United Nations ECE/CES/2020/27

Economic and Social Council

Distr.: General
20 April 2020

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixty-eighth plenary session
Geneva, 22-24 June 2020
Item 5 (b) of the provisional agenda
New roles for national statistical agencies and geospatial agencies in emerging national data ecosystems: Session 2: Experiences and results of concrete steps already taken by NSOs and the geospatial communities to modernize their role

Geo-enabling statistical production: from design phase to dissemination[1]

Note by Statistics Portugal

Summary

This document describes the experience of Statistics Portugal in geo-enabling statistical production. It provides an overview on specific projects and outputs developed based on geospatial data, analysis and tools and implemented across the different phases of the statistical production model, namely from the design phase up to dissemination.

Building upon these experiences, Statistics Portugal's involvement in pan-European forums and on national geospatial data production and usage, the paper also presents the main challenges associated from bringing geospatial data into statistical data production, followed by a set of recommendations on how to address them.

This document is presented to the Conference of European Statisticians seminar on "New roles for national statistical agencies and geospatial agencies in emerging national data ecosystems" for Session 2: "Experiences and results of concrete steps already taken by NSOs and the geospatial communities to modernize their role" for discussion.

[1] This document was scheduled for publication after the standard publication date owing to circumstances beyond the submitter's control.

GE.20-05835(E)

2005835

I. Introduction

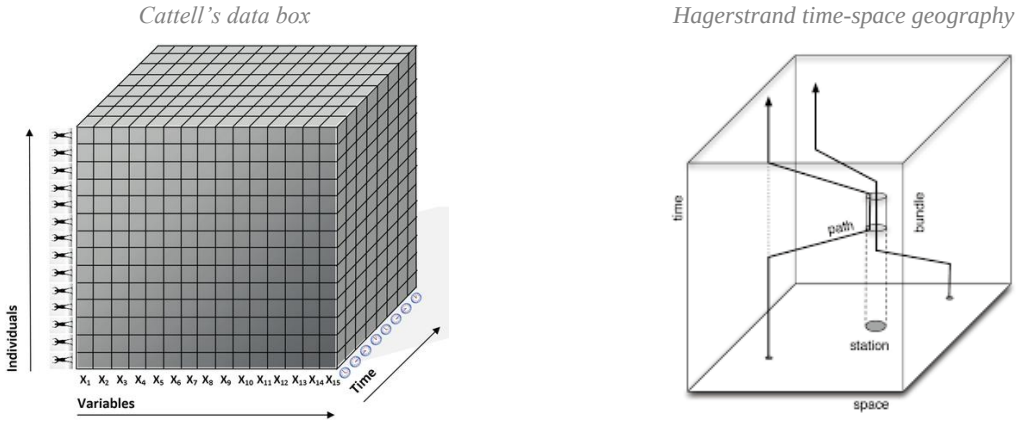
1. The paradigm of data production has been adapting to the changes operating in society, namely how, through technology, individuals, organizations and even other objects, interact with each other, leaving an increasing amount of digital traces. This digitalization of society means that increasingly movements, actions and transactions made are registered through some digital device or sensor making it possible to know WHAT is happening and WHEN it occurred, but also WHERE it is taking place.

2. *Space* (like time) is an essential component of statistical production. To address this fundamental data dimension, the use of geospatial data to properly capture the location element in the different phases of statistical production is essential – from the design phase, to data collection and management, up to the dissemination phase, to structure and map statistical results and to allow a territorial visual perception of data.

3. Within the scope of the statistical data production model, and having the Generic Statistical Business Process Model (GSBPM) as the background framework, *space* can have three critical dimensions (Cordeiro *et al.*, 2012): i) space is a fundamental dimension to *organize* data collection, storing, integration, analysis and dissemination of official statistics; ii) it becomes *context* meaningful as events captured at a specific territorial segmentation vary according to the territorial arrangements used to portray statistical results; and iii) it becomes itself statistical information as it *explains* and conditions the phenomenon at hand.

4. Data integration is on the verge of moving from a stovepipe model of statistical production to a horizontal and more flexible model of production that promotes a faster and higher quality response to emerging cross-cutting issues, including greater spatial granularity. Geospatial information plays a major role in this statistical production transformation, by allowing accurate data linkage and (spatial) data matching for the integration of different types of sources – from both public and private administrative information to big data and Earth Observations (EO). As a vision, it implies replacing the traditional data models, centred on the statistical project and on a specific population reference, by complex relational data models which integrate different thematic domains, based on the interaction of agents centred on their activities performed in space and time [Figure 1].

Figure 1
Traditional data model of a statistical project and theoretical scheme of agents relation in time and space



5. The combination of data types, ranging from traditional data sources, such as surveys, to administrative data and, more recently, to big data constitutes one of the key dimensions in data ecosystems. Adding value to data, by bringing together the expertise of statisticians, geospatial analysts, and data scientists, is essential to build an environment of statistical production that is able to track down the changes operating in society and provide official statistical data to monitor them. In terms of infrastructure that means to work in more flexible, digital and adaptive environments, by increasingly making use of open data, open source software, APIs, cloud storing systems, data hubs and shared platforms where location

attributes are central. Moving towards a more intensive and integrated use of administrative data and other types of data is at the core of Statistics Portugal's strategy to develop a National Data Infrastructure (NDI), where geospatial data, analysis and tools are playing a crucial role.

6. The integration of geospatial data into official statistics production model has shown to increase the value of the statistical information being produced and disseminated. Based on Statistics Portugal's experience of bringing geospatial data into the different phases of the statistical production model, the aim of this paper is to contribute to the discussion on the new roles for statistical and geospatial agencies in moving towards an integrated production approach. Using specific projects and outputs developed and disseminated by Statistics Portugal, this paper will reflect on the challenges and present recommendations for greater integration of geospatial information and tools within the statistical production chain.

