28 May 2018
ECE/CES/2018/4/Rev.1

English only

(A summary document carrying symbol ECE/CES/2018/4 is available in English, French and Russian on the CES 66th plenary session webpage)

Economic Commission for Europe

Conference of European Statisticians
Sixty-sixth plenary session
Geneva, 18–20 June 2018
Item 4 (d) of the provisional agenda
Use of registers and administrative data for population and housing censuses

Guidelines on the use of registers and administrative data for population and housing censuses

Prepared by the Task Force on register-based and combined censuses

Summary

The document presents the *Guidelines on the use of registers and administrative data for population and housing censuses*. The Guidelines are prepared by a Task Force composed of representatives of the Netherlands (Chair), Austria, Canada, Estonia, Germany, Ireland, Israel, Italy, New Zealand, Norway, Poland, Portugal, Republic of Korea, Slovenia, United Kingdom, United States, Eurostat, United Nations Food and Agriculture Organization (FAO), United Nations Population Fund, and UNECE.

The UNECE-Eurostat Group of Experts on Population and Housing Censuses discussed the draft guidelines at their meetings in September 2016 and October 2017. The CES Bureau reviewed the Guidelines at its February 2018 meeting, and asked the Secretariat to send it for electronic consultation to all CES members.

The Guidelines start by presenting some general definitions and discussing the essential features of population and housing censuses with regard to register-based and combined censuses (Ch. III). Then the transition from a traditional census to a register-based or combined census is discussed, covering advantages, necessary conditions and possible difficulties (Ch. IV). Other chapters are dedicated to a common framework for register-based and combined censuses (Ch. V), data sources and their quality (Ch. VI), linkage and transformation (Ch. VII), statistical registers (Ch. VIII), output quality (Ch. IX), approaches and case studies from different countries (Ch. X). Case studies from nine countries are presented as annexes. Finally, a glossary completes the document.

A draft version of the guidelines (ECE/CES/2018/4) was reviewed by CES members through an electronic consultation carried out in March/April 2018. The present revised version of the guidelines includes various amendments proposed by the Task Force in response to the comments received during the consultation.

In view of the support received through the electronic consultation, CES is invited to endorse the Guidelines.

Acknowledgements

The present document was prepared by the UNECE Task Force on Register-based and Combined Censuses, with contributions of the following members: Eric Schulte Nordholt (Netherlands, Chair of the task force), Adelheid Bauer (Austria), Julie Trépanier (Canada), Diana Beltadze (Estonia), Thomas Koerner (Germany), Stefan Dittrich (Germany), Stefan Schweinert-Albinus (Germany), John Dunne (Ireland), Yael Feinstein (Israel), Gerardo Gallo (Italy), Donatella Zindato (Italy), Christine Bycroft (New Zealand), Harald Utne (Norway), Janusz Dygaszewicz (Poland), Sandra Lagarto (Portugal), Danilo Dolenc (Slovenia), Becky Tinsley (United Kingdom), Vincent Mule (United States), Adam Wronski (Eurostat), David Thorogood (Eurostat), Sorina Vaju (Eurostat), and Paolo Valente (UNECE). The document was edited by Ian White. The contribution of all these experts is acknowledged with much appreciation.

Table of Contents

I. INTRODUCTION

I.1. Background

1. Between 2012 and 2015 the UNECE Steering Group on Population and Housing Censuses coordinated the preparation of the *Conference of European Statisticians (CES) Recommendations for the 2020 Censuses of Population and Housing*. The Steering Group managed the work of nine topic-related Task Forces established to prepare initial drafts of the various chapters of the Recommendations. The CES subsequently adopted the Recommendations in June 2015, and these are available both in electronic format on the UNECE website[1] and in printed form in English, French and Russian.

2. In October 2015, the CES Bureau conducted an in-depth review of the diversification of population census methodologies and sources, based on a paper by Finland and Turkey (ECE/CES/BUR/2015/OCT/3) and a note by UNECE (ECE/CES/BUR/2015/OCT/3Add.1). As an outcome of the review, the Bureau supported the preparation of new guidelines on the use of registers for population and housing censuses, and requested the Secretariat to prepare new terms of reference for the Steering Group on Population and Housing Censuses and for a Task Force on Register-Based and Combined Censuses (Report of the Bureau meeting: ECE/CES/BUR/2015/OCT/21).

3. A draft of the proposed new Guidelines produced by the Task Force was presented and discussed at the meeting of UNECE Experts on Population and Housing Censuses in Geneva in October 2017 and consequently revised in the light of comments made by countries at the meeting. This publication presents the Guidelines as subsequently agreed by the CES.

4. Before the guidelines are presented, the following section summarises the census methods adopted by countries in the UNECE region and their evolution over time.

I.2. Census methods in the UNECE region and their evolution over time

5. There are many different ways to conduct a population and housing census. For the sake of simplicity, this document summarises only the three main categories of census methods: the 'traditional' census, the 'register-based' census, and the 'combined' census. However, a more detailed discussion of the various census methodologies is given in the CES Recommendations.

6. The **traditional census** is here intended to mean a census based on the direct count of all individuals and the collection of information on their characteristics through the completion of census questionnaires, either in paper form or electronically. The information is collected in the field across the whole country in a relatively short period of time, normally no more than two weeks. Questionnaires can be completed either directly by the households (with delivery and collection of paper forms undertaken by enumerators, the postal service or other methods, or online in the case of electronic questionnaires), or by the enumerators during an interview of the household.

7. The traditional census has a number of disadvantages. First of all, it is a very complex and expensive operation, mainly due to the need to employ a large temporary workforce for the field data collection (enumerators, supervisors and managers), and to print, distribute, and process a very large number of forms. Moreover, in most countries there are increasing difficulties in enumerating certain population groups, particularly those characterized by high

[1] http://www.unece.org/publications/2020recomm.html

mobility and multiple residences, and an increasing reluctance of the respondents to be enumerated, for various reasons. Finally, the traditional census is normally conducted only every 10 years (because of its cost and complexity) and the results often only become available after a relatively long time after data collection, while many users would like to have timelier and more frequently updated information.

8. Some countries have addressed some of the disadvantages of the traditional census either by using sampling (where most households complete only a short form with basic information, while a sample completes a more detailed long form, thereby reducing the total amount of information collected and processed), or by facilitating an online self-response option, which may result in field cost savings and improved quality but requires very careful planning and implementation. Another approach is to spread the fieldwork over time and adopt sampling, as it is done in the 'rolling census' approach developed in France[2].

9. A totally different approach from the traditional census is the **register-based census** that was developed by the Nordic countries in the 1970s[3]. Denmark was the world's first country to conduct a fully register-based population and housing census in 1981. Under this approach there is no direct collection of data from the population, and the traditional enumeration is replaced by the use of administrative data held in various registers (population register, building/address register, social security register, etc.) through a matching process, normally making use of personal identification numbers. This approach permits the production of census data at a much reduced cost and with relatively limited manpower, once a good quality system of statistical registers has been established.

10. Since the 1990s, a number of other countries in Europe have developed innovative methods to conduct the census, combining the use of administrative data with a limited collection of data from a field enumeration of the population for specific variables. Under this approach, called a **combined census**, the field data collection can cover the whole population or just a sample. Often this approach is adopted in the transition from a traditional to a register-based census.

11. In the 2000 census round only few countries in the UNECE region[4] conducted a register-based or combined census (three and five countries respectively[4]) and the traditional census was still by far the most popular approach in the region (40 countries)[5]. However, in the 2010 round, there was a significant increase in the number of countries conducting a register-based census (from three to nine) or a combined census (from five to ten), and a corresponding decrease of the traditional census (from 40 to 34 countries)[4] (see Figure 1).

12. Based on information on tentative plans for the 2020 round, the trend of moving away from the traditional census continues: out of 48 UNECE countries for which information is available, 14 countries plan to conduct a register-based census (29 per cent), 12 countries are planning a combined census (25 per cent) and 22 countries are continuing with their traditional

[2] INSEE, France, The French rolling census, ten years after its launch. Paper submitted to the Meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, Geneva, 30 September to 3 October 2013 https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2013/census_meeting/24_E.pdf

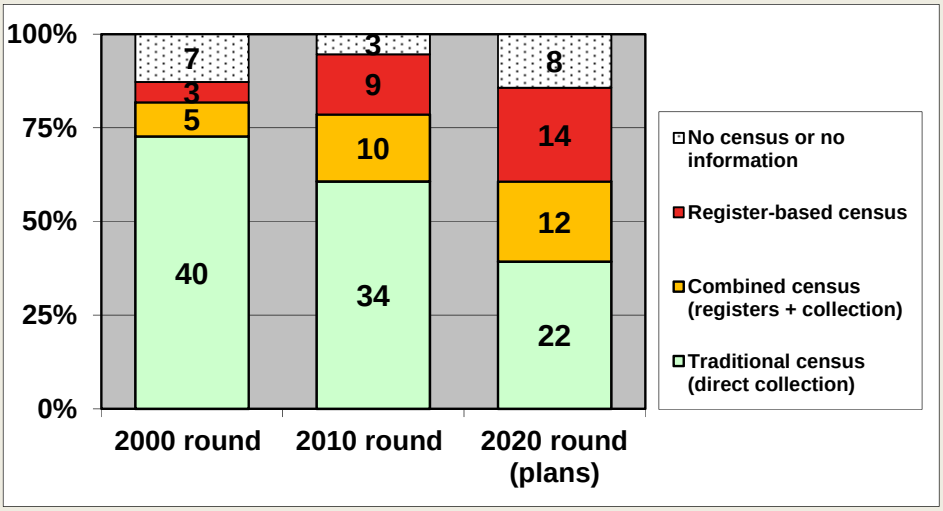
[3] See chapter 10 in: Register-Based Statistics in the Nordic Countries, UNECE, 2007 http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

[4] The UNECE region includes countries in Europe, North America, Central Asia, plus Turkey and Israel.

[5] Source: Valente, 2015, From the 2010 to the 2020 census round in the UNECE region – Plans by countries on census methodology and technology. Paper submitted to the Meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, Geneva, 30 September to 2 October 2015; http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2015/mtg1/UNECE_paper_Paolo_draft_0925_rev2.pdf

census (46 per cent). If only the 32 member countries of the EU (European Union) and EFTA (European Free Trade Association) are considered, then 13 countries plan a register-based census in the 2020 round (41 per cent), 9 countries a combined census (28 per cent), and just 10 countries will continue a traditional census (31 per cent).

Figure 1
Number of UNECE countries by census method in the 2000-2020 Census rounds[6]



Source: Valente, 2015.

II. SCOPE OF THE NEW UNECE GUIDELINES AND DEFINITIONS OF REGISTER-BASED AND COMBINED CENSUSES

13. The scope of these new UNECE Guidelines is not on traditional censuses, but on register-based and combined censuses. Therefore, only definitions of register-based and combined censuses are given. More information about traditional censuses can be found in both the global[7] and UNECE/CES Recommendations[8] for the 2020 census round. In these new UNECE Guidelines different kinds of registers (on persons and buildings) are noted with a focus on those used in censuses.

14. For some of the definitions we can refer to those presented in the UNECE publication *Register-based statistics in the Nordic countries*[9]. In para. 63 of that publication a ***register*** is defined as a systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up-to-date, correcting, or extending the register, that is, keeping track of any changes in the data describing the units and their attributes. ***Administrative data*** sources

[6] In 2006 Montenegro became independent and the number of UNECE countries increased from 55 to 56; at https://statswiki.unece.org/display/censuses/2020+Population+Census+Round an up-to-date list of country practices can be found.
[7] Principles and Recommendations for Population and Housing Censuses, Rev.3 (United Nations, 2017), see https://unstats.un.org/unsd/publication/seriesM/Series_M67Rev3en.pdf
[8] Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing (United Nations, 2015), see http://www.unece.org/publications/2020recomm.html
[9] http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

are data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. They include administrative registers (with a unique identifier) and possibly other administrative data without a unique identifier. ***Statistical registers*** are registers created for statistical purposes. They are typically created by transforming data from registers and/or other administrative data sources.

15. In some countries the term 'administrative data' is used as a synonym for 'register-based' data. In other countries a distinction is made between the two, and 'administrative data' is taken also to include administrative sources other than registers.

16. A register-based census system is built around a set of basic registers that contain comprehensive data on the units that are to be described in the population and housing census (see para. 123 of the CES Recommendations for the 2020 census round). Some register-based census countries miss some of the census variables in all of the available registers and choose to support their census with unit record data (microdata) from an already existing sample survey. All register-based census countries have in common the fact that no specifically designed census questionnaires are used to collect information about the population. Therefore, register-based censuses are in general much cheaper than combined censuses and especially so compared to traditional censuses.

17. In a combined census, statistics are created by using registers and other administrative sources, together with information from either sample field data or full field enumeration for selected variables (see paras. 52 and 116 of the CES Recommendations for the 2020 census round).

18. The remainder of the Guidelines are organised as follows. Chapter III describes the essential features of a census and how these may be met by register-based or combined censuses. Chapter IV describes a number of elements that need to be taken into account when planning a transition from a traditional census to a register-based or combined census. Chapter V outlines a common framework that describes the process of conducting these non-traditional censuses. Chapters VI - IX provide more details on the processes and methods associated with each aspect of the framework and the role that quality assurance plays at each stage. The final Chapter X, together with Annexes A-I, present case studies from a number of countries that have transitioned or plan to transition from a traditional full-enumeration census to a register-based or combined census.

