
OVERVIEW OF PORTUGUESE 2001 CENSUSES: CONCERNING DEVELOPMENT STRATEGY, ENUMERATION STRUCTURE, BASIC GEOGRAPHY UNITS AND USE OF ADMINISTRATIVE REGISTERS AS A SUPPORT CONTROL¹⁵

OS CENSOS 2001 EM PORTUGAL: PERSPECTIVA GLOBAL SOBRE A ESTRATÉGIA DE DESENVOLVIMENTO, A ESTRUTURA EXECUTIVA, AS UNIDADES GEOGRÁFICAS DE BASE E A UTILIZAÇÃO DE DADOS ADMINISTRATIVOS COMO ELEMENTOS DE CONTROLO

Autor: Fernando Casimiro

- Director do Gabinete dos Censos 2001¹⁶ - Instituto Nacional de Estatística

ABSTRACT:

- Population and Housing Censuses for 2001 (2001 Censuses) are being prepared following a development strategy that includes the analysis and approval of Global Programme and Dissemination Programme by High Statistical Council (HSC). Then, besides the historical and methodological framing, have been defined main objectives to get as statistical units and respective variables and modalities to be observed.

Enumeration structure should rely on local authorities (municipalities and parishes), with support and frame defined by National Statistical Institute (NSI), and on the identification of enumerator's area made with more precise cartography than that used in 1991. On other hand, electronic processing for questionnaires must be much more automatic than those occurred in 1991; must be used scanning, automatic and assisted coding for alphabetic fields; deterministic and probabilistic editing are being increased.

Basic geographic units (statistical sections and subsections) are defined over maps produced by Geographic Basis for Information Reference (BGRI); main objective for this project consists of digitising and updating boundaries of statistical sections and subsections used on '91 Censuses. Given the coincidence of methodological structure used by '91 and 2001 cartography, it should be possible to see the decade evolution on territory occupancy by human settlements and delimitation of localities.

Finally, administrative data on population and housing belonging to main administrative registers must be fully used to: create a warning system, to be used on the data collection phase which allows to estimate an acceptable size for data collected on each statistical unit; produce comparative analysis with 2001 Census data, when files are conceptually and methodologically closer.

¹⁵ Paper prepared for European Workshop on the Preparation Census Fieldwork – Rome, 12-14 April 1999

¹⁶ National Statistical Institute
Avenida António José De Almeida
P1000 Lisboa
Portugal
E-mail: fernando.casimiro@ine.pt

KEY-WORDS:

- *Population and Housing Censuses; '91 Censuses; 2001 Censuses; Global Programme; Dissemination Programme; scanning; administrative register.*

RESUMO:

- Os Recenseamentos da População e Habitação de 2001 (Censos 2001) estão a ser preparados seguindo uma estratégia de desenvolvimento que passou pela análise e aprovação do Programa Global e do Plano de Difusão no Conselho Superior de Estatística. Assim, para além do enquadramento histórico e metodológico destes recenseamentos, foram definidos os objectivos a atingir, bem como as unidades estatísticas, variáveis e modalidades a observar.

A estrutura de recolha de dados vai passar, sobretudo, por uma forte participação das autarquias locais nas várias tarefas e por uma identificação ainda mais precisa das áreas de trabalho de cada recenseador, através da utilização de cartografia de maior qualidade do que a utilizada nos Censos 91. Por outro lado, todo o tratamento electrónico dos questionários está previsto ser significativamente mais automatizado do que em 1991, nomeadamente através da utilização da leitura óptica e da codificação automática e assistida dos campos alfabéticos e incremento das validações determinísticas e probabilísticas.

As unidades geográficas de base vão ser definidas e delimitadas nas cartas a produzir no âmbito da Base Geográfica de Referência da Informação, cujo principal objectivo consiste na digitalização e actualização dos limites estatísticos e administrativos das secções e subsecções estatísticas utilizadas nos Censos 1991. Dada a coincidência entre a estrutura metodológica utilizada na preparação da cartografia de 1991 e 2001, vai ser possível analisar a evolução da ocupação do território pela habitação e delimitação dos lugares entre aquelas duas datas.

Finalmente, os dados sobre população e habitação existentes nos principais ficheiros administrativos vão ser amplamente utilizados com dois objectivos: constituição de um sistema de alerta, a utilizar durante as operações do terreno, que permita determinar um intervalo aceitável para a previsão dos dados a recolher sobre cada unidade estatística; produção de uma ampla análise comparativa dos dados dos Censos 2001 com os dados provenientes das fontes administrativas e que sejam passíveis de aproximação conceptual e metodológica.

PALAVRAS-CHAVE:

- *Recenseamentos da população e habitação; Censos 91; Censos 2001; Programa Global; Programa de Difusão; BGRI; BGRE; leitura óptica; ficheiro administrativo.*

1. INTRODUCTION

Portugal has been conducting two simultaneous censuses, population and housing, since 1970. So there are two census sets what means that the next round of censuses (2001 Censuses) corresponds to 14th Population Census and 4th Housing Census.