II. Bringing geospatial information into statistical data production

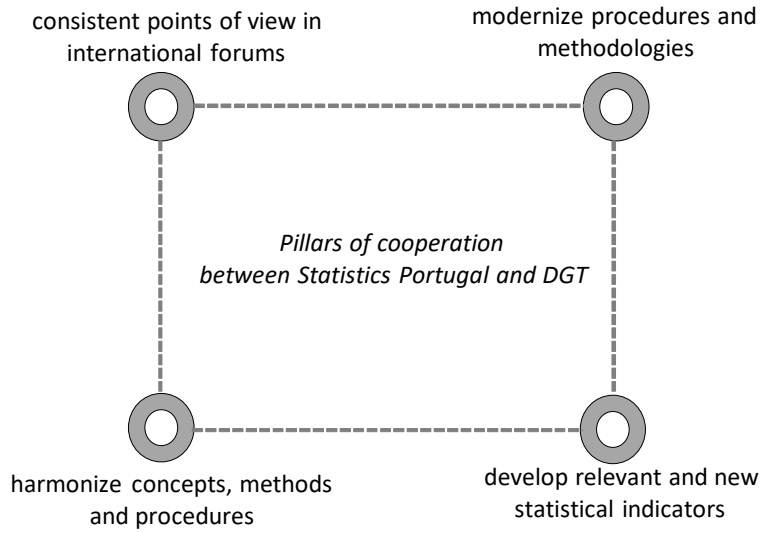
7. Geography has long been part of statistical production, especially to support the preparation and implementation of large statistical operations, such as Population and Housing Census. In Portugal we can refer back to the beginning of the previous century the use of maps to support the dissemination of official statistics. Coming to more recent days it is worth mentioning the use of cartography associated with the 2001 census. At that time, a "Geographic Information Referencing Base" (BGRI 2001) was developed based on Geographic Information Systems (GIS). For the 2011 Census round, an updated BGRI was created (BGRI 2011), which was an important tool to collect, for the first time, the x, y coordinates for all census buildings, and to establish a point-based database. This type of data was a crucial input to produce the 2011 Portuguese population grid, as part also of the European Statistical System (ESS) project GEOSTAT 2 and the dissemination of the 2011 European population grid – GEOSTAT 2011 grid dataset referenced to the 1 km^2 INSPIRE grid net (ETRS89-LAEA-1K).

8. The INSPIRE Directive (in force since May 2007) has also been playing an important role in harmonization of spatial data for relevant data themes and Statistics Portugal has been involved in five out of the 34 INSPIRE data themes, namely geographical names; buildings and addresses – which are central for the households register and to implement more data linking and data matching processes; statistical units; and population distribution and demography. In Portugal, the implementation of the INSPIRE Directive is coordinated by the Portuguese National Mapping and Cadastral Agency (NMCA), the Directorate-General for Territory (DGT).

9. As a way to increase interoperability between geospatial and statistical data, Statistics Portugal has been working closely with the Portuguese NMCA (DGT), and since 2015, has established a Memorandum of Understanding (MoU) that foresees four main pillars of cooperation, as shown in **Error! Reference source not found.**.

10. Besides contributing to broaden the scope of geographical and statistical integration within statistical indicators design and production, the MoU also provides a context for modernisation and harmonisation of concepts and methodologies, bearing in mind the need to meet the quality standards of statistical production.

Figure 2
Four pillars of cooperation between Statistics Portugal and the Directorate-General for Territory



11. One of the international forums in which Statistics Portugal, in articulation with the Portuguese NMCA (DGT), has been actively participating in is the UN-GGIM: Europe Working Group (WG) on Data Integration. This WG has been dedicated to geo-enabling the sustainable development indicators and in May 2019 has published the report, led by Statistics Portugal, on *The territorial dimension in SDG indicators: Geospatial data analysis and its integration with statistical data*. One of the main statistical outputs for Portugal, resulting from the work developed under the scope of this report was the calculation and dissemination of a *proxy* for SDG indicator 11.3.1 *Ratio of land consumption rate to population growth rate*, based on the Land Use and Land Cover Map (COS) produced by the Portuguese NMCA (DGT). As part of the WG 2019-2022 work plan, Statistics Portugal will continue to lead the task stream dedicated to geo-enable the SDG indicators, focusing on environment related SDG indicators and on the use of EO derived data.

12. The use remote sensing data for statistical purposes has a long history, especially for agricultural statistics (UNECE, 2019). In 2015, within the framework of the MoU, Statistics Portugal and DGT conducted a pilot study (ESS grant[2]) to explore remote sensing data and additional national data sources to produce land cover statistics at NUTS 3 level, as an alternative approach to LUCAS which is based on *in-situ* data collected by surveyors (Costa *et al.*, 2018). Presently, Statistics Portugal is also participating in an ESSnet on Big Data Work Package on EO, mainly on satellite data and aerial photography (Sentinel data) with the aim of defining a geospatial framework for data breakdown between statistical and geographical information, focusing on data availability and conditions of access relevant for statistical domains, such as agriculture, forestry or settlements enumeration.

13. Taking advantage of the increased and diversified use of GIS technology within statistical offices, Statistics Portugal's medium-term strategy focuses on the need to promote a greater interoperability between spatial and statistical data to support statistical production and to promote spatial and statistical integration to produce new statistical indicators, in a permanent effort to introduce the spatial perspective across the different phases of statistical production, as showcased by the following projects and outputs developed across the different phases of statistical production.

