UNITED NATIONS SECRETARIAT **ESA/STAT/AC.279/P6**
Department of Economic and Social Affairs **October 2013**
Statistics Division **English only**

United Nations Expert Group on the Integration of Statistical and Geospatial Information
First Meeting
New York, 30 October - 1 November 2013

Agenda: Item 6

Overview of the elements of the geospatial statistical framework of the European statistical system [1]

Prepared by
Eurostat

[1] This document is being produced without formal editing

EUROPEAN COMMISSION
EUROSTAT

Directorate E: Sectoral and regional statistics
Unit E-4: Regional statistics and geographical information

OVERVIEW OF THE ELEMENTS OF THE GEOSPATIAL STATISTICAL FRAMEWORK OF THE EUROPEAN STATISTICAL SYSTEM

Contact: Ekkehard PETRI, Telephone: +352 4301-36745, ekkehard.petri@ec.europa.eu

1. INTRODUCTION

In the current socio-economic context, marked by globalization and environmental issues, and affected by a financial crisis requiring structural changes in the European marketplace, a deeper improvement of the EU information capacity has become essential for EU policy making. In particular, for efficiency and cost benefit considerations, the territorial dimension of socio-economic phenomena needs to be better taken into account in order to define tailored, pinpoint policies.

In recent years the European Statistical System (ESS) has increasingly recognised the importance of the integration of statistical and geospatial information, e.g. for measuring the achievement of the EU targets of 'smart, sustainable, inclusive growth'. The EU's new high-level strategy to achieve these targets, Europe 2020, explicitly introduces a third dimension of cohesion - territorial cohesion – which requires greater territorial disaggregation of traditional statistics and a territorial perspective on cross cutting issues.

To move in this direction the ESS is convinced that the spatial dimension must be taken into account and embedded into the statistical production chain from its very beginning, beyond the traditional approach often related to administrative units. The spatial dimension, in fact, is already implicitly present in statistics, as a dimension affecting data collection, storage, analysis and dissemination, and as a means for creating different territorial aggregations. What is needed to improve the EU information capacity is a more in-depth and integrated approach, where spatial dimension, at the most appropriate level of accuracy, is associated at all stages of the statistical data production cycle. This will help to create the multipurpose information from multiple sources that are available, and integrate it into the common information system that our users need.

Against this background, it is the purpose of this paper to present the main elements of the current geospatial framework for statistics in Europe and to briefly introduce several initiatives and legal frameworks that may provide an input into the work of the UN-GGIM expert group from a European perspective. This will be complemented by national situation reports from several EU Member States.

2. SPATIAL REFERENCE SYSTEMS FOR STATISTICS IN EUROPE

2.1. Nomenclature of territorial units for statistics - NUTS

The Nomenclature of territorial units for statistics[1], abbreviated as NUTS (from the French Nomenclature des Unités Territoriales Statistiques) is a geographical nomenclature subdividing the territory of the European Union (EU) into regions at three different levels (NUTS 1, 2 and 3, respectively, moving from larger to smaller territorial units). Above NUTS 1 is the 'national' level of the Member State. The current version of NUTS (2010) subdivides the territory of the European Union and its 28 Member States into 98 NUTS 1 regions, 273 NUTS 2 regions and 1324 NUTS 3 regions. The NUTS is based on a European legal act[2]. NUTS areas aim to provide a single and coherent territorial breakdown for the compilation of EU regional statistics. The NUTS regions are constructed according to the following principles:

Principle 1: The NUTS regulation defines minimum and maximum population thresholds for the size of the NUTS regions;

Thus despite the aim of ensuring that regions of comparable size all appear at the same NUTS level, each level still contains regions which differ greatly in terms of population and area. For further details see Table 1 and Table 2 in the Annex.

Principle 2: NUTS favours administrative divisions (normative criterion);

For practical reasons the NUTS classification is based on the administrative divisions applied in the Member States that generally comprise two or three main regional levels. In the absence of a regional level it is aggregated from lower administrative units.

