United Nations Expert Group Meeting on
Contemporary Practices in Census Mapping and
Use of Geographical Information Systems
29 May-1 June 2007
United Nations, New York

Development and Use of Geocoding: Portugal approach [*]

Prepared by

Ana Santos
Departamento de Metodologia Estatística
Serviço de Geoinformação
Instituto Nacional de Estatistica Portugal

---

[*] This document is being reproduced without formal editing.

UNITED NATIONS SECRETARIAT
Department of Economic and Social Affairs
Statistics Division

INSTITUTO NACIONAL DE ESTATÍSTICA
PORTUGAL

Departamento de Metodologia Estatística
Serviço de Geoinformação

Census 2011 Mapping Portuguese Program

Abstract

The Portuguese National Statistical Institute (NSI) began in 2006 the construction of the 2011 Census maps.

In order to maintain the level of spatial desegregation of Census nomenclature and also to increase the availability of address data, the methodological aspects are still being studied.

Several questions related to the role of the spatial data infra-structure at NSI to support the activities of statistical and dissemination data are also under discussion. There is a strong need to develop a new component in the spatial data infra-structure to assure the treatment of administrative files, the geocoding of statistical units and the definition of a new master sample.

Contact Adress
Ana Santos
Av. António José de Almeida
1000-043 LISBOA
Tel.: + 351 218 426 244
E-mail: ana.msantos@ine.pt

INDEX

1 INTRODUCTION

The use of cartography as a mean of support for collecting information by Portuguese National Statistical Institute (NSI) began in 1981. At that time, only the major cities had census maps. For the 1991 Census, a full geographical base of the whole Portuguese territory in analogical format was developed – the "Geographic Spatial Referencing Base" **(BGRE 1991)** -. In 1995, NSI started the preparation of the supporting cartographical infra-structure to support the 14th General Population Census and the 4th General Housing Census – 2001 Census, named as "Geographic Information Referencing Base" **(BGRI 2001)** which is based on the Geographic Information System **(GIS).**

Since 2006, NSI has been preparing the supporting cartographical infra-structure Geographic Information Referencing Base **(BGRI 2011)** to implement the 15th General Population Census and the 5th General Housing Census – 2011 Census.

In this article, we describe the main steps which led to the digital mapping for 2001 Census, developed by NSI with GIS software of ESRI, and the strategy and the methodologies involved in the creation of BGRI 2011.

# 2 DESCRIPTION OF THE BGRI 2001

2.1 DEFINITION AND CONCEPTS

The **BGRI 2001** is a geographic referencing system based on cartographical or orthophotographical information in digital format. This system makes possible to divide each basic administrative unit, the "Parish", into smaller statistical areas – statistical sections and subsections, as shown in the scheme below.

Statistical Section

Territorial unit representing the continuous area of a single Parish with approximately 300 dwellings.

Statistical Subsection

Territorial unit which identifies the smallest homogeneous area, whether built-up or not, existing within the statistical section. It represents a block in urban areas, a location or part of a location in rural areas or residual areas which may or may not have dwellings.
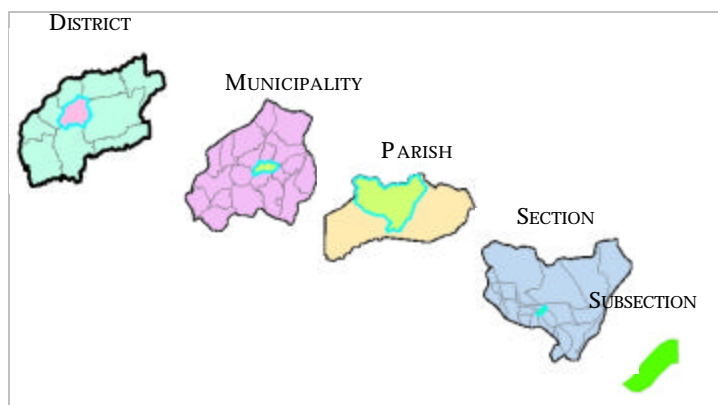
There was defined in 1991, a total of 13,705 statistical sections and 106,625 statistical subsections. In 2001 a total of 16,094 sections and 177,893 subsections was determined.

3   1997/2001 ACTIVITIES

The full digital Census mapping program was developed during the period 1997/2001. Each step and the decisions taken will be described in the next points.

