot be made available to users if feasible. Within the ESS, both concepts are anchored in SIMS (v2.0) by ESQRS (producer-oriented) and ESMS (user-oriented).

3. List of geospatial quality indicators

This chapter summarises the results on quality indicators related to geospatial aspects (processes, outputs) of work package 3 of GEOSTAT 4. One of the main objectives of this task was to identify quality indicators as possible elements of quality reporting and to relate the indicators to principles of GSGF. The presented indicators are the result of proposals of the member states of the consortium involved in WP 3.

The following Table 1 presents a list of indicators which have been discussed, tested and agreed upon. There is information regarding their definition's description, computation method, and purpose (benchmarking, internal or both). The benchmarking concerns a broader application of the indicator, enabling comparison between organisations and countries, while internal addresses a more restricted purpose related to organisational management.

Further on, the tables in chapter 4 show the assumed relations of the proposed indicators to the following breakdowns:

- Principles of GSGF
- Phases of GSBPM
- Quality dimensions as provided in the ESS Quality Framework
- Geospatial quality aspects as provided in ISO 19157 (quality of geospatial dataset)

Table 1: List of quality indicators (in alphabetical order)

Indicator (Usage)	Definition	Computation
ACCGeopr (benchmarking)	**Access to geospatial products** The indicator ACCGeopr measures the development of the number of accesses over time for a certain geospatial statistics product (can be a single chart, an atlas, an application bound to a certain statistical product, etc..) on the web.	Let u be a certain time unit, where u can be • calendar year • quarter • month Let n_u be the number of accesses in a certain time period u and n_{u+1} be the number of accesses during a subsequent time period u+1. then the indicator is defined as: $ACCGeopr = \frac{n_{u+1} - n_u}{n_u}$
CohereCheck (internal)	**Check of coherences** The indicator CohereCheck measures inconsistent and poorly defined geometric contours and is required to ensure data aggregation validation. It counts the ratio of cases where polygons are not strictly included in larger polygons.	Let S_12 be the surface of the intersection between 2 polygons, P1 of surface S_1 and P2 of surface S_2 with P1 theoretically included in P2. then the indicator is defined as: $CohereCheck = 1 - \frac{S_{12}}{S_1}$
CommonID (internal)	**Common IDs in registers** The indicator CommonID measures the rate of registers that have unique IDs out of all registers under analysis.	Let A be the number of registers with a unique ID and let B be total number of registers under analysis, then the indicator is defined as: $CommonID = \frac{A}{B} * 100\ \%$

Indicator (Usage)	Definition	Computation
CoorCheck (internal)	**Check of coordinates** The indicator CoorCheck measures the precision/correctness of geographic coordinates. It sets the number of spatial errors, i.e. geographic coordinates outside a specified geographical region (parcel, municipality, county...) in relation to the number of correct coordinates. E.g. businesses having a municipality code not corresponding to the municipality code in the spatial join.	The GIS function spatial join can be used to transfer the ID of the specified region to each coordinate. This result can be used to compare the region ID (e.g. municipality code) in a register with the region ID received by the spatial join. Let N be the total number of records and n be the number of records with no match, then the indicator is defined as: $CoorCheck = \frac{n}{N} * 100\ \%$ (0 ≤ CoorCheck ≤ 100)
GeoAdrReg (benchmarking)	**Availability of an address register** Do you have a national standard for address register information?	Choose one of these answers: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

Indicator (Usage)	Definition	Computation
GeoComplete (benchmarking)	**Completeness - rate of national administrative, statistical and functional geographies (NASFG)** The indicator GeoComplete measures the availability of national administrative, statistical and functional geographies (NASFG). The number of available geographies at certain time is divided by the number of geographies which should be available.	Let G be the number of relevant NASFG and G_{av} the number of available national administrative, statistical and functional geographies then the completeness rate GeoComplete is defined by: $GeoComplete = 100 * \frac{G_{av}}{G}\%$ GeoComplete can be calculated as well on a certain geographic level (for instance NUTS 3)
GeoDocDis (internal, benchmarking)	**Documentation for the dissemination of geospatial statistics** Do you have a standardized or documented method for the dissemination of geospatial statistics? For example, how to handle: - Open data or chargeable - Coordinate/reference system - File format/services for dissemination - Metadata standards - Disclosure standards	Choose one of these answers for each of the elements: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

Indicator (Usage)	Definition	Computation
GeoDocStandard (internal, benchmarking)	**Geometric quality documentation** Are there standards or documentation for the geometric quality of geospatial statistics regarding: • Statistics created with geodata that match statistics created by other registers (for example population by grid match population by register information) • Statistics by regional or local areas that match statistics on national level	Choose one of these answers: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant
GeoGL (internal)	**Georeferencing granularity level** The indicator GeoGL measures the granularity level of statistical units. A quantitative measure, based on 4 levels of geographic granularity, is calculated in order to assess the data file's quality with its geographic granularity. The aggregation method is a weighted summation. The weights are assigned to the level of granularity. A record of higher granularity is characterized by a higher weight.	We assign one of four geographic granularity levels to each unit record according to its geographic precision. The four levels are: a. Geographic coordinates (the most precise one; point-based data) b. Addresses, Building IDs etc. c. Area codes or other geographical classification (the least precise one) d. No geographical information at all Let n_a, n_b , n_c and n_d be the numbers of records characterized by levels a, b, c, d with $n_a+n_b+n_c+n_d=n$ then the indicator GeoGL is defined as: $GeoGL = 3 * \frac{n_a}{n} + 2 * \frac{n_b}{n} + 1 * \frac{n_c}{n} + 0 * \frac{n_d}{n}$ with $0 \leq GeoGL \leq 3$

Indicator (Usage)	Definition	Computation
GeoMC (internal)	**Rate of geographic misclassifications detected during a survey** The indicator GeoMC measures the rate of geographically misclassified survey units detected during a sample survey. The indicator can be calculated, whenever the sample units of a certain survey are geocoded. The indicator observes survey units, that cannot be reached due to wrong geographic coordinates (misplaced geographic objects).	Let N be the number of survey units, n_w be the number of units with wrong geoinformation, n_c be the number of units with correct geoinformation, and $N = n_w+n_c$ Then the indicator GeoMC in % is defined as: $GeoMC= \frac{n_w}{n_c} *100\%$
GeoMethDis (internal, benchmarking)	**Documentation of the dissemination methods of statistics as geospatial-based products** Do you have a standardized or documented method for the dissemination of statistics as geospatial-based products? (i.e. statistics presented on map). For example, how to handle: - Background map - Metadata and other relevant information - Disclosure standards	Choose one of these answers for each of the elements: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

Indicator (Usage)	Definition	Computation
GeoMethStandard (internal, benchmarking)	**Documentation of geospatial methods** Do you have a standardized or documented method for the production process of geospatial statistics? For example, how to handle: - Missing values - Multipolygons - Duplicates and overlaps - Enclaves of administrative regions - Area calculating method: Geodesic or planar - Standard for coordinate/reference system - Changes over time in geodata and/or administrative data - Matching data methods - Relevant controls when transferring between file formats - Coherent changes in output values, for example if the number of examined objects increases, their total area shall also increase.	Choose one of these answers for each of the elements: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

