UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

STATISTICAL OFFICE OF THE EUROPEAN UNION

Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses
(The Hague, The Netherlands, 10-11 May 2010)

Topic (iii) of the provisional agenda

FIT FOR THE CENSUS

Implementing a register-based census in Portugal:
changing the paradigm

Note by the Statistics Portugal

I.Introduction

1.	Presently, a number of conditions have come together to assess the possibility of using administrative registers as an alternative to Portugal's census design approach. The objective of this paper is to present and share Portugal's experience in the implementation of a step-by-step methodology for making the transition from a traditional to a register-based census.

2.	This paper focuses on the activities that have been taking place in preparation for the next Population and Housing Censuses, to be held in 2011. The paper starts by describing the main aspects and conditions that have been put together to implement a register-based census; this is followed by a presentation of Portugal's work programme and course of action regarding identification of the relevant registers, by focusing on the core registers regarding relevant statistical units (housing and population) and, more specifically, by addressing content coverage and assessing linkage possibilities; and finally, the main issues and challenges that Portugal will have to address and deal with are outlined.

II.Moving from a conventional census design to a register-based census in Portugal

3.	In Portugal, Population and Housing Censuses have been conducted following the traditional approach of extensive and exhaustive enumeration of statistical units, within a specific time frame and defined periodicity (every ten years), based on a comprehensive questionnaire that encompasses all the questions foreseen for housing, household, socio-demographic and socio-economic characterisation. However, it is widely consensual that conducting a traditional census corresponds to the most elaborate, complex and resource (both human and financial) consuming statistical operation undertaken by a given country.

4.   Portugal has been fostering policies regarding the modernisation of public administration services. One of the most visible consequences of this strategy was the creation of the SIMPLEX in 2006, a national governmental programme for administrative and legislative simplification that aims to facilitate the interaction of citizens and enterprises with public administration, and which reflects a will to improve public services in terms of optimisation and efficient use of resources. Within the scope of this programme, several measures have been taken, with the purpose of integrating services, establishing unique interlocutors or merging redundant forms or services, via investment in modern and state-of-the-art infrastructure and communication technologies.

5.   This current social and political context has been setting the tone for the preparation of the 2011 Population and Housing Censuses. An investment in the modernisation of collection methods and processes, through the use of information and communication technologies, has been made and, for the first time, a work programme has been put into motion regarding the use of administrative information for *changing the paradigm* in terms of Portugal's census design approach; and Statistics Portugal is dedicated to preparing and developing the foundations of a register-based census for the 2021 round, a commitment that has been assumed and specified in the 2011 Census Action Programme. In this regard, important decisions and measures have been put into practice to build up a positive political, social and institutional context to address this challenge.

A. Legal Base

6.   One important point that has resulted from other countries' experience regarding the use of administrative data for statistical (and census) purposes is the existence of specific legislation (see United Nations, 2007) granting access to national statistics institutes to administrative data on unit level with identifiers information. In Portugal, for the first time, the national statistics act (2008) clearly grants Statistics Portugal the right to access administrative information on unit level (corresponding to micro level data) with identification information and linkage possibilities for statistical purposes.

7.   The Portuguese national statistics act defines administrative data as all "data collected by entities of the public sector on natural and legal persons, including individual data, based on administrative procedures that normally have a primary goal that is not for statistical purposes" and specifically states that the use of administrative data should be extended to the production of official statistics. This fact has made it possible to consider administrative sources as an alternative mode of data collection for census purposes.

8.   Consequently, the 2011 Census legislation states that this census round should be the last one adopting a full traditional approach and that efforts should be conducted in order to implement an alternative census design, one that is less elaborate and expensive, in terms of human and financial resources, but also capable of generating census information more frequently than every ten years. Therefore, the 2011 Census legislation defines the following goal for the next census round – to prepare and to assess the possibility of making the transition from a conventional census design approach to an alternative census design based on administrative registers to produce cost efficient and more timely population and housing statistical information.

