

# Merging statistics and geospatial information, 2014 projects - Portugal

Statistics Explained

This article forms part of Eurostat's statistical report on *Merging statistics and geospatial information: 2019 edition*

Final reports July 2015, June 2016 and February 2017

## Problem

Statistics Portugal (INE) identified a weakness in relation to their business register statistics insofar as these were not integrated into the point-based component of their in-house spatial data infrastructure, thereby impacting on the production of statistics for small areas and the availability of spatially-enabled datasets.

## Objectives

The main aim of this project was to perform the harmonised geo-integration of statistics based on the statistical business register, by way of a spatially enabled and quality [U+2010]controlled point-based infrastructure. The objective was to make use of: i) the existing spatial data infrastructure, available statistical resources, official registries and other administrative data sources; ii) the potential offered by geographical information technologies; and, iii) institutional relations with the National Mapping Agency (NMA) and municipalities.

The specific objectives were:

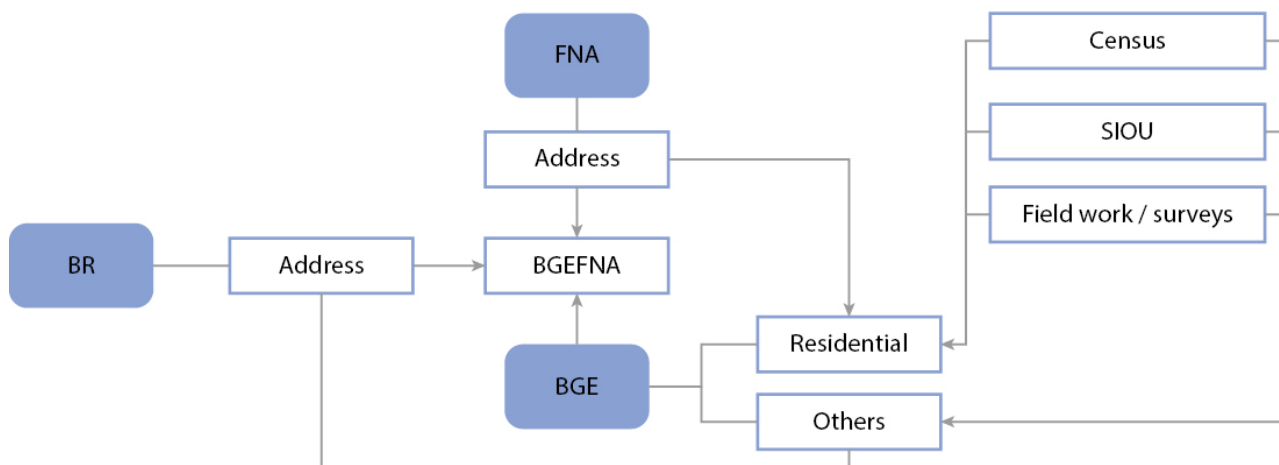
- to analyse the statistical business register concerning harmonisation of addresses and relevant attributes for geo [U+2010]processing;
- to define methodologies and technical solutions for geo-processing and data integration;
- to implement, validate, promote and disseminate a business register geographic database.

## Method

**Objectives 1 and 2** : analysis and methodology.

The starting point of the work was to analyse:

- the statistical business register (FUE);
- the national dwellings register (FNA) — the master file to which any address based system should be matched for the purpose of geo-referencing existing databases within the statistical office;
- the buildings geographic database (BGE) — this has a point-based coverage of buildings that can be used for residence and has been enriched to record other types of buildings with no residential capabilities and is dynamically maintained by municipalities through the indicators system of urban operations (SIOU) with respect to building permits and completed construction work.



**Figure 1: Inter-relationship of sources**

Analyses of these elements was carried out with the perspective of preparing the design and production of a business register geographic database (BRGD).

The common attribute of the three main databases was their address component. The quality assessment of the address component of the statistical business register (FUE) revealed the need to further enhance and harmonise it.