Indicator (Usage)	Definition	Computation
GeoMR (internal)	**Geocoding matching rate** The indicator GeoMR measures the matching process (geocoding) of a statistical input dataset containing indirect positions (typically addresses) to a source containing high-precision geographic references (i.e. geocodes – coordinates or small geographic area codes). Geocoding is the process of linking geocodes to each microdata unit record of the statistical input dataset. The indicator calculates a matching rate. It can be applied to set the target level in the production process.	Let A be a dataset containing n records of indirect positions and let K be a dataset containing geocodes. After a matching process a certain number k of the n records of A has been geocoded. The geocoding matching rate GeoMR in % is defined as $GeoMR = \frac{k}{n} * 100\%$
GeoRegister (benchmarking)	**Availability of a register of national geographies** Do you have a national standard for geodata with regional divisions (for example NUTS, LAU, grids)?	Choose one of these answers: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant
GeoStandards (internal, benchmarking)	**Documentation of geocoding quality** Are there standards for documenting the quality of the geocoding output?	Choose one of these answers: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

Indicator (Usage)	Definition	Computation
GeoTiming (internal)	**Currency/timingof geospatial data** The indicator GeoTiming measures the time gap between the release of geospatial data' new version (see metadata) and the moment when these versions are integrated within the statistical production process and spatial data infrastructure/ data warehouse. This indicator addresses the multiple versions of the geographies implemented within the statistical process, sometimes presenting major changes.	Let T_v be the time when a new version of geospatial data is released (mentioned in the metadata), and let T_i be the time when the new version is integrated in the statistical production process, then the indicator GeoTiming is defined as $GeoTiming = T_i - T_v$
GeoUniqueKey (internal)	**Geospatial objects unique IDs** Are there unique IDs for geospatial objects (building, address, geographies) in the input register? IDs may also be created during the geocoding process. Duplicates and other errors may have impacts on further usage.	Choose one of these answers: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

Indicator (Usage)	Definition	Computation
GIScount (benchmarking)	**Counting the number of georeferenced statistical subjects** The indicator GIScount measures the statistical subjects, that are georeferenced and included in geographic information systems. The statistical subjects available for being included in geographic information systems are counted.	Let S be the number of georeferenced statistical subjects and $GISsub_i$ be the number of subjects in certain geographic information system. Then the indicator GIScount is defined as: $GIScount = \sum_{I=1}^{S} GISsub_i$
OpenDatGeopr (benchmarking)	**Rate of geospatial statistics accessible as open data** The indicator OpenDatGeopr measures the rate of geospatial statistics (products and services) available as open data. The number of geospatial statistics available as open data is divided by the total number of geospatial statistics.	Let P_t be the total number of geospatial statistics of an NSI at a certain time t and POD_t the number of geospatial statistics available in full accordance to the open data principles at time t, then the rate of geospatial statistics accessible as open data at time t is defined as $OpenDatGeopr_t = \frac{POD_t}{P_t}$
StandardDocumentation (internal, bechmarking)	**Documentation of the quality of input geodata** Are there standards or documentation for the geometric quality of the input geodata, regarding: • Overlapping geometries • Duplicates • Missing data • Number of objects that match the number of objects in attribute data and metadata • Geographical location matches attribute value of location (for example NUTS or LAU code)	Choose one of these answers for each of the elements: ❑ Yes= Good ❑ In progress= Acceptable ❑ No= Action required ❑ Not relevant

4. The quality indicators and their relation to GSGF, GSBPM and ISO elements

One objective of quality indicators is to assess processes and outputs by assigning them to the quality criteria laid down in the definition of quality relevant for the ESS. However, there are more standards and frameworks that are relevant for geospatial processes. First of all, the Global Statistical Geospatial Framework[3] (GSGF) as a common high-level framework for the statistical and the geospatial community provides an underlying mechanism for integrating statistical and geospatial information.

The leading role of the GSGF justifies not only assigning the quality indicators to the principles of GSGF, but also considering this framework as a backbone when it comes to cross references to other standards. Consequently, the following tables show which elements of three other important frameworks the indicators were assigned to. The following standards were considered:

- **Quality definition for the ESS**
  As already mentioned in chapter 2 the definition of quality as provided by article 12 of EU regulation 223 sets out the main criteria for quality reporting and the assessment of processes and outputs for official European statistics. When it comes to quality reporting users are familiar with the criteria and the indicators developed by GEOSTAT 4 can be used in relevant concepts of the Single Integrated Metadata Structure (SIMS – see chapter 6) which is also defined as a non-legal binding standard within the ESS.

3 GEOSTAT 4 (2021). GSGF Europe. GEOSTAT 4 and Eurostat
which was based on the original GSGF by UN-GGIM
UNSC and UN-GGIM (2019). The Global Statistical Geospatial Framework. The GSGF.
(https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf)

- **GSBPM**

  The Generic Statistical Business Process Model (GSBPM) is a means to describe statistics production in a general and process-oriented way. It is used both within and between statistical offices as a common basis for work with statistics production in different ways, such as quality, efficiency, standardisation, and process-orientation. GSBPM is a non-legal binding standard within the ESS. GEOSTAT developed a process map where geospatial processes were mapped to GSPBM phases and their subprocesses. The mapping was the fundament to be able to assign the indicators finally to the GSBPM phases. The main objective for relating quality indicators to GSBPM is to make use of them in order to be able to assess process quality.

- **ISO 19157**

  ISO refers to the International Organization for Standardization. This organization develops and publishes worldwide technical, industrial and commercial standards. ISO 19157 establishes the principles for describing the quality of geographic data.

a. Quality indicators assigned to GSGF principles and Quality Dimensions

Table 2: Quality indicators assigned to GSGF principles and Quality Dimensions

	Quality Dimension					
GSGF principles	Relevance	Accuracy	Timeliness and Punctuality	Accessibility and Clarity	Comparability	Coherence
Principle 1	GeoMR	GeoGL CoorCheck StandardDocumentation GeoAdrRegister	GeoTiming			CohereCheck
Principle 2		GeoMC GeoStandards			GeoMethStandard	
Principle 3		GeoDocStandard		GeoRegister	GeoComplete	
Principle 4		GeoMR				CommonID GeoUniqueKey
Principle 5	GIScount			ACCGeop OpenDatGeop GeoMethDis GeoDocDis		

b. Quality indicators assigned to GSGF principles and GSBPM production phases

Table 3: Quality indicators assigned to GSGF principles and GSBPM production phases

