**How geospatial information adds value to existing sub-national data and territorial typologies**

Valeriya Angelova-Tosheva, Eurostat, [Valeriya.ANGELOVA-TOSHEVA@ec.europa.eu](mailto:Valeriya.ANGELOVA-TOSHEVA@ec.europa.eu)

Hannes Reuter, Eurostat, [Hannes.REUTER@ec.europa.eu](mailto:Hannes.REUTER@ec.europa.eu)

**Abstract**

*Eurostat aims to provide service-oriented, high-quality statistics for Europe to, for example, support policy decisions. With that goal in mind, the regional statistics and geographical information team not only provides methodological support to policy-makers, researchers and the general public to better understand the existing data but also offers new ways of meeting data requirements in an innovative and flexible manner. By combining Geographic information with in-depth statistical knowledge, requests such as the transformation of data available by official territorial typologies into data by a customised territorial type can be accomplished. Two recent examples show a) how the number of schools, kindergardens, post offices, sport clubs and bars as social meeting points in individual rural regions per 100 000 inhabitants have been derived and b) how for various demographics and socio-economic data by maritime ports in Europe have been derived. Geospatial information and statistical information combined proved essential to support policy decisions in the impact assessment of the Common Agricultural Policy domain as well as providing reliable statistical estimates for the blue growth.*

**Keywords:** GIS, sub-national data, user needs, territorial typologies, reengineering existing data, alternative data sources, service orientation

**1. Problem statement**

Eurostat's mission is to provide high-quality statistics for understanding Europe and enabling better decisions to be made. Therefore, the regional statistics and geographical information team focuses on regional statistics, territorial typologies and city indicators. Eurostat's database (EuroBase) contains 750 sub-national datasets. These are data available at the NUTS 2 and NUTS 3 level and data on official territorial typologies such as cities and their Functional Urban Areas, Degree of Urbanisation, urban, rural, intermediate regions, coastal-non-coastal regions, coastal areas and metropolitan regions as defined in the TERCET Regulation amending the NUTS Regulation[[1]](#footnote-1) plus the additional typologies of border, island, outermost and mountain regions that are not included in the TERCET.

In our daily work, we are experiencing a high number of user requests and new emerging data needs. Sometimes, these are for unique or unusual sub-national indicators of interest that are outwit the scope of the existing data collections at Eurostat and the National Statistical Institutes (NSIs). Furthermore, users often need data on customised typologies complementing the official, available ones.

To satisfy these requests, Eurostat is now producing smart statistics on demand by applying spatial estimation techniques on existing data, preparing customised territorial aggregates based on geolocation and searching for alternative data sources.

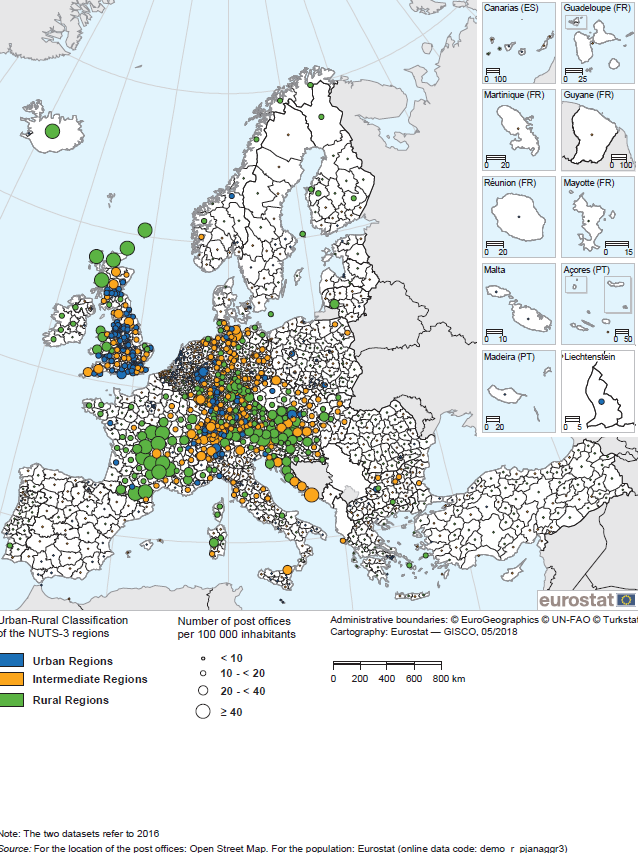
This article aims at outlining the practically unlimited power of the Geographical Information Systems (GIS) for data reengineering, which, combined with methodological awareness of subnational statistics, becomes an innovative service offered by Eurostat to inform society and support political decision making. The sections below describe two recent examples of how geospatial information has been used in order to satisfy ad-hoc data requests.