Principle 3: NUTS favours general geographical units;

General geographical units are normally more suitable for any given indicator than geographical units specific to certain fields of activity. NUTS is used for socio-economic analyses of the regions, comparison of regions, for measuring social cohesion and for deciding on the eligibility for aid from the EU funds at various NUTS levels.

The NUTS is the recognised framework for the collection, development and harmonisation of EU regional statistics and is referenced in many EU legal acts, not only on statistics. As such the NUTS is a core element of the EU political system.

[1] http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Nomenclature_of_territorial_units_for_statistics_(NUTS)

[2] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:154:0001:0041:EN:PDF

2.2. Local Administrative Units – LAU

Below the NUTS classification the local administrative units, abbreviated as LAUs[3] form a system for dividing up the economic territory of the European Union (EU) for the purpose of statistics at local level. They have been set up by Eurostat and they are compatible with NUTS.

At local level, two levels of LAU have been defined:

- The upper level (LAU1) is defined for most, but not all, of the countries.
- The lower level (LAU2) consists of municipalities or equivalent units in the EU Member States.

Every year between 700 and 2000 of the ~122 000 LAUs in the EU are changed.

2.3. Derived territorial classifications and typologies

In addition to the purely administrative territorial classifications NUTS and LAU, derived classifications[4] have been introduced in the ESS with a more functional focus. However at present these classifications still use the LAU2 and NUTS3 as their building blocks:

- Classification of LAUs according to their degree of urbanisation[5];
- Rural-Urban typology[6] of NUTS3 regions;
- Classification of cities and their surroundings zones using LAU2 or NUTS;
- Coastal areas based on LAU2s.
- Urban classifications and city definitions including metropolitan regions

Most of these classifications are based on population densities and/or a distance measure to a geographical feature. Some already use a 1km² population grid as the central auxiliary dataset for the classification. The territorial classification are used as one element in the design or analysis of surveys, e.g. in the design of the Labour Force Survey or the Survey on Income and Living Conditions.

[3] http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/local_administrative_units

[4] http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Regional_typologies_overview

[5] http://epp.eurostat.ec.europa.eu/portal/page/portal/degree_urbanisation/introduction

[6] http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Urban-rural_typology

2.4. TERCET – a legal framework for territorial classifications and typologies for statistics in the EU

The European Commission is proposing to create a new legal framework for territorial classifications and typologies for statistics using a three tier structure: (1) basic act, (2) delegated acts and (3) methodological notes. The territorial classifications to be covered are: NUTS, LAU, and the derived territorial typologies, as listed above. As many of these definitions rely on population distribution at grid level, the new legal framework may give legal authority to population grids. The process has been launched recently and will take several years to be completed. The discussions and conclusions of the UN-GGIM expert group on the statistical geospatial framework will certainly be relevant for this process.

2.5. Criticism of territorial classifications and typologies for statistical purposes based on administrative units

In the EU the approach to administrative areas used for statistical purposes varies a lot between Member States. While in some countries the system has remained stable for decades, other countries introduce frequent changes into their system to meet demands from politics, e.g. for efficiency gains in public administration or to reflect demographic trends. These changes make it necessary that the NUTS system is revised every three to four years, and statistics have to be backdated. This instability of the territorial building blocks for statistics makes time series analysis difficult if not impossible for many types of statistics and as such limits its use for analysis or spatial planning.

Also, the extreme difference in area and population of territorial units (see Table 1 and Table 2) make any type of spatial analysis and comparison difficult, as e.g. vast but almost empty areas in the north of Scandinavia are compared with small but densely populated megacities.

In essence, the current system is considered to be biased towards the accounting and reporting use case of statistics, as requested by governments or responsible administrations while the spatial analysis and spatial planning use case are less well served. Since several years, spatial analysts in National Statistical Institutes (NSI), researchers and spatial planners have been suggesting to complement the current territorial system for official statistics with a neutral, hierarchical and stable system.