B. Identification numbers / Identifiers

9.   The unique identification number (the person number in the case of Denmark) was a crucial element for the success of the use of register-based information for statistical purposes in many of the Nordic Countries that currently have a register-based census. As Wismer (2003) notes this was not only important for the linkage of information, but also because it allowed for the standardisation and normalisation of other crucial information for achieving a reliable register, such as: address codes, municipality and parish codes, or nationality codes.

10.   In Portugal, the situation on this matter is rather different and more challenging, as there are several identification numbers used for different administrative purposes, for example: i) the ID number (Civil Register); ii) the tax identification number; iii) the social security number; iv) the national health system number; v) different foreigners' identification numbers, depending on the type of visa. However, in 2007 an important measure was established, within the context of the SIMPLEX programme, which represents a significant advancement regarding the possibility of linking information from different administrative registers – the creation and implementation of the Citizen Card[1].

11.   The Citizen Card holds, in just one citizenship identification document, information that used to be separated over four different documents: i) the ID number; ii) the tax identification number; iii) the social security number; and iv) the national health system number. Portuguese citizens are obliged to ask for the Citizen Card, when the previous ID document has expired and up until now around 2 million have been issued, around 20% of the total population. The card can be issued for those holding Portuguese nationality and also for those under the scope of the Treaty of Friendship, Cooperation and Consultation between the Federative Republic of Brazil and the Portuguese Republic.

12.   Besides being a step forward regarding the possibility of linkage between different administrative registers, the creation of the Citizen Card also enabled for better standardisation, normalization and updating of crucial elements, namely those regarding address information, zip codes, municipality and parish codes, which will permit associating the location of an individual to a specific geographical unit level. However, there are still specific groups of the population not encompassed within this register. These include, for example, foreigners and children below the age of six. The Citizen Card is a mandatory document for all national citizens, living in Portugal or abroad, only from the age of 6 or as soon as this identification document is required for any administrative or legal act. Therefore, these two segments[2] of the population are not within the scope of this register and information must be collected from other administrative registers, namely through the Foreigners and Borders' Office register and birth registrations' registers.

C. Articulation and cooperation with register-keepers

13.   The process of integration of register-based information for statistical purposes means that statistical offices are no longer depending exclusively on their data collection methods. A third party

---

[1] Official website - www.cartaodecidadao.pt
[2] Population below the age of 6, represents around 6% of the total population (2001 Population and Housing Censuses) and the foreign population represents around 4% (Foreigners and Borders Office, 2008).

becomes involved in this process, namely the official and public entities responsible for keeping the registers.

14.   Administrative registers hold a different philosophy in terms of information management, as they are collected primarily not for statistical purposes. Administrative data and statistical data are created and used for different purposes, with different roles being attributed to individual objects (Wallgren and Wallgren, 2007). That is, in administrative registers, information results from administrative actions and measures that take place within the scope of the register-keeper's own activities and objectives and, thus, conclusions and decisions are drawn from individual objects. Information on a specific object is retrieved and relevant individually. On the other hand, within a statistical system, conclusions are drawn from the analysis carried out and for this reason objects are not interesting individually, only their aggregated outputs.

15.   In this vein, the pre-requisites of data processing procedures in terms of standardisation and normalisation regarding, for example, codes or classifications, are different for administrative and statistical purposes. Therefore, in order to maximise the use of administrative registers, it is important to work collaboratively with register-keepers not only to establish good articulation and cooperation procedures, which are important to guarantee the continuity of a register-based system, but also to improve registers' quality for statistical purposes.

16.   In this sense, Portugal has been investing in the articulation with register-keepers, by first trying to make them aware of the change of paradigm this process will entail and how crucial their role is for the success of its implementation, as they become the main providers of information for census purposes; and secondly by defining specific interlocutors to contact with and discuss the issues regarding each administrative register. Although experiences from other countries indicate that few adjustments and extra data collection have been made by administrative authorities (see United Nations, 2007), Portugal is committed to a bidirectional collaborative work, i.e., that register-keepers may also benefit from the work carried out by Statistics Portugal, regarding quality improvement procedures, for example. The idea is to address this collaborative work as a win-win situation for both parts and, in this context, the national statistical act provides the key foundation and constitutes a useful instrument.