III. ESSENTIAL FEATURES OF A POPULATION AND HOUSING CENSUS

19. The essential features of a population and housing census were originally defined by the International Conference of Statisticians as early as 1853 in Brussels, and by incorporating these features, countries have been able to carry out censuses that have been internationally comparable in terms of methodology and quality over time. Nowadays, these five essential features have been redefined and highlighted by the CES[10] with the aim of ensuring the coherence of census data gathered in different countries with different levels of technical development and different cultures. Adopting all these features – regardless of the methodology of data collection – enables NSIs (National Statistical Institutes) to collect population data of

[10] Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing (United Nations, 2015), see http://www.unece.org/publications/2020recomm.html, paragraphs 23-28.

internationally comparable quality that allows making decisions and forecasts on the population development.

20. The five essential features of a census are:
- Individual enumeration;
- Simultaneity;
- Universality (within a precisely defined territory of a country);
- Small area data;
- Defined periodicity.

How register-based and combined population and housing censuses can be designed to satisfy each of these features is discussed in the following sections.

III.1. Individual enumeration

21. The principle of individual enumeration is a fundamental feature for any census of population. Traditionally, this has been effected by providing questionnaires that ask questions of each individual within a household. In the case of register-based censuses a different approach is adopted where the data are taken from administrative registers. In such circumstances it is important that each census unit (i.e. individual, household, or dwelling) has a special, uniquely identified, record in the registers used. Then the registers become a useful source for the census. In the case of a combined census only some of the variables are derived from administrative data sources, and then this same approach is used for those variables.

22. If a single identifier for a particular unit does not exist across a range of registers, it is necessary to create a new statistical identifier (based on a group of identifying variables) to link the variables held in the respective registers, and to check carefully its quality (for errors and uniqueness).

23. Sometimes it is necessary to create the necessary census variable using information from several administrative registers and composing special algorithms for its calculation. This is possible if the units in all these registers are uniquely identified by the same identifier. In this case the variable created in such a way should be uniquely identified as well and saved in a statistical register. In the event that not all units are uniquely identified by the same identifier it may be possible to create a new statistical identifier as explained in the next paragraph.

24. The basic counting units of a population and housing census include not only persons, but also households, families, and dwellings (whether occupied or vacant). All of these units need identification, but there is no need to use five different identification variables. The minimal necessary identifiers are the one for persons (person ID) and the one for dwellings (dwelling ID). These IDs must be linked with each other (a dwelling ID is assigned to each person) and for each occupied dwelling the list of person IDs of people living in it must be given. The dwelling ID makes use of the address code, which may contain also spatial coordinates.

25. Information about households is usually collected on the basis of the housekeeping concept[11] by those countries conducting a traditional census. This definition can be achieved through asking questions on a survey or census, but is more challenging for countries conducting a register-based census. Many such countries instead use the household-dwelling concept, which considers all persons living in the same housing unit to be members of the same

[11] See paras. 768-769 of http://www.unece.org/publications/2020recomm.html

household. While adopting this definition has minimal impact on the total number of private households, it can have a larger impact for certain household types, such as one-person households. This bias in the number of private households and in the estimated structure of the household types depends on the traditions of the country and on living conditions. These challenges for register-based countries also extend to construction of families within households using relationship information.

26. In some countries (such as, Slovenia, see Annex E) a household register exists. The existence of such a register eases the organisation of a register-based census, especially when household IDs are included in the register. Then there is accurate information available about which person ID belongs to which household ID. A household register might therefore improve the quality of a register-based or combined census significantly. However, the situation of Slovenia is an exceptional one. Ireland is researching the potential of using a decision tree algorithm to determine relations between people in the same dwelling so that the current housekeeping definition of the household can be continued (see Annex A). As part of the work to understand the impact of transitioning to its so-called Administrative Data Census (see Annex G), the Office for National Statistics (ONS) in the United Kingdom is currently exploring the potential impact on users of changing from a housekeeping concept to a household-dwelling concept .

27. Sometimes it is useful to use identification codes for other units such as enterprises and organisations. If these are linked with person IDs and dwelling IDs they form a helpful tool for deriving other statistics, such as on commuting between place of residence and place of work.

III.2. Simultaneity

28. The fixed census moment is the condition defining the simultaneity of the census data. Traditionally, to ensure this condition, the enumeration is carried out over a very short time frame, ideally during one day only. Though most modern day enumerations are conducted over a two-three week period, all the data collected should refer to a specified reference period. This essential feature should be respected also in the case of a register-based or combined census.

29. If the administrative registers in use are regularly updated, then it is necessary to fix the census period and to take the data from all registers with reference to this period. Sometimes the registers are updated regularly at some specific date such as the beginning of a year, and then it is possible to use this date as the census period, and the simultaneity of the census is guaranteed.

30. In the case of a combined census it is important that the census reference period mentioned in the questionnaires and the reference period of the information taken from the registers are the same or as close to each other as possible.

31. When several administrative registers are used in the census, it is important that all data taken from them have the same reference period. Usually, census variables derived via special algorithms take some time to calculate; hence those census variables are only ready for publishing sometime after the census period. For some specific variables in combined or register-based censuses different reference periods are defined for the particular administrative purpose of the register. Demographic data can normally be taken from population registers at the beginning of the year. However, labour force data might be more relevant somewhat earlier in the year - as fewer people tend to be in employment around Christmas and New Year’s Eve. For some administrative registers, it may not be possible to have Census Day as a reference

day. For example, education registers often have relevant education data (e.g. referenced to a day early in the academic year) that may differ from a chosen Census Day. In such cases, the NSI may wish to take education data with a reference day as close as possible to Census Day as a compromise.

III.3. Universality (within a precisely defined territory of a country)

32. To ensure the universality of the traditional census the questionnaires used in the enumeration process are the same for all households and individual persons. If there are questionnaires in different languages, it is important to check if their content and the meaning of all questions is exactly the same.

33. If the administrative registers used in the census are common for the whole country and all population groups, the condition may be regarded as being met. However, if there are different administrative data in different areas or for different population groups (such as an urban population register and a rural population register or if different administrative data are held in different cities), then it is necessary to analyse the possible discrepancies between the different administrative sources and find a way to define common census variables using these different administrative sources. In this case plausible results can be derived from these newly defined variables (via an appropriate algorithm) in a statistical register.

III.4. Small area data

34. Providing a rich wealth of information for small geographic areas and small population sub-groups (generically referred to here as ‘small area data’) is a key objective for a census of any kind, as there is generally no other single source of comparable data.

35. In the context of register-based or combined censuses, small area census data can be derived from administrative data providing the coverage is high, preferably covering the whole population. If there are some small areas that are poorly covered, resulting in the administrative data lacking some information (and thus showing poor universality), it will be necessary to seek to improve the administrative dataset before it can be used as a source for the census. Such poor coverage is likely to be a problem for the every-day usage of the administrative data, and so it should be improved anyway. Improving the statistical register can sometimes be done by adding information from another source, providing that linkage through common IDs is possible.

36. Sometimes there might be special administrative data for some small areas (particularly for some small groups of people). Then it will be necessary to combine these different administrative sources (see section III.2.). If the result is satisfactory, this combination is useable. In the case of a combined census it is also possible to supplement the lack of information in administrative sources with a survey that, if necessary, may use different data collection methodologies, such as doorstep or telephone interviews, or self-completion with paper or online questionnaires, to suit different areas. However, where a sample survey is used, problems can still arise with the coverage of small area data. In such circumstances, users’ needs regarding the required level of detail of the census outputs should be taken into account before any decision is made on the sample size of the census survey.

III.5. Defined periodicity

37. Nowadays, censuses are generally organised worldwide on a ten-year cycle. The United Nations recommend that countries conduct at least one census every ten years (between 2015 and 2024 for the 2020 census round). To meet European Union requirements, member countries were required (by an EU Regulation) to conduct a census in 2011, and will similarly have to conduct the next one in 2021. However, some countries (such as Australia, Canada, Ireland, New Zealand, and Slovenia) have shorter periods between censuses. The same underlying ten-yearly cycle should be kept in all censuses, regardless the methodology to provide international comparability. If a five-year period has been used, then, for EU member states, one of the two census years of the country should coincide with the census year fixed by Eurostat.

38. An advantage of a register-based census is, however, the opportunity to conduct the census more often than the usual ten-year cycle, as register data are permanently available and more regularly updated. It is advisable also to prepare census software in such a way that it is permanently ready for using for any reference date. Then the periodicity between censuses can be ten, five, or two years, or censuses could even be conducted annually. Yearly updates of demographic data are a key objective in many European countries that now adopt different census methodologies.

39. It is also possible to produce some census updates with a shortened list of variables (but long enough to meet users' requirements). This could save resources. From this it follows also that in countries where a regular decennial census will be continued using a combined methodology, some updates with a shortened list of variables can more easily be done more often if the variables on the shortened list can be derived from administrative data sources.

III.6. Conclusions of the features defined by the CES

40. From the above it can be concluded that if a country has a good system of administrative registers that is consistent, easily useable and of high quality (that means, all units are uniquely identified using a common identifier), or if statistical registers can be built with equivalent quality from administrative data sources, then it is possible to organise a population and housing census that satisfies all required CES features.

41. If the list of administrative data sources cannot achieve the required quality for the whole range of census variables, it may still be possible and useful to derive census variables using a combined or register-based methodology with a shortened list of variables (including those covered in administrative registers) collected in between regular censuses.

IV. CONSIDERATIONS WHEN TRANSITIONING FROM A TRADITIONAL CENSUS TO A REGISTER-BASED OR COMBINED CENSUS

42. The decision to move to a combined or register-based census is motivated because of the several advantages to be gained. However, the move needs to be carefully managed, and there are necessary conditions relating to data, technology, legal and stakeholder issues for a successful transition from a traditional to a combined or register-based census. This transition may therefore give rise to some challenges. These are set out in section IV.2 and can present significant obstacles for some countries.

IV.1. Advantages

43. Conducting a combined or register-based census has a number of advantages and opportunities that are described below.

IV.1.1. Lower per capita costs

44. Traditional censuses are very expensive. In many countries that conduct a traditional census it is common that the census costs are equivalent to about two annual budgets of the NSI. It is understandable, therefore, that governments put pressure on the institutes to cut these expenses, especially when other data sources are available.

45. If a combined census is conducted with full field enumeration for some of the variables, the cost savings made by shortening the census questionnaires are partly lost again in the effort required for combining the information from the field with administrative data sources. While the savings may be modest, the approach may still be preferable, particularly where the transformation process allows more data from administrative sources to be used in future censuses.

46. If a combined census is conducted without full field enumeration for selected variables much larger savings can be achieved. Practice shows that introducing a combined census could lead to reducing the costs by 22 per cent compared to a traditional census[12]. Moreover, if a register-based census is conducted no surveying is required specifically for the census and large savings can be expected. Several countries have proved that on average 98 per cent of the costs of a traditional census can be saved this way[13]. However, one should realise that such census cost savings can only be reached once the necessary access to the appropriate registers has been established.

47. Some countries have switched from a traditional census to a register-based census in one census cycle. If this is attempted, then all the costs of making the change must be met within the decennial period. More usually, however, such a move is done in several stages, often by first adopting an intermediate combined census approach before moving to a register-based census. The cost of the change can then be spread over two or three census cycles.

48. Moving to a combined approach and, especially, to a register-based approach contributes to a more cost-effective census. It is clear that meeting government-imposed budgetary constraints provides an incentive for such moves, even if registers are incomplete or of insufficient quality to be used as sources for the census. In such cases, however, it should be made clear from the beginning that a country should not attempt a move immediately and should continue conducting some form of traditional census. However, even when countries continue to do so, innovations making greater use of administrative data could help the NSI work more efficiently. It helps if the relevant public authorities make administration data sources available to the statistical institutes to produce proto-type register-based statistics. The government itself can help a great deal both by removing any legal barriers to data sharing and by subsidising the transition.

[12] Calculation based on PPP information in Table 7.2 of http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf
[13] Calculation based on PPP information in Table 7.2 of http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf

IV.1.2. Quicker to conduct

49. In countries with an established register-based statistical system, the total production time to conduct a register-based census is much shorter than any other kind of census, due principally to the fact that no field enumeration has to be conducted. This, of course, does not hold for combined censuses in which some fieldwork is still necessary. Without fieldwork, the census may generally require less time in the planning stages (thus having the advantages of saving both time and money) while maintaining - or even improving - the time taken for the delivery of output. However, one has to realise that the first time a register-based census is conducted it may take more time than in later census rounds as census planning has to be set up anew.

IV.1.3. Fewer problems with non-response and reduced response burden

50. With no need to conduct any field enumeration, fewer problems with non-response and a reduction in response burden to zero can be expected if only, or mostly, administrative data sources are used, and providing these are comprehensive and cover the whole population. With ever-increasing non-response rates being reported in international censuses and surveys this is going to be a more and more important aspect. Even if the move is to a combined census, the response burden on the population will also be lower, in particular if only sample field information is collected.

IV.1.4. Possibility of a continuous census

51. The more administrative data a country uses in its census, the better the possibilities of an annual or even more frequent census. In theory, with good quality administrative data updated outputs could be produced on a daily basis. Such real-time censuses may be something for the future, but more regular than decennial data are now becoming increasingly expected by users. As information from other sources can provide such real-time information, censuses will be expected to keep pace. As a by-product, it is easier for the NSI to keep the knowledge and IT-infrastructure up-to-date if annual census updates are produced.

IV.1.5. Better cooperation between units within the NSI

52. In some NSIs the different divisions or directorates are often structured to perform in isolated silos without too many contacts with each other. If an NSI moves to a register-based statistical system there is potential for this silo structure to be abandoned. By moving from a survey-based statistical organisation towards a register-based system, the traditional one-to-one relationship between sources and statistics is replaced by an m to n inter-relationship across all statistical branches. A better cooperation of units within the NSI thus becomes essential. Moreover, by better integrating statistics, the coherence of the statistical framework within the NSI is improved.