In fact Portugal only has started with an independent housing census and fully accomplished with the international recommendations, in 1970 with the first Housing Census.

Since 1981, under the perspective to join to the European Community and having realised that spring is the best period of the year to collect census data, because of the weather conditions and population's movement at Christmas Holidays, we changed the census day from the end of the year finishing in zero to the Spring on the year finishing in one.

Population and Housing Censuses starts with preparing a Global Programme which includes the description of the main purposes to reach with these censuses, the identification and definition of each statistical unit to be observed and the respective variables (topics) and modalities and some of the main support instruments like cartography, data control and evaluation, legislation, publicity campaign and time schedule for each activity in the project.

Parallel to writing Global Programme is designed the Dissemination Programme which includes the strategy to disseminate census data and identifies each support and form to be used for. One of the main features of this Programme is the tabulation programme and the lowest level of disaggregation for each table in the version available and the version for publication.

The responsibility to prepare the first version of the Global and Dissemination Programmes belongs to the organisational unit in charge of Population and Housing Censuses, which is, for next censuses, 2001 Censuses Bureau.

We should say that these are the standardised steps to start the preparation of census and are similar to any other statistical project designed for data collection.

2. DEVELOPMENT STRATEGY - WHAT HAVE BEEN HAPPENING FOR 2001 POPULATION AND HOUSING CENSUSES?

2001 Census Bureau has been created as a specific unit of National Statistical Institute (NSI) in the beginning of 1998, and one of the first tasks was to prepare the referred Programmes and the proposal for specific legislation. These three

“instruments” constitutes the basis to start the analysis with main representatives of users, belonging to High Statistical Council (HSC).

In the HSC are represented every central government ministries, employers unions, trade unions, local authorities, universities, autonomous regions and consumers representatives.

2.1. LEGISLATION

According to statistical legislation every statistical projects and respective legislation wherever it exists, must be analysed by that Council which created a special section to follow 2001 Censuses until the end; this special section should be closed with the approval of final report of the statistical operation.

The first act for that special statistical section of the High Statistical Council was the analysis of specific legislation proposed by NSI in order to be after approved by Central Government and Parliament. Even Central Government needs a specific authorisation from Parliament in order to approve this legislation, because it includes regulations to regional and local authorities for which only Parliament has constitutional power to make legislation.

Specific legislation is needed for these censuses due to the assigned responsibility to regional and local authorities on data collection and delimitation of administrative and statistical boundaries. Legal responsibility to data collection belongs to local authorities with the support of NSI, and this is a way to get a strong involvement of these authorities on this task. On the other hand, some administrative boundaries are not so easy to find in the territory when deciding if a building belongs to a specific parish or not.

Now that proposal for legislation is under the responsibility of Central Government in consultation with Regional Governments, Local Authorities Associations and Personal Data Protection Committee. NSI expects to have the final approval of this legislation by the end of the first half on this year.

The main reason to have ready the legislation with this anticipation to 2001 comes from the need to update the administrative and statistical boundaries in accordance with local authorities; this boundaries are being digitised by NSI, as we can see in topics 2 and 3 of this workshop.

2.2. GLOBAL PROGRAMME

Global Programme for 2001 Censuses has been prepared according to five main principles:

- Maintain the census data sets;
- Comply with the international recommendations specially for core variables;
- Satisfy, as far as possible, new user's needs;
- Anticipate, as soon as possible, the date to dispose final data;

- Produce and disseminate consistent quality indicators, which should allow users to know and fully accept the coverage and content rates for these censuses.

2.2.1. CENSUS DATA SETS

Analysis of Global Programme has been recently finished in the section of HSC. Concerning the census data sets, mainly those coming from 1981 and 1991, was decided to keep almost all of them, in spite of changes made in some variables related mainly with building and family statistical units. In the building they were added some variables related with earthquake risk and the availability of urban solid waste collection.

For dwelling was decided to add two new variables: period of the renting contract, if dwelling is rented, and types of heating. Data on period of renting contract allows a better understanding of rented housing because rules on housing turnover are different depending on the date of this contract. In spite of type of heating is a core recommended variable, Portugal has not collected this kind of data due to the existing mild climate and a low frequency of housing heating. However nowadays there is a growing number of dwellings equipped with systems to keep an inside steady temperature which represents a new standard of housing conditions.

In the family, changes have been made mainly due to the new “statute” of children which allows a child to belong to parent’s family independently of having or not been married before.

For resident persons, only two main changes have been decided: to draw back variables related to date of last and first marriages and number of children born alive, asked to women older than 12 years, and to change the minimum age limit from 12 to 15 years old to be economic active population. Main reasons for first change are related to the fact that Family and Fertility Survey was recently conducted and this kind of data can be produced in a more deeply way through that survey than it could be in a census. The change of minimum age limit to be economic active is due to legal and in fact improvement of a minimum nine years school leaving and the legal minimum age of 16 years old to start working. On the other way population census is not now the best adequate statistical solution to find out people working below that age limit.