## III. Developing and implementing a step-by-step methodology

17.   In the preparation for the next 2011 Census round, a work programme has been defined to specifically address the challenge of making a transition to a register-based census for the 2021 round. This work programme was made available in the Census 2011 Action Programme.

18.   A transition such as this means that the entire paradigm revolving around Population and Housing Censuses operations in Portugal will be radically transformed, and a significant change like this represents a lengthy, elaborate and persistent commitment. It took Finland 20 years to produce a fully register-based census in 1990 and Norway and Sweden will be able to accomplish this transition for the next 2011 Census round (Tonder, 2008). Although these countries have now reached a point where all Census variables can be produced based on administrative registers, this was a gradual transition process with different administrative registers being used for census purposes at different points in time. For example, the population register in Sweden was first created

in 1967 and used for census purposes eight years later, in 1975; and only more recently was Sweden able to create the dwellings register (2008) that will be used in the 2011 Census round.

19.   Following on from these countries' experiences, actions have been prepared so as to implement a step-by-step methodology, where specific statistical units have been defined with the purpose of identifying the administrative registers that can account for census variables. The objective is to introduce register-based variables progressively for census purposes, after their quality and reliability have been assured.

20.   Two main topics regarding relevant statistical units have been identified around which a specific work programme has been defined: the first is concerned with housing topics register and the second with the creation of a population register.

21.   The work programme, and its specific actions, has been outlined as follows (see chart 1):

    a) Identifying relevant administrative registers and articulation with register-keepers
    b) Assessing the quality of registers (coverage, content and identifiers) and their potential use to generate census variables;
    c) Defining a stable and reliable group of housing and population variables to compare with the 2011 census results;
    d) Assess, measure and understand possible differences and deviations so as to consolidate administrative registers' information.

Chart 1 – Work programme for implementing a step-by-step methodology
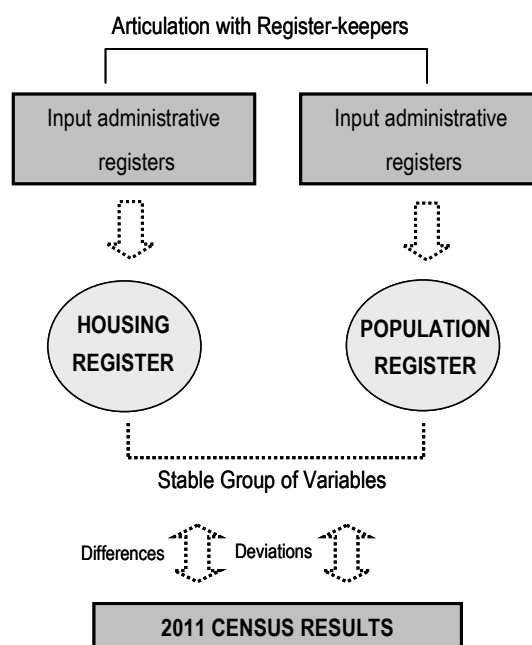
IV. The Registers

22.   A necessary condition for the implementation of a register-based system for census purposes is the existence of comprehensive and reliable registers for administrative needs. According to the work programme defined, actions have initially been directed to the identification of relevant administrative registers and to the articulation with register-keepers so that data is made available to proceed with the next stages.

23.   A general principle for articulating and processing administrative data is to combine different sources. As emphasized by Wallgren and Wallgren (2007), the combination of variables from different sources can be used to achieve a richer content, but also as a way to validate and cross-check information. Coverage and editing possibilities may also be increased if multiple sources are used and compared.