	GSBPM production phases			
GSGF principles	4. Collect	5.Process	6. Analyze	7. Disseminate
Principle 1	GeoMR GeoGL CoorCheck CohereCheck StandardDocumentation GeoAdrRegister	GeoTiming		
Principle 2	GeoMC GeoStandards	GeoMethStandard GeoDocStandard		
Principle 3	GeoRegister	GeoComplete		
Principle 4		CommonID GeoUniqueKey		
Principle 5	GIScount			ACCGeop OpenDatGeop GIScount GeoMethDis GeoDocDis

c. Quality indicators assigned to GSGF principles and ISO 19157 elements

Table 4: Quality indicators assigned to GSGF principles and ISO 19157 elements

	ISO 19157 elements					
GSGF principles	Completeness	Logical consistency	Positional accuracy	Thematic accuracy	Temporal quality	Usability
Principle 1	GeoMR	CoorCheck CohereCheck	GeoGL		GeoTiming	StandardDocumentation GeoAdrRegister
Principle 2		GeoMC				GeoStandards GeoMethStandard
Principle 3	GeoComplete		GeoDocStandard			
Principle 4		GeoMR CommonID GeoUniqueKey				
Principle 5						GIScount ACCGeop OpenDatGeop GeoMethDis GeoDocDis

5. Recommendations for quality reporting on geospatial aspects

In this chapter, some concepts and sub-concepts of the Single Integrated Metadata Structure (SIMS) relevant for geospatial outputs and processing are presented according to the European Statistical System (ESS) handbook for quality and metadata reports — 2020 edition[4]. For each concept and sub-concept, the relevance to geospatial content was evaluated and added where found appropriate. An excerpt of the relevant concepts is included in the Annex. It shows:

- The SIMS nomenclature
- The definition of the concept or sub-concept
- The guidelines as they are provided in the handbook
- Geospatial content: Recommendations of GEOSTAT considering which contents should or at least could be reported in the sub-concept when it comes to geospatial outputs or geospatial processing.

Numerous concepts/sub-concepts relevant for geospatial outputs/processing were identified and recommendations given based on the premise that there is some use of geospatial data along the production process of a certain statistical product.

It should be noted that there is no distinction here between concepts and sub-concepts included in producer or user-oriented quality reports. Refer to SIMS v2.0 for information on this aspect.

Each table is followed by short notes about what is generally seen as relevant quality elements for each concept. These notes are included as summary here as follows.

The recommendations for the geospatial contents can be used to highlight the growing importance of geospatial outputs by providing metainformation in a quality report. The modular structure of SIMS allows it not only to enhance the reports which are prescribed for numerous European statistical products but also to extract and produce special tailor-made geospatial reports for internal and external purposes.

[4] "European Statistical System (ESS) handbook for quality and metadata reports — 2020 edition"

S01 Contact

The general intention of S01 is to provide contact information for readers of the quality report in order to be able to have a possibility for any inquiries. Since the geospatial part of product is bound to expertise outside the subject matter units it is recommended to include contact information in the corresponding concepts. As a minimum requirement GIS contact should be provided in the sub concepts S01.03 "Contact name" and S01.4 "Contact person function". It could also be very useful to include a contact person of the National Geospatial Information Agency.

S02 Metadata update

This concept contains information about update procedures of the whole quality report. There are no specific characteristics related to geospatial activities.

S03 Statistical presentation

Statistical presentation refers to the disseminated data which can be displayed to users as tables, graphs or maps etc. It comprises three parts:

- Data description, classification system and sector coverage
- Statistical concepts, definitions, units and populations
- Reference area, time coverage and base period.

How far geospatial aspects deserve to be included into the sub-concepts, depends on the distinction for the geospatial part of the product. Statistical units for instance might not differ for the geo-processing, but if there is only a certain share of variables being processed for geospatial outputs, it should be apparent under this concept.

S04 Unit of measure

There could be a statement that the units of measure have been used for geospatial outputs consistently.

S05 Reference period

There could be a statement that the reference period has been respected for geospatial outputs consistently.

S06 Institutional mandate

Legal acts and relevant agreements for sharing data are important basic elements for producing official statistics in general and especially for data collection. Regarding geospatial activities, it is of interest if there are special legal acts which should be mentioned in a quality report.

S07 Confidentiality

Confidentiality is an important issue when disseminating official statistics. Conveying information about the handling of confidential cells serves the purpose to enhance trust in official statistics of users and respondents. Regarding geospatial outputs, the reporting should focus on the issue in how far confidentiality rules are obeyed in geospatial outputs.

S08 Release policy

The release policy is usually defined at institutional level. There are no specialities regarding geospatial processing and outputs.

S09 Frequency of dissemination

If there is a divergence for the geospatial part of the disseminated outputs, it should be stated here.

S10 Accessibility and clarity

Accessibility and clarity are one of the most significant quality dimensions. In SIMS this is divided in the concepts related to the access environment via various channels (S10.1 to S10.5.) and two concepts (S10.6. and S10.7) which are dedicated to methodical documentation and quality information. Since there is no specific element here for geospatial outputs the sub-concept S10.5 "other" has to be used to provide a summary on the geospatial output channels relevant for the statistical product. For metadata, corresponding issues can be described accordingly in S10.6 and S10.7.

S11 Quality management

In the geospatial community the quality management procedures are laid down in various ISO norms (e.g. ISO 19157). That means that if the quality management system of the statistical authorities also considers such frameworks for geospatial processing, there should be a corresponding reference to that.

S12 Relevance

There are users who are mainly interested in geospatial outputs. Consequently, there might be user needs which are very specific to this part of the output portfolio for a certain statistical product. Given that, the reporting on concept S12 should consider the geospatial part sufficiently.

S13 Accuracy and reliability

Accuracy can be defined as the degree or closeness to which the information on a map matches the values in the real world. In GIS data, accuracy can be referred to a geographic position, but also to an attribute, or conceptual accuracy. Precision refers to how exact the description of data is. Attribute accuracy indicates the attribute attached to the point, line and polygon features of the spatial database, which shows how reliable and reasonably correct or free from bias these are. The reporting should concentrate on the issues of non-sampling errors here mostly related to measurement errors, coverage errors and processing errors.

S14 Timeliness and Punctuality

Timeliness of geospatial output normally depends a lot on the availability of input data. For reporting on quality, geospatial issues should only be mentioned when shortcomings in timeliness for geospatial outputs are caused by reasons located in the geospatial processing. Regarding punctuality, the planning for geospatial outputs is normally embedded in the project plan compiled for the statistical product as a whole.

S15 Coherence and comparability

To be able to compare results is one of the basic elements for users, in order to take full advantage of statistical outputs. Geospatial output presents most of the time regions (mainly related to a national territory) augmented by statistical characteristics. Therefore, for geospatial outputs, information should be available, if there are concerns regarding comparability over regions. Another possible issue is comparability over time, which is relevant when the geospatial presentations include a time component.

Coherence goes mainly to divergences between domains. It seems that this is not very relevant for geospatial outputs.

S16 Cost and burden

Response burden is an issue for surveys and not related to the geospatial part of a statistical product. However, accounting systems of statistical authorities should allow the derivation of the costs relevant for the production of geospatial outputs.