The methodology for the BRGD was established, defining the processes for data integration, data flows for maintaining the BRGD, the data structure required, and the technical specifications for building the planned applications. Concerning data integration, a variety of issues were assessed, including, among others, a harmonisation procedure for addresses and other required attributes, and a process for dealing with multipurpose buildings. The technical specification included, among others, the system architecture, main structure and specific data model for the BRGD, as well as the validation platform for municipalities to verify the accuracy and the completeness of the initial BRGD and tools for internal visualisation and for the external geoportal, with spatial query capabilities. The proposed applications were designed in accordance with existing technological resources in order to benefit from a common infrastructure supporting the BRGD, the validation platform and the dissemination applications, as well as the entire in-house spatial data infrastructure.

**Objective 3** : implementation, validation, promotion and dissemination.

The steps in this objective were to:

- build a partial test version of the BRGD following the methodology developed under the first action;
- develop a validation platform and perform the validation;
- develop an internal application for dissemination;
- develop an external application for dissemination.

Three target subsets of units were selected to test filling the BRGD: the units needed for the consumer price index sample; all units (regardless of activity) in a particular region (Oeste); and all units (regardless of region) in a particular activity (Subclass 68 322). The main problems faced in matching units were related to the quality of the address, for example because it was wrong, had changed, had been misspelled, was in a business area, or was written differently — various approaches were implemented to improve the matching process.

A validation platform (GeoBGE) was developed as an extranet application with restricted access rights (permission). Various cartographic elements from multiple sources were integrated into the application as basemaps (for example Bing maps and Open Street map data) to facilitate the location and recognition of the geography within different perspectives for the territory.

The validation of the results obtained through the test implementation step relies on the long-term effective relationship between the statistical office and municipalities to make use of their extensive knowledge of local entities. The validation was done, using the GeoBGE platform, for one municipality (Torres Vedras) within the Oeste

region. The municipality was asked to validate records where a match had been achieved, looking specifically at the location (point coordinates), address and economic activity of each record. Proposals by the municipality to make changes were then checked by the statistical office.

The development of the internal application (Geoplaneamento) was designed to assist the work of the methodology unit and the data collection and analysis unit within the statistical office. The internal application comprises not only administrative and statistical units (such as small statistical areas, NUTS, the location of the surveyors), but also includes the primary sampling units for samples referenced to households in a point-based approach supported by the BRGD. Some additional features, such as viewing Google Street View images or locators for household codes and addresses are included to help identify households.

The external applications aimed to give users direct access to statistical information. An example was the GeoEscolas application which used BRGD records for schools. The application provided information related to the surrounding area of a specific primary or secondary school. Spatial queries can be made of a 200, 500, 1 000 or 2 000 metre radius to obtain statistics related to buildings, dwellings, families, population (total and by age group).

## Results

### Objectives 1 and 2 :

The analyses and methodological developments related to data integration gathered all the relevant information to launch the activity to implement the development of the BRGD. Furthermore, the specifications for various applications were defined.