24.   In this vein, and similarly to other countries, Portugal has privileged the combination of different sources so as to increase coverage, editing possibilities and validation procedures. This principle has been complemented with an extensive collection of information on the registers' content (metainformation) in terms of concepts, the type of forms used, and information updating procedures, which are important for a precise definition of units and time references (e.g. Borchsenius, 2000; Tonder, 2008).

25.   Specifically, two stages of intra and inter validation have been identified (see chart 2). The first stage corresponds to an independent assessment of the consistency and reliability of each administrative register regarding coverage, content, identifiers and quality. The second stage corresponds to an evaluation of linkage possibilities and of data cross-checking, which will also allow for identification of duplicates, within the same statistical units and between registers of different statistical units. The main purpose is to create a register-based system that articulates registers within a statistical unit and between statistical units. The final expected outcome is the creation of a statistical register that will be used to produce housing and population census statistics.

26.   The following points provide an overview of the relevant administrative registers that have been identified regarding housing and population information. Specifically, for each topic a core register has been identified, corresponding to the most extensive register available taking into consideration coverage, content and identifiers information, which will interlock with other registers so as to cover as many variables as possible regarding relevant statistical units.

A. Housing Register

27.   Concerning housing topics, a main administrative register has been identified – the Housing Taxation Register, from the Directorate-General for Taxation of the Ministry of Finance. This register includes statements of registration or updating of urban property delivered to the Finance departments. All properties (buildings or parts) that pay tax are covered by this administrative act. This register covers most of the mandatory variables for the 2011 Census round, regarding geographical location, building and housing characteristics (see table 1).

28.   Two other registers have also been considered, namely the Personal Income Tax Register and the Energy Register from Energy Portugal for additional information on geographical identification (address information and codes).

Table 1 – Variables in Housing Administrative Registers

Variables	Housing Taxation Register	Personal Income Tax Register	Energy register
Tax Identification number	x	x	
Address	x	x	x
ZIP Code	x	x	x
Municipal and Parish Code	x	x	x
Geographical Coordinates	x		
Type of building	x		
Type of use	x		
Number of living quarters	x		
Year or period of construction	x		
Type of living quarters	x		
Water supply system	x		
Occupancy status	x		
Toilet facilities	x		
Bathing facilities	x		
Number of rooms	x		
Useful floor space	x		
Ownership	x		

Limitations

29.   Preliminary analyses indicate some limitations regarding harmonization and standardization of registers and high rates of missing data regarding some variables (e.g., different codes and labels used for the same categories and descriptions, high rates of missing and invalid information). Although the variables that have been made available need extensive editing before they can be used for statistical purposes, only one of the core topics variables could not be obtained from the administrative sources: type of heating. Regarding coverage, the registers are not exhaustive on their own and a maximization of their potential will result from their combination. For this process, it is important to stabilize a matching key for registers' linkage and efforts have been directed towards geographical identification.

Linking

30.   In order to move from an administrative to a statistical register, it is necessary to build an information system for storing, updating and use of administrative data for housing topics. This system will be based on one essential point which is geographic identification. The identification key, which has not been defined yet, may be an address (full address information and codes) or a geographic coordinate (X, Y). On this specific subject, for the 2011 Census round a methodological system is being put forward that will allow for all buildings to have an X and a Y coordinate. The correct definition of an identification key of the living areas becomes even more important when linking housing registers with the population registers. The linkage between these two registers will guarantee reliability and subsequent use for data management, analysis and dissemination (see chart 2).

B. Population Register

31.   Concerning population variables, a core register has also been identified as the most comprehensive in terms of coverage – the Civil Register, from the Registers and Notary Institute. The civil database contains information on all citizens holding Portuguese nationality regarding the main demographic variables: sex, age, filiation, marital status, country/place of birth, place of usual residence, and address information. This administrative register corresponds to the ID number that has been included in the Citizen Card, along with three other numbers, as mentioned above.