3.1   COOPERATION STRATEGY

To create the BGRI 2001, synergies were promoted between the Portuguese National Statistical Institute and major institutional producers and users of geographic data: the *National Mapping Agency*[1], the *Portuguese Postal Services*[2] and, in particular, the *Army Geographical Institute*.[3]

At the regional and local level of administration, the participation of Municipal Associations (AM's) and Municipalities in the BGRI 2001 construction process is also noteworthy.

3.2   CONSTRUCTION PHASES

The creation of the *BGRI 2001* was essentially based on updating the 1991 census mapping, the *BGRE 1991*, while maintaining the same philosophy of territorial division. This process was developed in accordance with the following stages:

3.2.1 Construction of the digital BGRE 1991

The first phase of this project consisted of digitalizing the *BGRE 1991.* The digital BGRE 1991 was created at the expense of cartography, and, to a certain extent, non-homogeneous coverage in paper form. The cartography of the 1991 Census consisted of a number of documents which were not properly mapping documents, they did not have captions (what made reading more difficult), some of them were not scaled or had inadequate scales, while others were only sketches.

For each Municipality, adequate cartography was chosen based on its date of acquisition and accuracy. Mapping information using different scales, of different types and even with different reference systems was used, guaranteeing however the same projection and the same ellipsoid through projections operations.

3.2.2 Editing BGRE 1991 (preliminary BGRI 2001)

Creation of BGRI 2001 was based on updating the administrative and statistical boundaries representing the sections and subsections created in the BGRE 1991 and on the association of housing unit files, to redefine the statistical breakdown into sections and subsections. The process considered two components: graphic updating and alphanumeric updating:

Graphic updating

This includes the correction of the statistical boundaries of existing Parishes, resetting Parishes boundaries created subsequently to the 1991 Censuses, and defining new subsections representing new blocks, new locations or the geographic expansion of existing localities as well as their possible extinction.

_____

[1] Instituto Geográfico Português (IGP)
[2] CTT Correios de Portugal
[3] Instituto Geográfico do Exército (IGeoE)

Alphanumeric updating of BGRE 1991

This includes the redefinition of the existing statistical breakdown into sections and subsections on the basis of 300 dwellings per statistical section. Priority was given to using administrative housing unit files as an alternative to carry out collection in the field.

3.2.3 Process of Validating Data

In order to get validated data by the Municipal Authorities, a local process was developed between the regional directions of NSI and the Municipalities.

The validation process of the preliminary version of the BGRI 2001, meant verifying the proposed administrative and statistical boundaries, which led to a new stage: the joint analysis with a representative of the Municipality and with each Parish Municipality Chairmen (PJF), in order to ensure their commitment to accept the preliminary version of BGRI 2001.

All the following information was verified and ratified:
- the administrative boundaries of each Parish
- the identification of localities (place, names and perimeters)
- the model of statistical breakdown proposed
- the graphic updating
- the alphanumeric updating

3.2.4 Final Data

All changes and corrections made during the local validation were included in the final version of the BGRI 2001 for the output of maps supporting the collection of 2001 Census data.

3.3 MAJOR PROBLEMS

The major problems detected in these program activities, were:
- the inexistence of updated and standard geographical data of the whole country;
- different versions of administrative boundaries – **Municipalities (NUTS 4) and Parishes (NUTS 5) –** and difficulties defining both localities name and boundaries;
- inexistence of update data about the number of households in each statistical section, for the purpose of size definition;
- the large dimension of the editing process and maps plotting.

Also, the difficulty in importing and checking the quality of local edited data, when the municipalities had assumed the responsibility to create the BGRI 2001.

4   THE 2007 TECHNOLOGICAL INFRA-STRUCTURE

In the 1990s the NSI, like many others organizations all over the world, started implementing a GIS. The system was based on single users workstations and aimed at establishing a digital representation of its geographical information, namely the one from the "BGRE 1991" and the "BGRI 2001".

At the time, a reference data server for single users workstations, where all the editing work was carried out, was seen as the best solution to store data.

The geographical information was represented by means of different elements, such as dots, lines and polygons in format files of the GIS software used, namely ArcInfo coverages and shapefiles.

In 2002, the architecture of NSI GIS changed into a centralized configuration system based on a centralized data base of orthophoto maps and vectorial data – geodatabase in ArcSDE and SQLServer.

At the present, NSI GIS geodatabase is in ArcSDE and ORACLE. It contains reference data and data which constitute the Infra-structure of Geographical Reference **(IRG)**, namely its component of polygonal nature corresponding/related to the Census units of 1991 and 2001.