### Objective 3 :

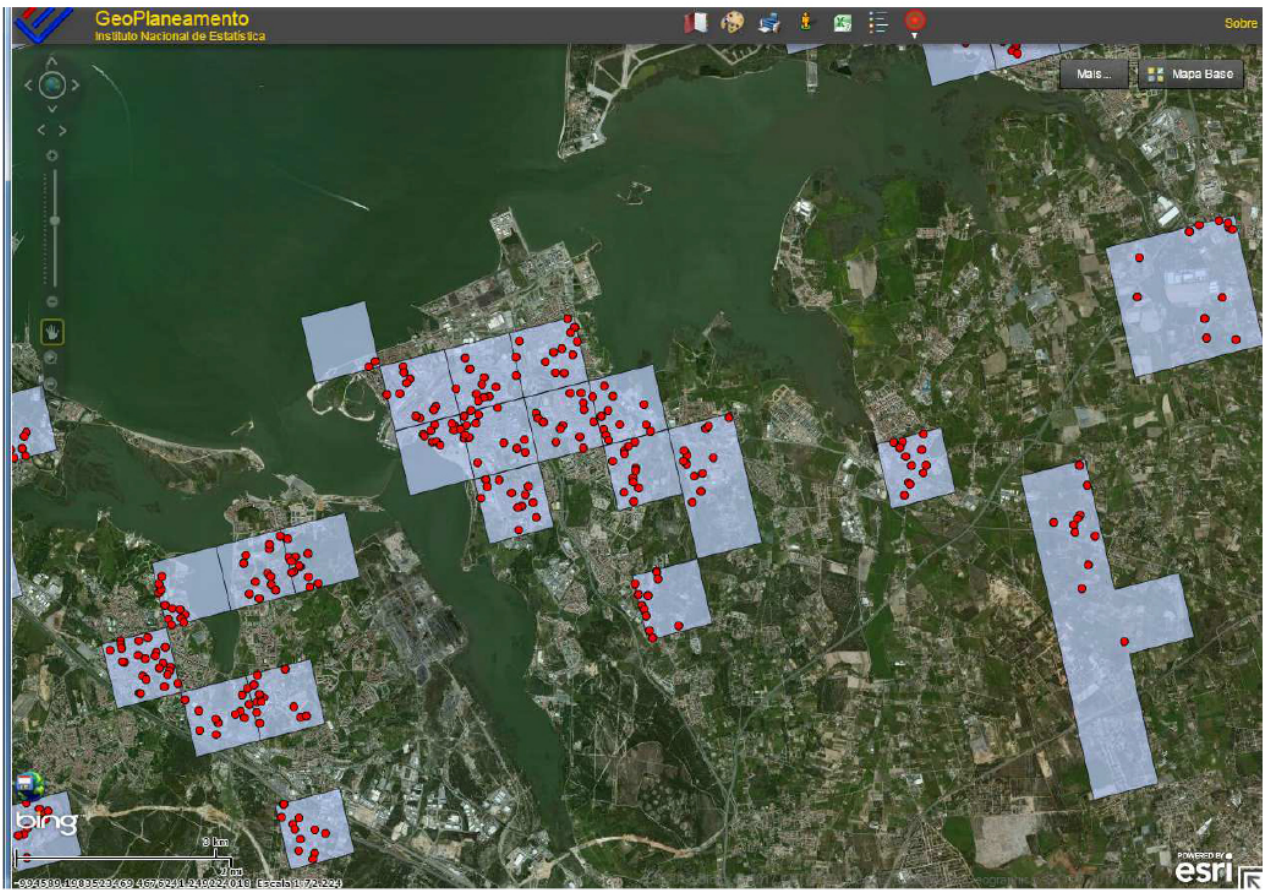
It was concluded that the statistical business register (FUE) requires further harmonisation for address components; its accuracy might also be improved through the process of building the BRGD. The efficiency of the complete address as the locator to match records was enhanced by use of a common structure created to relate to the various datasets. However, the complete address is insufficient for geo-referencing the statistical business register (FUE) and a step-by-step approach based on elements of the address was valuable to assist and complete the process. Experience was gained assessing the usefulness of various locators to match records.

The complete implementation of BRGD requires prior investment to improve the quality of addresses in the statistical business register (FUE). The statistical office adopted an internal regulation specifying the data model to be used for addresses in all databases and systems, the purpose of which is to improve the quality of addresses in the business register (FUE) and the national dwellings register (FNA).

The validation platform (GeoBGE) created for municipalities to validate the BRGD is a mapcentric fully interactive web application. Users can navigate by zooming in and out or dragging. They can identify data by querying the attributes of specific geographic objects. GeoBGE is customised for data editing and validation.

For the municipality of Torres Vedras, 90.3 % of records were validated with no changes, while for 0.2 % of the records changes were proposed to all three attributes (location, address and activity); the remaining 9.5 % of records had changes proposed for one or two of the attributes, most commonly concerning a change to their address. Overall, 98.0 % of matched records were confirmed in terms of their location, 91.0 % in terms of their address and 98.6 % in terms of the economic activity.