32.   To complement the information in this core register, other administrative registers have also been considered, namely:
- Personal Income tax register, which holds potential information for household operationalization;
- Employment registers from the Institute for Employment and Professional Training, for socio-economic variables
- Social Security Register from the Ministry of Labour and Social Solidarity, for socio-economic variables;
- Education Registers from the Office of Education Statistics and Planning of the Ministry of Education and the Office of Planning, Strategy, Evaluation and International Relations of the Ministry of Science, Technology and Higher Education, for education variables
- Foreigners register, managed by the Foreigners and Borders Office for characterization of the foreign population living in Portugal.

33.   Variables available across these different administrative registers are presented in table 2. Most of the main variables are potentially available, but information regarding family relations and family nucleus (size and type) appears to be one of the most challenging variables.

Table 2 – Variables in Population Administrative Registers

Variables	Civil Register	Social Security Register	Employment Registers	Education Registers	Personal Income and Taxation Register	Foreigners Register
ID number	x	x	x	x	x	
Social Security number		x	x	x		x
Tax Identification number			x		x	x
Name	x					
Address	x	x	x	x	x	x
Place of usual residence	x	x			x	x
Housing status						
Age	x	x	x	x	x	x
Sex	x	x	x	x	x	x
Marital status	x	x				x
Country/place of birth	x	x				x
Country of citizenship	x	x				x
Ever resided abroad and year of arrival in Portugal						
Educational attainment			x	x		
Current activity status		x	x			
Occupation			x			
Status in employment		x	x			
Branch of economic activity			x		x	
Location of place of work or school			x	x		
Relationship to reference person of household					x	
Identification of the spouse/partner in a consensual (de facto) union					x	
Identification of father	x					
Identification of mother	x					

Limitations

34.   Although several limitations in terms of conceptual correspondence are expected, the main challenge Statistics Portugal has been facing regarding population administrative registers is access to individualised and identifiable information on unit level regarding some groups of variables. In addition, moving from a traditional design approach to a register-based census represents changing not only the *locus of control* in terms of data holders, but also a cultural change regarding data collected for administrative purposes. In this regard, register-keepers have been reluctant to share their registers, basing their arguments on different interpretations of the Portuguese Statistical Act. In fact, all register-keepers refused to cooperate at first and only after access was granted from the Portuguese Data Protection Authority (CNPD) were some of the registers made available to Statistics Portugal, namely the Civil Register and the Social Security Register. Nevertheless, there are still some groups of variables, namely those related to individual socio-economic situation (e.g. status in employment, occupation, branch of economic activity, etc.), that are still under discussion by the Portuguese Data Protection Authority, since they are considered to fall under the category of sensitive information. Therefore, special efforts have been undertaken in this regard as the involvement of register-keepers is fundamental for this transition, which is expected to be an on-going process.

35.   A first attempt for family relations operationalization was made using the Income and Taxation Register, which revealed several limitations, as there is no way, up until now, of linking and aggregating persons to the same income tax return based on the register that was made available. Moreover, a closer look at the concept of household applied to taxation purposes showed that there are substantial differences to the one used in census operations. Therefore, to address this issue other possibilities are being considered, namely through address information / codes.

36.   On the topic of education, relevant limitations have also been found, namely regarding data on tertiary education. Specifically, it was reported that data on 1st and 2nd level of tertiary education are anonymised (no variables of identification are kept within the administrative register). This has been the method adopted by the Ministry and the universities, which are the data providers.

37.   Finally, analyses of the administrative registers made available until now show a relatively high percentage of missing data for those variables that correspond to identification numbers (e.g. 50% for the ID number in the Social Security register).

Linking

38.   In Portugal, it is not possible, at least in the immediate future, to implement a system of identification with a unique number, and several numbers will have to be combined in the implementation of a statistical system based on administrative data. This, as previously mentioned, adds an element of complexity to the task of linking different registers relating to the same statistical unit and between different statistical units. In this vein, linkage will only be accomplished through the combination of different primary keys (see chart 2).