2.2.2. COMPLY WITH THE INTERNATIONAL RECOMMENDATIONS

Every core topics of international recommendations should be observed in each statistical unit, with only one exception: legal marital status. In fact, since 1981 we use the prevalence of “de facto marital status” if there is difference to the legal one. Main reason for that is the objective to collect data which allows to analyse the “the facto” situation of couples without any kind of unacceptable questions on behalf of enumerator. On the other hand, during the questionnaires checking process enumerator is not faced with inconsistencies between marital status and family relationship among

members of same household or family, especially if there is more than one family in the household.

Over and above the core topics, are observed in these censuses 26 of 52 non-core topics internationally recommended, plus 11 topics which correspond to specific national needs. From these national 11 topics, 6 belong to the building and 5 to the dwelling.

2.2.3. SATISFY NEW USER'S NEEDS

A lot of new user's needs have been identified during the analysis of 2001 Censuses Programme in the HSC.

Users are usually expecting from population and housing censuses the opportunity to get ready a substantial amount of data that would answer all their statistical needs. This situation concerns an increasing number of local and regional users, which are not satisfied with data coming from surveys, because of their sampling error when dealing with data for regional levels below NUTS II.

Even sometimes their needs are not covered by statistical information system, what helps to press census to respond those data needs.

Some of the topics raised during this discussion are: accessibility and safety in the building; state of repair of water, electricity, gas and swage disposal networks in the building and in the dwelling; closed condominiums; urban quality of life concerning green areas, social settlements, parking areas, etc.; statistical data on professional training and same sex couples. Otherwise, it was suggested to the HSC that some of the existing surveys should reduce the time lag between each edition, and other surveys conducted on the basis of a not regular periodicity should carry on a regular one.

As referred before, new topics accepted are mainly related to earthquake risk on buildings and two variables for dwellings (type of heating and period of the renting contract).

2.2.4. ANTICIPATE, AS SOON AS POSSIBLE, THE DATE TO DISPOSAL FINAL DATA

One of the main constraints for census data is time lag between census day and the moment when final data are ready to be used by customers.

Last censuses only had final data available two years and a half past census day what has meant a lack of refreshes for some important data. Because the census data processing is a heavy task with a lot of temporary workers and a sophisticated system of editing, two ways are under development to shorten this elapsed time:

- Complete scanning for questionnaires;
- Deeply development of C91 system, which has been used for the first time in '91Census and allows the automatic and assisted coding of alphabetic answers.

Now, four basic questionnaires are designed to be used by scanner and results from the first test are really encouraging. We have tried to combine the easy filling for respondents with technical conditions needed for scanning and for the moment there were no negative reactions. If there is a successful combination, this means that we can spend about 3 months doing what in the recent past has been done by 15 months.

C91 system was developed to perform two main objectives:

- Assume that an alphabetic description written accordingly to respective classification should be coded automatically;
- When description needs to be coded by an operator, the system assumes the decision taken by the operator and should automatically repeat it when the same situation occurs.

From the experience of '91 population census we have ready dictionaries with coded descriptions, written by population and enumerators, for occupations, industry, countries, municipalities and field of study. We expect these dictionaries can significantly shorten time needed to code these alphabetic fields.

2.2.5. PRODUCE AND DISSEMINATE CONSISTENT QUALITY INDICATORS, WHICH SHOULD ALLOW USERS TO KNOW AND FULLY ACCEPT THE COVERAGE AND CONTENT RATES FOR THESE CENSUSES

Data from '91 Censuses have been very surprising even for opposite reasons. While resident population was about 4,5% below the estimated, increasing only 0.3% for the decade, dwellings had an increasing rate of 22%, including dwellings occupied as usual residence with 10% increasing for the decade too. Coverage rate measured by post enumeration survey was 99% for resident population and 99,4% for dwellings. However we feel that users have assumed the suspicion that undercoverage was higher than 1% for population.

So, given that important surprise, we are convinced that the 2001 Censuses will be subject to very careful observation on behalf of their main users.

In a way, despite the fact that ten years will have gone by in the meantime, the 2001 censuses data will end up being an important factor of evaluation of the '91 Census given that no significant and unexpected demographic "accident" has taken place nor is foreseen to take place in the country's demographic evolution process.

Besides the usual controls applied on fieldwork, we have scheduled a control and evaluation system with two main elements:

- "Warning system", based on a set of indicators coming from demographic surplus, electoral roll, geographic information system supported by digitised maps with black cells (buildings), scholar population, post addresses and electricity customers, for each possible and most disaggregated level; this system should allow us to have an expected figure for each covered statistical unit at the respective lowest level; if that figure is significantly not attained or is surpassed, it must be investigated a comprehensive reason for that;

- Post Enumeration Survey (PES), with coverage and content purposes, for a sample of each NUTS II region - Nomenclature of Territorial Units for Statistics; PES have to be the quality benchmark of 2001'Population an Housing Censuses; to assure the guaranty of independence on the meaning of final results of PES, we foresee a partnership with an independent, external and prestigious scientific organisation.

We hope these measures can play an effective role on assuring users that 2001 quality indicators should be out of question.