IV.1.6. More time and resources for innovations

53. Innovations are crucial for the long-term perspectives of an NSI. As introducing statistical registers in the process of producing official statistics saves both time and money, it becomes easier to innovate. The resources saved could, instead, be used to stimulate administrative and technological innovations so that the processes of data production remain up-to-date.

IV.1.7. More flexible and responsive to new information requirements

54. If all data are stored appropriately, not only can the regular statistics be produced more frequently, but there is potential for the creation of new statistics to meet changing user needs. The NSI can then become more flexible and responsive to new information needs and increase its value to society. Although this may not be an aim in itself, it can lead to a greater level of user satisfaction.

IV.2. Necessary conditions for a successful transition to a register-based or combined census

55. If a country wants to move to a combined or register-based census a number of conditions are necessary before information from administrative registers (and other sources) can successfully be integrated to create the underlying statistical register. A number of these necessary conditions can present some challenges to the NSI, and these are discussed in the following paragraphs.

IV.2.1. Legal base

56. Whatever type of census an NSI conducts, it should be within an established legal framework. To be able to conduct a combined or register-based census, in particular, there must be legal provisions that prescribe the access to, and protection of, the administrative data. Such a legal base is normally enshrined in a Statistics or Census Act.

57. The NSI must have legal authority to access the relevant administrative data, ideally free of charge, from whatever public authority sources, preferably including personal identifiers. To avoid legal uncertainty or dispute, it should be stipulated that the right of access applies except in legal cases pertaining to the protection of public order or the security of the country. A further key issue that should be addressed in legislation is the legal provision for the NSI to have some influence or authority in the creation, revision or deletion of those administrative data that are to be used in the statistical registers.

58. In turn, the NSI must have a legal obligation to protect the confidentiality of the administrative data it obtains, and to adhere to the 'one-way traffic' only principle, except under specific circumstances mentioned in the legislation. Indeed, the relevant legislation could do more by generally prohibiting other data controllers having access to data held on the NSI's statistical registers.

59. In certain countries, legal requirements may constrain how a census can be conducted[14]. In some countries, the NSI first started exploring administrative data sources and, thereafter, found a legal base to make register-based statistics possible. In other countries, the legal base was first established and, thereafter, register-based statistics were produced and published. To gain experience in moving progressively to a register-based census it is often simpler for NSIs to start with producing register-based statistics covering just a selection of those variables collected in a traditional census, though it should be noted that the legal obstacles to overcome may be no fewer.

60. It is always the case that moving to a census methodology where administrative data sources play a role needs careful preparation including, in particular, pilot studies. NSIs should realise that once (part of) the fieldwork operation for the census is abandoned, reinitiating it

[14] Particularly where representation in the national legislature depends on census results.

becomes rather difficult. After the passage of time, the knowledge of how to conduct a traditional census is lost, and especially so in the case of a register-based census where no fieldwork is undertaken at all.

61. Legislation on access to administrative data may need to be supported by policies and directives that are internal to the NSI and that translate legislative requirements and central government policies and directives into requirements and responsibilities for the managers and employees of the NSI.

IV.2.2. Public approval

62. While the law might give a legal licence for the NSI to a combined or register-based census, public approval is also necessary to ensure that such a census is acceptable. This might be more difficult to achieve than establishing the legal base. While in some countries people may get the impression that 'big brother is watching you' in the course of a traditional census, in some other countries using and linking administrative data collected for non-statistical purposes may be seen to be even more intrusive as the public has no control at all over the information about them that is to be disclosed.

63. In a traditional census, privacy concerns may lead to lower response rates or the deliberate giving of wrong information. It is getting more and more difficult to correct for such unit and item non-response. So, on the one hand the public might prefer a situation where fewer questions are asked if the equivalent information is already available. On the other hand, part of the population may prefer to answer census questionnaires directly instead of having their information taken from, and combined with, several administrative sources.

64. In a register-based or combined census people may feel uneasy about, or even object to, information from different administrative data sources is reused and linked in a census. It may not be clear to them that in the census the information is only used for statistical purposes. If no more census forms have to be filled in and only registers are used for the census, the public will generally be less aware that a census is being conducted. However, the absence of any public reaction should not be misinterpreted as public approval.

65. It is desirable, therefore, to prepare for possible specific questions on privacy, confidentiality and security issues in conducting a register-based or combined census. In combined censuses a discussion can be expected about which variables are to be included on the census questionnaires and which variables are to be derived from administrative data sources.

IV.2.3. Stakeholder approval

66. Stakeholders - or, more specifically, data users - typically want each census to provide at least the same level of detail of information as in the previous census. However, this is not always possible when the census methodology changes.

67. It is important to inform and consult stakeholders beforehand. Users in particular can become critical if their expectations are not met. Even well-informed users can become very critical if they believe that they are going to lose some of the information that they had access to in the previous census. However, it is usually not possible to satisfy all users, and disappointment among some of them is often unavoidable when adopting a new census methodology.

68. It is important, therefore, to have a communication strategy for stakeholder engagement that should encompass some, or all, of the following goals:

- Create a transparent environment concerning the plans of the NSI;
- Assure users that their requirements will be taken into consideration;
- Inform stakeholders of the benefits of using administrative data and demonstrate that the information will continue to be kept secure;
- Strengthen partnerships with the stakeholders so that the NSI can benefit from outside expertise;
- Make stakeholders part of a successful transition to a new census approach.

69. Openness and the clear identification of new opportunities and benefits for the stakeholders will help to gain their approval. It is particularly important that there should be adequate consultation on any change in the provision of those statistics that have financial consequences for stakeholders (such as transfers of money to municipalities).

IV.2.4. Cooperation between the NSI and other authorities

70. Good cooperation between the NSI and other (mainly government) authorities is vital in using administrative data sources in the census. The NSI needs to know when microdata (the administrative unit records) and the accompanying metadata can be made available before any register-based statistics can be produced. In a combined census, and even more so in a register-based census, the NSI is heavily dependent on administrative data holders to comply with their agreed or legal obligation to provide good quality data on time. If data holders fail to deliver, it is usually the NSI that is held responsible for the failure to publish census statistics on time.

71. It is vital to inform administrative data holders how important their data are for the NSI and how their data are to be used. Additional to a legal base (see subsection IV.2.1.) and good contacts with other authorities, signing cooperation contracts or service level agreements could help in supporting the census process. In theory, administrative data sources from non-governmental authorities could also be used in the census, but this often creates privacy and data quality concerns and involves commercial considerations; private sector data are more often than not only acquired at a substantial cost to the NSI.

IV.2.5. Comprehensive and reliable statistical register system

72. A statistical register system that is comprehensive and reliable (in that it contains accurate and timely data) is essential for conducting a combined or register-based census. Administrative data sources, including administrative registers such as a population register, are not normally set up for statistical purposes such as conducting a census. A transformation process is therefore necessary in order to create a reliable statistical register system (see Chapter VII).

73. To assure the use of register-based statistics in official statistics it is important to have good working relations with administrative data holders. Conditional on the provisions of the legal base, and if the administrative bodies are conducive to it, in some countries there is also a potential to improve the relations between the data holders and the NSI by introducing new, or extend existing, register-based statistics, such as longitudinal studies to evaluate policy implementation. Of course, there has to be contact between the NSI and the relevant administrative data holders whenever an administrative data source is introduced as a new source. However, permanent contacts, facilitated for example via account managers, are vital to

keep both the administrative data holders aware of the important role that their data play and the NSI informed about any changes to the microdata and metadata they receive. Only with regular contact between the NSI and the administrative data holders can the success of register-based statistics be maintained.

74. In many register-based census countries the system of administrative data sources is used by many different public authorities. The more users this system has, the better the quality one can expect. In using such a system for the census it is the quality of the resulting statistical register rather than the quality of the underlying administrative data sources that counts: are the data of good enough quality on which to base reliable census outputs?

IV.2.6. Unified identification system

75. A unified identification system across different administrative data sources greatly facilitates register-based censuses. It is preferable to have unique ID-numbers at the unit record level that are common across all registers. For countries where unique ID numbers for persons do not exist, the ability to link data efficiently and accurately is a particular challenge.

IV.2.7. Knowledge of administrative sources

76. When a country wants to move from a traditional census towards a combined or register-based census, building up a wide-ranging knowledge of the data held in administrative sources is important before actually making the move. Although building up knowledge can be effected in stages well before the census planning, the effort needed to make this process successful should not be underestimated. Many lessons about failures and successes can be learnt from countries conducting combined or register-based censuses, but the national context should always be taken into consideration. It is never advisable for an NSI to simply adopt the methodology of another country when setting up a combined or register-based census. However, by learning from the experiences of others the transformation period can be shortened drastically.

IV.2.8. Transparency

77. If there are planned moves to a different census methodology, it is good practice to be transparent and share with stakeholders information on plans and tests as much as possible. As discussed in section IV.2.3 above, it is particularly important to inform users of any decision to move towards a register-based census as such an important change in methodology is likely to have an impact on the content and availability of output. Transparency and openness facilitate external review and feed-back on the new processes.

IV.3. Difficulties that may arise

78. Despite the advantages noted in section IV.1 above, conducting a combined or register-based census has a number of disadvantages and risks that are described below.

IV.3.1. Dependency on public authorities

79. In moving to a combined or register-based census, the NSI becomes heavily dependent on the public authorities holding the administrative records being used. NSIs have to realise that, for such authorities, the production of statistics is not a core activity to which they would

normally give priority. For the NSI, any failure or shortcomings in the administrative registers will affect the quality of the derived official statistics, for which it must take responsibility.

IV.3.2. Differences in concepts and definitions

80. Registers and other administrative data sources often adopt different concepts and definitions of population-related variables than those that generally apply in traditional censuses. NSIs should be aware that such differences may exist and decide whether these differences are acceptable when moving from a traditional to a combined or register-based census. What may, in one country, be considered an acceptable difference when assessing the balance between the continuity and coherence of the resulting statistics and the reduction in field costs, may be considered unacceptable to users elsewhere. NSIs should weigh up the balance before deciding whether they are willing to pay this price when moving towards a register-based census or a combined census without full field enumeration for selected variables. Sometimes, original definitions and concepts can be approximated rather accurately by derivations from different sources or by editing information from newly acquired census sources. However, this is not always the case and the NSI should then weigh up the balance between the acceptability of the differences and the costs of continuing full field enumeration for selected variables.

IV.3.3. Timeliness of administrative registers

81. Public authorities responsible for maintaining administrative registers do not hold the data for statistical purposes, and, as a result, will have other priorities that could cause delays in the delivery of the relevant administrative data and metadata to the NSI. This can cause issues for the NSI regarding the timeliness of their register-based statistics, particularly where the timeliness of the delivery of data from different sources varies considerably.

IV.3.4. Different reference periods

82. A particular problem that NSIs encounter when moving towards a combined or register-based census is that different sources of administrative data often have different reference dates. Sometimes a source gives the option of distinguishing clearly between reference dates and dates of events, but this good practice does not always apply. If these problems cannot be resolved sufficiently, the risk is that not all sources will be harmonised to the same reference date. Then the question ‘What is an acceptable difference in reference dates?’ arises. However, the answer to this is dependent on the variable concerned. Some variables are rather stable over time and then a small difference in reference date is normally not a problem. Large differences in reference dates are always unwanted. Finally, it is relevant to realise that also in the case of census questionnaires not all information may in practice refer to the single reference date of the census. Especially in case the census information is given on a moment further away from the reference date recall effects may play a role and respondents do not always give the answers specific to the census reference date.

IV.3.5. Privacy and security concerns

83. Using administrative data for purposes other than those for which the information was originally obtained inevitably leads to privacy and security concerns. These concerns often relate to the linkage of personal data from different sources. In some countries the legal framework has been specifically adapted to provide for this, suggesting that there is public

approval (or at least acceptance) in the use of administrative data in official statistics. For other countries such a consensus has not yet been achieved.

IV.3.6. Difficulty in identifying sub-populations

84. In a census it is important to ensure universality. However, the range and detail of outputs in a register-based census will be limited to those variables that can be derived from existing sources. These may not all relate to the entire population. Moreover, even for those countries that use sample surveys to collect data on information not available in administrative sources, it is sometimes difficult, or even impossible, to produce accurate outputs for small areas or specific sub-groups of the population because of the size of the sample population.

IV.3.7. Keeping knowledge and IT infrastructure up-to-date

85. In countries that conduct census projects with large gaps between them, it may be difficult to retain staff within the NSI with the necessary experience and expertise to keep the knowledge and IT infrastructure up-to-date during the inter-censal period. However, when yearly census updates are introduced this difficulty is minimised.

