

Geocoding workplaces in Statistics Portugal

The National Dwellings Register (FNA) is the master file in which any address-based system must be matched for the purpose of geocoding existing databases at Statistics Portugal, specifically through its related Buildings *Address Database* subset, which is the key for pinpointing the Business Register records.

The methodology for building the Geographic Database of the Business Register (BRGD) considers the address as the key element to directly or indirectly match the records with the existing Buildings Geographical Database (BGE), by following a step-by-step approach based on different locators capable of sequentially pinpoint the Business Register records.

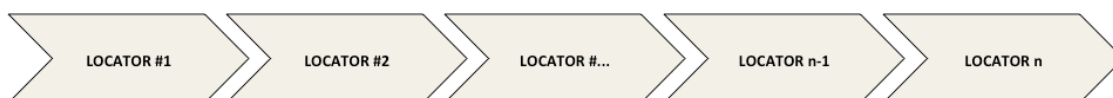


Figure 1: Illustration of the step-by-step approach to sequentially pinpoint the Business Register records by means of a set of locators progressively less precise in terms of spatial accuracy from Locator #1 to Locator n.

The generic methodological approach relies on performing address-matching routines for the Business Register directly over the Buildings Geographical Database (BGE), which is synchronised with the National Dwellings Register (FNA), being the locators progressively less precise in terms spatial accuracy from Locator #1 to Locator n.

Due to specific aspects revealed by the analysis of the Business Register, some adjustments are considered, namely the use of locators retrieved from the Road Segments Database, which is indirectly related with the Buildings Geographical Database (BGE) through spatial attributes (e.g., topology), or the edition of specific fields to improve the results obtained by a given locator.

The locators used to pinpoint the Business Register records are described in table below.

Locator	Description
Locator #1	Complete address composed by type of road, name, number, 7 digit postal code
Locator #2	Used over the Road Segments Database in order to overcome discrepancies in the address
Locator #...	Uses the locality name and the 4 digit postal code
Locator n-1	Based on the 7 digit postal code, which is a linear structure used to code each block façade composed by the CP4 and 3 additional digits
Locator n	Based on the 4 digit postal code, which is a polygonal structure used to code each postal distribution area

Table 1: List of locators sequentially used to pinpoint the Business Register records

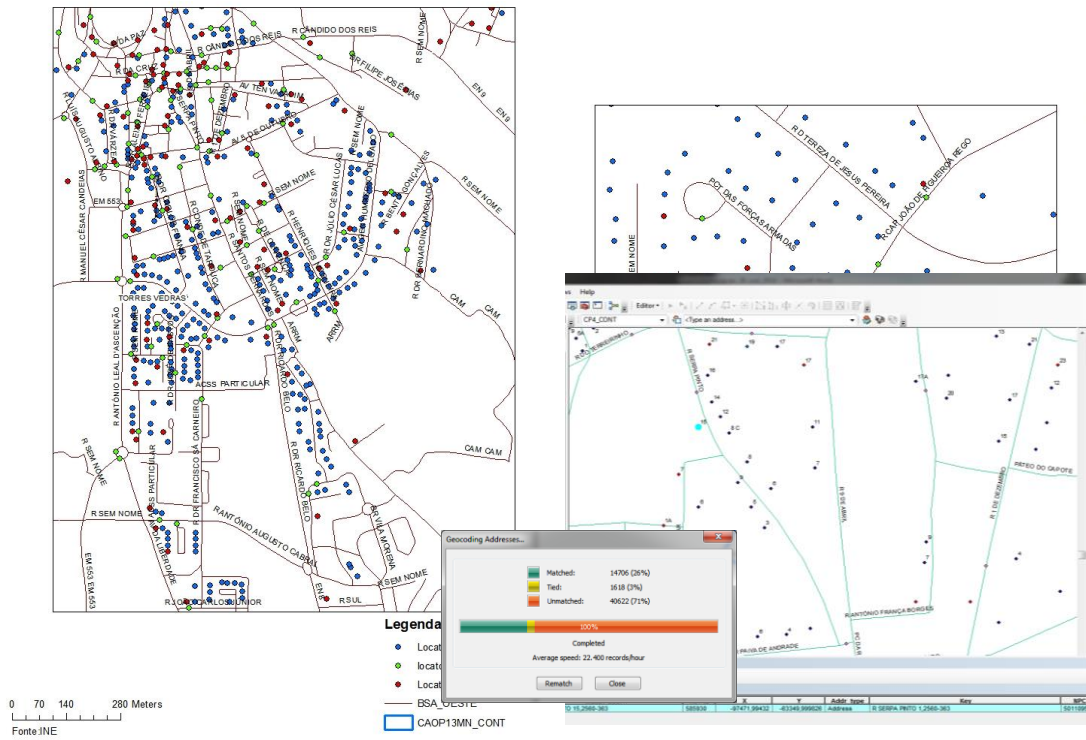


Figure 2: A different mix of locators has been used for the cases processed.

Further work

Currently, Statistics Portugal is testing a commercial routing and georeferencing service to make network analysis and distance calculations for survey purposes and to assign geographic coordinates to business registers and administrative data sources to support statistical production process. These efforts come from the aim to increase quality of the outcomes from these services.