Output areas based on equal population counts represent a step in the right direction but do not overcome some of the intrinsic shortcomings of statistics based on administrative units such as instability over time or the variety in area. Hence their focus on population numbers may still unnecessarily limit their usage for various purposes and statistical products and such a system may not be useful as a means to integrate data from various sources and for various thematic areas.

3. STATISTICAL GRIDS

To respond to this criticism of territorial frameworks based on administrative units or other irregular tessellation systems, and to meet the requirements of researchers and analysts, grid statistics have been proposed as an alternative and gained wide recognition in the EU as a complementary system for disseminating statistics. Starting in the Scandinavian countries in the 1970es several countries have added grid statistics to their product portfolio, mainly in the area of population statistics.

The latest housing and population census in 2011 has been a catalyst for this trend, as many NSIs used point based location frameworks, in most cases address and building registers during the preparation and execution of the Census. Preserving the location element with the statistical information now allows them to aggregate the data to any type of area including grid cells.

3.1. General advantages of grid statistics

Statistics based on grids overcome many of the shortcomings of statistics collected in the traditional way on the basis of administrative units or other irregular tessellation systems, with regard to the planning and analysis use case. Grid statistics are particularly useful in an international, cross-boundary situation like in Europe where data from many countries with a range of administrative and statistical traditions, and from various sources have to be joined together. Statistical grid data help to better describe the spatial extent and magnitude of phenomena, study their causes, plan actions to improve our socio-economic and environmental conditions and measure the achievement of targets.

- Grids are stable over time and hence independent of the frequent administrative changes. They are perfect for time series.

- It is easy to construct a hierarchical system of grids, where the size of grid cell is in proportion to the size of the study area and the resolution of the phenomenon, ranging from local to global. The transition from point based microdata to grid data becomes a continuum.

- All grids have the same size, allowing for easy comparisons;

- Grid cells, provided they have a suitable size, can be joined together to create virtually any functional output area for statistics based on objective criteria, e.g. to statistics for mountain areas, school districts, coastal areas, river basin districts, … This is usually impossible with other territorial systems and makes the grid system extremely valuable for analysis and spatial planning.

- Grid data with suitable cell size corresponding to the extent and magnitude of the phenomena avoid the dilution effect of too large building blocks.

- Socioeconomic statistics at grid level are easily linked with scientific data and data from spatial modelling, which are often created based on grid cells or other finite elements, e.g.

meteorological data, biodiversity information, or data from climate modelling. This is particularly important for some of the most pressing challenges for sustainable development such as mitigating climate change risks, or containing flooding risk.

A survey among users and producers of small area statistics in Europe by European NSIs has revealed that grid statistics represent a good compromise between user and producer requirements[7].

3.2. GEOSTAT project

To expand and accelerate this trend to grid statistics, Eurostat has launched the GEOSTAT action in cooperation with the European Forum for Geostatistics (EFGS)[8]. The goal of the first project, GEOSTAT 1 is to create a 1km² population grid dataset of the 2011 census for the EU. Through a series of three projects Eurostat and NSI will draft guidelines for the creation of national population grids that are suitable for the wide range of national production systems in NSIs, and eventually produce national grid datasets based these guidelines with comparable quality. In the first phase, the project has already produced a prototype population grid, GEOSTAT 2006[9]. This grid dataset combines national population grids from 12 European Countries with a downscaled population grid for another 18 countries, using a spatial distribution model of LAU2 population densities.

The demand for this prototype has been relatively high and the data have been used already in some of the definitions of the derived territorial classifications. Using a population grid helped to take account of the actual spatial distribution of population in LAU2 and NUTS areas and helped to avoid the dilution effect of average density in large area administrative units with a highly concentrated population.

The final GEOSTAT 2011 grid dataset constructed from ~20 national grid datasets from census information will become available at the end of 2014. The dataset will be completed with downscaled data for the missing countries. A first release with lower participation will be released beginning of 2014.