**2. Use case one: Emerging data needs from DG Agriculture and Rural Development met by using geolocation and alternative data sources**

One of the policy-making partners and stakeholders of Eurostat is the Directorate General for Agriculture and Rural Development (DG AGRI). DG AGRI makes extensive use of Eurostat's data by Degree of Urbanisation (DEGURBA), Local Administrative Units (LAU) based typology)[[2]](#footnote-2) and by Urban-rural regions (NUTS 3 based typology)[[3]](#footnote-3), all legally recognised in the TERCET, for designing, negotiating, implementing, monitoring and evaluating the Common Agricultural Policy of the European Union. Whilst deeply analysing the development of the rural territories in Europe, policy analysts from DG AGRI have intuitively captured a negative phenomenon, which however still needs to be statistically proven. The hypothesis is that people living in rural regions lack access to social meeting points which leads to isolation, or drives people to search for different services in a neighbouring, larger residential area, hence there is an impact on their quality of life. Therefore, DG AGRI asked Eurostat to provide them with data on the number of schools, kindergartens, post offices, sport clubs and bars in rural regions. Since there is currently no official data source for this at Eurostat, some geospatial techniques, combined with alternative data sources, needed to be applied in order to obtain the requested data.

Eurostat provided the following, straightforward and efficient solution making use of the geodata available from volunteered geographic information (VGI). Firstly, the GIS object's categories of interest all over Europe were selected from the OpenStreetMap (OSM) database and exported in a GIS environment as a resulting layer. Secondly, the resulting point features layer were overlaid by geolocation with the layer containing the NUTS 3 boundaries[[4]](#footnote-4) and aggregated by individual NUTS regions. After these simple GIS operations, the data obtained were matched with the latest urban-rural classification of the NUTS 3 regions (based on the 2013 NUTS version). Finally, since the figures were very small for some NUTS 3 regions, Eurostat processed the data by normalising it by 100 000 of inhabitants living in every single NUTS 3 region. Map 1 shows the results that have been obtained.

Map 1 Are the people living in rural regions lacking access to social meeting points?



Several papers have analysed the data quality and completeness of OSM data. Examples are work executed by Barrington-Leigh & Millard-Ball (2017) for road networks and Jacksone et al. (2013) for points of interest, concluding that for the European extent, the quality is comparable to authoritative or proprietary data sources. A disadvantage of using the OSM data for addressing similar data requests is that the data quality and completeness might vary by country, object type and region and this therefore needs to be assessed and documented.

Graph 1. Distribution of the NUTS 3 regions by category of access to post offices

We observe that in the European Union, an average of 9.3 post offices are available per 100 000 inhabitants with a standard deviation of 8.8 with a range from zero in several Outermost regions to 108 post offices per 100 000 inhabitants in the region UKM66 (Shetland Islands). For the investigated regions with zero values, this indicates that some data issues might exist - either not yet added to the map by the community (topological issue) or wrongly attributed (semantic issue). Upon inspecting these maps, the higher number of postal offices per 100 000 inhabitants on islands seems plausible, whereas some of the regions in southern France seem to have a surprisingly large number of postal office per 100 000 inhabitants and therefore this requires further investigation. Regarding the urban-rural dimension, it appears that access to post offices is something that is generally country defined. However, in the north of the UK and in Central Europe we can observe a 'rural paradox' since the rural regions there have more post offices per 100 000 inhabitants than the urban and intermediate regions do.