2.3. DISSEMINATION PROGRAMME

Dissemination Programme has been designed with the description of every census products associated to a strategy to disseminate their data.

Because of the long time period between each census, users expect census data faster and faster, some of them even in the census day. Something like political elections!

In fact that is not exactly possible but we must use every alternative to give users some data as soon as possible. For answering these expectations we had designed tabulation programme using several steps to dispose census data according to data processing programme and on an increasing reliability of respective data. It means that provisional data are much more reliable than preliminary ones, because checking process is more exhaustive than that used for preliminary data.

So, 2001 censuses data tabulation have been organised in three steps:

- **Preliminary data**, produced on the basis of administrative controls; when a enumerator ends the data collection in his statistical section he must count the questionnaires in order to be paid for that; these counts will be the input for the preliminary figures about population, households, dwellings and buildings;
- **Provisional data**, produced on the basis of files getting out of first level editing rules; provisional data are made of eight different tables covering every primary statistical unit (building, dwelling, household and person) and variables used for these tables are those ones which are not subject to strong editing rules:
 - Type of building, main use, number of dwellings, availability of solid wastes collection and period of construction, **for buildings**;
 - Type of living quarter, occupancy status, electricity, water supply system, toilet facilities, bathing facilities, type of sewage disposal system **for dwellings**;
 - Private households by size, **for households**;
 - Sex, age group, de facto marital status, literacy, school attendance, educational attainment, **for persons**.
- **Final data**, after ending every electronic processing steps.

Final data available for users will correspond to four different kinds of products:

- **Tables** belonging to the tabulation plan (105 basic tables), which should be ready up to the lowest level of disaggregation (statistical subsection which corresponds to 108.000 for all over the country); on the other hand, each basic table must have several versions to be published or only available to paper or electronic consultation;
- **A file (“ficheiro-síntese”)** with the most important aggregated counts for each statistical subsection (108.000); this file is made of 91 different counts for each statistical subsection; those counts could be aggregated for any upper level (statistical section, locality, parish, municipality and every NUTS regions – Nomenclature of Territorial Units for Statistics);
- **A CD-Rom** including a long set of census data (from 1864 up to 2001) and a specific Geographical Information System for data from 1991 and 2001 censuses; these two last censuses have a comparable and digitised cartography, which allows the evaluative analysis on the territory occupation;
- **A central file**, which would allow users to make their own tabulation using Internet network.

Statistical secret is the limit to the users access to census data.

High Statistical Council approved Dissemination Programme in the end of March this year.

2.4. PREPARATION PHASES

Preparation phases include cartography, tests of questionnaires and pilot survey.

2.4.1. CARTOGRAPHY

From '91 Censuses we have ready the cartography support, which consists on the division of whole territory belonging to each parish (the lowest administrative unit) into statistical sections (enumeration areas); and each statistical section is divided into statistical subsections. Every supports of this cartography were on paper and we are changing them to a digitised basis.

More details about cartography are in chapter on topic “Basic geographic units”.

2.4.2. QUESTIONNAIRES TESTS

Two questionnaire tests have been scheduled with the following main objectives:

- Evaluate the public reaction to the questionnaires content;
- Reach the best way to combine, in the questionnaires, technical conditions for scanning and easy questions to self-enumeration;
- Evaluate scanning performance;
- Test the remuneration system.

First test was conducted on last October and main conclusions are:

- Questionnaire design with two columns is more acceptable; this means that models used on previous censuses still remain more acceptable than those using only one column;
- They were not detected strong negative reactions to content of questionnaires;
- Self-enumeration rate depends more on person's availability to fill the questionnaires than to question's difficulty;
- Questionnaires designed under technical conditions for scanning are accepted by population;
- Scanning performance was very encouraging, albeit the number of questionnaires used have been low (a sample of 500 for each type of questionnaire).

In April this year, a new questionnaire test is foreseen with a larger sample (about 35000 people), corresponding to a group of complete parishes located in every 7 NUTS II regions.

2.4.3. PILOT SURVEY

Pilot survey should be done one year before census day what means March 2000. Its main objective is to make a complete rehearsal of the final census operations.

3. ENUMERATION STRUCTURE, CONTENTS, PRODUCTION PROCESS (ENUMERATION UNITS, NEW DATA CAPTURE TECHNIQUES, EDITING AND IMPUTATION)

As we have seen before, 2001 Censuses cover population and housing for which we use 5 statistical units: building, dwelling, household, family and person. While one of them (family) is derived, all other units should be observed on the basis of a specific questionnaire. Family unit is derived from the existing family relationships between members of respective household (child/parent/spouse), asked for everyone in the household questionnaire.

Fieldwork is mainly composed by two phases:

- Identification of each building and dwelling units to be observed and delivery of dwelling and individual questionnaires; in this phase we ask population to carefully read the questionnaires and respective instructions and to fill in them on census day; usually the delivery of questionnaires starts about two weeks before census day;
- Data collection, which starts on census day.