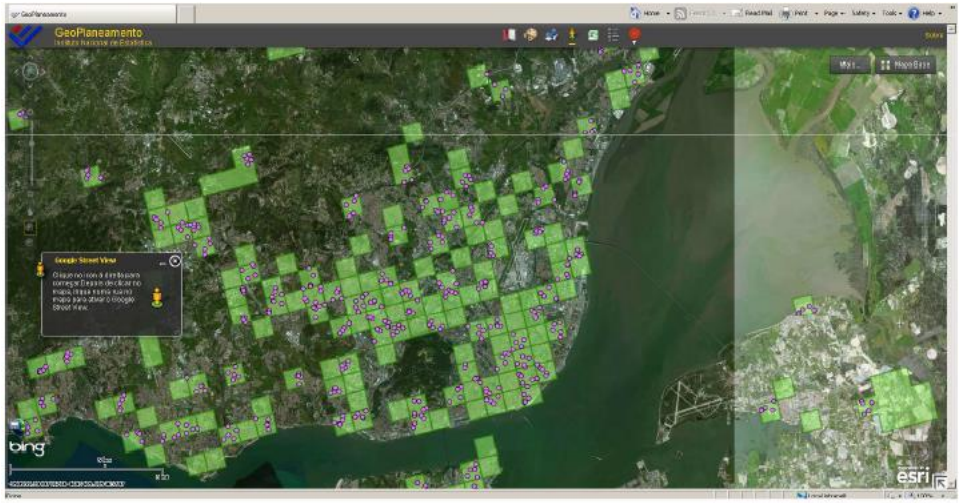
[2] EUROSTAT/Contract No: 08441.2015.002-2015.724 - Provision of Harmonised land cover/land use information: LUCAS and national systems.

A. Using spatial sampling design

14. Under the scope of a strategy to increase the modernization and efficiency of statistical production, through its methodological and technological development, Statistics Portugal has put into practice a new methodology to define sampling frames and sample design. Taking advantage of the georeferenced information (x, y coordinates) for all the 2011 Census buildings, a National Dwellings File has been defined to support the sampling process for household surveys, regularly updated through administrative data. An important geospatial instrument has also been integrated in this process, the European 1 km^2 grid (INSPIRE grid net ETRS89-LAEA-1K) as a new reference for PSU (Primary Sampling Unit) selection.

15. Usually, sampling selection follows a stratified and multi-stage sampling scheme, in which the primary sampling units (PSUs), geographically constituted by one or more contiguous cells of the 1 km^2 [Figure 2], are systematically selected with a probability proportional to the size of the number of dwellings of usual residence; the secondary sampling units (SSUs) are systematically selected within the units of the first step. All the PSU of sampling frames for surveys with rotations must include roads.

Figure 2
Example of grid cells selection to define PSUs



16. Using this spatial sampling design has allowed to reduce the intra-cluster correlation coefficient (which measures the similarity of statistical units) associated with selecting dwellings in "segments".

17. A georeferenced sampling frame has shown to improve the accuracy of estimates. The more the sampling design selects individuals geographically distant from one another, the more the estimation will be precise for a spatially auto-correlated variable (Favre-Martinoz *et al.*, 2018). Additionally, in case of face-to-face interviews knowing the location of the statistical units sampled makes it easier to identify them in the field and to manage interviewers' locations during the fieldwork. Maintaining the underlying point-based data update is crucial to increase the efficiency of the spatial sample design process, as well as of data collection.

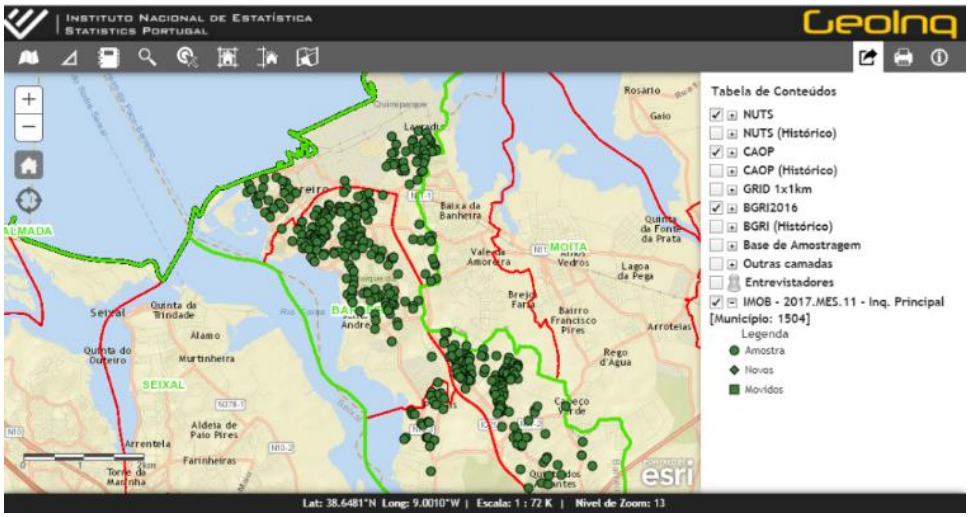
B. Increasing efficiency in data collection management with geospatial tools

18. One key area of statistical production refers to data collection. Developing procedures and tools that make it easier for survey respondents to provide information, while at the same trying to reduce response burden, is an important goal. But working on solutions that make it easier for interviewers to conduct their work in the most productive way possible is also a fundamental dimension to increase efficiency in data collection.