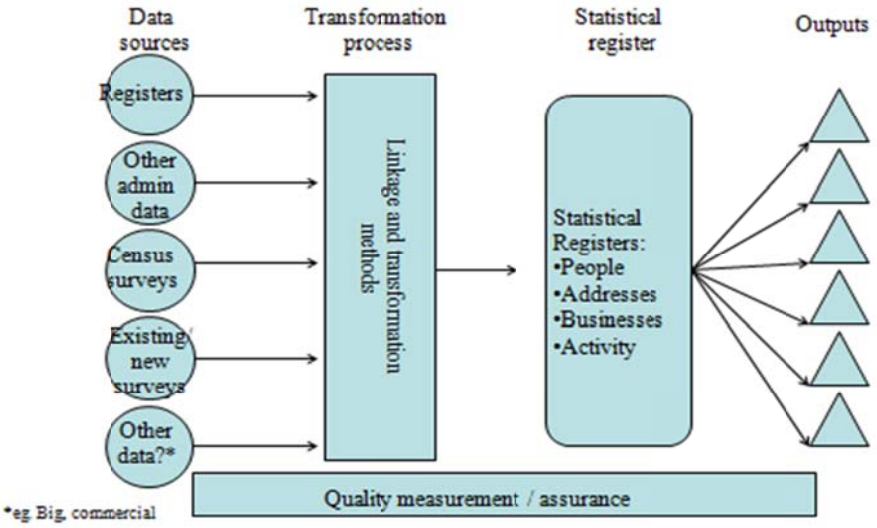
IV.3.8. Diminishing interest

86. In countries where censuses are carried out using questionnaires, not only the users but the general public itself will be interested in knowing the results. However, in register-based census countries, where people no longer complete census forms, there is often a decline in public interest in census results. Many people will not even be aware that a census has been taken and, as a consequence, national interest in the census is greatly reduced. Users will still have an interest in the statistical outputs, though evidence from register-based census countries suggests that their interest in the choice of original sources and the methodology used to produce the census data diminishes over time. This is also due to the fact that in those countries other outputs are often much quicker available than census results.

V. COMMON FRAMEWORK FOR REGISTER-BASED AND COMBINED CENSUSES

87. As noted in Chapter 1 of these Guidelines, increasingly more countries are moving towards register-based or combined censuses. However, the methods and processes that each country may take to deliver such a census can vary. It is therefore helpful to consider a common framework that can be applied, showing the key stages required – as outlined in Figure 2.

Figure 2: Common framework for register-based and combined censuses



88. This framework is divided into five key stages:

- Identifying data sources;
- Transformation process;
- Constructing statistical registers;
- Disseminating outputs;
- Quality measurement/assurance.

The remainder of this chapter describes briefly each of these stages.

V.1. Identifying data sources

89. So far register-based censuses have been conducted based on population registers, but more recently countries that do not have an established national population register are exploring how they may combine other administrative sources to create an equivalent statistical population register. Countries typically use a range of other sources to improve the quality or range of outputs that can be produced from a register-based census. This might include administrative data, sample surveys and other data sources such as big data or commercial data.

90. For example, in its work to explore the potential to move to an Administrative Data Census (see Annex G), the Office for National Statistics (ONS) in the United Kingdom has created a Statistical Population Dataset (SPD)[15] from which it would produce census-type outputs. The SPD aims to produce a single, coherent dataset that forms the basis for estimating the size of the usually resident population. It is produced by linking records across multiple administrative data sources and applying a set of inclusion and distribution rules, to good effect. Similar work is currently conducted in the Central Statistics Office of Poland (CSO) (see Annex C).

91. When carrying out a population and housing census using numerous data sources, including extensive use of administrative registers, an important issue to consider is the

[15] For the latest methodology, see https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20

possibility to integrate sources. The easiest way for linking different data sources is using unique identifiers for persons and for addresses (as noted in section III.1). In some countries there are also other unique identifiers - for enterprises, for example - that are very useful for data linking. Having identifiers constituting integrating variables allows for the use of data from various sources. However, not having those identifiers should not necessarily constitute a barrier to the integration of numerous data sources. The process of combining data from several sources relating to the same units can be carried out using various deterministic and stochastic methods. In the case of a deterministic method the key variable combining all the sets is an identifier occurring in all sources. In the case of a stochastic method the appropriate identifier ought to be created using attribute information that is common throughout the combined sources.

92. The range of sources that might be included to meet the essential features outlined in Chapter III is likely to differ for each country. For those conducting a combined census, the key difference is the inclusion of census data from either sample field data for selected variables or full field enumeration for selected variables.

V.2. Transformation process

93. As data in population registers and other administrative sources are not primarily collected for the purpose of a census, a transformation process is required to produce a reliable statistical register system. For countries that use such data as part of their register-based or combined census, there is usually a need to link data from a range of sources and transform these linked data into statistical registers.

94. The first step is to link the different data sources that are to be used. In countries where a unified identification system is available (that is, where all relevant registers and data sources use the same unique identifier for people - preferably ID numbers), this process is relatively straightforward and the resulting error in the linking process is minimised.

95. This is more challenging for countries that do not have such a unified identification system, or that are using administrative data sources without a unique identifier. However, sometimes this can be achieved by linking through a set of identifiers (such as name, sex, date of birth, numerical address) resulting in matches that can then be used to produce outputs of similar accuracy to those produced through linking registers that do contain a unique identifier. Some methods are discussed briefly in Chapter VII. Countries in this situation may consider assigning a unique identifier as part of the creation of the statistical register to ensure that data can be integrated and used effectively. If such successful links (that is, links that are of good quality and thus not burdened by too many errors) are not possible, it is likely to be impracticable for a country to move to a combined or register-based census, but nevertheless the administrative data may then be used for benchmarking or quality assurance purposes (see Annex A for an example in Ireland).

96. Once the data have been linked, further transformation may be required to create or improve the quality of the statistical register. In Chapter II, we defined administrative data sources as "*data holdings that contain information collected primarily for administrative (not research or statistical) purposes*". For countries that use such data as part of their register-based or combined census, there is usually a need to transform the linked data into statistical registers. There are two main aims of building statistical registers in NSIs. The first is data cleaning and editing; the second is forming the census variables using special algorithms. Sometimes, to meet this second aim, data from different registers can be effectively combined

in a statistical register. This process will vary between countries, and may include adding information from additional data sources (through linkage), and by carrying out some statistical procedures. Such examples might include cleaning, deleting erroneous values, resolving discrepancies between sources (for example, address information may differ across a range of sources), editing and coding data, investigating and resolving missing values (possibly through imputation), and selecting records that meet the population group of interest (for example, those resident on Census day).

97. Statistical registers are necessary for all NSIs conducting register-based censuses, but the ability to transform multiple administrative data sources into a statistical register is often a key challenge for NSIs without population registers. There are some variables that should be derived using the information from different registers and applying complex algorithms. All these procedures are much more effective using statistical registers.

V.3. Constructing statistical registers

98. A comprehensive and reliable system of statistical registers is essential in order to conduct a combined or register-based census. Usually, it is useful to have a system consisting of several registers that can be connected through links between the identifiers on each of the primary register units showing how these units are related. Wallgren and Wallgren (2014) refers to the creation of four separate linked base statistical registers:

- Population register – a register of residents of the country. This may exist in many versions relating to: (a) the current population, (b) the population at a specified point in time (such as 31 December), (c) all changes during a specified period (such as a calendar year), and (d) the population that is continuously present for a specified period (such as over a calendar year);
- Address/dwelling register – a register of addresses/dwellings;
- Business register – a register of businesses;
- Activity register – a register that holds information about residents' different activities. This register usually consists of three sections: (a) employment or job activities, (b) study activities, and (c) other activities relating to the labour market (for example spells of unemployment, military service, benefits and pensions).

99. In the Estonian registers' system, for example, there are four registers, but their content is a little different:

- Population register – a register of residents of the country;
- Address/dwelling register – a register of addresses/dwellings;
- Business register – a register of businesses;
- Farm register – a register of agricultural households.

All the registers are linked with different identifiers (person ID, address ID, enterprise ID and farm ID). Estonia is transitioning from a combined to a register-based census (see Annex B).

100. The Polish experience shows that it is strongly recommended that the address/dwelling register should include a geographic component for spatial location of each dwelling (building) with highest possible precision of x-y coordinates. This would specifically allow the use of GIS technology to support the main stages of census field operations and subsequent spatial analysis. Poland conducts a combined census (see Annex C).

V.4. Disseminating outputs

101. An essential part of transitioning to a register-based census is demonstrating the ability to produce a range of standard census-type statistics with associated measures of quality before the traditional census is abandoned. By having an effective and compatible system of registers (including statistical registers) it is possible to disseminate census data periodically with shorter time intervals than ten years. As previous noted, annual census-like tables then become possible, and indeed already exist in the Netherlands and a number of countries.

V.5. Quality measurement/assurance

102. Quality measurement and assurance should be undertaken throughout all stages of the framework, though the extent and methods for conducting quality measurement and assurance processes will vary between countries.

103. In a broad sense, these processes can be broken down into input quality (the quality of the input data), process quality (changes in quality as each process is added), and output quality (the quality of the resulting statistics). Some countries may want to add extra quality measurement processes, for example the case study by England and Wales (see Annex G) describes an extension of this framework to include a Population Coverage Survey not only to measure the level of coverage but also to adjust the output data for under-coverage. Countries that are looking to transition to a register-based or combined census may wish to conduct a quality evaluation of the new model compared to the old one to ensure that a transition is viable (see the England and Wales case study).

104. The remaining chapters describe processes and methods and the role that quality assurance plays at each stage.

VI. DATA SOURCES AND THEIR QUALITY

105. The quality of data used in the process of compiling census statistics strongly affects the quality of the statistical output products. Thus, the quality of data from administrative registers and other administrative sources is a key element that should be considered in the decision-making process on the use of administrative data in the production of statistics. Therefore, it is necessary to prepare and implement a standard method of assessing the quality of administrative data as potential sources for the census. The degree of effective integration of data from different administrative sources is an important factor in making such an assessment.

106. The quality of administrative data is usually difficult to measure due to their complexity and multi-dimensionality, and indeed, many factors affecting quality are non-measurable. The methodology of assessing such quality is based on numerous aspects, criteria and indicators. It may take the form of a checklist or a survey (in the event that there is no other standard to test against) and should encompass information about the general characteristics of the register (such as the level of national coverage, frequency of updating and means of access, as well as information about the variables held (such as definitions and identifiers). Key aspects of the quality assessment of the register include: availability, clarity, usefulness, relevance and consistency, and the cost of accessing the register. The key aspects of the assessment of the quality of the data held in the register include: accuracy and comparability.

107. New statistical methods based increasingly on the use of administrative registers and other administrative data sources also require new methods of quality assurance. The quality of

data from such sources is another key element to be considered in the decision-making process to move to the use of administrative data for statistical purposes.

108. Firstly, the assessment of the quality of administrative data may be used to evaluate the usability of the data source. To this end a set of indicators should be developed by which each source may be analysed. On the basis of these indicators the NSI can decide whether or not to use a specific source. Secondly, the quality of administrative data has an effect on the quality of census outputs. Indicators on the data sources may be integrated in a quality framework that assesses the input (the administrative sources), the process quality and the product quality of a register-based census. The ESS.VIP ADMIN[16] project aims to provide appropriate methods for quality assurance and facilitate access.

109. Such information about the administrative data should provide answers to the following questions:

- How are data compiled and for what reason?
- Is there a legal obligation?
- What is the target population?
- Are data regularly updated?
- Are there plausibility checks?
- How are variables defined and are the definitions comparable with statistical concepts?

110. In a first step, information from the relevant administrative authority is required. This can be achieved by studying handbooks, forms and supporting documents. It is also recommended that persons at the administrative authority who are responsible for maintaining the source should be consulted. A standard checklist or questionnaire should be used for measuring pre-defined dimensions of the quality of the sources and the metadata.

111. In a second step, the microdata from the source (total number of records, missing values, values out of range, duplicates, number of records without a key) should be analysed. At this stage, linking records from the source with survey data at the unit level (if such linkage is allowed and possible) may answer additional questions, for example on the comparability of variables or on the timeliness of administrative information.

112. Thus, three kinds of quality can be studied further:

- Quality of sources relating to:

  The supplier: confidence and reliability of the data holder (for example with regard to the punctuality of data delivery), effectiveness of contact and communication with the supplier (such as whether it is good/bad, periodic/sporadic);
  Relevance: administrative purpose of the attribute in question (such as, is there only a legal basis or an intrinsic self-interest in recording and maintaining the data?);
  Privacy and security: the mode of data transmission, level of data decryption;
  Delivery: supply agreement, legal obligation, interval of periodic data delivery

[16] ESS.VIP ADMIN (European Statistical System Vision Implementation project on Administrative data sources) is a project running between 2015 and 2019. The project aims to facilitate the use of administrative sources across the ESS, by improving the access to administrative sources, by improving methodological knowledge needed for integrating administrative data in statistical production and by providing tools for assessing the quality of outputs based on administrative sources. It will support Member States in implementing these theoretical outcomes in concrete statistical areas; for more information, see https://ec.europa.eu/eurostat/cros/content/essvip-admin-administrative-data-sources_en.

(monthly/quarterly/annually/sporadically), date of delivery, costs;
Procedures: data collection and maintenance.

- Quality of the metadata relating to:

  Clarity: clear structure and content of data and metadata, well organised and precise metadata, documentation of changes over time;
  Data treatment: description of data management, consistency checks.

- Quality of data covering:

  Technical checks: technical input checks, data format checks (for example, readability of the data file), compliance of the data to the metadata definitions;
  Accuracy: degree of precision, degree of certainty for faithful data records (such as the presence of implausible values and ineligible records;
  Completeness: definition of register population (indicating the level of under- or over-coverage), missing values for variables;
  Time-related dimension: updating process carried out by the supplier, changes in concepts, definitions and coverage, whether or not there is a cut-off date for the purposes of continuity and coherence of historical series;
  Clarity: clear structure and content of data;
  Integrability: extent to which the data source is capable of undergoing integration or of being integrated into the statistical system (at the unit level and for each variable);
  Comparability: data definitions compatible with those of the NSI, aggregation level sufficient for the statistical purpose;
  Unique keys: existence of a unique identifier at the micro level, and, if there are multiple sources, the ability to interlink should be facilitated (through a common identifier).

113. A quality framework can typically be used to study different potential sources and to decide which sources are viable for use in the census. Tools to enable a systematic evaluation of the quality of administrative data sources have been developed, for example in the Netherlands and in Austria[17] as summarised in Box 3.