3.1. CONTENT AND FORM OF THE QUESTIONNAIRE FOR EACH STATISTICAL UNIT

3.1.1. BUILDING

Building questionnaire is an A4 format with the respective questions occupying only the front page. This questionnaire consists of 14 questions, over and above the geographical identification, concerning the following topics: type of building, main use, number of floors, ground floor building characteristics, relative position to other neighbour buildings and to the respective block, number of dwellings, availability of solid wastes collection, period of construction, structural materials used, materials used to cover external walls, type of roof and respective materials used for and repair needs on the structure, on the roof and on the external walls.

Only 2 of the 14 questions ask for numerically fulfilment; all others ask for mark fulfilment.

Only enumerator must fulfil this questionnaire.

3.1.2. DWELLING

Dwelling questionnaire is a double A4 format with the respective questions occupying only first two pages (front and back pages). Pages 3 and 4 consist of instructions to be used for self-fulfilment by a household's person and remain as protection cover for all questionnaires belonging to the respective dwelling. By the questionnaire preparation for scanning, pages 3 and 4 must be detached using an existing prick of.

The content of this questionnaire consists of 16 questions, over and above the geographical identification, concerning the following topics: type of living quarter, occupancy status, electricity, water supply system, toilet facilities, bathing facilities, type of sewage disposal system, kitchen, type of heating, number of rooms, financial loan costs for owner-occupied dwellings, rent form, period of renting contract, amount size of rent and type of ownership.

Only one of the above 16 questions is filled numerically; using marks fills every other one.

3.1.3. HOUSEHOLD

Household questionnaire is an A4 format with the respective questions occupying front and back pages. Must be filled in only by enumerator during the questionnaire collection phase.

This questionnaire consists in a list of persons belonging to a household, irrespective of being residents or only temporarily presents, and for each person we ask for: name, relationship to the representative (head) of household, identification of spouse (if living in same household), identification of father and/or mother for persons without spouse and/or without their own children living in same household.

The capacity of this questionnaire goes up to 36 persons a household.

Name is an alphabetical field and other fields are numerical.

3.1.4. PERSON

Individual questionnaire is an A4 format with the respective questions occupying front and back pages. Must be filled in by respondents or by enumerators during the questionnaire collection phase.

The content of this questionnaire consists of 30 questions covering the following topics, over and above the geographical identification: name, sex, place of usual residence, place where found at time of census, date of birth, de facto marital status, place of birth, country of citizenship, place of residence one year prior to census day, place of residence five years prior to census day, literacy, school attendance, educational attainment, educational qualifications, field of study for university qualifications, place of work, length of journey to work or school, mode of transport journey to work or school, main source of livelihood, current activity status, time usually worked, occupation, main tasks on main occupation, status in employment, industry, number of persons working in the enterprise, religion (as free answer).

For the extreme chance, there are 8 alphabetical fields, because:

- Some of these fields should not be answered due to the age limit;
- Other ones accept marks for the expected majority of respondents;
- Only personal name, field of study for those persons having a completed university level, occupation and industry must be filled in always on alphabetical.

Date of birth and geographical identification are numerical and all the rest of individual questionnaire fields must be filled in by marks.

3.2. QUESTIONNAIRES ORGANISATION AFTER FIELD WORK

At the end of fieldwork, questionnaires must be organised according to hierarchical order of each statistical unit:

- Every one of the individual questionnaires belonging to the same household must be ordered, on a complete sequence, from 1 to N into the household, with the household questionnaire over all of them;
- The complete collection of questionnaires belonging to each household should be put inside the respective dwelling questionnaire, which forms a cover to the respective household and individual questionnaires; if there is more than one household in a dwelling, the households must be ordered from 1 to N into the dwelling and kept inside the dwelling questionnaire according to this order;
- All the dwelling questionnaires (with the respective household and individual questionnaires inside) belonging to one building are ordered from 1 to N into the building with respective building questionnaire over them;
- For each building there is an auxiliary building cover where are kept inside all the respective questionnaires according to order we have seen before; this building cover has the function of enumerator's report book too;
- Because buildings are numbered from 1 to N into the statistical subsection, all the auxiliary building covers, with the respective questionnaires inside, are numbered according to the respective building number and kept inside an auxiliary statistical subsection cover, ordered according to building number;
- All the auxiliary statistical subsection covers, with the respective content, are kept inside a box with the identification of statistical section.

3.3. DATA CAPTURE

Data capture should be made by using scanning. Test results of April 1999 must constitute the benchmark to take decisions on the number and location of scanners and respective infrastructure, namely the number of workstations including ICR processing and editing stations.

Scanning should be made for 4 questionnaires, as seen before, plus 2 auxiliary sheets: one for statistical subsection identification; another one for transcription of data belonging to people present but not resident in collective dwellings.

Scanning will cover all the content of questionnaires, which means the use of recognition for marks, numerical and alphabetical fields.

Results from the last October's test were positively surprising on performance attainment; only one recognition engine has been used and software can be easily improved for better results:

- Marks do not constitute a major problem for this technology;

- For numerals recognition rates were sized by 86,7 and 97,8%, because numeric fields of questionnaires have filling in different quality levels;
- Alphabetic fields have recognition rates sized between 64,1 and 90,6%.