S17 Data revision

Data revisions are either planned ones, following a fixed revision schedule, or conducted on an occasional basis (for instance when substantial errors have to be corrected). Changes in the input data coming from the statistical product (survey for instance) are not related to the geospatial outputs. However, when the revision results in a substantial change of incoming geodata, because the former have been erroneous, a quality report should take note about such an incident.

S18 Statistical processing

The processing for geospatial outputs focuses on linking unit data to geocoded data and assigning geographic characteristics. Finally, there are aggregation processes in order to construct maps and other geospatial output. Considering that the most important reporting concept is S18.5.

Another focus should be laid on the collection of data as well as the validation of certain input sources specific for geographic processes.

S19 Comment

Supplementary descriptive text which can be attached to data or metadata. Anything that did not fit anywhere else.

ANNEX

Recommendations for geospatial quality reporting based on the SIMS concepts

In this annex, some concepts and sub-concepts of the Single Integrated Metadata Structure (SIMS) relevant for geospatial outputs and processing are presented according to the European Statistical System (ESS) handbook for quality and metadata reports — 2020 edition[5]. For each concept and sub-concept, the relevance to geospatial content was evaluated and added where found appropriate.
For each concept and sub-concept, the table shows:

- The SIMS nomenclature
- The definition of the concept or sub-concept
- The guidelines as they are provided in the handbook
- Geospatial content: Recommendations of GEOSTAT considering which contents should or at least could be reported in the sub-concept when it comes to geospatial outputs or geospatial processing.

Numerous concepts/sub-concepts relevant for geospatial outputs/processing were identified and recommendations given based on the premise that there is some use of geospatial data along the production process of a certain statistical product.

It should be noted that there is no distinction here between concepts and sub-concepts included in producer or user-oriented quality reports. Refer to SIMS v2.0 for information on this aspect.

Each table is followed by short notes about what is generally seen as relevant quality elements for each concept.

The recommendations for the geospatial contents can be used to highlight the growing importance of geospatial outputs by providing metainformation in a quality report. The modular structure of SIMS allows it not only to enhance the reports which are prescribed for numerous European statistical products but also to extract and produce special tailor-made geospatial reports for internal and external purposes.

[5] "European Statistical System (ESS) handbook for quality and metadata reports — 2020 edition"

S01 Contact

SIMS	Concept name	Definition	Guidelines	Geospatial content
S01	Contact	Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.	See sub concepts	See sub concepts
S01.1	Contact organisation	The name of the organisation of the contact points for the data or metadata.	Provide the full name (not just code name) of the organisation responsible for the process and outputs (data and metadata) that are the subject of the report.	Normally this is the statistical authority (mainly the NSI). To name an additional contact for the geospatial part makes only sense if the work is done outside the NSI (NGIA – National Geospatial Information Agencies).
S01.2	Contact organisation unit	An addressable subdivision of an organisation.	Provide the full name of the organisational unit responsible. The name can include a unit number.	This is the organisational unit in charge for the product.
S01.3	Contact name	The names of the contact points for the data or metadata.	Provide the first and last names of the contact point(s). If more than one name is provided, the main contact should be indicated. If the author of the report is different from the person(s) responsible for process and its outputs, provide this name also.	This is the name of the contact point in charge for the product.
S01.4	Contact person function	The area of technical responsibility of the contact, such as "methodology", "database management" or "dissemination".	Provide the title(s) and area(s) of responsibility of the person(s) indicated as contact(s), for example Senior Research Assistant, Economics Division	Since the geospatial part is specific a point “Geospatial Information” should appear in this concept.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S01.5	Contact mail address	The postal address of the contact points for the data or metadata.	Provide the postal address(es) of the person(s) indicated as contacts.	This can also include one person responsible for the geospatial part (NSI or NGIA – National Geospatial Information Agency).
S01.6	Contact email address	E-mail address of the contact points for the data or metadata	Provide the email address(es) of the person(s) indicated as contacts. The address(es) can be (an) individual e-mail address(es) or a mailbox in the organisation to which the person(s) has (have) access.	This can also include one person responsible for the geospatial part (NSI or NGIA – National Geospatial Information Agency).
S01.7	Contact phone number	The telephone number of the contact points for the data or metadata.	Provide the telephone number(s) of the person(s) indicated as contacts.	This can also include one person responsible for the geospatial part (NSI or NGIA – National Geospatial Information Agency).
S01.8	Contact fax number	Fax number of the contact points for the data or metadata.	Provide the fax number(s) of the person(s) indicated as contacts.	This can also include one person responsible for the geospatial part (NSI or NGIA – National Geospatial Information Agency).

The general intention of S01 is to provide contact information for readers of the quality report in order to be able to have a possibility for any inquiries. Since the geospatial part of product can be bound to expertise outside the subject matter units it is recommended to include contact information in the corresponding concepts. As a minimum requirement GIS contact should be provided in the sub concepts S01.03 “Contact name” and S01.4 “Contact person function”. It could also be very useful to include a contact person of the National Geospatial Information Agency.

S03 Statistical presentation

SIMS	Concept name	Definition	Guidelines	Geospatial content
S03	Statistical presentation	Description of the disseminated data which can be displayed to users as tables, graphs or maps	See sub-concepts	See sub-concepts
S03.2	Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common.	List all classifications and breakdowns that are used in the data (with their detailed names) and provide links (if publicly available). Explain deviations, if any, from ESS or international standards. *European level:* Provide an overview of national deviations from ESS and/or international standards.	Classifications which are in use for geospatial data presentations for instance in maps, charts etc. should be listed here. It concerns as well the quality of the thematic attributes and of the classifications of features and their relationships as well as the classification correctness.
S03.3	Sector coverage	Main economic or other sectors covered by the statistics.	List the main economic or other sectors covered by the data and the size classes used, for example, size classes based on number of employees. *European level:* Provide a summary of differences in the main economic or other sectors covered by national data and the size classes used.	If feasible, economic sectors, size classes and other important breakdown variables, which are in use for geospatial data presentations for instance in maps charts etc., should be listed here.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S03.4	Statistical concepts and definitions	Statistical characteristics of statistical observations, variables.	Define and describe briefly the main statistical variables that have been observed or derived. Indicate their types. Indicate discrepancies, if any, from the ESS or international standards. Note that any difference between these variables and the variables desired by users is a relevance issue and is discussed in S.12. *European level:* Summarise the national discrepancies from the ESS and/or international standards	There should be a list of variables, which are used for geospatial presentations. For instance • income, • number of schools, • number of employed persons etc. classified by regional entities.
S03.5	Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	Define the type of statistical unit about which data are collected, e.g. enterprise, kind of activity unit, local unit, private household, dwelling, person, import transaction. If there is more than one type of unit, define each type. *European level:* Summarise the differences in units used at national level.	The statistical units are those which are surveyed or enhanced by characteristics in any other data collection process. According to EFGS principle 2 processes of linking and storing high-precision geographic references to each statistical unit should be implemented.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S03.6	Statistical population	The total membership or population or "universe" of a defined class of people, objects or events.	Define the target population of statistical units for which information is sought. Note that a difference between the target population and the ideal population desired by users is a relevance issue and is discussed in S12; and the difference between target population and the survey population is a coverage issue and is discussed in S13.3. If there is more than one type of statistical population, define each type. *European level:* Summarise the differences in statistical populations used at national level	Refers to the statistical product as a whole, so there is normally no additional need for reporting for the geospatial part. Only if certain parts of the population are excluded from geospatial presentations, it should be reported.
S03.7	Reference area	The country or geographic area to which the measured statistical phenomenon relates.	Describe the country, the regions, the districts, or the other geographical aggregates, to which the data refer. Identify any specific exclusions in the data disseminated. If coverage includes overseas territories this should be stated, and they should be specified. *European level:* Describe the geographical area covered by the data disseminated, e.g., EU Member states, EU regions, USA, Japan, or aggregates such as EU, EEA).	Describe to which geographic or other regional entities the geospatial presentations are referring to.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S03.8	Time coverage	The length of time for which data are available.	State the time period(s) covered by the data, e.g. first quarter 2018, or quarters 2015-2018, or year 2018, or years 1985-2018. Note that any issues concerning comparability over time are discussed in S15	If the geospatial presentations contain a time component, it should be reported how it is implemented.
S03.9	Base period	The period of time used as the base of an index number, or to which a constant series refers.	Note that this concept applies only to certain types of outputs, such as indexes. State the base period, for example, year 2000. Indicate base period update time frame and date of next update.	If indices are the subject of geospatial presentations, it should be indicated that the base period is apparent in all geospatial presentations.