Geoplaneamento has been used by the data collection unit as it provides the exact location of those dwellings that were surveyed, supporting more efficient fieldwork planning and the identification of the closest surveyors available to collect data for specific dwellings. The methodology unit uses Geoplaneamento to analyse the dispersion and location of samples and the sampling frame for each survey.



**Figure 2: GeoPlaneamento**

The external web application was developed to be used by teachers and students and is freely available through the statistical office's website to promote statistical literacy.

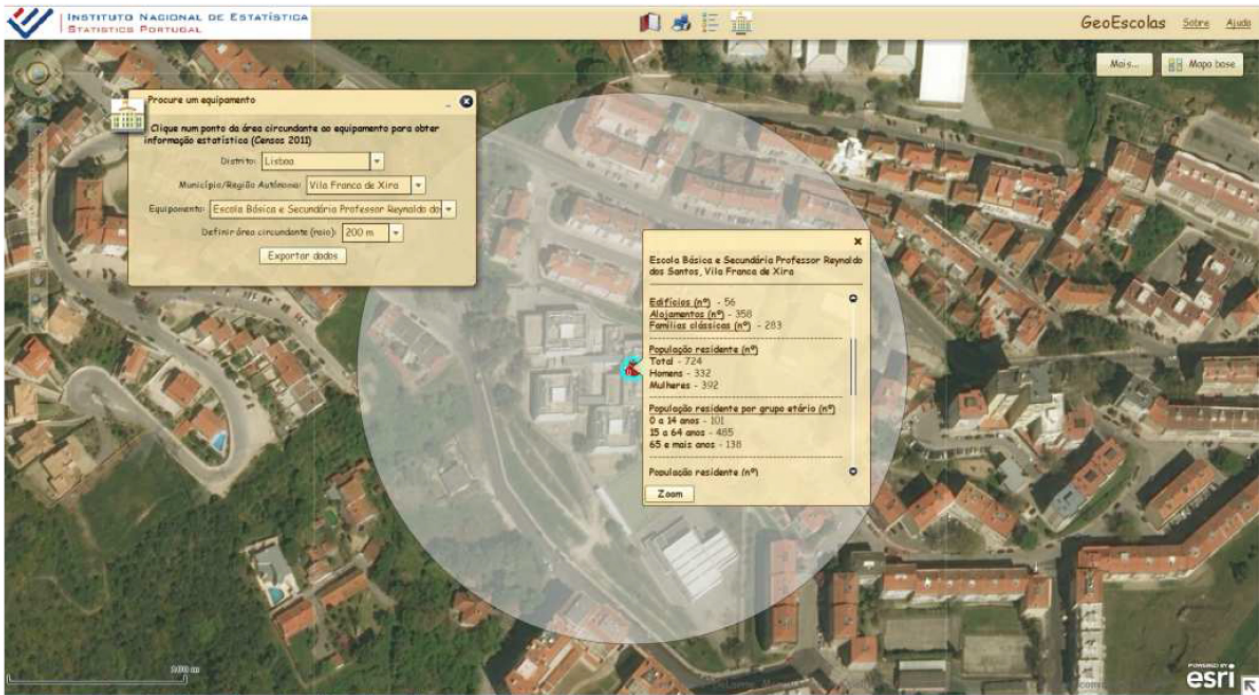


Figure 3: GeoPlaneamento

## Methodology

- Final report — analysis and methodology
- Final report — implementation
- Final report — validation and dissemination

## External links

- [Statistical office](#)
- [2011 census \(mapping\) portal](#)
- [GeoEscolas](#)

View this article online at [https://ec.europa.eu/eurostat/statistics-explained/index.php/Merging\\_statistics\\_and\\_geospatial\\_information,\\_2014\\_projects\\_-\\_Portugal](https://ec.europa.eu/eurostat/statistics-explained/index.php/Merging_statistics_and_geospatial_information,_2014_projects_-_Portugal)