So we feel that using an improved technology (several engines, carpet system, updated scanners), which is on the way for the next test (April this year), recognition results could be much better than those obtained in the past test.

3.4. EDITING, CODING AND IMPUTATION

Editing, coding and imputation are tasks, which we are trying to automate to the maximum under the objective of reducing elapsed time between data collection and final results. To do that and according to the experience of last censuses (1981 and 1991), we are developing the imputation system based on cold and hot deck rules (deterministic and probabilistic), and the coding system called C91, which allows automatic and assisted coding for alphabetic fields.

For the first editing phase (data just coming from scanning) main rules concern recognition quality, questionnaires hierarchical order and blank questions that can be filled in.

The existing file was built with alphabetical descriptions made in the individual questionnaire of '91 census and coded in that time, which covers about 450.000 different respondent self-made and enumerators written descriptions for each one of occupation and industry variables. So, with this file we expect to save a lot of time doing this task. Recent developments of this system are concentrated on descriptions with 5 or more frequencies (about 30.000 for each of those files), which represent 80% of respective statistical universe.

So, the processing data sequence is: scanning, recognition, first editing phase (to solve recognition problems and other questions that need immediate access to questionnaires), coding (occupation, industry, field of study for university qualification, nationality, place of prior residence), second editing phase (mainly incoherences between questions), third editing phase, data specialising and final data tabulation.

Third editing phase must be done with cold deck and hot deck rules (deterministic and probabilistic) on a totally automated way. By this editing phase we foresee also to check counts of each primary variable and eventually correct any missing inconsistency, after automatic rules processing.

Data specialising concerns the production of a file, for each statistical unit, including primary answers and derived classifications; for example, socio-economic group classification or age for each person.

No imputation for missing statistical units is foreseen.

4. BASIC GEOGRAPHIC UNITS

Geographic units are those coming from the Geographic Basis for Information Reference (BGRI). This geographic basis was built up for '91 Censuses and consists on the division of the whole territory of each lowest administrative unit (parish) into statistical sections and subsections with a delimitation supported on the best available cartography. The whole statistical subsection belongs to a unique statistical section and a whole statistical section belongs to a unique parish.

So, the statistical organisation of the Portuguese territory is:

- Portugal
 - NUTS I (3 regions)
 - NUTS II (7 regions)
 - NUTS III (30 regions)
 - NUTS IV (308 municipalities)
 - NUTS V ($\cong 4240$ parishes)
 - Statistical sections ($\cong 14000$)
 - Statistical subsections ($\cong 108000$)

Statistical section is defined as a continuous area belonging to a unique parish and having about 300 dwellings; however each parish must have at least one statistical section irrespective of having less than 300 dwellings. Statistical subsection correspond more to a “homogenised” portion of territory than a number of dwellings, what means that one statistical subsection may have from 0 to 300 dwellings.

In fact a block is always a statistical subsection irrespective of having or not dwellings, because constitutes a “homogenised” portion of territory. The same happens to a small locality, which constitutes at least one statistical subsection; if a locality could be divided into several statistical subsections, so the respective locality results can be the sum of every statistical subsection belonging to that locality.

The complete identification sequence for each statistical subsection is:

- District – 2 digits
 - Municipality – 2 digits into the District
 - Parish – 2 digits into the Municipality
 - Statistical section – 3 digits into the Parish
 - Statistical subsection – 2 digits into the Statistical Section
 - + 4 digits for Locality to which subsection belongs

Digital map identification for each statistical subsection is made of 15-digit sequence, which include administrative and statistical boundaries, locality name and delimitation. Enumerators only use first 11-digit sequence, because updating process of BGRI makes correspondence between locality and statistical subsection.

District was the prior administrative division of territory, used by statistics, and it corresponds to a group of municipalities. NUTS classification, at levels I, II and III, uses municipality as base unit to constitute those regions; so it is always possible to change from District/Municipality to NUTS/Municipality classification. The reason why District/Municipality classification is kept belongs to less number of digits used (4 against 8 on NUTS).

While in 1991 the geographic basis was supported by maps hand designed on the basis of several root maps and scales, for 2001 it was decided to update and digitise the BGRI over a national and standardised cartography according to a sequence of steps. This work is done using the ArcInfo software with the future purpose of building up a Geographical Information System and a basis for street tracks.

Fieldwork organisation, as to statistical census data from 2001 censuses, should be prepared according to areas and regions supported by this geographic infrastructure.

4.1. UPDATING BGRI

BGRI updating is a combined process among National Statistical Institute (NSI) and the most important national cartography producers, with co-operation of municipalities, which have the responsibility for local planning. However, final responsibility and property of BGRI belongs to NSI.

One of the followed principles by this updating process concerns the preoccupation that the final result of the 2001 BGRI delimitation be compatible with the '91 version, as much as possible, in a way to allow evolution analysis between the two moments of the decade. So it will be possible to see how territory occupation has changed by ten years.

Comparison process is made starting with the statistical subsection. If there is no change on the respective delimitation between 1991 and 2001, comparison can be made fully at this level; if 91 delimitation has been changed by the updating process, a minimal group of statistical subsections, with delimitation changes, must be added up to find out the minimal area, which can be compared into correspondent boundaries.