Building on the success of GEOSTAT 1, Eurostat is planning to launch a long term initiative to create a point based geospatial statistical framework for the next census and to define population grids as one of the standard products of population statistics. A joint working group of NSIs and National Mapping and Cadastral Agencies (NMCA) will also work on the definition of a broader concept and an implementation plan to link a range of socio-economic and environmental statistics to a location and this way

---

[7] GEOSTAT 1A final report: http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/publications/geostat_population_grid_report

[8] European Forum for GeoStatistics is a European association of experts from NSIs, other public organisations, research, and users in the field of georeferenced statistics and GIS. Activities are mainly concentrated on the development of the best practices in the production of geostatistics in Europe. See http://www.efgs.info/ for more information.

[9] http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_grids

contribute to the information capacity needed to achieve the EU sustainable development targets.

3.3. Limitations of grid statistics

It is important to note that grid statistics are easy to produce from geocoded microdata from full enumerations like in censuses, or from registers or other administrative data sources containing virtually all entities. However achieving statistical representativeness for small areas from survey data is extremely costly. This means that for surveys, grid cell sizes would have to be enormous and reach 10s or even 100s of km, thus devaluing the grid advantages.

'Big Data' may have the potential to overcome these limitations, and may help to reach the high spatial resolution necessary for grid statistics. Eurostat will launch projects to assess this potential.

On the dissemination side the correct handling of disclosure control in any small area including grids quickly becomes an issue. This is very relevant in population statistics and census information where crossing just a few population characteristics in a small area with low population quickly leads to the risk of disclosure of individuals. Within the ESS a discussion is taking place on the sensitivity of various population characteristics and on suitable disclosure control thresholds.

4. INSPIRE

The spatial properties of the areas used for statistics, called "statistical units" are regulated under the INSPIRE directive[10]. Statistical units are defined as spatial features or "units for dissemination or use of statistical information". For the definition of the data model of statistical units[11], geographical experts worked closely together with experts from NSIs and Eurostat to make sure that the statistical requirements are sufficiently taken into consideration.

In essence statistical units can have vector or grid representation and are defined with their spatial characteristics by way of reference to other, basic feature types like administrative units, geographical grid systems, or area management zones that are also regulated under INSPIRE.

Statistical units have a hierarchical structure with a number of levels representing different area sizes, like the NUTS. Another important property that is regulated under INSPIRE is the temporal evolution of the geometry of statistical units that needs to be traceable (splits, merges, boundary shift).

In addition to the characteristics of spatial features, INSPIRE also provides standards for metadata and for a web-based infrastructure for viewing, discovery and download of geographical information. INSPIRE metadata is restricted to the spatial aspects of features and mainly serves the discovery use case.

[10] http://inspire.jrc.ec.europa.eu/

[11] http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SU_v3.0rc3.pdf

In statistics, a parallel infrastructure has emerged with the SDMX standard as a foundation where Eurostat is one of the main sponsors. The interoperability of the two infrastructures and standards will be one of the main challenges for any statistical geospatial framework.

5. Cooperation between NSIs and NMCAs

In Europe NSIs and NMCAs are usually separate organisational entities with a varying degree of cooperation at national level. In the ESS the cooperation between the two major providers of public sector information has been addressed with the so called GISCO working group, organised by Eurostat. The group has mainly been working on the usage of geospatial information and GIS for statistical purposes, e.g. during the development of INSPIRE and in the preparation of the recent census, but recently the focus has shifted towards a more integrated and comprehensive approach to both types of information.

6. Conclusions

The ESS is based on the NUTS as the principal territorial classification for official European statistics. The rooting of the NUTS in the administrative divisions of Member States is deep, and many European legal acts refer to the NUTS. Also the notion of a region in the public opinion may be very strong for many of the NUTS areas and has to be respected. The NUTS will therefore remain the first and main pillar of the territorial classification for statistics.