Box 3: Quality framework in the Netherlands and Austria

The Netherlands

A quality framework can be used as a procedure to determine the quality of data sources in a systematic, objective, and standardized way. For this purpose, Statistics Netherlands has developed a quality framework that distinguishes three different views on quality, namely the Source view, the Metadata view, and the Data view. The Source view focuses on quality aspects essential for the delivery of the data source, whereas the Metadata view focuses on the metadata aspects of the data source. In the Data view, technical- and accuracy-related aspects of data quality are studied. The quality framework developed in the Netherlands has been used in many different projects including its 2011 Census.

Austria

The Austrian framework for assessing the quality of administrative data was developed for the

[17] See e.g. http://www.cbs.nl/nl-NL/menu/methoden/onderzoek-methoden/discussionpapers/archief/2009/2009-42-x10-pub.htm and http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_16_Austria.pdf

2011 register-based census. In every stage of the data processing a quality-indicator was derived for each attribute. Even though the framework was developed around the register-based census, it was designed for general applicability. Due to the modular design, every step of the framework could be applied individually. The approach for the assessment of administrative data was inspired by work from other NSIs (Daas, Ossen, Vis-Visschers, & Arends-Tóth, 2009) and relied on four quality-related hyperdimensions (Documentation, Pre-processing, External Sources and Imputations) which aim to measure the quality at three production stages (raw administrative data; the combined dataset, that is the integration of registers; and the final dataset, that is after data editing and imputations).

At the **raw data level**, three hyperdimensions were studied: documentation, pre-processing and external sources. **Documentation** describes quality-related processes as well as the documentation of the data (metadata) at the administrative authorities. The degree of confidence and reliability of the data source keeper was monitored by the use of a questionnaire containing several open and scored questions. The open questions collected information of general interest, such as the timeliness of data delivery. Scored questions measured for example:

- "Data History" (such as whether or not changes over time are stored and information relating to Census Day are available);
- "Definitions" (such as whether or not data definitions are comparable to those of the NSI?);
- "Administrative Purpose" (such as whether or not the topic is relevant for the data source keeper and there is a legal basis for the topic in the administrative data source);
- "Data Treatment" (such as how quickly changes of a topic are recorded in the administrative data source, whether or not data are verified by the data source keeper, for example by requesting documents or identity cards, and whether or not technical checks and consistency checks between attributes are carried out by the data source keeper).

Pre-processing refers to the proportion of data records that cannot be used, such as those without a unique identifier, or with no-information (item non-response), or where values are out of range. **External Source** compares the administrative data source with another source, for example the Labour Force Survey, by matching individual records and computing the share of consistent observations per variable and administrative data source. By combining the three hyperdimensions, a quality indicator for each variable for each administrative data source was calculated (for example for citizenship from the Central Population Register, for citizenship from the Unemployment Register, etc.).

114. The European Statistical System is undertaking a comprehensive project on improving the use of administrative data sources. One of the work areas of the ESS.VIP ADMIN aims to develop and promote quality measures for evaluating the quality of administrative data and of the statistical outputs that use a combination of sources, among which are administrative sources. Those tasks are covered by the ESSnet on quality of multi-source statistics[18], where eight European NSIs are involved. The broad objectives are: to gather existing knowledge on quality assessment and reporting and review it critically; to provide up-to-date guidelines on quality assessment for the purposes of statistical production (input, output and frames for social statistics); to develop indicators for measuring the quality of output based on multiple sources and a methodology for reporting on the quality of such output; and to produce recommendations for updating the ESS Standard and the ESS Handbook for Quality Reports.

[18] The on-going work and the progressive deliverables of the ESSnet can be found on the CROS portal: https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en.

In particular, the project has reviewed current practices on assessing the quality of administrative sources, has tested some approaches, and has produced a recommended checklist for input quality (quality of administrative data). This is, of course, relevant for the census if administrative sources are candidates for its input. Further work in the ADMIN project will include the production of quality guidelines when administrative sources are integrated in the statistical production. Particular examples concerning the decision to use some administrative source in the census could be discussed. The work is planned to finish in 2019.

VII. LINKAGE AND TRANSFORMATION

115. There is likely to be a significant difference in the quality of the linkage between different administrative data sources with and without unique keys. If there is a unique identifier in all or most of the records, the linkage becomes relatively easy and the level of successful linkage is normally high (although the quality of the variables in the statistical register should still be measured). However, if such a unique identifier is absent and cannot be constructed, the quality of the linkage should be measured, and the impact on the resulting outputs should be assessed[19].

116. As an example, due to the lack of unique identifiers across administrative data sources, the UK's Office for National Statistics (ONS) has spent a number of years in developing and refining methods to overcome this challenge and progressing its work to explore the potential for moving to an Administrative Data Census. The method currently being used can be described in two stages:

- Deterministic method using match-keys to link records across the administrative sources. Match-keys are created by combining key identifying variables (or parts of them) such as name, sex, date of birth and postcode. The same set of match-keys is produced for each dataset. If the match-keys are the same on each source, a link is made.
- Probabilistic method. This approach identifies links between records in two datasets by comparing and quantifying the relative similarity of records (for example, giving a similarity score). The main difference from the deterministic matching stage is that probabilistic matching does not require record values to be identical between the two records ('fuzzy matching')[20].

117. At the time of the preparation of these Guidelines, further work had been carried out to refine these methods, including the development of a statistical spine to help resolve multiple matches across three or more datasets[21]. The links to reports provided in this, and the previous, paragraph also describe what, by 2017, had been done to quality assure these developing linkage methods. The methods showed promise, but further refinements were needed to obtain the high-quality linkage that is required to produce robust estimates of the population.

118. Process quality is not straightforward in its definition. Its elements encompass:

- **Best methods** comprise sound methodology (including adequate tools, procedures and expertise) and appropriate statistical procedures implemented from data collection to data validation.
- **Cost effectiveness**: resources are used effectively.

[19] For more detailed methodological information see UNECE-HLG MOS Data Integration Guide (https://statswiki.unece.org/display/DI/Guide+to+Data+Integration+for+Official+Statistics).
[20] More detail about these methods can be found in http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-matching-anonymous-data--m9-.pdf.
[21] More detail about these methods can be found in https://www.ons.gov.uk/census/censustransformation programme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20.

- **Low response burden**: Whenever possible, administrative data sources should be used (the response burden will then be zero). However, if collecting some data from respondents is necessary, the resulting burden should be proportionate to the needs of the users and should not be excessive for the respondents. The NSI monitors the response burden and sets targets for its reduction over time.

VIII. STATISTICAL REGISTERS

119. Using integrated data from various sources in a population and housing census is becoming the increasingly adopted approach. In countries with administrative registers using unique identifiers for unit records, it is possible to distinguish basic registers and supplementary registers. As noted and defined in section V.3, the key registers enabling a statistical description of units in the census (persons, households and dwellings) include: population registers, housing registers (or dwelling or address registers), business registers and activity registers. Linking information from the available data at the unit record level makes it possible to determine those characteristics of the population. However, for the purposes of the census it may be necessary to integrate data not only from administrative registers but also from ongoing research referring to a specific population and its features (such as a longitudinal study of a census-based sample), results from the previous censuses, and information accumulated in a statistical sampling frame. The integrated dataset from these sources assumes the form of a statistical register (a database that can be used in the further process of collecting and compiling data) for the purposes of a census.

120. As noted at section V.1, the ONS and the CSO Poland have produced a Statistical Population Dataset (SPD) by linking administrative data sources. The ONS, in particular, linked four administrative data sources:

- patients registered with a general practitioner;
- people who have a National Insurance Number;
- students who are registered on a Higher Education course;
- pupils attending state schools.

Once these data sources have been linked, a series of rules are applied to produce the SPD to make decisions about which records refer to the usually resident population. Additional 'activity' (or 'signs of life') data are used to make decisions about the usually resident location of a record, where conflicts are seen between the different administrative data sources[22]. The SPD is then used to produce estimates about the size of the population. To understand the quality of these population estimates based on administrative data, comparisons are made with census and official population estimates. These comparisons show promising results[23].

[22] More details about the ONS methods for constructing a SPD can be found in https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20.

[23] See https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation.

IX. OUTPUT QUALITY

121. Census outputs are produced for a wide range of users and, as discussed in Chapter IV, both stakeholder and more general public approval of such outputs is necessary for a successful transition. The use of administrative data can provide the opportunity to increase the range and periodicity of statistical outputs. It can also facilitate longitudinal studies. However, as a result, there may be some changes to both definitions of variables and output classifications. The impact of these changes on the statistical outputs should be investigated and explained to users.

122. Regardless of the data collection methodology, assessing the quality of the output of census data has always been an important and necessary task. There are several different ways and methods to assess the quality of statistics, including the quality of census output. Assessing the quality of a census that makes use of a new methodology is especially important, as it provides relevant information on the reliability of the new census results, and how the quality may differ from the results of previous censuses. Some of these different approaches of assessing quality are described in this chapter. It is always useful if more than one method is adopted.

123. The quality of the output of any census can be determined by the assessment of product quality, by coverage studies, in quality reports, and through quality committees. In considering quality and confidentiality aspects, the question is how the quality of the output can be measured. For example, where sample surveys are used, minimal cell frequencies should be defined to determine whether or not the produced estimates are accurate enough to publish. On the other hand, where a country only relies on registers or a complete field enumeration, confidentiality rules should be adopted to prevent disclosure of personal information. In the census context, quality reports (in which results can be compared with other output, such as from a Labour Force Survey for example) one typically finds that differences regarding demographic variables are relatively small while differences in economic variables (such as current activity status) can be relatively large. It should be noted, however, that as censuses are more commonly carried out on a ten-year cycle, assessments of census quality as published in quality reports may sometimes become out of date by the time of the next census, and may then not be specifically relevant to any new census methodology.

124. As noted in the previous chapter, the ONS has compared their population estimates based on administrative data with census and official population estimates, and has used a set of quality standards to better understand their quality[24]. In 2017, ONS produced a method for independently assessing the quality of the Statistical Population Datasets during the research phase of this work (for example, a method for creating a confidence interval around the SPD estimates). In the long-term, ONS plans to run an annual Population Coverage Survey which would be used to measure and adjust for coverage errors in the SPD (in a similar way to its Census Coverage Survey)[25]. This would be supported by other quality assurance processes, for example demographic analysis and comparisons with other data sources.

[24] See Appendix A in http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-options-report-2--o2-.pdf.
[25] More information about this work is described in http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011--producing-population-estimates-using-administrative-data--in-theory--m9-.pdf.

IX.1. Product quality

125. The quality in official statistics can be described on the basis of product quality and process quality as previously described. Based on these two kinds of measures, a wide range of quality indicators has been developed at the level of the European Statistical System (ESS) and the United Nations Statistical Commission, and some of those indicators are required in quality reporting in the ESS. The following paragraphs describe product quality in this context.

126. Product quality comprises:

- **Relevance** refers to the degree to which statistics meet current and potential needs of the users. It refers to whether or not all the statistics that are needed are in fact produced, and the extent to which concepts used (including definitions and classifications) reflect user needs. The three more common approaches to assessing relevance are:
  - to ask users directly in either a user needs survey (before the census) or a user satisfaction survey (after the census);
  - to install stakeholders and client feedback mechanisms;
  - by analysing the uses made of the data in a census use study.

  The level of compliance with international recommendation can provide some measures of relevance at the global level.

- **Accuracy** refers to the closeness of estimates to the unknown 'true' values. The major components are:
  - assessment of coverage;
  - unit and item non-response and the methods used for dealing with incorrect values (edits) and missing values (imputations);
  - sampling errors.

  Usually, handbooks on quality assessment provide a set of indicators for measuring accuracy.

- **Timeliness** refers to the period between the availability of the information and the event or phenomenon it describes. Common indicators are the time lags between the end of the census reference period on the one hand and the dates of first release and final results on the other hand. There is usually a trade-off between accuracy and timeliness where a preference for results as early as possible is met at the expense of accuracy. However, is it often the case that different users will have different views on the balance between the two.

- **Punctuality** refers to the delay between the date of the release of the results and the target date (the date by which the data should have been delivered according to some official release calendar, or laid down by legislation, or previously agreed with users).

- **Accessibility** and **clarity** refer to the conditions and modalities by which users can obtain, use and interpret data (interpretability) determined by:
  - the range of publications (or tabulations) that are made available;
  - the level of access to databases;
  - whether or not appropriate metadata including information on the quality of data is available.

- **Comparability** refers to the degree to which statistics are comparable between geographic areas and over time (between censuses).

- **Coherence** refers to the degree to which the census data are reliably combined in different ways and for various uses with statistical information from other sources within a broader framework. When originating from different sources, and in particular from statistical surveys using different methodologies, statistics are often not completely identical, but show differences in results due to different definitions, classifications and methodological standards. There are several areas where the assessment of coherence is regularly conducted:
  - between provisional and final statistics;
  - between annual and short-term statistics;
  - between statistics from the same socio-economic domain;
  - between survey statistics and national accounts.

127. The Austrian framework, referred to in Box 3 above, describes how the quality of all attributes of interest for their register-based census is computed (see Box 4).

Box 4 Austrian quality framework

In the processing stage, the entire information from the registers is combined to the **central database** which covers all attributes of interest for the register-based census. At this level, a quality indicator for each attribute across all administrative data sources is computed. If a variable is only derived from one administrative data source, then the quality of this attribute on raw data level is the same as in the central database. If several administrative data sources are combined in order to derive a variable (for example current activity status) or to establish the most plausible value (for example, where there is conflicting information on date of birth in two or more administrative data sources), then the quality indicator is calculated. This is done by using the Dempster-Shafer theory in order to combine quality indicators from different data sources. In addition, a comparison with an external source (for example, the Labour Force Survey) is carried out. In the last step of the data processing, missing values in the central database are imputed. For the assessment of the data quality in the **final dataset**, the quality indicator for **Imputation** is computed.