4.1.1 STEP ONE

First step corresponds to digitisation of delimitation of every statistical section and subsection belonging to '91 BGRI, over the national and standardised cartography (1/25000 and 1/10000). However final support file must be ready over cartography on 1/25000 scale.

Because administrative boundaries always belong to a section and subsection delimitation, it means that this digitisation also allows having a full scale digitised map with all kinds of boundaries.

4.1.2. STEP TWO

By this step we mean the editing of new territory cover with a proposal for an updated delimitation of statistical sections and subsections, which should be confirmed and updated again during the next phase. Inserting '91 BGRI delimitation over updated cartography it provokes immediately a value added to BGRI, because it is possible only by that to see new settlements on the territory.

4.1.3. STEP THREE

This step corresponds to a local validation and updating new delimitation made in the step two for statistical sections and subsections, in a paper support. Specialised people, who belong to National Statistical Institute, make this validation with the municipalities and parish's collaboration.

4.1.4. STEP FOUR

Fourth step deals on editing, over the digitised version of step two, updating corrections made by step three.

4.1.5. STEP FIVE

After corrections made on step four there are produced the final maps in a sized scale between 1/2500 and 1/5000, which may constitute the support to data collection by enumerators.

For each statistical section are produced two maps:

- **Parish Panoramic**, which permits to see delimitation of all statistical sections belonging to the parish; with this map, enumerator can easily see where is his section and the respective boundaries and codes for numeric identification;
- **Statistical Section Panoramic**, which permits to see delimitation of all statistical subsections belonging to respective section; with this map, enumerator can easily see where is each subsection, the respective boundaries and codes for numeric identification, which will be used in the 2001 Census questionnaires.

4.2. GEOGRAPHICAL FIELDWORK ORGANISATION

Fieldwork organisation is based on that geographical basis. This means that each statistical section is assigned to one enumerator, who will receive the two maps

referred above to support and identify each smaller area (statistical subsection) in his section.

For the parish level there is a co-ordinator, who will be the responsible by all the respective enumerators, even for controlling data collection quality; parishes with more than 6 statistical sections should have sub-coordinators, each one controlling 5-7 enumerators.

At the municipality level there will be a municipality delegate, chosen by the municipality, whom is responsible by all the municipality work; for larger municipalities it can be more than one municipality delegate.

At regional level (NUTS II), responsibility for 2001 Censuses belongs to NSI's Regional Directorates, which will indicate regional delegates.

Training process is designed according to this administrative and geographical breakdown but must end at the municipality level. So there is a starting training phase given by national co-ordination to regional co-ordinators; regional co-ordinators do training to municipality delegates, and these last ones must train parish co-ordinators and enumerators.

4.3. FINAL DATA BREAKDOWNS

Final data may use every geographic breakdowns listed before, and all the tables belonging to tabulation plan must be ready down to statistical subsection, which is the lowest level for data disaggregation. However there are some limitations for current use: majority of scheduled tables should be ready for immediate use only down to parish level; for those disaggregation levels each demand should be evaluated to see if there are or not questions concerning statistical secrecy.

4.4. CONFIDENTIALITY AND DATA PROTECTION

For census purposes we have assumed that figures below 3 units could not be under statistical secrecy down to parish level. Below that level (statistical section and subsection) was prepared, since 1981, a specific product called "Ficheiro-Síntese", which permits to have ready data for aggregated variables, for which have been assumed not being under statistical secrecy.

In the near future, probably something has to be changed on this matter, because concerns of population on private life are growing up more and more. Thus, we are studding it on reaching the best way to satisfy users preserving individual confidentiality.

5. USE OF ADMINISTRATIVE REGISTERS AND OTHER SOURCES AS SUPPORT TO FIELDWORK AND QUALITY CONTROL

Population and Housing Censuses are statistical operations, which should constitute a benchmark for statistical data on this subject.

In Portugal, existing administrative registers are used mainly for administrative purposes and often don't comply with statistical concepts and they have important updating problems.

One of the most used administrative registers for statistical purposes is civil registration of births and decease, which allow natural balance, and there is no problem concerning statistical concepts and data quality.

For migration movements, there is only an administrative register for immigration of non-nationals; so it means that immigration of nationals and illegal people is not under administrative control. On the other hand, emigration is almost all out of administrative control, even because the most important part of them goes to countries of European Union.

However, for population estimates, net migration is the other one component that is not so feasible as natural balance; so, concerning net migration, final estimates for each decade are built up with census data.

We may ask why migration trends are not enough measured by population surveys, as for Labour Force Survey or any other one. In fact, for a country like Portugal with a strong migration movement, to evaluate the immigration, including return migration is not so difficult; however evaluation of emigration is not so easy because there is no anyone in respective dwellings to answer survey questions. Then, future european migration statistics wouldn't be excused from using an interchange administrative model that permits to know, to the origin country, that a citizen is asking for a residence authorisation in another country.