19. At Statistics Portugal, interviewers regularly faced difficulties in locating their sample housing units in household surveys as they could only rely on tables with address information, name and contacts of the household representative. A geospatial web tool, custom-designed to respond to the needs of statistical data production, was implemented, within the scope of integrating geospatial data into the official statistics' production model.

20. The GeoINQ web application [**Error! Reference source not found.**] was developed by Statistics Portugal in partnership with ESRI using an API for ArcGIS environment. The tool integrates point-based data for households of sampling frames and a set of relevant background geospatial layers (NUTS, Administrative Units, 1 km^2 grid, BGRI) and base maps, including the orthophotomaps from the Portuguese NMCA (DGT).

Figure 4
GeoINO web application



21. With GeoINQ interviewers can easily identify the precise (x, y) location of dwellings and have access to associated data. GeoINQ runs on mobile devices and users can only access those features and geographical layers compatible with their user profile.

22. GeoINQ is fully integrated with other systems developed at Statistics Portugal, in particular with the global interview survey management system (SIGINQ-IE). Therefore, besides interviewers, other internal users make use of this web application to meet their needs on data management and analysis, namely to analyse the geographical dispersion and overlap of samples on national territory within the process of spatial sampling design, as described in the previous section; and to support and manage interviewers in their fieldwork, including sample allocation. Maintaining the underlying geospatial data updated is, in this context, fundamental to keep benefiting from the useful features associated with this type of geospatial tools supporting statistical data production.

C. Implementing geo-solutions to capture challenging variables

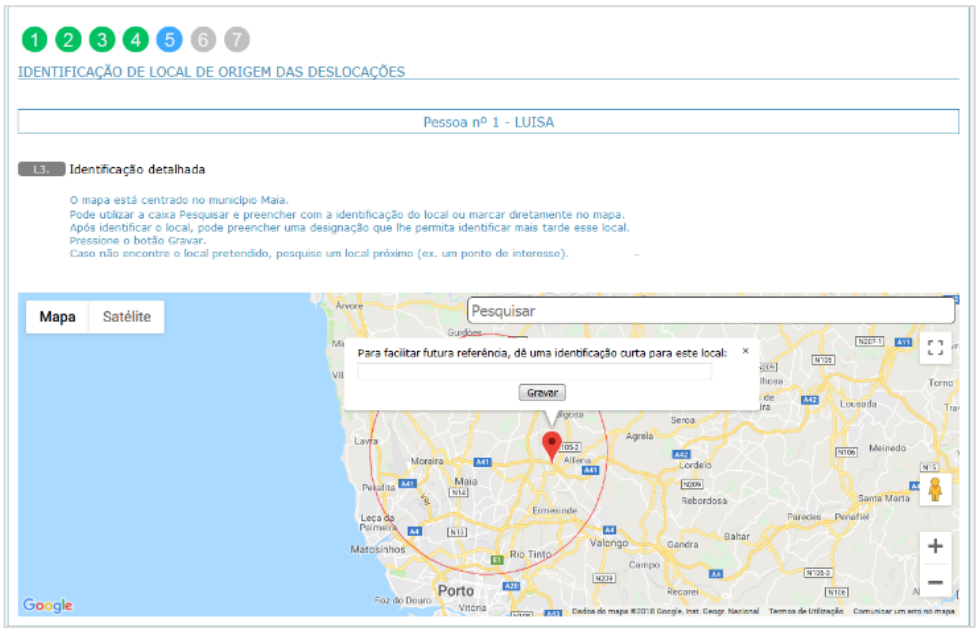
23. In 2017, Statistics Portugal conducted a survey on mobility in the two Portuguese metropolitan areas – the Metropolitan Area of Oporto and the Metropolitan Area of Lisbon. Based on a stratified and multiphasic random sample, which considered homogeneous areas of accessibility to transport, a mix-mode data collection approach was followed, by combining Computer Assisted Web Interview (CAWI) and CAPI (Computer Assisted Personal Interview).

24. The aim of the survey was to characterise the movements (not limited to commuting) of the resident population (6-84 years old) in the two metropolitan areas, which involved being able to capture points of origin and destination for each trip during a specific day of the week, as well as other dimensions in order to understand how people move, how often they travel, how much time they spend moving, where they go and to do what. The main challenge associated with designing a web-survey to meet this aim was to come up with a

way people could easily register their movements during the day and find/pinpoint the locations where they went to.

25. Instead of descriptive reports, an innovative solution was implemented using Google Maps. The maps were used to capture travel destinations with the same functions people are used to finding in Google Maps, as well as location circles based on the centroid of the municipality to the farthest point to help people navigate the different locations [see **Error! Reference source not found.**].

Figure 5
Example of the response screen for identifying locations on the Survey on mobility in metropolitan areas



26. Nevertheless, outsourcing services for statistical purposes is not exempt from an assessment of their basic assumptions in order to ensure that they meet the quality criteria for statistical production. This assessment may be more limited for commercial bases and products. In addition, it also implies being dependent on external services with limited capacity for intervention and being subject to changes that may direct or indirectly affect implemented statistical production processes.

D. Producing statistical indicators to monitor SDG at the territorial level

27. Recently, the 2030 Sustainable Development Agenda (United Nations, 2015) and the definition of 17 Sustainable Development Goals (SDGs) to be monitored by 232 indications have emphasized the importance of geographical disaggregation of data (such as, urban *vs.* rural), along with other segmentations, in order to cope with the motto of *leaving no one behind*. At the European level, an indicator set has been established to measure progress towards the SDGs in an EU context (Eurostat, 2019). Statistics Portugal has put together the information available for Portugal according to the global SDG monitoring framework[3]. Since 2018, an annual report has also been published (e.g., INE, 2019) with a brief analysis of the performance of each available indicator (from 2010 up to the most recent year), including data with geographical breakdown at regional (NUTS 2 and 3) and municipality level.