4.1   DATA INFRA-STRUCTURE – GEODATABASE

The geographical information in NSI, is stored in a centralized Data Base of images and vectorial data. In NSI GIS geodatabase, information is divided into 6 logical groups, as follows:

- Reference Data
- Administrative Units
- Nuts
- Census Units
- Streets
- Buildings

4.1.1 Reference Data

This is a logical group quite heterogeneous made up of Topographical Cartography from the IGeoE, on a scale of 1:25 000, Orthophoto maps from IGeoE, and IGP, and Forest General Services[4], and Regional Geography and Cadastral Services[5], on a scale of 1:10 000, and also a set of cartography belonging to Municipalities and Municipal Associations, which scales vary between 1:1 000 and 1:10 000.

4.1.2 Administrative Units

Administrative units correspond to the Portuguese administrative division of *parishes*, *municipalities* and *districts* in a polygonal structure.

4.1.3 NUTS Units

This group corresponds to the category of territorial units for European statistical purposes – NUTS, and also consisting in polygons.

---

[4] Direcção Geral dos Recursos Florestais
[5] Direcção Regional de Geografia e Cadastro da Região Autónoma da Madeira

4.1.4 Census Units

This logical group is formed by localities, sections and subsections of BGRE 1991 and BGRI 2001.

4.1.5 Streets

This logical group is not a real Street Segments data support (BSA) for geocoding purposes. It consists in a set of BGRI 2001 boundaries polygons with linked street names, and it is not yet complete since it still doesn't include the whole Portuguese territory. It is a linear level.

4.1.6 Buildings

This logical group is formed by the *Buildings Geographical Base* (BGE). Like *Streets* group this is another unit for collecting information and setting up addresses as geo-references.

This layer is not yet complete. At the moment, there are collected, about 300 000 buildings, corresponding to those observed during 2001 Census in the Alentejo region and also to the buildings geo-referenced in the course of the Master Sample Updating Operation.

As far as the BGRI 2011 is concerned, the Geographical Data Model (GDM), presents a generic structure of geographical data and the relationships between charts for the following set of geographical units described above, namely:

- Census Units (Statistical Section and Subsection)
- Localities
- Administrative Units (Districts, Municipalities and Parishes)
- NUTS
- Streets
- Buildings
- Structural elements (water lines, streets, railways …)

The Geographical Data Model (GDM) still includes topological fundamental rules for the definition of the BGRI as a polygonal entity, constituted from BSA, defined as a linear entity. The topological relationships will make possible to formalize a geographical hierarchy model, establishing a relationship between the smaller statistical units (statistical Sections and Subsections), Administrative and NUTS Nomenclatures.

The scheme below shows the features of the NSI spatial data infra-structure

4.2   ESRI SOFTWARE

The NSI has acquired, since 1995, ESRI software products, represented in the following figure.
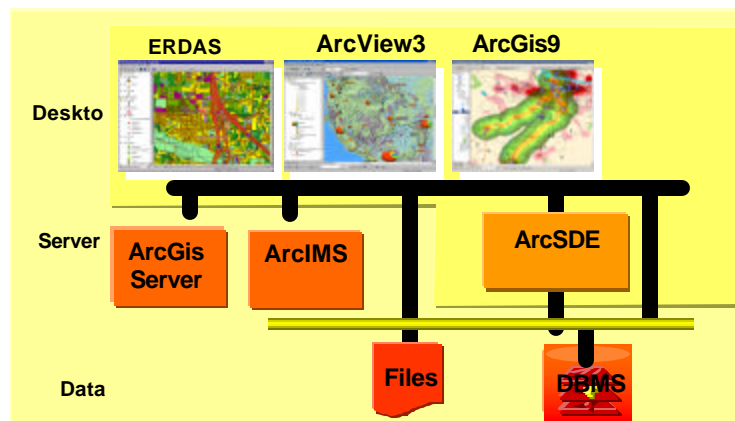
5   2011 CENSUS MAPS

5.1   GOALS

The main goal underlying the creation of new census cartography involves the following aspects:

- Exhaustive updating of the geographical, administrative and statistical boundaries;

- Updating the contents of statistical sections (enumeration districts), related to the number of dwellings.

This set of goals ensures the production of a mapping support identical to the one used during the 2001 Census. This is the reason why NSI decided to carry out an updating operation concerning the BGRI which will focus on the improvement of several aspects, such as the positional exactitude, consistency of sources and the improvement of territory partition into small statistical areas.