Chart 2 – Integration and articulation of registers
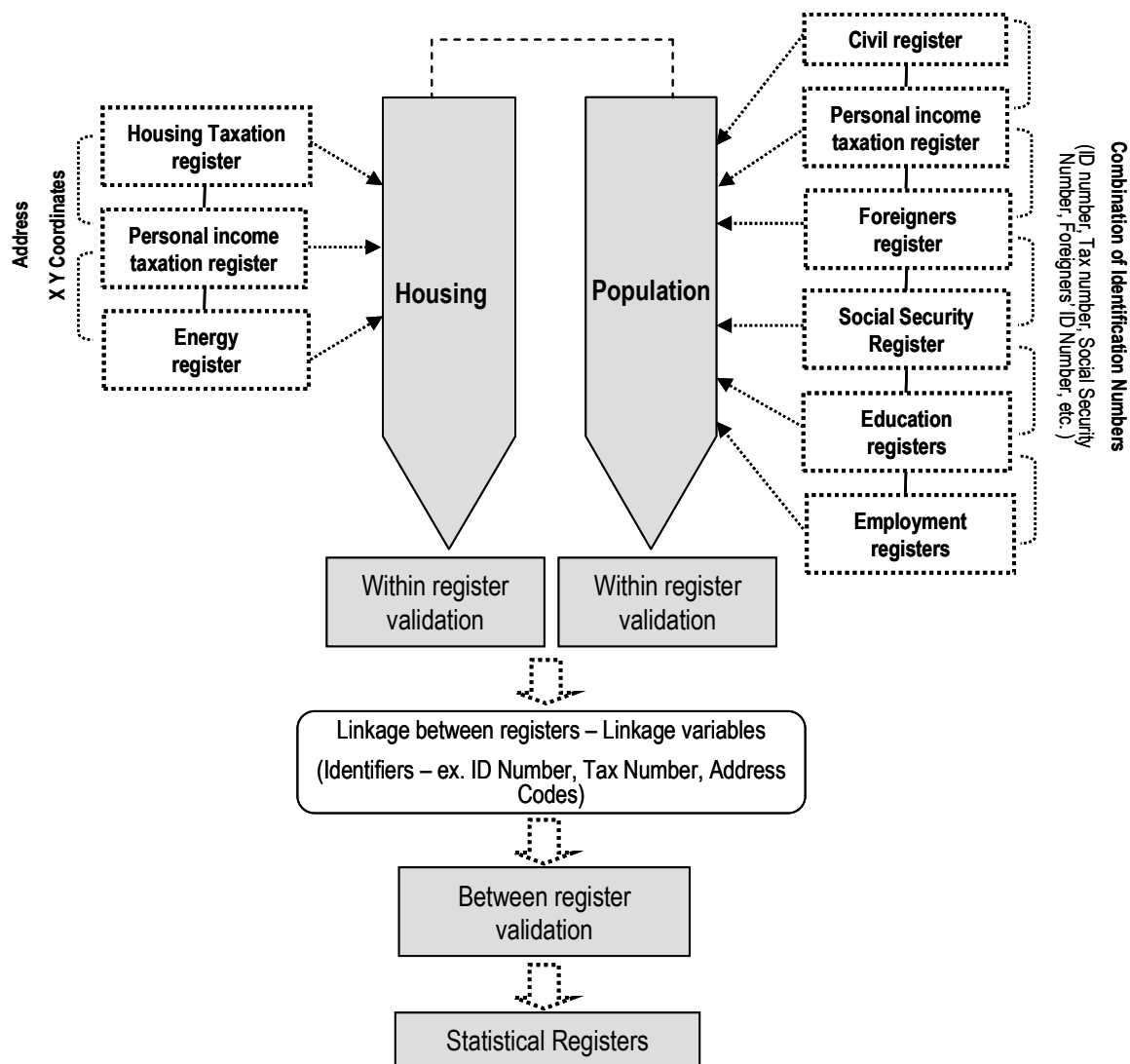
V.Main challenges and issues Portugal will have to address

39.   The work that has been undertaken, in order to lay the foundations of a register-based system that can account for census variables, has come across several challenges that Portugal will have to address and deal with in order to move forward.