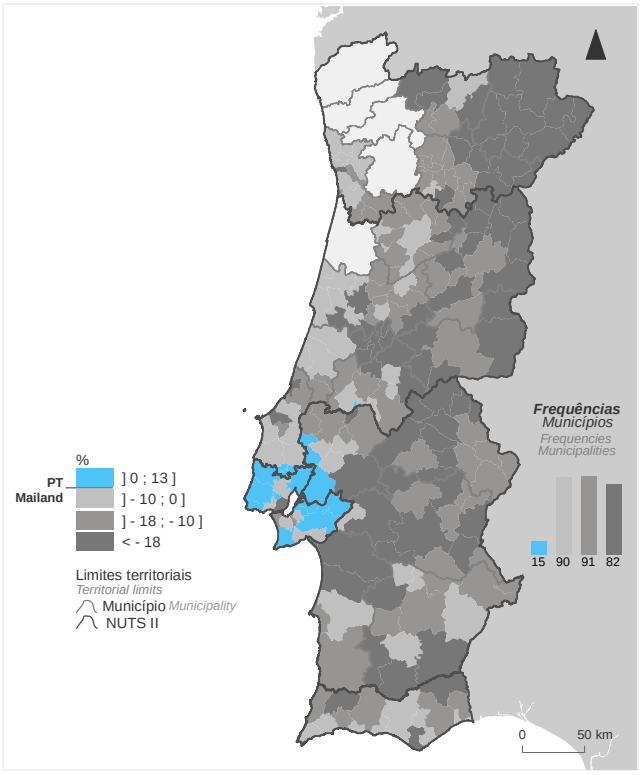
[3] A dedicated section to the SDGs has been published at Statistics Portugal website.

28. In the case of Portugal, some SDGs have a lower coverage of statistical indicators, especially if the monitoring framework includes tier II and tier III indicators[4]. Therefore, trying to increase the scope of SDG indicators available, particularly at the territorial level, has been a relevant task tackled by Statistics Portugal. Specifically, progress has been made to increase the scope of information for the monitoring of Goal 11 on sustainable cities and communities, resulting from the integration of geospatial and statistical data and geospatial analysis.

29. In 2018, Statistics Portugal published a new set of Land Use and Land Cover Statistics (LCLUStats) based on the Land Use and Land Cover Map (COS) produced by the Portuguese NMCA (DGT) using photo interpretation of orthorectified aerial images. The LCLUStats includes the calculation at municipality level of a *proxy* to SDG 11.3.1 tier II indicator (*ratio of land consumption rate to population growth rate*) based on the Land Use Efficiency (LUE) formula (Corbane *et al*., 2017) as proposed by the Joint Research Centre (JRC),. The LUE combines data from COS and from the annual resident population estimates for the reference years of COS - 2010 and 2015. The results are normalized for a ten year period.

30. The result for Portugal's mainland, for the period 2010-2015, was -10%. Only 15 municipalities, mainly located in the Metropolitan Area of Lisboa, scored positive LUE values, i.e., an increment of population faster than the increase of artificial land. A group of 90 municipalities, located mainly in the coastal area of Norte and Centro regions, scored a decrease on the LUE, but still less significant than the average value for Portugal's mainland (-10%) [**Error! Reference source not found.**].

Figure 6
LUE by municipality 2015



[4] At the global level, indicators have been classified according to three tier system regarding data availability and established methodology; i) tier I indicators have an established methodology and data are already widely available; ii) tier II indicators have an established methodology but data are not easily available; and iii) tier III indicators have not yet an internationally agreed established methodology.

31. As this constitutes the first statistical operation disseminated by Statistics Portugal based on a geospatial data source and on its integration with statistical data, its dissemination comprised a few challenges in order to accommodate geospatial data and analysis according to the standard statistical methodological document, which describes all the procedures, concepts and classifications associated with a statistical operation.

E. Using open source geospatial tools to measure accessibility to services

32. Accessibility to services is a relevant dimension to measure people's well-being and quality of life, which have become important dimensions of assessment at policy level in order to better capture the progress of society and of people's living conditions (e.g., OECD's *How's life initiative*). The 2030 UN agenda for sustainable development also emphasises accessibility as a relevant dimension to monitor Goal 11 on sustainable cities and communities and has included an indicator on accessibility to public transport for its monitoring, but that has been defined as a tier II indicator, meaning that a methodology has been established to calculate this indicator, but data are not easily available.

33. Under the scope of a European Statistical System (ESS) grants on sub-national statistics (Urban Audit, 2017-2019[5]), Statistics Portugal has developed a task dedicated to increase the knowledge on measuring accessibility indicators. The task focused on accessibility to schools and experimental measures of territorial and population coverage were calculated by considering walking and car distances from the school location isochrones of time, defined between 5 and 40 minutes with time intervals of 5 minutes [Figure 3]. These service areas were calculated using open source data and software, namely *Open Street Map* (OSM) navigation network through *Open Route Service* (ORS) plug-in in Quantum GIS environment. The proportion of territorial (surface area) and population (point-based 2011 Census data) covered by schools was calculated for different territorial units, including at grid [**Error! Reference source not found.**] and municipality level [Figure 4].