For the NSI, to enrich its data infra-structure would imply the improvement of the quality of the streets logical group with street names and building numbers, from sources already available, such as the municipalities and commercial enterprises of street networks. This is a possibility under study.

This is an asset which offers obvious advantages to the Census operation, since it enables the identification of the majority of addresses visited by the enumerators, and may also become extremely useful, not only to the Census but also to the whole NSI.

Apart from this, a street network with addresses would make possible the diffusion, the visualisation and prints of considerable interest and comparative analysis of all files used at NSI, as long as they contain an address, and the beginning of an infra-structure of greater flexibility in the creation of sample bases.

5.2   STRATEGY

For the acquisition of updated reference cartography and for the development of the process of editing BGRI 2001 boundaries or street network files, NSI should privilege a cooperation strategy with official entities producing or owning geographical information.

5.2.1 Acquisition of updated Reference Cartography

- Obtaining the new national orthophoto maps series from IGP and the annual updates;
- Establishing protocols with Municipalities and Municipal Associations to increase the access to municipal and regional cartography of large scale.

5.2.2 Updating the Geographical Reference Infrastructure (IRG)

- To increase partnership between NSI and the Municipalities or Municipal Associations, hence assuring the updating of BGRI 2001;
- The appropriation of administrative sources of information, for size definition of the geographical areas where data is collected;
- The assessment of the possibility of using street network commercial files.

5.3   METHODOLOGIES

There are three key procedures that NSI has adopted to create the 2011 Census maps:

- The use of official administrative boundaries;
- The use of orthophotomaps to perform the editing process;
- The study of the possibility of using external commercial or municipality data to implement a network street file.

The scheme below identifies the BGRI 2001 updating stages and the construction of a street files network to obtain the 2011 BGRI. The editing process will be developed using

_____

ArcGis Server. This is an important change in the technological platform that requires the increase of NSI expertise.
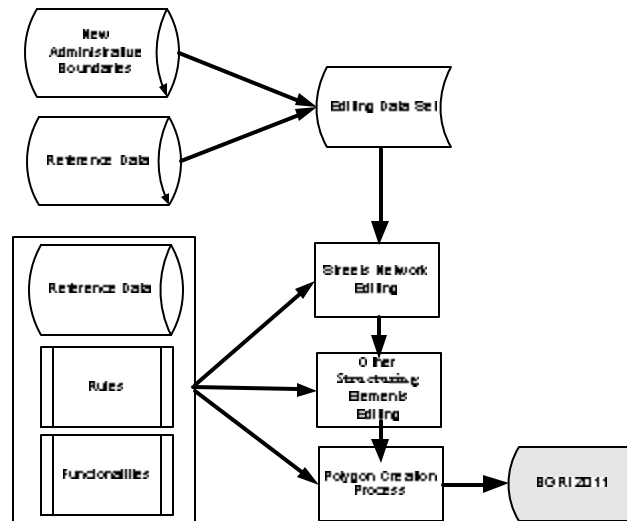
Construction schema of BGRI 2011

First step: to apply the new administrative boundaries to the BGRI 2001. Since 2001 19 new parishes were created. This official data is being provided by the Portuguese Geographical Institute (IGP);

Second step: the editing process combines the use of the streets commercial network or municipality files with structuring elements from the NSI dabase (localities, water lines, ...) in order to obtain the linear features to construct polygons.

Third step: the polygon definition and codification to obtain the enumeration districts.

All the editing process will be done with large scale orthophotomaps as reference data, to adjust and create the new boundaries or polygons.

NSI is planning to start the editing process next July, using the web ArcGis Server application. In the meantime NSI must decide to take advantage of the opportunity to use the commercial street network files. The tests are partially done and have the following objectives:

- To feed a street base for NSI, totally integrated in the BGRI, or being part of it;
- Redesigning the statistical division in areas of greater dynamic in the national territory;
- Helping to adjust, together with the orthophotomaps, the positioning of the lines which form the polygon;
- Maintaining the streets segment base and names updated.

In spite of the need to do the editing and classification work, related to the:

- Existence of geometrically wrongly positioned arcs;
- Out of date data (digitalization of new arcs);
- Inexistence or deficient classification of arches with attribute addresses,

the result of the tests already effectuated let us conclude that the technical viability of the use of commercial data, as nuclear data for the updating of the geographical reference infra-structure, is possible.

The use of commercial data enables NSI to minimize the effort of digitalization of new arcs and polygons, and increase the availability of address attributes existing in IRG.