128. The quality of any census is judged by the quality of its output. The common demands of census output are fixed in quality guidelines that are common for all types of censuses. Three levels of quality should be considered:

- quality of single census variables, measured by all quality criteria;
- quality of hypercubes (high-dimensional census tables with typically four or more dimensions) assessed using the quality measures of the census variables;
- quality of the census population (in terms of coverage).

These are each discussed in the following paragraphs and section IX.2.

IX.1.1. Quality of a single census variable

129. The quality of a single census variable depends on

- the coverage of the administrative data variable(s) used for deriving the census variable;
- the accuracy of the administrative data variable(s) used for deriving the census variable;

- the adequacy of the software used for deriving the census variable;
- technical or human errors.

130. If contextual checks of the variable have been made through comparisons with other sources, then the quality characteristics to be measured are the:

- number of missing values;
- number of errors or outliers.

Usually, as the outliers can be considered in a way similar to missing values, it is common to concentrate mainly on missing values. In defining quality standards, it is reasonable to define the maximal ratio of missing values allowed for excellent, good and satisfactory quality of a census variable. These standards are purely subjective, as no indication is given of the respective level of each quality dimension.

IX.1.2. Quality of a census hypercube

131. The quality of a census hypercube depends on

- the quality of all relevant census variables;
- the adequacy of the cube-generating software.

132. A common rule is that if the hypercube includes k variables and these variables have correspondingly m_1, m_2, …, m_k missing values, then the number of missing values M in the k-dimensional hypercube is determined by the following inequalities:

$$\max(m_1, m_2,\ldots, m_k) \leq M \leq m_1 + m_2 + \ldots + m_k.$$

That means, the number of missing values in a hypercube is, in general, larger than for any single variable. And also, the higher the dimension of a hypercube is, in general, the more missing values it contains. Reasonable quality measures for hypercubes should be defined. One possibility is to use the following measure:

$$M = \{\sum m(i)\}/k,$$

Where m(i), for i=1, 2, …, k, are the quality marks of the census variables included in the hypercube with k dimensions, and M is the quality mark for the hypercube. It is also advisable to add some additional condition: the quality mark of a hypercube cannot have a high score if at least one of the m(i) has a very low mark.

133. As noted above, in determining quality standards it is reasonable to define the maximal ratio of missing values allowed for rating the quality of a census hypercube as excellent, good or satisfactory. The levels of quality of single variables and the hypercubes itself must be consistent, that is, if all variables have excellent quality, then the hypercube also has excellent quality. Conversely, a hypercube having some marginals of poor quality cannot have a good or excellent quality rating.

134. Any quality report of the census should also describe the broad methodology of Statistical Disclosure Control (SDC) and data protection transformations applied by the NSI to protect confidentiality. The resulting consequences of such applications on data quality (such as the level of loss of accuracy) should be reported. However, countries should be wary of providing too much detail of the SDC methodology as this in itself would pose a risk to disclosure.

IX.2. Coverage

135. Achieving full coverage of the population is one of the major challenges in carrying out any census. In the case of a traditional census the most common problem is under-coverage, and particularly so nowadays when people are generally more mobile and where they may have more than one residential address making them difficult to enumerate. Another problem is that some people value their privacy very highly and may prefer not to provide their data through enumerators. Also, there might be particular reasons for some people (in the case of illegal or unregistered migrants for example) to want to hide from public authorities. Keeping a low official profile may also be the reason why some people do not have records in administrative registers, resulting in a similar under-coverage in register-based and combined censuses.

136. In the case of a register-based census there is also the issue of over-coverage of registers to consider. If, for example, people have not officially declared their emigration, their records may be kept unchanged in administrative registers, and consequently be the cause of over-coverage in the census. If registers are not of sufficient enough quality in this respect, a possible means of avoiding coverage errors is to create a residency index on the basis of the records held in multiple registers in order to determine a so-called 'signs of life' score. The approach is to define for all possible residents the 'sign of life' as a binary score (with a value 0 or 1) for each record in each register. Using these signs of life as explanatory variables, it is possible to build a model forecasting the size of under- and over-coverage[26]. The score should be recalculated annually to define the population size more accurately.

137. There are several options to check the coverage errors in the case of a combined census. One is to use the methodology commonly adopted for traditional censuses (for the field enumeration component), by organising a post-enumeration survey and estimating both over- and under-coverage statistically. The 'signs of life' approach, comparing records from different registers, can then be adopted to provide information on the level of under- and over-coverage of the register-based component of the census.

138. As previously noted, ONS plans to run an annual Population Coverage Survey which would be used to measure and adjust for coverage errors on the SPD (in a similar way to its traditional Census Coverage Survey). At the time of the preparation of these Guidelines (in 2017), ONS began to test such a survey, with the aim of running a full-scale pilot in 2020[27]. This would then be supported by other quality assurance processes, for example demographic analysis and comparisons with other data sources.

139. The Duel System Estimation (DSE) is used by the Central Statistical Office in Ireland to research coverage issues in population estimates based on administration data. Population size is estimated by applying capture-recapture methods to two independent population lists. The first list is provided by an annual Sign Of Life (SOL) register, based on eleven administrative data sources and vital statistics, that aims to capture activity across all ages. Further information

[26] See e.g. Tiit, E.-M. (2012) Estimated undercoverage of census 2011. Quarterly Bulletin of Statistics Estonia, 4, 12, 110-119 and Tiit, E.-M., Meres, K., Vähi, M. (2012) Estimation of census population of census 2011. Quarterly Bulletin of Statistics Estonia, 3, 12, 79-108.

[27] More information about this work is described in http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011--producing-population-estimates-using-administrative-data--in-theory--m9-.pdf.

on this method can be found in the presentation 'Compiling population estimates from administrative data' given at the 5th Administrative Data Seminar in 2016[28].

140. Research into this method includes:

- Investigating the make-up of the SOL register;
  - inclusion/exclusion of different data sources;
  - inclusion/exclusion of individuals based on level of activity, for example, to exclude if number of weeks worked is less than 20 and the individual does not show up in the other data sources;
- Investigating potential sources of the second list;
  - a second register-based list (driver licence renewal for adults);
  - a second survey-based list (a quarterly household survey likely to continue into the future).

141. The Estonian assessment of the census quality can, as an example, be found in Box 5.

Box 5 Assessment of the census quality using administrative databases in Estonia

Problem

The main characteristic of the census quality is **coverage**, that is the ratio c/n of census population size c and factual population size n. If the ratio is bigger than 1, then the census has net over-coverage, which means that some people have been enumerated who do not belong to the factual population and their number exceeds the number of people that have not been enumerated. Here the reason might be that some people have been enumerated more than once. When the ratio c/n is less than 1, then the census has net under-coverage, which means that part of the factual population has not been enumerated and its size exceeds the number of those who have been erroneously enumerated.

Nowadays, **under-coverage** arises from several reasons: people are very mobile and it may happen that some of them are not at home at the census moment, while some people value their privacy so highly, that they do not want to disclose their personal information and avoid the census.

If all people living in the country have identification codes (ID-numbers) then the **over-coverage** caused by duplicates can easily be avoided.

Traditional methodology

The assessment of the census coverage has been for a long time a serious task for statisticians. One possible solution is **comparison of the census population with the current population statistics**. This method is convenient in times when the population is not effected by any big changes in natural or migration increase, but requires a high level of compliance with civil registration. This method has been used in Estonia several times and very good results were achieved in 1934 and in 1970.

Another method relies on **post-enumeration**. When using this methodology, an independent sample survey is carried out very shortly after the census asking some of the same census questions. Now the number of people $n(1)$ – enumerated in both the census and the survey,

[28] See http://www.cso.ie/en/csolatestnews/eventsconferencesseminars/administrativedataseminars/5thadministrativedataseminar/.

$n(2)$ – enumerated in the census but not in the survey, $n(3)$ – enumerated in the survey but not in the census and $n(4)$ – not enumerated in either the census or the survey, can be used to assess the share of people not enumerated by the census, although belonging to the population. In spite of the fact that this method is quite popular, it has several problems – it uses the assumption that missing in the census or in the survey is random. If there are several non-random reasons for any under-enumeration in either the census or the survey, then the method might not give a sufficiently accurate answer.

New methodology on the basis of existing administrative data sets

If a country has a set (system) of reliable registers or other administrative data sets, then it is possible to use them for checking the coverage errors of a census.

Testing under-coverage using signs of life

With this aim Statistics Estonia regards the set of all people who might live in the country (be residents). The following groups belong to this set (**super-population**): all people enumerated, all people who lived in the country in the last census (or earlier) and who have emigrated, and all people at the lists of some administrative registers of the country.

For all people from the super-population their **Signs Of Life** (SOL) are found in the following way. Suppose in the country there are K different registers. If a person j has been at least once active (has a record) in the register i during the last year, then he has a sign of life $E(i,j)$ of value 1; otherwise he has $E(i,j)$ =0. Signs of life defined in such a way are binary variables characterising all persons from the super-population.

The next step is to create **test-populations** consisting of **'confident residents'** and **'confident non-residents'**. The population of confident residents consists of people enumerated as Estonian residents who were also Estonian residents according to the Population Register. The confident non-residents were people not enumerated as Estonian residents and belonging to the Estonian Population Register as non-residents. For assessing the probability of being resident, use was made of test populations and a lognormal model was created with SOLs as explanatory variables. The model was used for all persons belonging to the super-population and not enumerated. Some of the people not enumerated had a high probability of being a resident (these were people having many SOLs demonstrating their activity in Estonia during the last year). Non-enumerated people, who had according to the model, a high probability of being residents were thought to be under-covered and were added to the census population in future population counts. The share of the estimated under-coverage was about 2 per cent and the assessed model error was less than 5 per cent of the set of people in question.

Similarly, it is possible to find in the census population those persons who do not belong to the factual population, and thus represent the over-coverage.

IX.3. Quality and confidentiality

142. Ensuring data quality while protecting the confidentiality of personal information are in some sense conflicting aims, but, nevertheless, both of them are essential aspects of census output. It is important to understand that it is difficult, if not impossible, to publish completely accurate outputs (especially for small areas) while, at the same time, adopting sufficient levels of disclosure control to ensure the protection of confidentiality. It should be noted that other

publications, in particular those based on the same administrative data as the census, have to be taken into account when adopting these sufficient levels.

IX.4. Comparison of census output with surveys

143. A commonly adopted method of assessing the accuracy of census data is by comparing the output results with data from national surveys (and registers in case other derivations have been made than in the census). This approach is particularly useful in the case of a register-based census.

144. Generally, the larger the sample size of a survey, the more reliable will be the result of the check. In this context the LFS is a good choice for assessing the census quality. However, there are also several other issues to be taken into account - for example, the definitions adopted in the two data sources and the reference dates may be different. Also, as no surveys can cover the whole range of census variables, only some of the census variables can be assessed in this way. And last, but not least, if there are any differences it will not always be possible to demonstrate which data are the more accurate.

145. In spite of all these difficulties, comparison of census results with survey results is valuable and should, wherever possible, always be attempted. The results should then be analysed thoroughly to explain the reasons for any differences discovered.

IX.5. Quality reports

146. A quality report is a document that contains the information on all steps of a quality check: the comparisons and analyses of the results, including also the explanations of any differences between census and survey data. If explanations exist as to why the output of a register-based census (or the part of the combined census where variables come from administrative data) differs from the results of surveys, then there might be a need for additional steps:

- the quality of the statistical register should be checked using alternative methods;
- the quality of the survey should be checked through a comparison with other surveys;
- the software (including algorithms) used for creating the census variables from administrative data should be checked;
- a check should be made as to whether or not the differences were caused by the applied confidentiality measures;
- the possibility of errors in all procedures should be checked.

The results of all such checks should be reflected in the quality report that could then be used as an input for the following census subject to the qualification noted at the beginning of Chapter IX that assessments of census quality as published in quality reports may sometimes become out of date by the time of the next census.

IX.6. Quality review panels

147. The quality check of a register-based or combined census should be made on several levels that might include the following:

- **Internal quality committees** formed by the NSI and consisting of census team members or other people who worked with the census team. Such a committee would check the quality of all variables and tabulations and might be responsible for the preparation of the quality report.

- **Audits based on quality guidelines** that might be carried out by NSI experts outside the census team.
- **External audits (to check privacy or peer reviews)** performed by independent experts from outside of the NSI, possibly from the academic community or even another country.

148. Finally, the ESS.VIP ADMIN project referred to in Chapter VI should be mentioned again. The goal of this project is to help statisticians to make wider and better use of administrative sources in the production of official statistics. It is done by addressing the most typical challenges faced in the use of these sources: limited access to data, the lack of quality of sources, methodological issues related to the processing of data and the integration of several sources. It also aims to ensure that the statistics produced using administrative data are comparable and are of sufficient quality by providing tools for assessing the quality of outputs based on such administrative sources.

X. APPROACHES AND CASE STUDIES FROM DIFFERENT COUNTRIES

149. A general overview of the approaches taken by countries to overcome some of the difficulties in moving to a register-based or combined census is given in this section. Some case studies that share specific country experiences are set out in Annexes A-I.

X.1. Technical approaches

150. Although a common framework was described in Chapter V, each country should develop its own path for implementation. It is beneficial for NSIs to share their experiences with other countries, but it is often not possible for a particular NSI to simply copy the practice of another country if it takes account, as it should, of the national context. Some of the technical approaches that have been used in different countries are summarised in this chapter.