Figure 3
Service areas of basic education institutions between 5-40 minutes walking distance

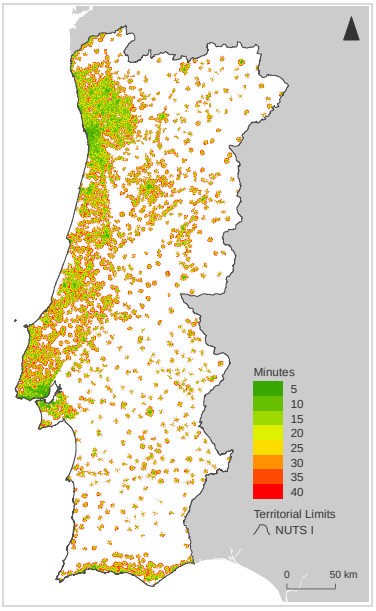
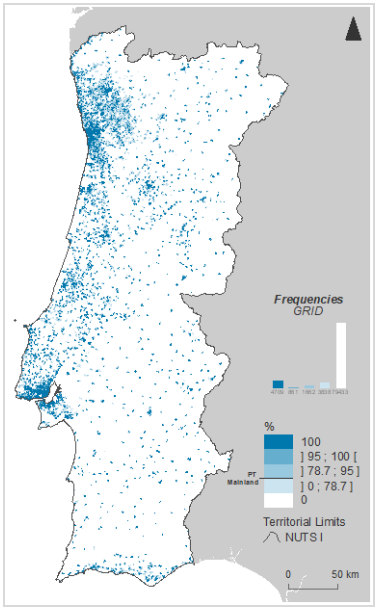


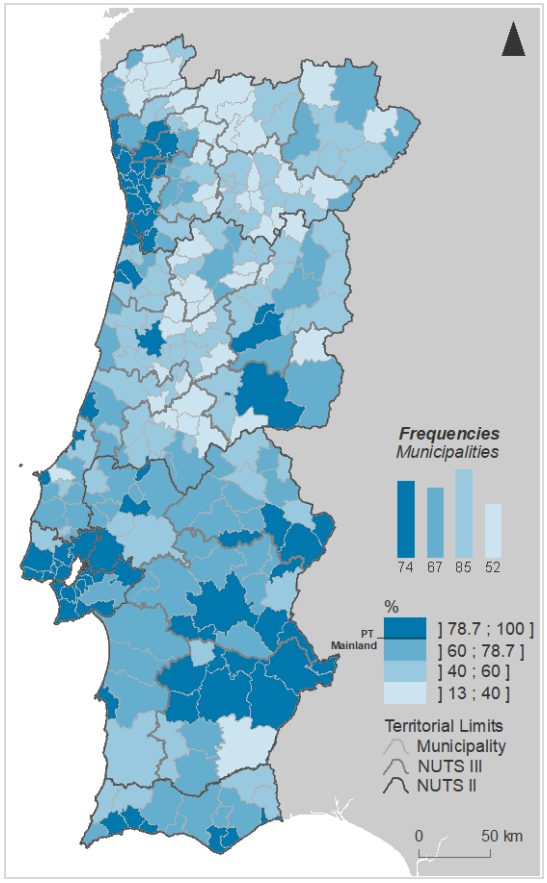
Figure 8
Population coverage of basic education institutions at 15 minutes walking distance by 1 km² grid



[5] EUROSTAT/Contract No: 08142.2017.002-2017.432 - Data collection for sub-national statistics (mainly cities).

Figure 4

Population coverage of basic education institutions at 15 minutes walking distance by municipality



34. Given the experimental nature of these accessibility indicators, and aiming at determining data quality, a comparative analysis was carried out. Some results, for the same origins and destinations, were compared with other available solutions, and it was possible to observe that walking distances seem to be more robust than the ones by car. Therefore, and despite the fact that the use of open source GIS data and tools made it possible to overcome the absence of an updated official navigation network for the context of Portugal, it is important to benchmark the results obtained with other sources in order to assess the consistency and robustness of the results obtained, aiming at producing official accessibility statistical indicators.

F. Creating geo-based data visualization tools

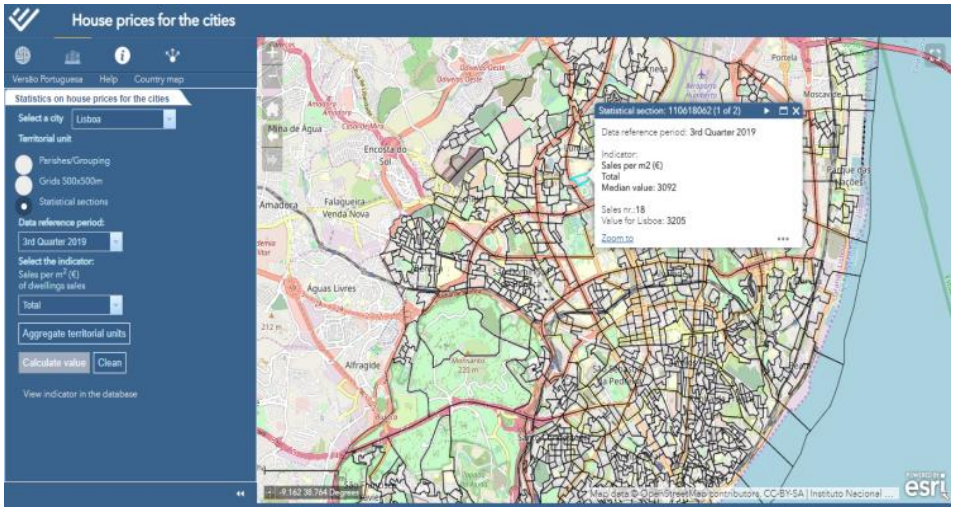
35. Following the international financial and economic crisis, there has been an increasing need for territorial information on housing prices to monitor the changes that have been taking place in the housing market in Portugal. At the EU level, Eurostat has also been working with Member-States to develop statistical tools for the analysis of the evolution of the real estate market, namely housing (Eurostat, 2018).