Concerning main population registers (Electoral Register and Civil Identification Register), there is a strong feeling that a lot of people living outside Portugal keep their legal residence in Portugal due to several reasons.

On the other hand, if census data are not according to yearly population estimates users often suspect from census data, what represents a real challenge to population census to explain reasons why these data are feasible.

Housing data are not so difficult to control because buildings and dwellings are much more "stable" and easy to find than people are.

So, to avoid surprises and suspicions like those occurred in '91 Census, we are designing a control and evaluating programme that uses administrative data to check every step on the collection phase and to give to the users a global analysis about data odds due to different sources.

5.1. WARNING SYSTEM

As we have seen before (chapter 2.2.5), this system combines data from different sources to estimate an acceptable size for each statistical unit, at parish level. Those estimated sizes should be used to define if data collected by 2001 census fieldwork are acceptable or not; if not a special supervising process should be unchained to check census data.

Electoral Register for population estimates, postal domestic addresses and private electricity customers, both for dwelling estimates, should play an important role on this system.

This warning system must concentrate on a reduced number of administrative sources to avoid a large conflict data range. On the other hand, accuracy levels among administrative registers are very much different and they cover different populations, which are not complementary to each other. So we concentrate on those registers, which are most used by local and regional authorities to check census data.

5.1.1. FOR POPULATION

Electoral Register has been updated recently, but includes every one living outside, for whom there is no any other electoral registration. So every one, who has changed his "de facto" residence to elsewhere and did not update the new electoral address, still remains as officially resident where the electoral registration belongs. This occurs mainly with emigrants that are living and working outside but they do not change the electoral registration address, because they have temporary working contracts or even because they do not want to change their "official" residence from Portugal; sometimes they keep more than one legal residence because there is no any way to check it.

So, even after the last updating process of electoral register, there is a difference for about more 10% population with 18 or more years of age on electoral register than NSI population estimate has. Then using electoral register's figures at the parish level to estimate higher size limit for population older than 18 or more years, we are quite sure that only few parishes could be outside this checking process, even because not enrolled population, as elector is not expressive.

For people below 18 years old, school population and birth registration must be used to estimate this age group at parish level. However school population enrolled at the compulsory and the secondary education is not disaggregated to parish, but only by municipality, what implies the statistical partition by parish into the respective municipality.

Another one main concern is younger people; children belonging to first two years of age often are forgotten to be enumerated. A special attention should be asked to enumerators and on the publicity campaign to enumerate younger children

5.1.2. FOR HOUSING UNITS

Housing units are more “stable” than people and data from '91 Census have not been so surprising as to those for population. So we can start with data of last '91 Housing Census adjusted to building licenses issued by local authorities during last decade.

Otherwise housing units stock could be checked with electricity private consumers register, which assure updated data for every active consumer. Nevertheless major problem on this register concerns the disaggregation to parish: albeit data at municipality level are accurate, there is an expressive number of consumers that have no information on which parish they belong. So we have to use a statistical method to adjust this data, at least, to the smaller administrative areas.

Another source, not exactly an administrative one, refers to data on postal addresses: under the basis of an agreement with PTT company we can access to counts of private postal addresses made by postmen, some of them disaggregated to statistical subsection in the urban areas.

5.2. COMPARATIVE ANALYSIS

Many users make comparative analysis of census data with data belonging to other sources, most of them with different concepts and time reference periods.

To help users and anticipate criticism to census data we decided to make a broad comparative analysis of main comparable data sources reminding where are differences and their causes.

Data belonging to electoral register, pupils and students, employees from yearly administrative declaration and legally resident foreigners are the most used to comparative analysis with census data. All these data could be used only after data processing once is really difficult to check every individual characteristic by data collection phase.

Electoral and legally resident foreigner's registers have similar problems: people with "de facto" residence different from the registered one. Because main source of foreign immigration to Portugal are African ex-colonies there are two sources of errors when considering these registers: people that keeps their legally residence in Portugal but in fact is living outside due to business or any other reason; people "de facto" living in Portugal, but not enrolled in foreigners register (illegal ones). Because census questions don't ask about legally situations, only Post Enumeration Survey content errors allow adjusting this data for comparative analysis.

Employees from yearly administrative declaration (Quadros de Pessoal) is an administrative source used by Labour Ministry to control people belonging to enterprises and the respective establishments. They have been used to produce some labour statistics on wages, time worked, size of enterprises, etc, and was based on the

identification of each employee belonging to the enterprise in October each year. Besides there was some lack of coverage, in the comparative analysis to '91census data we realise that there was a strong consistency for employees belonging to enterprises with 10 or more people and the industry classified from C to K, according to NACE (Rev.1); this is the enterprises group with more consistent coverage by "Quadros de Pessoal". For this group of enterprises corresponding to about 1,6 million people, for a total of 2,1 million employees covered by this administrative source, the difference between '91Census data and "Quadros de Pessoal" was 1,4% more people in "Quadros de Pessoal". Regarding the estimated coverage error of -1% for resident population on census data, we can assume that a strong level of comparison consistency should be expected for these two completely independent sources.