A future topic for more sophisticated geospatial analysis at Eurostat could be an estimation of the access of people living in rural grid cells to social meeting points, measured by distance.

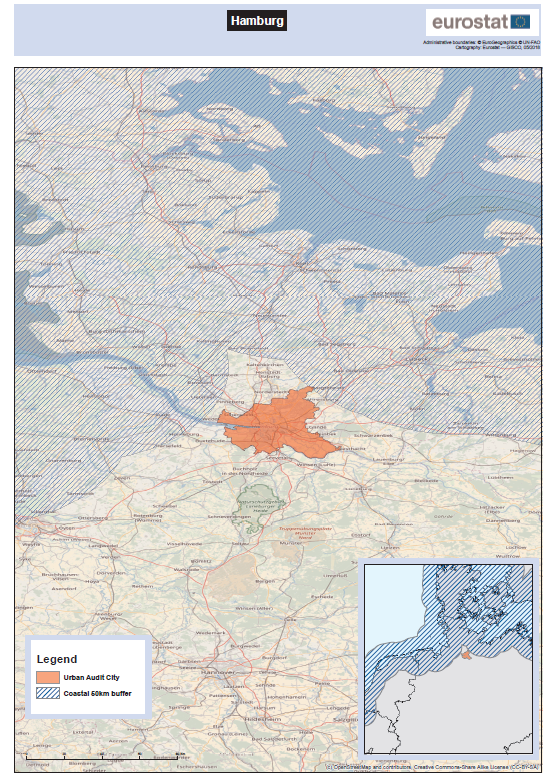
**3. Use case two: DG** Maritime Affairs and Fisheries' **data request for a customised territorial typology**

Another key partner and stakeholder of Eurostat is the Directorate General for Maritime Affairs and Fisheries (DG MARE). From 2018 onwards, DG MARE will regularly monitor the progress and developments in the EU Blue Economy and will produce an Annual economic report on the subject. The Blue Economy encompasses all the sectoral and cross-sectoral economic activities related to oceans, seas and coasts. Ports, as multi-activity nodes, play a crucial role in the development of the established and emerging maritime sectors. That is why DG MARE requested additional data on the largest maritime ports/coastal cities.

In the domain of maritime transport, Eurostat provided data on all ports in a country, main ports, top twenty or top five ports. DG MARE, as one of the main users of these data, asked for more detailed data on maritime ports.

Before finding a solution to this request, the situation at Eurostat was the following. The TERCET regulation did not contain any specific provisions for coastal cities and a new specific data collection on ports was not foreseen then as the existing one was well established with a legal basis[[5]](#footnote-5). Eurostat maintained a list of ports, published in an Official Journal[[6]](#footnote-6). This list also included inland water ports. Furthermore, there was no firm definition of a "maritime port" either in Eurostat or DG MARE. However, a geodataset 'Ports, 2013 – Transport Networks', containing a point feature class with the location of 2 440 pan-European ports was available at GISCO (the Geographic Information System of the Commission)[[7]](#footnote-7). However, only geographical information was available in this dataset and it was unable to distinguish in a systematic way between maritime and inland ports. The complexity of defining what a maritime port is could be is explained through the following empirical evidence. In the case of fjord or delta coast types, cities that are undoubtedly linked to the sea in terms of economy, biodiversity and perception of the people living there can be classified as non-maritime due to the specificity of the coastline. One of the best examples to explain the issue is the case the port of Hamburg. Map 2 presents the situation that if we draw even a 50 kilometres buffer around the city polygon of Hamburg, then it would not be classified as a coastal city due to the factors described above.

Map 2. The case of Hamburg



To overcome all the limitations mentioned above, Eurostat proposed a solution based on the integration of different geographical layers, reprocessing of existing data from various domains other than the maritime transport unit and reuse of the official territorial typology of coastal areas[[8]](#footnote-8).