36. In 2017, Statistics Portugal began the dissemination of quarterly statistics on house prices at local level based on geo-referenced administrative tax data, namely the Municipal Property Transfer Tax (from where the transaction prices are obtained) and the Municipal Property Tax (from where identifying characteristics of the transacted dwelling are obtained, including x, y coordinates and the smallest Local Administrative Units (LAU) - parishes) provided by the Portuguese Tax and Customs Authority.

37. Besides the regular dissemination of statistical indicators according to common territorial units (NUTS, municipality and parish level), Statistics Portugal also aimed at providing a tool that would allow users to browse information according to a more detailed geography at the local level. The solution was found through a geo-based data visualisation tool that provides users with the possibility of customizing their search on house prices based on different geographies. Data for the web application only includes registers with valid x, y coordinates, after a validation procedure has been conducted and complementary information from the Portuguese Energy Agency (ADENE) has been linked (using the 'Tax Authority dwelling code' variable) with IMT and IMI data. Geo-coordinates and LAU coding are also validated based on the Official Administrative Map of Portugal (CAOP).

38. The 'House prices – Cities' tool was developed using an API (JavaScript) for ArcGIS environment and is compatible with mobile devices. This web application tool allows to search for median prices of dwellings sales (€/m^2) for the seven Portuguese cities with more than 100 thousand inhabitants – Lisboa, Porto, Vila Nova de Gaia, Amadora, Braga, Funchal and Coimbra [Figure 5].

Figure 5
House prices for cities web application tool - Lisboa



Source: Statistics Portugal, House price statisics at local level

39. Users can browse and customize their data selection by parish level, statistical section (Census 2011 geography) and by a 500m x 500m grid. For statistical sections and grids, results refer to a minimum of seven transactions.

40. The house prices web application is one of the most consulted products of Statistics Portugal, which is indicative of its usefulness and responsiveness to users' needs. Given the relevance of x, y coordinates for data tabulation at city level, the implementation of validation procedures, including consistency with administrative division units and the use of auxiliary data sources (ADENE – National Agency for Energy) are essential to increase the scope of data availability and to ensure data quality. Additionally, the level of data granularity implies a careful assessment of the data reliability, and the median was taken as the parameter of reference for the dissemination of house prices at local level to better cope with highly asymmetric distributions and confidentiality issues raised by the possibility of custom data selection according to territorial arrangements defined by users.

III. Challenges and recommendations

41. Building on the previous examples and on Statistics Portugal's overall experience in bringing geospatial information into statistical data production, a number of challenges,

associated to the statistical principles as defined by the *European Statistics Code of Practice*[6], which could also be relevant for the context of other countries and for the global European context, are presented below, followed by recommendations on how to address them.

A. On meeting the statistical principles on Commitment to Quality (principle 4)

42. The use of geospatial data, analysis and tools, and its integration with statistical data, has been opening up the possibilities of deriving new relevant information to address cross-cutting issues and to respond to global challenges as is the case of the SDGs monitoring framework. Nevertheless, data quality must be assured when making use of non-official data sources and tools, whether they are commercially based (as in the example presented of using a Google Maps API to capture locations and calculate distances for the Survey on mobility) or open source (as in the case of OSM data and ORS tool for QGIS to calculate indicators of accessibility to schools). Testing for data stability consistency and reliability, by carrying out a comprehensive metadata report and by benchmarking results are, in this context, essential steps. Furthermore, the use of these geo-based analytical tools and sources highlight the convenience of having well documented and certified official geospatial data and tools to produce statistical results.

B. On meeting the statistical principle on Sound Methodology (principle 7)

43. The range of geospatial information within the scope of statistical operations is not limited to geospatial data collected by Statistics Portugal or by the NMCA (DGT). Several other public administration entities produce relevant geospatial data as a result of pursuing their activities. Nevertheless, different methodological approaches come into play in this regard, which hinder and compromise data compatibility and interoperability. This is the case, for example, of the point-based data used for *House prices at local level* (based on data from the Portuguese Tax Authority) and the georeferenced 2011 Census data on buildings (produced by Statistics Portugal) which are not compatible, neither on coding systems or on geo-referencing standards. Coordination on this regard is thus essential at National and European levels.

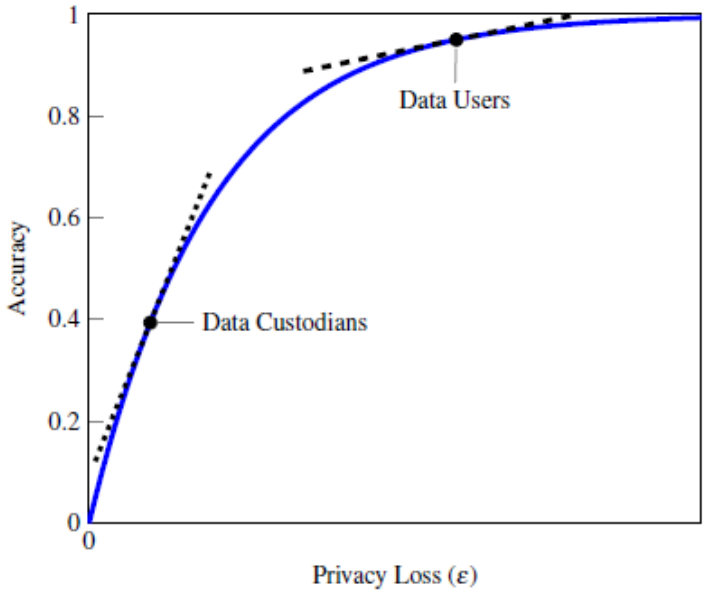
C. On meeting the statistical principle of Statistical Confidentiality and Data Protection (principle 11)

44. Increasing data granularity and the production and dissemination of data according to high-detailed level geographies, including the possibility of selecting specific territorial arrangements, as in the case of geo-based visualization tool for dissemination of house prices statistics at local level (House Prices – Cities), constitutes a challenge in maintaining data confidentiality. A critical assessment must be put into practice in order to guarantee data protection, while trying to meet users' data needs [Figure 11].