151. As has previously been noted, a number of countries have already moved to a register-based census. Early adoption of this data collection methodology has only been facilitated by the availability of pre-existing population registers. Examples include Denmark and Finland. Other countries have made the transition more recently. Austria completed a register-based census with a mix of population registers and other administrative sources in 2011. Estonia and Poland both conducted a combined census in 2011 including on-line self-enumeration. Other countries, such as the United Kingdom and Canada, are working towards a combined census system in the longer term, but still have a mix of challenges to overcome.

152. Countries can put a safeguard system in place to monitor register quality, methodology and technological capacity during the transition phase. For example, England and Wales currently (at the time of the preparation of these Guidelines) has developed a traffic light system to indicate progress, and Estonia has the quality requirements for register holders.

153. Many aspects will be extremely dependent on external factors, such as the legal base and the approval of stakeholders, but the ability to transform multiple administrative data sources into a statistical register is often the main in-house challenge for NSIs.

154. Where statistical registers have already been used, the number of administrative sources and base registers involved has varied from country to country. Austria, for example, used eight base registers for its 2011 census but used other registers to assess quality. Estonia integrated 22 sources to develop its statistical registers, while Poland integrated as many as 28

administrative data sources and three non-governmental sources (out of some 300 potential data sources originally assessed) into a statistical register called the Master Record.

155. Integration of the administrative data sources means linking the information for people and dwellings to obtain comprehensive coverage of the population. A key difference between countries is whether or not unique identifiers are available which allow deterministic matching. The main challenge in some countries is the linking of data records from different registers without unique identifiers. In such cases, probability-based, or fuzzy, matching over several fields can be attempted.

156. Generally, when they are to be used for census purposes, administrative records go through an anonymization process to protect the identity of individuals where direct identifiers, such as a personal identification number (PIN), and indirect identifiers are removed from the register and replaced with a proxy. An example of a unique identifier is the Personal Public Service Number (PPSN) used in Ireland. There, the CSO uses a protected identifier key linked to the PPSN by an external third-party organisation. Austria, which does not have a unique person identifier, developed a way around this by using branch-specific personal identification numbers, bPINs, for the different administrative branches. These bPINs can be linked to the central population register. In this case the linking is done by the Austrian Data Protection Authority (see AnnexD). In addition to identifying individuals a ‘sign of life’ filter is applied to indicate whether the person is a usual resident or a resident at any particular time.

157. While small area statistics may be derived from knowing the geographic area where a person (or household) lives, in order to derive household variables, dwellings must be identified. In some countries, dwellings may be identified and positioned by unique building identifiers such as Eircode in Ireland and the Unique Property Reference Number (UPRN) in Great Britain (England, Wales and Scotland). Some other countries have established new building and dwelling registers. Without a unique building identifier linkage can be achieved via automated address matching. Specific commercial software is available. Some countries (such as Canada) develop their own in-house system. Address matching is a complex rules-based process that needs frequent up-dating. The address needs to be matched to addresses in other registers, then to persons, and then be geo-coded to a location. An ‘Address Standard’ was introduced in Estonia, for example, to facilitate such address matching. Rules should be put in place for resolving any conflicting or ambiguous information, such as a person linked to more than one address or for multiple-household dwellings.

158. Quality checks and standards, that will support the process after the traditional census is no longer available as a benchmark, should be developed and rigorously tested. Surveys to assess quality and imputations or modelling to produce robust outputs are further necessary steps that are typically involved in the transformation of administrative data to a statistical register.

X.2. Specific country experiences

159. To illustrate how transitions can take place and what kind of practical problems countries may have to overcome, some specific country experiences are presented in more detail in the Annexes A-I. These examples are based on the experiences of countries that have moved from a traditional census recently, or plan to do so in the future. It should be noted, however, that these case studies reflect the position that was current in each country at the time of the preparation of these Guidelines (mid 2017). Changes in methodology and processes may have since taken place. Nevertheless, other countries can benefit from these experiences.

- Case study Ireland (see Annex A)
- Case study Estonia (see Annex B)
- Case study Poland (see Annex C)
- Case study Austria (see Annex D)
- Case study Slovenia (see Annex E)
- Case study Portugal (see Annex F)
- Case study England and Wales (see Annex G)
- Case study Italy (see Annex H)
- Case study Germany (see Annex I)

ANNEX A

IRELAND CASE STUDY

This note is based on a paper *The Irish Statistical System and The Emerging Census Opportunity* (Dunne, 2015). presented at the New Techniques and Technologies for Statistics (NTTS) Conference held in Brussels in 2015 (Dunne, 2015).

Environment

1. The Central Statistics Office (CSO), Ireland has made significant progress with developing administrative data sources for statistical purposes (MacFeely and Dunne, 2014). This work is underpinned by developing a national data infrastructure (NDI) in Ireland. In summary, NDI is a conceptual framework promoting more efficient use of data across the Irish public sector. It promotes and advocates the use of common official identifiers for persons, business and property across all official systems (following the Nordic model). A central population register with up-to-date information on where everybody lives is not envisaged.

2. There exists a person identification number that is generally used across all person-based administrative systems (schools, welfare, employment, etc.) that is assigned at birth or when a person enters the country to live. This identification number is stored on a master file or administrative register. This register is maintained and updated by the Department of Social Protection (DSP – the government ministry that oversees social welfare administration); however, there is no requirement for all public-sector organisations to provide updates to the DSP.

3. Recently, a project has been undertaken to summarise each person's annual activity on key public administration systems with a simple yes/no indicator in a Person Activity Register (PAR). The PAR employs a Protected Identifier Key (PIK) to reduce privacy risk while preserving the linkage possibilities over time and across data sources. Key administrative data sources include births, children's benefit, education (early, primary, secondary, higher and further), employment, unemployment, occupational pensions and social welfare (including state pension). The purpose of the PAR is threefold:

- to enable longitudinal analysis of population cohorts across different administration systems;
- to explore population structures over time; and
- to provide a summary master key with respect to different administrative data sources in order to examine the feasibility of different potential projects.

4. The potential to provide for longitudinal analysis of specific population cohorts has proven to be of significant value in promoting the concept of "joined-up government needs joined-up data" across the Irish public sector. In the strict sense of the definition of the term, the PAR is not a register but is simply a statistical population dataset (SPD).

5. There are a number of official identification numbers in use for businesses on the main tax, employment and company registration systems. These numbers are linked and available on the CSO business register for exploiting data sources for statistical purposes. The tax authorities also have these numbers linked. Currently, it is planned to expand the use of official identification numbers for businesses across the public sector.

6. A linked employer/employee statistical file has been developed based on employer and employee tax returns. This file captures every paid employment in the State and facilitates the linkage of business- and person-based pillars of the NDI.

7. A new postcode system has recently been rolled out in Ireland. The postcode uniquely identifies each letter-box in the State. At present, the use of the postcode is not mandatory. It will likely take a number of years for this postcode system to become a unique reference for properties available on all official systems. In Ireland, over 30 per cent of address strings are not unique and are only distinguishable by using a person's name and the postman's local knowledge of who lives where to deliver the mail. The postcode system and its underlying register of properties is the basis of a sampling frame and master address frame for social surveys and the 2011 and 2016 Census operations.

8. The obvious gaps in data sources for conducting a census using administrative data relate to population coverage and where the population resides. In addition to continually striving for better quality data, these gaps are being addressed by developing transformation processes - the second phase of the framework - where administrative data is transformed into a statistical register by techniques such as modelling and linkage.

Census background

9. Ireland typically has high population movements and migration flows that necessitates conducting a census every five years (4.6 million persons living in 1.6 million houses in 2011). Ireland typically carries out a traditional *de facto* census in that approximately 4,000 enumerators hand deliver a census form to every household to be filled in by the head of household with respect to every person present on census night. The forms are then collected by the enumerators to be compiled and collated centrally by the CSO. The 2016 experience found that it is becoming harder to contact each household.

Future Census intentions

10. It is not known how feasible a census in Ireland that is based on administrative data sources and existing surveys would be. But if Ireland were to conduct such a census it might be undertaken along the following lines:

> Step 1: Create an up-to-date master address file that would also collate building characteristics from previous censuses and administrative data holdings to assist in the housing component of the census. There may also be a possibility to identify the occupancy status (vacant/occupied) of houses based on the records of utility companies.
>
> Step 2. Identify all persons interacting with public administration systems for the census year and summarise this on a register against a person identification number. This is the person activity register (PAR) referred to earlier. The PAR will contain a list of all persons that are active in a given year (referred to elsewhere in these Guidelines as the 'signs of life' approach). The key administrative data sources in the Irish context will include births, children's benefit payments, primary school database of enrolments, higher and further education sources of enrolments and awards, social welfare, employment, including self-employment data sources from tax authorities, persons registered on property rental leases, persons claiming medical benefits (or registering in a given year), social welfare and occupational/state pensions.

Step 3. Use statistical methods to correct for under-coverage and attach a correction factor to each record. The statistical methods may rely on existing sample surveys or another second independent source. One such method is currently being explored where driver licence renewals are used as a second independent source to identify under-coverage in the person activity register using capture-recapture methods.

Step 4. Allocate each identified person with an address/postcode from the master address file. Use a suitable decision-tree algorithm that is capable of incorporating situations where more than one address is identified with a person in administrative data sources. Household relationships identified in administrative data sources (such as children's benefit) may also feed into this algorithm.

Step 5. Form household relationships using relationships identified in administrative data sources and persons identified as living at the same house.

Step 6. Estimate and include attributes for each person on the PAR using existing surveys, administrative data sources and appropriate methodologies.

Step 7. All census outputs are now compiled from the PAR as updated by the above steps.

The Census 2021 will more than likely be used to test some aspects of conducting an administrative data based Census. Further information on methods used and applied to compile population estimates from administrative data sources are available in Zhang and Dunne, 2017. The book chapter also describes an extension of DSE methodology called Trimmed DSE that can be used to look for overcount or erroneous records in a DSE system.

References

Dunne, J (2015). *The Irish Statistical System and The Emerging Census Opportunity*. NTTS 2015. *Statistical Journal of the IAOS*, Vol 31, No. 3, pp. 391-400, 2015. This is currently available at https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji915

MacFeely, S and Dunne, J (2014). 'Joining Up Public Service Information: The Rationale for A National Data Infrastructure'. *Administration*, Vol. 61, No. 4 (2014), pp. 93–107. https://unstats.un.org/unsd/trade/events/2014/india/background/Forum_Vol_61_4.pdf

Zhang, LC and Dunne, J (2017) 'Trimmed Dual System Estimation' in Capture-Recapture Methods for the Social and Medical Sciences pp. 237 – 258, Chapman & Hall/CRC.

ANNEX B

ESTONIA CASE STUDY

The move from a traditional census

1. During the 2000s, Statistics Estonia worked very productively on the systematic development of registers, including identifying the data in all personal registers on the basis of personal identification codes and linking them with the Estonian X-Road system[29] to facilitate exchange of information. However, an initial analysis indicated that Estonia's registers were not at that time ready for successfully conducting the census in 2011. The main reasons were:

a) Comparison with data from the Labour Force Survey from 2007 up to 2015 indicated that at least 20 per cent of the addresses specified in the Population Register were not the actual places of residence of the people concerned;
b) the Education Information System only contains data on young people (general and higher education diplomas from 2000 onwards; vocational and other certificates from a later date);
c) no register contained information on the occupations of persons;
d) addresses were recorded differently in different registers, with a variable degree of specificity, making the data incompatible;
e) registers had been used only for a short period of time and their quality and adequacy had not been verified;
f) the consistency of definitions used by the different registers and information technological compatibility of registers had not been analysed.

2. Consequently, in 2009 a combined census methodology was approved for the Census 2011 comprising a number of elements with the aim of optimising the use of registers and the option for self-response, and making the census as paperless as possible:

a) *Coordination of data sources*. Previously created data sources (registers) were used together with Internet and face-to-face interviews. In the 2011 Census, registers were used in three ways: as a tool for preparing the census (preparation of work lists and census sheets), pre-filling of questionnaires, and supplementation of census results in the event of missing data. The information on educational studies of enumerated permanent residents was taken from EHIS (the Information System of Education) and the corresponding question was not included in the questionnaires.
b) *Combined survey methodology*. Unlike previous Estonian censuses, self-completed questionnaires were used in 2011 to complement interviews. This required the preparation of extensive training instructions for enumerators, and the provision of comprehensive guidance for the persons being enumerated.
c) *Combined data collection methodology.* All previous censuses in Estonia have been conducted using paper questionnaires, or census forms. The data of persons enumerated by the census – the answers to the census questions – were entered on these paper or cardboard sheets by enumerators, using a special machine-readable pencil. Two new technologies, that were demanding much training, were introduced in the PHC 2011: self-completion of questionnaires via the Internet and entry of answers directly onto laptop computers during census interviews by enumerators. However, the option of

[29] The X-road is a system that facilitates citizen's interactions with the State through the use of electronic solutions. Via X-road it is possible to get personal information from administrative registers (subject to data protection restrictions), but the system also enables the transfer of data from the registers to Statistics Estonia.

using paper questionnaires was kept as a back-up for emergency situations. The option of telephone interviews was also planned for particularly exceptional circumstances, especially in cases where access to households was extremely difficult (such as on small remote islands). In practice both these emergency modes of interviewing were used in less than 1 per cent of the cases.

3. In general, the combined census went well in Estonia. Two thirds (67 per cent) of the persons to be enumerated used self-enumeration via the Internet and the remaining third were interviewed face-to-face (Statistics Estonia, 2014).