One of the traditional data collections of Eurostat is on cities. It is based on grants- and gentlemen’s agreements with the National Statistical Institutes (NSIs). The geographical scope of the City data collection is more than 900 cities and their Functional Urban Areas (FUAs) (as defined in the TERCET regulation). Data for 197 different variables in the domains of demography, social aspects, economic aspects, training and education, environment, tourism and culture are available in EuroBase[[9]](#footnote-9). Some of these cities are coastal cities/maritime ports. This customised typology of coastal cities needs to be defined in order to enable DG MARE to reuse the wealth of data available in our city database.

This goal has been achieved through a simple GIS operation. In a GIS environment, the list of the cities having at least 50% of their surface area in a zone covered by coastal areas (LAU-2) were exported and provided to DG MARE together with the labels and codes used in the city database. The result is that we have 255 coastal cities in Europe. However, since the city data collection is based on grant and gentlemen's agreements, this also contains the risk that data for part, or all of a specific city and/or for a particular reference year might not be available. For instance, city population data coverage is close to 100%. However, the 2015 data for the cities in Sweden, Norway, Ireland, France, Greece and Denmark are still missing from Eurostat's database. For 2015 for the 204 coastal cities reported currently in the database, we observe more than 46 million people which is around 9% of the European population.

While being a technically simple solution, it shows the advantage of different specialists working together under a common goal of overcoming obstacles.

**4. Conclusion**

These two examples show how GIS can support data reengineering and the production of customised territorial typologies in order to address specific user requests in a simple, yet efficient way. The combination of inside knowledge of statistical data and geospatial information has proven to be essential to support policy decisions in the impact assessment of the Common Agricultural Policy domain as well as providing reliable statistical estimates for the blue growth.

**5. References**

Barrington-Leigh C, Millard-Ball A (2017) The world’s user-generated road map is more than 80% complete. PLoS ONE 12(8): e0180698. <https://doi.org/10.1371/journal.pone.0180698>

Jackson SP, Mullen W, Agouris P, Crooks A, Croitoru A, Stefanidis A. Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. ISPRS International Journal of Geo-Information. 2013;2(2):507–530.

**5. Acknowledgements**

We are grateful for assistance from Michael Harrop and Jorge Calvo and for helpful suggestions from Teodora Brandmueller of Eurostat's Unit E4 - Regional Statistics and geographical information.

1. Regulation (EU) 2017/2391 of the European Parliament and of the Council of 12 December 2017 amending Regulation (EC) No 1059/2003 as regards the territorial typologies (Tercet): <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R2391> [↑](#footnote-ref-1)
2. Link to the data by Degree of Urbanisation in Eurostat's Database: <http://ec.europa.eu/eurostat/web/degree-of-urbanisation/data/database> [↑](#footnote-ref-2)
3. Link to the data by Urban-rural typology in t Eurostat's Database: <http://ec.europa.eu/eurostat/web/rural-development/data/database> [↑](#footnote-ref-3)
4. http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts#nuts13 [↑](#footnote-ref-4)
5. The statistics on maritime transport are collected within Directive 2009/42/EC and Commission Decision 2008/861/EC, as amended by Commission Decision 2010/216/EU of the European Parliament and of the Council of 14 April 2010, by Regulation 1090/2010 of the European Parliament and of the Council of 24 November 2010 and by Commission Delegated Decision 2012/186/EU of 3 February 2012. [↑](#footnote-ref-5)
6. COMMISSION DECISION of 29 October 2008 on rules for implementing Council Directive 95/64/EC on statistical returns in respect of carriage of goods and passengers by sea (2008/861/EC) [↑](#footnote-ref-6)
7. <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/transport-networks> [↑](#footnote-ref-7)
8. <http://ec.europa.eu/eurostat/web/maritime-policy-indicators/methodology> [↑](#footnote-ref-8)
9. <http://ec.europa.eu/eurostat/web/cities/data/database> [↑](#footnote-ref-9)