Statistical presentation refers to the disseminated data which can be displayed to users as tables, graphs or maps etc. It comprises three parts

- Data description, classification system and sector coverage,
- Statistical concepts, definitions, units and populations,
- Reference area, time coverage and base period.

How far geospatial aspects deserve to be included into the sub-concepts, depends on the distinction for the geospatial part of the product. Statistical units for instance might not differ for the geo-processing, but if there is only a certain share of variables being processed for geospatial outputs, it should be apparent under this concept.

S04 Unit of measure

SIMS	Concept name	Definition	Guidelines	Geospatial content
S04	Unit of measure	The unit in which the data values are measured.	The data usually involves several units of measure depending upon the variables. Examples are: Euro, national currency, number of persons, and rate per 100,000 inhabitants. The magnitude (e.g., thousand, million) of numerical units should be included.	There could be a statement that the units of measure have been used for geospatial outputs consistently.

There could be a statement that the units of measure have been used for geospatial outputs consistently.

S05 Reference period

SIMS	Concept name	Definition	Guidelines	Geospatial content
S05	Reference period	The period of time or point in time to which the measured observation is intended to refer.	The value of a variable refers to a specific time period (for example, the last week of a month, a month, a fiscal year, a calendar year, or several calendar years), or to a point in time (for example, a specific day, or the last day of a month). The variables in a dataset may refer to more than one reference period. All reference periods should be stated. Note that the difference, if any, between the target reference period(s) and the actual reference period(s) is an accuracy issue and should be discussed in S.13.3. Note that if the survey population does not include all the units in the target population for the specified reference period, this is a coverage issue and should be discussed in S.13.3. *European level:* Summarise differences in reference period across countries.	There could be a statement that the reference period has been respected for geospatial outputs consistently.

There could be a statement that the reference period has been respected for geospatial outputs consistently.

S06 Institutional mandate

SIMS	Concept name	Definition	Guidelines	Geospatial content
S06	Institutional mandate	Law, set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics.	See sub-concepts	See sub-concepts
S06.1	Legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics.	State the national legal acts and/or other reporting agreements, including EU legal acts, the implementation of EU directives. *European level* State the legal base or other agreement, for example, the EU legal act, or ESS Five-Year-Program, that underpins the reporting obligations on countries.	Any legal acts demanding and supporting the production and dissemination of geospatial outputs, should be listed here
S06.2	Data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between data producing agencies within the national statistical system.	Any agreements facilitating data sharing (for instance with mapping agencies) relevant for geospatial outputs, should be listed here.

SIMS	Concept name	Definition	Guidelines	Geospatial content
			European level: Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between international data producing agencies, for example, a Eurostat data collection or production that is in common with the OECD or the UN.	

Legal acts and relevant agreements for sharing data are important basic elements for producing official statistics in general and especially for data collection. Regarding geospatial activities, it is of interest if there are special legal acts which should be mentioned in a quality report.

S07 Confidentiality

SIMS	Concept name	Definition	Guidelines	Geospatial content
S07	Confidentiality	A property of data indicating the extent to which their unauthorised disclosure could be prejudicial or harmful to the interest of the source or other relevant parties.	See sub-concepts	See sub-concepts
S07.2	Confidentiality - data treatment	Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	For aggregate outputs: Provide the rules that define a confidential cell; Describe the procedures for detecting confidential cells (primary confidentiality) and checking for residual disclosure (derivation or secondary confidentiality); Describe the procedures for reducing the risk of disclosure by treating confidential cells, for example by perturbation, controlled rounding, cell suppression, or cell aggregation. For micro-level outputs: Describe the procedures that are used in protecting confidentiality. *European level*: Summarise the commonalties and differences in national approaches.	Provide an explanation on how far confidentiality rules are obeyed in geospatial outputs.

Confidentiality is an important issue when disseminating official statistics. Conveying information about the handling of confidential cells serves the purpose to enhance trust in official statistics of users and respondents. Regarding geospatial outputs, the reporting should focus on the issue in how far confidentiality rules are obeyed in geospatial outputs.

S09 Frequency of dissemination

SIMS	Concept name	Definition	Guidelines	Geospatial content
S09	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	State the frequency with which the data are disseminated, e.g. monthly, quarterly, yearly. The frequency can also be expressed by using a code from the harmonised ESS code list so long as this is considered to be easily understandable by users	If there is a divergence for the geospatial part of the disseminated outputs, it should be stated here.

If there is a divergence in the frequency for the geospatial part of the disseminated outputs, it should be stated here.

S10 Accessibility and clarity

SIMS	Concept name	Definition	Guidelines	Geospatial content
S10	Accessibility and clarity	The conditions and modalities by which users can access, use and interpret data.	See sub-concepts	See sub-concepts
S10.4	Micro-data access	Information on whether micro-data are also disseminated	State whether the data are accessible in micro-data form, e.g. for researchers. If so, cross reference the micro-data confidentiality rules in S7.	Depending on confidentiality rules location information might be included and mentioned.
S10.5	Other	References to the most important other data dissemination done.	Describe any other important dissemination mechanisms, for example policy papers, within outputs produced by other statistical processes. Summarise the accessibility and clarity of the data associated with the various dissemination formats, including relevant results from user surveys, and the effects of pricing policies and confidentiality provisions. Describe dissemination of data to Eurostat and other international organisations, and internal dissemination.	A full picture of the geospatial outputs should be provided here. This includes: • Dissemination platforms • Dissemination channels • Dissemination formats • Quality indicators when available (see also ACCGeopr of quality indicator product) Refer to aspects listed in ISO 19157 quality of geospatial dataset – External quality: • Fitness for use: degree of agreement between data characteristics (i.e., internal quality) and the explicit and/or implicit needs of a user for a given application in a given context.