40.   As already mentioned, most of the administrative registers are not created and used for statistical purposes and therefore, how information is collected, processed and managed does not meet the criteria needed to achieve the required statistical standards. This was one of the first issues that emerged from the preliminary analyses conducted. The quality of registers has been identified as relatively limited, namely regarding standardisation, normalisation and consistency of information. Improving their quality requires an extensive process of editing that will have to be carried out in close articulation with register-keepers. In this regard, standard editing procedures are

being defined and tested, such as the application being used for normalization of address descriptions and codes.

41.   Closely linked to this topic is the issue of unique identification of units. Information concerning identification numbers also has quality deficits, namely regarding the rate of missing data. Efforts are being put into practice in order to work with register-keepers to assess how this situation can be improved. Yet again, it is important to stress that the fact that Portugal has several identification numbers used for different administrative acts constitutes an element of complexity in the process of articulation and linkage of information from different administrative registers. Although the Citizen Card has been created, and this has represented an important advancement regarding the possibility of linking information, it did not represent the merging of the information within a unique and integrated administrative register.

42.   However, the most important topic in this regard is the limitations and constraints that have been emerging regarding access to individualised and identifiable data that address specific groups of variables, namely those concerning socio-economic characterization. The way in which Statistics Portugal will cooperate with register-keepers and other entities that have a say on these issues is a crucial element for moving forward towards a register-based Census from 2021 onwards. Additionally, a strong political will, in the sense of establishing this process as a common goal and of empowering institutional involvement and implication, is equally important and decisive.

43.   The different stages of identification of relevant administrative registers and of assessment of their potential to produce census-like information has also provided an initial insight on the topics that will be difficult to measure based on administrative information. One of these topics is the issues concerning family relations. Not only is their operationalization not yet possible, conceptually there are several differences to be addressed, such as the census variables rooted within a more sociological conceptualisation of family and family relations. Another challenging topic to measure has to do with education variables. Although administrative registers in Portugal are kept regarding non-tertiary and tertiary education, unit level information is not universally linked to available identifiers, making the use of that information more difficult for statistical purposes. To overcome this situation a change in practice will have to be accomplished and information regarding students' identification, which is already collected, should be made available.

VI. Final Remarks

44.   The work programme that has been defined and the actions that are currently in motion will determine how and to what extent in the next 2021 Census round the transition to a register-based census will be possible in Portugal. The issue as to whether it will be possible to achieve a full transition or whether a combined method will have to be considered (data partly obtained from administrative registers, in addition to field enumeration or sample surveys) will depend on how some of the limitations identified are surpassed. Nevertheless, important steps have been taken to move forward and to create a set of political and institutional conditions for *changing the paradigm* in terms of the census design approach. The challenge resides in making the most of the positive context that is being created with regard to the possibility of using individualised administrative information.

References

Borchsenius, L. (2000). *From a conventional to a register-based Census of population.* Paper
    presented at the INSEE-Eurostat Seminar on the Censuses after 2001, 20-21 November, Paris.
Tonder, J-K. (2008). *The register-based statistical system: Preconditions and processes.* Paper
    presented at the International Association for Official Statistics Conference, 14-18 October,
    Shanghai.
United Nations (2007). *Register-based statistics in the Nordic Countries: Review of best-practices
    with focus on population and social statistics.* Geneva: United Nations.
Wallgren, A. & Wallgren, B. (2007). *Register-based statistics: Administrative Data for Statistical
    Purposes.* London: Wiley.
Wismer, K. (2003). *Use of registers on social statistics in Denmark.* Paper presented at the United
    Nations Expert Group Meeting on Setting the Scope of Social Statistics, 6-9 May, New York.

*****