Eurostat and the ESS aim to create a second pillar for certain types of official European statistics and this way support better the planning and analysis use case. This additional system will be based on point data from administrative sources and registers sitting at the national level. These national grid datasets will be assembled to European datasets and the target resolution is $1km^2$ at the European level.

Introducing a third system of output areas based on equal population numbers or more generally on equal numbers of statistical entities, which is independent from the LAU will be a challenge in the European context, due to the solid foundation of the statistical system on the NUTS and the strong link between the LAU and the NUTS. The situation for small areas potentially useful for statistics varies a lot between countries depending on their size, population, administrative systems, and traditions. The national statistical systems in Europe therefore might resist to setting up another system that is somewhat similar to the LAU system but not necessarily compatible with it.

Eurostat welcomes the explicit reference to the 2020/2021 Round of censuses in the Terms of Reference as a core element of the work of the expert group. Demographic changes and migration represent two of the major challenges for the European societies, and responding to these challenges require the best possible census information we are able to create. The ESS plans to make location an integral part of the next census.

From a European perspective the interoperability of statistical and geospatial information, of the related information infrastructures, metadata standards, and quality documentation, and the easy linkage of the two types of data are key requirements from users for any geospatial statistical framework.

Both NSIs and NMCAs consider UN-GGIM as an opportunity to intensify the cooperation and to address together the various issues that will be discussed in the expert group. Initial efforts will be necessary to develop a common vision for the two information areas and to align essential concepts, e.g. for metadata.

ANNEX

Table 1: Population ranges for the various NUTS levels

LEVEL	MINIMUM	MAXIMUM
NUTS 1	3 million	7 million
NUTS 2	800 000	3 million
NUTS 3	150 000	800 000

Table 2: Area of the NUTS regions (km^2)

	NUTS 1 Average	NUTS 1 Minimum	NUTS 1 Maximum	NUTS 2 Average	NUTS 2 Minimum	NUTS 2 Maximum	NUTS 3 Average	NUTS 3 Minimum	NUTS 3 Maximum
EU-27	45400	161	336837	16310	13	226775	3400	13	106012
BE	10176	161	16844	2775	161	4440	694	101	1592
BG	55501	42672	68330	18500	14487	22365	3964	1349	7748
CZ	78867	78867	78867	9858	496	17618	5633	496	11015
DK	43098	43098	43098	8620	2561	13124	3918	180	8720
DE	22319	404	70552	9398	404	29480	867	36	3058
EE	45227	45227	45227	45227	45227	45227	8686	3364	15533
IE	69797	69797	69797	34899	33252	36545	8725	921	14286
GR	32989	3808	56792	10151	2307	19147	2587	356	5461
ES	72284	7447	215322	26631	13	94226	8576	13	21766
FR	70315	12012	145645	24340	1128	83534	6328	105	83534
IT	60278	49801	73224	14352	3263	25711	2740	212	7400
CY	9250	9250	9250	9250	9250	9250	9250	9250	9250
LV	64559	64559	64559	64559	64559	64559	10760	303	15246
LT	65300	65300	65300	65300	65300	65300	6530	4350	9731
LU	2586	2586	2586	2586	2586	2586	2586	2586	2586
HU	31009	6916	49499	13290	6916	18337	4651	525	8445
MT	316	316	316	316	316	316	158	69	247
NL	10386	7291	11893	3462	1449	5749	1039	128	3437
AT	27960	23563	34377	9320	415	19186	2397	415	4613
PL	52114	27521	74860	19543	9412	35559	4708	261	12090
PT	30697	801	88967	13156	801	31551	3070	801	8543
RO	59598	36274	72612	29799	1821	36850	5676	238	8697
SI	20273	20273	20273	10137	8061	12212	1689	264	2675
SK	49035	49035	49035	12259	2053	16263	6129	2053	9455
FI	169210	1582	336837	67684	1582	226775	17812	1582	98984
SE	147116	48203	312052	55168	6789	165296	21017	3055	106012
UK	20274	1574	77932	6574	319	40423	1750	35	14238