SIMS	Concept name	Definition	Guidelines	Geospatial content
				• Usability: based on user's requirements, all internal quality elements of ISO 19157 may be used to evaluate usability.
S10.6	Documentation on methodology	Descriptive text and references to methodological documents available.	List national reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process.	Provide an overview about documentation of geospatial outputs. Special metadata files related for instance to INSPIRE should be referred to.
S10.7	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	List relevant quality related documents, for example, other quality reports, studies. Cross reference to descriptions of quality procedures in other chapters, especially S.13. *European level:* Summarise availability of national quality reports	Describe to which geographic or other regional entities the geospatial presentations are referring to.

Accessibility and clarity is one of the most significant quality dimensions. In SIMS this is divided into the concepts related to the access environment via various channels (S10.1 to S10.5.) and two concepts (S10.6. and S10.7) which are dedicated to methodical documentation and quality information. Since there is no specific element here for geospatial outputs the sub-concept S10.5 "other" has to be used to provide a summary on the geospatial output channels relevant for the statistical product. For metadata, corresponding issues can be described in S10.6 and S10.7.

S11 Quality management

SIMS	Concept name	Definition	Guidelines	Geospatial content
S11	Quality management	Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.	See sub-concepts	See sub-concepts
S11.1	Quality assurance	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Describe the procedures (such as use of a general quality management system based on EFQM or ISO 9000 series) to promote general quality management principles in the organisation. Describe the quality assurance framework used to implement statistical quality principles. Describe the quality assurance procedures specifically applied to the statistical process for which the report is being prepared, for example training courses, process monitoring, benchmarking, assessments, and use of best practices. Include descriptions of all forms of quality assessment procedures (such as user satisfaction survey, self-assessment, peer review, compliance monitoring, audit, labelling, certification) and when they most recently took place. Describe any ongoing or planned improvements in quality assurance procedures	It should be described what is done for the geospatial part of the product. This could include a reference to: ISO 19157 quality of geospatial dataset (internal and external quality).

SIMS	Concept name	Definition	Guidelines	Geospatial content
S11.2	Quality assessment	Overall assessment of data quality, based on standard quality	Summarise the results of the most recent quality assessments and cross reference to the chapters in the report where the results are presented in more detail.	If there exists any special assessment about the geospatial part, describe it here. This could include a reference to: ISO 19157 quality of geospatial dataset -Internal quality: • Completeness: presence or absence of features, their attributes and relationships (Omission or commission, cf. overcoverage, undercoverage - frame errors in statistics). • Logical consistency: degree of adherence to logical rules of data structure, attribution and relationships (Conceptual consistency, domain consistency, format consistency, topological consistency). • Positional accuracy: accuracy of the position of features within a spatial reference system (Absolute accuracy, relative accuracy, gridded data position accuracy.) • Thematic accuracy: quality of the thematic attributes and of the classifications of features and their relationships, (Classification correctness, non-quantitative attribute correctness, quantitative attribute accuracy.)

SIMS	Concept name	Definition	Guidelines	Geospatial content
				• Temporal quality: quality of the temporal attributes and temporal relationships of features, (Accuracy of time measurement, temporal consistency, temporal validity.) ISO 19157 quality of geospatial dataset - External quality: • Fitness for use: degree of agreement between data characteristics (i.e., internal quality) and the explicit and/or implicit needs of a user for a given application in a given context. • Usability: based on user's requirements, all internal quality elements of ISO 19157 may be used to evaluate usability.

In the geospatial community the quality management procedures are laid down in various ISO norms (e.g. ISO 19157). That means that if the quality management system of the statistical authorities also considers such frameworks for geospatial processing, there should be a corresponding reference to that.

S12 Relevance

SIMS	Concept name	Definition	Guidelines	Geospatial content
S12	Relevance	Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.	See sub-concepts	See sub-concepts
S12.1	User needs	Description of users and their respective needs with respect to the statistical data.	Provide: a classification of users, also indicating their relative importance; an indication of the uses for which users want the statistical outputs; an assessment of the key outputs desired by different categories of users and any shortcomings in outputs for important users; information on unmet user needs and any plans to satisfy them in the future; and details regarding those quality components which do not meet user requirements.	List here: • User groups which have specific interest in geospatial outputs. This could include e.g., Open data communities. • Which kind of geospatial outputs are requested (maps, graphs, charts etc.). • Any unmet specific geospatial user needs. Refer to aspects listed in ISO 19157 quality of geospatial dataset – External quality: • Fitness for use: degree of agreement between data characteristics (i.e., internal quality) and the explicit and/or implicit needs of a user for a given application in a given context. • Usability: based on user's requirements, all internal quality elements of ISO 19157 may be used to evaluate usability.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S12.2	User satisfaction	Measures to determine user satisfaction	Describe how, and how often, the views and opinions of the users are collected, for example by user satisfaction surveys or other user consultations. State how often such investigations are conducted and when the most recent took place. Present the key results from the recent investigations. Present view of user satisfaction over time, in the form of a user satisfaction index if available.	Are there specific measures in order to determine user satisfaction for the geospatial output? For instance: the possibility of assessing geospatial output should be included in a regular user satisfaction survey.
S12.3	Completeness	The extent to which all statistics that are needed are available.	Provide qualitative information on the extent to which content requirements in relevant legislation, regulations and guidelines are met. Where such requirements are not fully met, reasons for this should be provided. Provide information on the extent to which user needs related to content are satisfied.	Are there variables which are not used to produce geospatial outputs and are there reasons not to do so? Refer to the user needs formulated in S12.2

There are users who are mainly interested in geospatial outputs. Consequently, there might be user needs which are very specific to this part of the output portfolio for a certain statistical product. Given that, the reporting on concept S12 should consider the geospatial part sufficiently.

S13 Accuracy and reliability

SIMS	Concept name	Definition	Guidelines	Geospatial content
S13	Accuracy and reliability	Accuracy of data is the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.	See sub-concepts	See sub-concepts

SIMS	Concept name	Definition	Guidelines	Geospatial content
S13.1	Overall accuracy	Assessment of accuracy, linked to a certain data set or domain, which is summarizing the various components	Describe the main sources of random and systematic errors in the statistical outputs and provide a summary assessment of all errors with special focus on the impact on key estimates. The bias assessment can be in quantitative or qualitative terms, or both, and may be expressed as bias risk. It should reflect the producer's best current understanding (sign and order of magnitude) and include actions taken to reduce bias. *European level:* Provide a summary picture of accuracy across countries. The emphasis placed on various types of errors should depend upon the error profile of the respective process. For repetitive processes, describe how accuracy is developing over time and what efforts are underway to improve accuracy from an ESS perspective	Provide a summary here. Take into account the specific contents provided by the following sub-concepts.