5.4 PROBLEMS /DIFFICULTIES:

5.4.1 The Codification System

The Codification System used by NSI to classify the statistical subsections of the 1991 BGRE and 2001 BGRI, presents a numerical sequenciality and the non use of codes when changes occur to the boundaries of statistical subsections or sections.

The identification of the 1991 BGRE and 2001 BGRI subsections by the user of Census data, became easier due to this codification System.

Subsection
Section
Parish
Municipality
District

11062303403
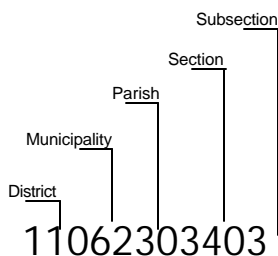
Given to the fact that the identification system attributes two digits when numbering the subsection within the section and the definition of a considerable number of new statistical subsections is expected, it is possible to conclude that this codification system is no longer valid.

5.4.2 HISTORIC

Census 1991 / Census 2001

Due to the differences between BGRE 1991 and BGRI 2001 brought about by urban growth, administrative boundary lines changes, the growing or decreasing number of housing units, it is not possible to compare or study the evolution of both bases at a subsection level.
The need to overcome this problem in order to assure the comparability of produced statistical information to support Census and to define geographical levels- statistical section and subsection – led to the creation of a minimum common base (BMC).

From a technical point of view the solution found consists of aggregating the statistical subsections, that had suffered changes until their delimitations in 1991 BGRE and 2001 BGRI match.

At an alphanumeric level, a new field called "b_comum" is added to the structure table, to classify the aggregations of polygons. This field will assume the same value as in the field 2001 BGRI if no updating is done in 2001 BGRI. If changes are made, it means that the common base is formed by a group of subsections.

Then these polygons will be given the Parish value (dtccfr). The counting procedure will begin with value 90001, for example 02100890001. 90001 is given as a value to prevent code duplication and better distinguish the polygons belonging to aggregations.
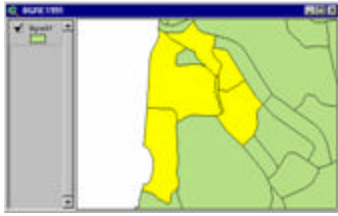
Figure 6 – Configuration of the BGRE 1991 polygons, in a Parish in Lisbon

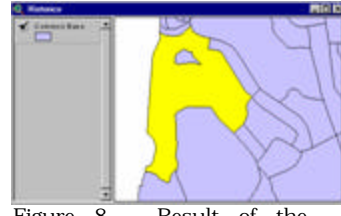Figure 7 – Configuration of the BGRI polygons, in a Parish in Lisbon

Figure 8 – Result of the aggregation of subsections

As we can see in pictures 6 and 7, the delimitation of the 2001 BGRI underwent changes in relation to the BGRI of 1991. The result of aggregating the subsections is shown in picture 8.

Census 2011

We should point out that to maintain the historic is now more complicated because three Census moments must be interrelated, in a situation of non adjustment of the limits of previous bases.

This fact should be subject to analysis since the current codification system may become no longer valid as well as the expected results concerning the dimension of the Minimum Common Base.

That is why it is important to evaluate:

- the Minimum Common Base viability as a tool which will allow to ensure the historic among the geographical units of the three census moments;
- the possibility to ensure the historic through the geo-reference of buildings in Census 2011.

6   FINAL NOTES

The implementation of a digital to 2001 Census Mapping program performed by NSI in the 1990's cost about 22 % of the Census total cost of the operation.

This infra-structure of a strictly polygon nature concerning to small statistical areas is used for data collection and data dissemination. This hierarchy level enables the user to define any superior level.

This infra-structure is the starting point to create 2011 census maps, a process started by NSI in 2006.

However, the new needs of NSI at production levels and dissemination processes, and the existence of new technology and data, justified an analysis of the role of the spatial infra-structure data at NSI. Henceforth, NSI decided to discuss the subject with the departments involved, to analyze the cost/benefits of implementing a new component in that infra-structure in order to perform a new set of activities, related to the production and dissemination, extra census mapping for example:

- the definition of a new 2011 Master Sample
- the treatment of registered administrative files
- the increase of the statistical units files geocoding
- the increase of the dissemination of geographical data
- the preparation of the spatial data infra-structure to enable initialization in 2021 the transition to a new census model

_____

It is important that this decision is taken because usually Censuses favour the
development of this type of infra-structure.