45. On the other hand, increasing geospatial content within statistical data production may have a positive impact on statistical disclosure control methods and procedures in order to guarantee confidentiality.

[6] The Code (2017 revised edition) "has 16 principles concerning the institutional environment, statistical processes and statistical outputs. The Code aims to ensure that statistics produced within the European Statistical System (ESS) are relevant, timely and accurate, and that they comply with the principles of professional independence, impartiality and objectivity" (https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice).

Figure 11
The cost benefit between loss of privacy and information detail



D. On meeting the statistical principle of Coherence and Comparability (principle 14)

46. The availability of national geospatial data sources, and their integration with statistical data, provides an opportunity for countries to have statistical indicators and national typologies, with a higher territorial breakdown, that are relevant for the formulation and monitoring of territory-based policies. This, however, may imply conceptual and methodological differences from the regulation framework established by the European Statistical System (ESS) for a specific domain, which may compromise, in some cases, comparability with other countries. For example, national LCLUStats provide relevant detailed data, namely up to municipality level, on land use and land cover status and changes to inform national regional planning policies. These are derived from the national Land Use and Land Cover Map (COS), which relies on a different methodology than the one being used by the EU in the LUCAS Survey to provide harmonised and comparable statistics on land use and land cover for EU regions, but only up to NUTS 2 level.

Recommendations to geo-enable statistical production[7]

Harmonise common geospatial data themes at the European level, having in mind core data features for spatial analysis and data integration for statistical purposes (e.g. metadata, scales, attributes, accuracy) and following UN-GGIM: Europe core data recommendations, complementing INSPIRE data specifications by defining the priorities on the core content in order to fulfil user needs and address the SDGs.

Implement common key geospatial data themes, such as Buildings, Addresses, Land Use and Land Cover, Cadastral data, Transport networks, as authoritative data at the European level, with NMCAs assuming a relevant coordination role at the national level.

Ensure availability and access to geospatial data sources and tools for geospatial data processing, analysis and visualization at the European level as a way to geo-enable statistical production in a harmonized and consistent way across the Member States.

Increase harmonization and interoperability of geospatial data produced by national agencies under the scope of the definition and implementation of a National Spatial Data Strategy, bearing in mind the requirements for statistical data production.

Expand communication and articulation between geospatial data producers, statistical offices, data scientists and researchers to leverage National Spatial Data Infrastructure and geospatial and statistical data integration.

[7] These recommendations benefit from the discussions within UN-GGIM: Europe Working Group on Data Integration and specifically from their outputs (UNGGIM: Europe, 2019a and 2019b).

IV. References

Abowd, J. M., Schumtte, I. M., Sexton, W. & Vilhuber, L. (2019). Why the economics profession must actively participate in the privacy protection debate. American Economic Association Papers and Proceedings, 109: 397-402.

Corbane, C., Politis, P., Siragusa, A., Kemper, T. & Pesaresi, M. (2017). LUE User Guide: A tool to calculate the Land Use Efficiency and the SDG 11.3 indicator with the Global Human Settlement Layer. Luxembourg: Publications Office of the European Union.

Cordeiro, H., Vala, F. & Santos, A. (2012). Spatial Data Infrastructure for statistical production: challenges and opportunities. Paper presented at 98th Directors General of the National Statistical Institutes (DGINS) Conference, Prague, 24-25 September.

Costa, H., Almeida, D., Vala, F., Marcelino, F., & Caetano, M. (2018). Land cover mapping from remotely sensed and auxiliary data for harmonized official statistics. ISPRS International Journal of Geo-Information, 7(4), 1-21. [157]. DOI: 10.3390/ijgi7040157.

Eurostat (2019). Sustainable development in the European Union: Monitoring report on the progress towards the SDGs on an EU context. Luxembourg: Publications Office of the EU.

Eurostat (2018). Housing price statistics - house price index. Statistics explained, available at http://ec.europa.eu/eurostat/statisticsexplained/index.php/Housing_price_statistics_-_house_price_index.

Favre-Martinoz, C., Fontaine, M., Le Gleut, R. & Loonis, V. (2018). Spatial sampling. In V. Loonis and M.P Bellefon, Handbook of Spatial Analysis: Theory and Application with R (pp. 255-276). Montrouge: INSEE.

INE – Instituto Nacional de Estatística (2019). Sustainable Development Goals - Indicators for Portugal. 2030 Agenda. Lisboa: INE.

UN-GGIM: Europe (2019a). The territorial dimension in SDG indicators: geospatial data analysis and its integration with statistical data. Lisboa: INE.

UN-GGIM: Europe (2019b). The integration of statistical and geospatial information — a call for political action in Europe. Luxembourg: Publications Office of the EU.

UNECE (2019). In-depth review of satellite imagery / earth observation technology in official statistics. Prepared by Canada and Mexico for 67th plenary session, Geneva, 26-28 June.

United Nations (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Resolution A/RES/70/1 adopted by the General Assembly on 25 September 2015.