S13.2	Sampling error	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	State whether sampling error is relevant. If probability sampling is used: for user reports, provide the range of variation of the A1 indicator among key variables at user report level of detail; for producer reports, provide the range of variation of the A1 indicator among key variables at producer report level of detail; indicate the impact of sampling error on the overall accuracy of the results; state how the calculation of sampling error is affected by adjustments for nonresponse, misclassifications and other sources of uncertainty, such as outlier treatment. If non-probability sampling is used, provide an assessment of representativeness, a motivation for the invoked model for estimation and risk of sampling bias. *European level:* If probability sampling is used: present sampling errors for key estimates across countries; indicate which country to country differences are significant and which are not; for a repetitive survey, describe at least broadly the trends in sampling error over time provide sampling errors for ESS level estimates.	Sampling errors are in principle not relevant for geospatial outputs. However, if there are very specific issues (for instance sampling errors are also visualized in geospatial output), then it should be stated here.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S13.3	Non-sampling errors	Error in survey estimates which cannot be attributed to sampling fluctuations.	Summarise the most important aspects of coverage, measurement, non-response, processing and model assumption errors. Discuss the corresponding bias risks and actions undertaken to reduce them. *European level:* Provide a summary of the above across countries.	Provide a summary of S13.3.1 to S13.3.5 and highlight the most important aspects.
S13.3.1	coverage error	Divergence between the survey population and the target population	Provide information on the frame and its sources. Provide an assessment, whenever possible quantitative, of overcoverage and undercoverage, including an evaluation of the bias risks associated with the latter. Describe actions taken for reduction of undercoverage and associated bias risks. *European level:* Provide an overall picture of coverage across countries. This is often best done in the form of tables with the important coverage aspects, country by country.	Frame errors (related to over coverage and under coverage) can have an impact on the output quality. Therefore, it should be described what frames are in use for the extraction of information related to the geospatial processing.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S13.3.2	Measurement error	Measurement errors are errors that occur during data collection and cause recorded values of variables to be different from the true ones	The main sources of measurement error should be reported and assessed. Their description should be accompanied by any available analysis, otherwise by the producer's best knowledge. Where available and relevant describe: identification and general assessment of the main sources of measurement error; efforts made in questionnaire design and testing, information on interviewer training and other work on error prevention; results of assessments based on comparisons with external data, re-interviews or experiments; results of indirect analysis, for example, of the editing phase; and actions taken to correct measurement errors . *European level:* Where measurement errors are important as a single source of error provide a comparative summary across countries. Otherwise include them within overall accuracy in S.13.1.	Coordinates and geographic characteristics are the most important elements for the geospatial processes. How reliable are the processes and data sources for collecting this information?

SIMS	Concept name	Definition	Guidelines	Geospatial content
S13.3.4	Processing error	The error in final data collection process results arising from the faulty implementation of correctly planned implementation methods.	If processing errors are significant, identify the main issues regarding them. Present an analysis of processing errors, where available, otherwise a qualitative assessment. Report their extent, and impact on the outputs, of the most significant types of error. Include descriptions of linking and coding errors, if applicable. Where mistakes relating to programming or publishing have occurred, corrective measures taken as well as actions for avoiding them in the future should be reported. *European level:* Provide a summary across countries of processing errors.	GSBPM (and the geospatial view of GSBPM) The GSBPM can be used as a template for the description of all aspects of a statistical process. Describe what measures are taken to prevent processing errors and the most relevant process steps. The most important sources for errors are linking of statistical data and various geocoding processes. If other relevant processes are identified as potential sources for errors, this should be described here.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S13.3.5	Model assumption error	Error due to domain specific models needed to define the target of estimation.	Describe process specific models, for example, as needed to define the target of estimation itself. Provide an assessment of the validity of each model. (Descriptions of models used in treatment of specific sources of error should be presented in the section dealing with those errors.) *European level:* Where different models are used across countries, provide a comparative overview and discuss their validity and the likely effects of the differences.	Describe models specifically for the production of geospatial outputs. For instance, urban/rural classification.

Accuracy can be defined as the degree or closeness to which the information on a map matches the values in the real world. In GIS data, accuracy can refer to a geographic position, but also to an attribute, or conceptual accuracy. Precision refers to how exact the description of data is. Attribute accuracy indicates the attribute attached to the point, line and polygon features of the spatial database, which shows how reliable and reasonably correct or free from bias these are. The reporting should concentrate on the issues of non-sampling errors here mostly related to measurement errors, coverage errors and processing errors.

S14 Timeliness and Punctuality

SIMS	Concept name	Definition	Guidelines	Geospatial content
S14	Timeliness and Punctuality	See sub-concepts	See sub-concepts	See sub-concepts
S14.1	Timeliness	Length of time between data availability and the event or phenomenon the data describing.	Outline the reasons for the time lag. Outline efforts to reduce time lag in future. *European level:* For reports only published at European level do the above. Otherwise summarise the above across countries.	State here, if there are specific issues causing a further delay in the geospatial outputs (example: non-availability of geo input data like geometries).
S14.2	Punctuality	Time lag between the actual delivery of the data and the target date when it should have been delivered.	Report only for annual or more frequent releases. If a release schedule was made available to users and/or specified in a regulation; • provide the percentage of releases delivered on time, based on scheduled release dates, over a specified period and/or set of outputs. • in the event of any non-punctual releases, explain the reasons and outline efforts to improve punctuality. In the absence of a release schedule, explain why there is no schedule and indicate what efforts will be made to make one available in the future.	List here the project plan for the geospatial part of the product. Normally geospatial output is not required for delivery to Eurostat. If there is a delay for geospatial output, it should be mentioned explicitly here. Provide as well the reasons that caused the delay.

SIMS	Concept name	Definition	Guidelines	Geospatial content
			European level For outputs first published at European level, report as above. For outputs first published at country level, • state the agreed time frame for delivery of national data and the actual delivery dates; • summarise punctuality across countries	

Timeliness of geospatial output normally depends a lot on the availability of input data. For reporting on quality, geospatial issues should only be mentioned when shortcomings in timeliness for geospatial outputs are caused by reasons located in the geospatial processing. Regarding punctuality, the planning for geospatial outputs is normally embedded in the project plan compiled for the statistical product as a whole.

S15 Coherence and comparability

SIMS	Concept name	Definition	Guidelines	Geospatial content
S15	Coherence and comparability	Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics.	See sub-concepts	See sub-concepts
S15.1	Comparability – geographical	The extent to which statistics are comparable between geographical areas.	Describe any problems of comparability between regions of the country. The reasons for the problems should be described and as well an assessment (preferably quantitative) of the possible effect on the output values. Give information on discrepancies from the ESS/ international concepts, definitions, with reference to other chapters for more details. *European level:* Focus on factors that affect the comparability between countries. Analyse asymmetries in statistical mirror flows where possible.	Describe how you take into account potential comparability problems between regions and areas. It should be clear to the user, when using geospatial outputs, if there are problems for comparing the results.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S15.2	Comparability – over time	The extent to which statistics are comparable or reconcilable over time.	Provide information on possible limitations in the use of data for comparisons over time. Distinguish three broad possibilities. There have been no changes, in which case this should be reported. 1. There have been some changes but not enough to warrant the designation of a break in series. 2. There have been sufficient changes to warrant the designation of a break in series	If there is a time component in the geospatial outputs, it should be described how far this is relevant.

To be able to compare results is one of the basic elements for users, in order to take full advantage of statistical outputs. Geospatial output presents most of the time regions (mainly related to a national territory) augmented by statistical characteristics. Therefore, for geospatial outputs, information should be available, if there are concerns regarding comparability over regions. Another possible issue, is comparability over time, which is relevant when the geospatial presentations also have a time component.

Coherence goes mainly to divergences between domains. It seems that this is not very relevant for geospatial outputs.

S16 Cost and burden

SIMS	Concept name	Definition	Guidelines	Geospatial content
S16	Cost and burden	Cost associated with the collection and production of a statistical product and burden on respondents.	Cost: Provide annual operational costs of the process, with breakdown by major cost component. Describe recent efforts to improve efficiency and comment on the extent to which information and communication technology is used. *European level:* Describe recent initiatives and efforts to improve efficiency at the European level. Burden: Provide an estimate of the respondent burden imposed by the process. Describe all the means taken to minimise burden. *European level:* Describe recent initiatives and efforts to minimise burden at the European level.	Response burden is usually not an issue related to geospatial processing or outputs. However regarding costs, there could be a statement reflecting the share of costs in order to produce geospatial outputs.

Response burden is an issue for surveys and usually not related to the geospatial part of a statistical product. However, accounting systems of statistical authorities should allow the derivation of the costs relevant for the production of geospatial outputs.

S17 Data revision

SIMS	Concept name	Definition	Guidelines	Geospatial content
S17	Data revision	Any change in a value of a statistic released to the public.	See sub-concepts	See sub-concepts
S17.1	Data revision – policy	Policy aimed at ensuring the transparency of disseminated data, whereby preliminary data are compiled that are later revised	Describe the data revision policy applicable to data output from the statistical process being reported. In so far as they are relevant to the process being reported, summarise the general procedures for treatment of planned revisions, benchmark revisions, unplanned revisions, and revisions due to conceptual and/or methodological changes. *European level:* Describe the data revision policy and procedures at European level.	The revision policy is provided by the statistical product described. There are no specialities regarding geospatial outputs. However, if an occasional revision takes place, geospatial outputs should take in revised data. However, if a revision of geodata has to be conducted (for instance error in coordinates), it has to be reported.

Data revisions are either planed ones, following a fixed revision schedule, or conducted on an occasional basis (for instance when substantial errors have to be corrected). Changes in the input data coming from the statistical product (survey for instance) are not related to the geospatial outputs. However, when the revision results in a substantial change of incoming geodata, because the former have been erroneous, a quality report should take note about such an incident.

S18 Statistical processing

SIMS	Concept name	Definition	Guidelines	Geospatial content
S18	Statistical processing	See sub-concepts	See sub-concepts	See sub-concepts
S18.1	Source data	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Indicate if the data are based on a survey, administrative data, multiple data sources, or macro-aggregates. In the event of multisource or macro-aggregate processes describe each source. For each survey source, report the survey population, cross referencing the description of the target population presented in S.03.6, and summarise the sample design. For each dataset from an administrative source, summarise the source, its primary purpose, and the most important data items acquired. *European level:* Provide an overview of the sources used across countries	Describe the data sources needed for geospatial processing. This includes: • Geometries • Input data for geocoding • Input data for geographic features

SIMS	Concept name	Definition	Guidelines	Geospatial content
S18.3	Data collection	Systematic process of gathering data for official statistics	For each source of survey data: • describe the method(s) used to gather data from respondents; • annex or hyperlink the questionnaires(s). For each source of administrative data: • describe the acquisition process and how it was tested. For all sources: • describe the types of checks applied at the time of data entry. *European level:* Provide a summary of the commonalities and differences in the collection methods, questionnaires and checks used in different countries.	Describe the environment from which the source data can be accessed (register, database, other form). Describe if there are special data collection processes related to the geospatial part. Are there special collection procedures like remote sensing, GPS, photogrammetry etc.? Describe the methods of collecting and using administrative data needed for geospatial processing. Describe how an effective data management of source data is implemented. (GSGF Requirement 2.1). Describe the functioning of the geocoding process.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S18.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of statistical results.	Describe the procedures for checking and validating the source data and how the results are monitored and used. Describe the procedures for validating the aggregate output data (statistics) after compilation, including checking coverage and response rates, and comparing with data for previous cycles and with expectations. List other output datasets to which the data relate and outline the procedures for identifying inconsistencies between the output data and these other datasets. *European level*: Provide a summary of the commonalities and differences in the validation methods used by countries.	Matching of geocoded data and unit record data is one of the most decisive process steps. Describe how the point-of-entry validation in collection of administrative or statistical data is implemented (Requirement 2.5 of GSGF). Describe any other validation processes in various process steps related to geospatial processing.

SIMS	Concept name	Definition	Guidelines	Geospatial content
S18.5	Data compilation	Operations performed on data to derive new information according to a given set of rules.	Describe the procedures for imputation, the most common reasons for imputation and imputation rates within each of the main strata. Describe the likely impact of imputation. Describe the procedures to derive new variables and to calculate aggregates and complex statistics. Describe the procedures for adjustment for nonresponse and the corrections to the design weights to account for differences in response rates. Describe the calculation of design weights, including calibration (if used). Describe the procedures for combining input data from different sources.	Describe the process of how to link geocode data to unit record data. Describe how to assess the quality and list the quality indicators you use. List possible indicators If you use imputation for some geodata, describe the methods used here. Describe also the aggregation processes relevant for producing maps.

The processing for geospatial outputs focuses on linking unit data to geocoded data and assigning geographic characteristics. Finally, there are aggregation processes in order to construct maps and other geospatial output. Considering that the most important reporting concept is S18.5.

Another focus should be laid on the collection of data as well as the validation of certain input sources specific for geographic processes.

S19 Comment

SIMS	Concept name	Definition	Guidelines	Geospatial content
S19	Statistical processing	Supplementary descriptive text which can be attached to data or metadata.	Provide any information • that is pertinent to the report but does not fit under any of the other concepts; or • to repeat key issues; or • to refer to annexes that might be attached to the report.	Anything for geospatial processing and/or geospatial outputs that did not fit anywhere else.

Supplementary descriptive text which can be attached to data or metadata. Anything that did not fit anywhere else.