Preparing for the register-based censuses in Estonia, 2011-2013

4. By 2010 – even before the combined census of 2011 - preparations had started for a register-based census in 2021. Here, the experience of those countries that had already conducted a register-based census had been taken into account, and a project (REGREL) was initiated to develop the transition to a register-based methodology, the first stage of which was extensive analysis (see paragraphs 6-8 below) which began in autumn 2010 and was completed in September 2013 (Puur et al, 2013).

5. The REGREL methodology project (of which 80 per cent was funded by the European Social Fund) was a partnership between Statistics Estonia, the Estonian Institute for Population Studies (at Tallinn University) and the consultancy firm AS Ernst & Young Baltic. The analysis was carried out by a few dozen scientists and experts from the University of Tartu and from Tallinn University, by lawyers and by analysts from Statistics Estonia. A very important role was played by the representatives of databases and registers, who took an active part in the process (Puur et al, 2013).

6. The analysis was carried out in two parts:

- meta-analysis of obligatory PHC characteristics; and
- detailed analysis of characteristics that required data quality analysis (as indicated by the meta-analysis).

7. In addition to these analyses the team also made other preparations for register-based censuses:

- legal analysis;
- preparation of methodological guidelines for the creation of a census glossary; and
- analysis of international experience and practice.

8. All in all, the project team analysed the data held in nearly 20 registers. One of the most significant outcomes of the methodology project was the network of main registers and databases for REGREL (containing data on the EU's mandatory characteristics as prescribed in EU Regulation no. 763/2008).

9. The results of the REGREL methodology project showed that there is still much to do to prepare for register-based censuses. Statistics Estonia will manage and coordinate these activities. However, much of the development work is being done by those stakeholders outside Statistics Estonia responsible for maintaining the registers to resolve the shortages, problems and bottlenecks in the Estonian register system.

10. At the same time that the REGREL project was being developed a further important part of the preparatory work involved developing the software solutions for transporting the necessary

data from registers, creating the census characteristics and saving them in the Statistics Estonia database. Also, the necessary proposals were made to make changes in laws with the aim of accessing the administrative information necessary for creating the census variables. A further major task was their preparation and realisation of algorithms for creating census variables from the administrative data.

11. Regular collaboration with the holders of the administrative registers was vital, and the quality of register's data was checked and suggestions made to improve the quality where necessary.

Census pilot in 2014

12. The main goal of the REGREL pilot census was testing of the production system with selected EU mandatory census characteristics, namely: place of usual residence; sex; age; legal marital status; country of birth; place of birth; citizenship; relationships between household members; and level of education. After the pilot census its results were analysed, and the emerging problems were considered in the plans of the following years.

Action plan during the period of 2016-2020

13. The main preparatory work for the register-based census during this period related to the following tasks:

- data acquisition from registers (contracts, description of the data set, checks on data quality and the acquisition procedure);
- formation of census characteristics, programming of the necessary rules;
- testing the statistical system as a whole;
- testing statistical registers system and filling it with data 2015-2018; and
- analysing the needs and expectations of potential users.

14. Statistics Estonia developed a set of legal and organisational measures to improve the quality, timeliness and coverage of the administrative registers necessary for the census. This set of measures was submitted to the relevant ministries. Currently, the registers do not hold data to derive all the required census characteristics for the entire population, and it is also unclear whether or not all register data are updated regularly enough.

2016 Census trial

15. The period of trial census activities was 2 January – 8 December 2016 and the trial census population consisted of:

- the entire usual resident population of Estonia; and
- all conventional dwellings regardless of occupancy, and occupied non-conventional dwellings located in Estonia.

The objective was to practice conducting the register-based population and housing census adopting Eurostat's recommendations and quality requirements as far as was possible. For this, data had to be acquired from 24 national databases, to be followed by processing and analysis.

16. On the basis of census characteristics derived from register-based data using the specially created algorithms, it is planned that a series of hypercubes will be created (part of the hypercubes demanded by Eurostat under the terms of the EU Regulation).

17. The subsequent quality check will consist of three parts:

- the quality of all census characteristics will be checked in terms of the five key quality criteria;
- the quality of all hypercubes (and at the same time marginal cubes) will be checked taking into account the quality of the characteristics included in the cubes; and
- a comparison of household and family structures when the two different household definitions are used: the housekeeping-based concept (used more commonly in the case of traditional censuses) and the dwelling-based concept (adopted in register-based censuses).

18. A quality report will be created where all quality measures and issues will be identified. Where the data from registers is subject to quality issues, the relevant data holders will be informed, and the possible solutions will be discussed.

19. During the next census trial, scheduled in 2019, the preparations for the register-based census will be focused on the datasets (improving the problems found in earlier stages) and data flow. Also, the effect of confidentiality measures on data quality will be analysed in more detail.

Summary

20. In Estonia preparations are being made for a register-based census in 2021 in the course of which data will be captured from various databases adopting measures for maintaining data protection and statistical security. The requirements being developed for databases will be sufficient for ensuring the interoperability of state information systems if:

- all the relevant data are submitted with metadata, including classification codes;
- capture and data updates take place via the X-Road system;
- data are presented in XML format and the description of data will be submitted by the creator of the X-Way service as XSD and updates include the time of presentation.

21. The primary data in the databases are required to meet the quality requirements in order to guarantee that census objectives are fulfilled. Quality will be indicated by the following:

(a) The coverage of the registers needs to be at least 97 per cent for the population and 95 per cent for single characteristics;
(b) Some 95 per cent of the data need to be linked with the classifications registered in RIHA (the Information System Authority of Estonia); and
(c) All residents and foreign citizens need to be assigned a unique identification code.

References

Puur A, Sakkeus L and Aben S (2013). *Development of a register-based population and housing census (REGREL) methodology. Project Final Report*. Ernst & Young Baltic AS, Estonian Institute of Demography, TU. Tallinn (Available only in Estonian).

Statistics Estonia (2014). *Annual Report 2013*. Downloadable at http://www.stat.ee/annual-report.

ANNEX C

POLAND CASE STUDY

Introduction to the Polish Population and Housing Census 2011

1. The National Census of Population and Housing (NSP 2011) conducted in Poland in 2011 by the Central Statistical Office (CSO) was designed and implemented with the application of a mixed model approach, that is using data from administrative registers (covering base demographic variables for the whole population) and data obtained directly from respondents through two surveys (one covering a 20 per cent national sample using a long-form, and another covering the other 80 per cent using a short form) collected with the use of electronic questionnaires. As a result, the use of paper questionnaires was eliminated altogether.

2. The legislation for the 2011 Census (the National Census Act 2011) stipulated that public administration systems should be used as widely as possible for the purposes of census, meaning that information from such administrative sources should be used to prepare and update an address and housing register (followed by preparation of an address and housing frame for samples to be used in the sample survey), as well as providing a source for the census data itself. Data not included in the public administration information system or where data were of insufficient quality to be used for the census were collected directly from respondents. This method was considered to be safer and more effective, taking into consideration the present level of development of administrative sources, their quality, and the degree of advancement of methodological work concerning the estimation and imputation of missing data.

The use of administrative sources

3. The necessity to use data from administrative systems in Polish statistics resulted from:
- economic reasons – in particular the demand for greater efficiency, the minimisation of the costs of the production of statistics, and reducing the burden on respondents;
- the risk of an increased non-response in statistical surveys, including the censuses; and
- an extensive development of public administration IT systems taking advantage of advanced technologies.

4. Implementing a census based on administrative and non-administrative data systems has brought numerous benefits, including:
- an effective use of administrative and non-administrative sources;
- reduced costs;
- reduced public burden;
- an improvement in data security;
- a guarantee of surveys harmonisation through the use of common identifiers;
- the prospect of providing future census data annually;
- the availability of data for any level of territorial disaggregation,
- an improved ability to identify double entry errors (over-counting);
- the creation of a micro-database supporting indirect estimation – modelling at the unit level;
- an improvement in estimation for small areas; and
- an improvement in the coherence and reliability of statistical data.

5. The decision to use data from administrative registers for the census required an in-depth review of the range of information that would be available from these sources. An analysis of all the sources and variables of potential use for the census was carried out. To facilitate this the metadata were obtained for approximately 300 administrative registers, of which the 30 most useful were selected. For each of these registers separate records were opened and all variables from these sources were subjected to a utility analysis. The variables were evaluated with regard to their conformity, in terms of definitions and classification, with the dictionaries existing in Polish and EU statistics. Appropriate weights were determined, both for the variables and administrative registers from which these variables came, taking into consideration their utility and quality. An assessment of the quality and utility of variables from different registers formed the basis for developing the rules for merging data, and for the estimation and imputation in the operational base of microdata created. The result of this work was invaluable knowledge of the utility of each register and potential for integrating those different registers that the statistical service had at its disposal.

6. In the event, the CSO used 28 sources from central and local government, and from outside public administration such as registers of building administrators, housing co-operatives, power distribution plants and telecommunication operators. The administrators of all these databases were approached with regard to the use of their information for the purposes of the census.

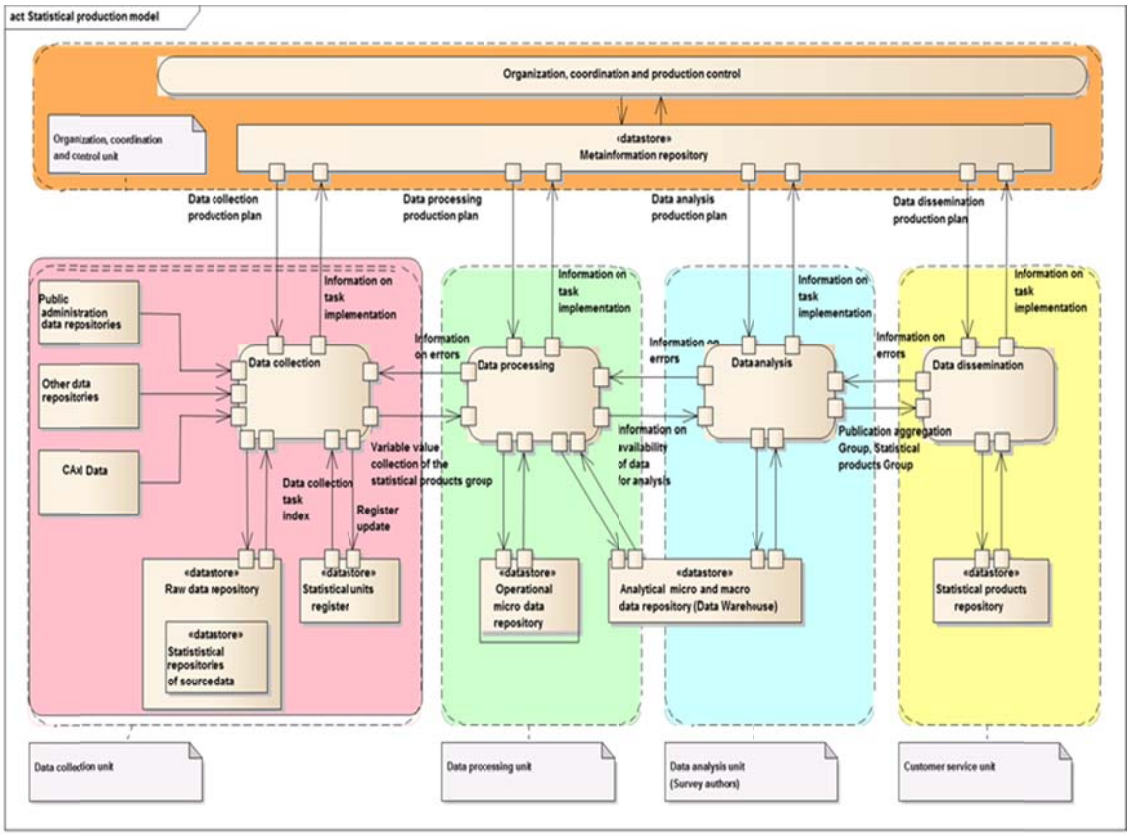
7. To enable the administrators to transfer data from their various systems (including, for example, from over 2,500 local government offices) via tele-transmission, the CSO constructed an electronic platform for data collection and processing, together with a net-based application for a direct data transfer via electronic means in a secure connection.

8. The unit record data obtained from registers were converted into statistical registers, being subject to the simultaneous process of cleaning, de-duplication and standardisation of data. The process was carried out in a DQS SAS environment. At the same time, metadata were collected on the quality of the input data obtained from registers, the applied cleaning procedures, and the final quality obtained after applying the DQS procedures. Data from the administrative sources converted to the statistical register were then used to derive the Master Record - the set of variables derived from the registers that was included on the census forms in order to be verified (confirmed or updated) by the respondents.

Quality evaluation of data from administrative sources

9. The quality of the public administration data systems was difficult to measure due to the complexity and multi-faceted issues, such as the absence of any possibility to use a single synthetic indicator. Hence, the assessment of quality was based on many indicators.

10. There were three elements of quality assessment: evaluation of the raw data sources; systems, data sets after a transformation (that is after adjusting them for use in the census), and statistical products; and resulting data. With regard to the sets from administrative data sources, quality assessment of raw data sets provided by the administrators and sets after a transformation, (that is, after adjusting them for use in the census) was carried out. Quality measurement was undertaken in all stages, in all processes of development of administrative data, and in the integration of data from different sources. The scheme below shows a graphical representation of the processes in the census, including quality assessment.



11. A 'meta-information repository' was created to collect methodological, technical and operational metadata. It enabled the monitoring of, and ensured process quality control of, all the census processes from data collection through to dissemination of outputs.

Other data collection methods used in the census

12. Poland was one of the first countries in the world that prepared a totally innovative method, consisting of using several of the most modern techniques simultaneously, for collecting census data. Apart from the use of public administration registers, three field data collection methods were adopted, referred to, collectively, as CAxI:

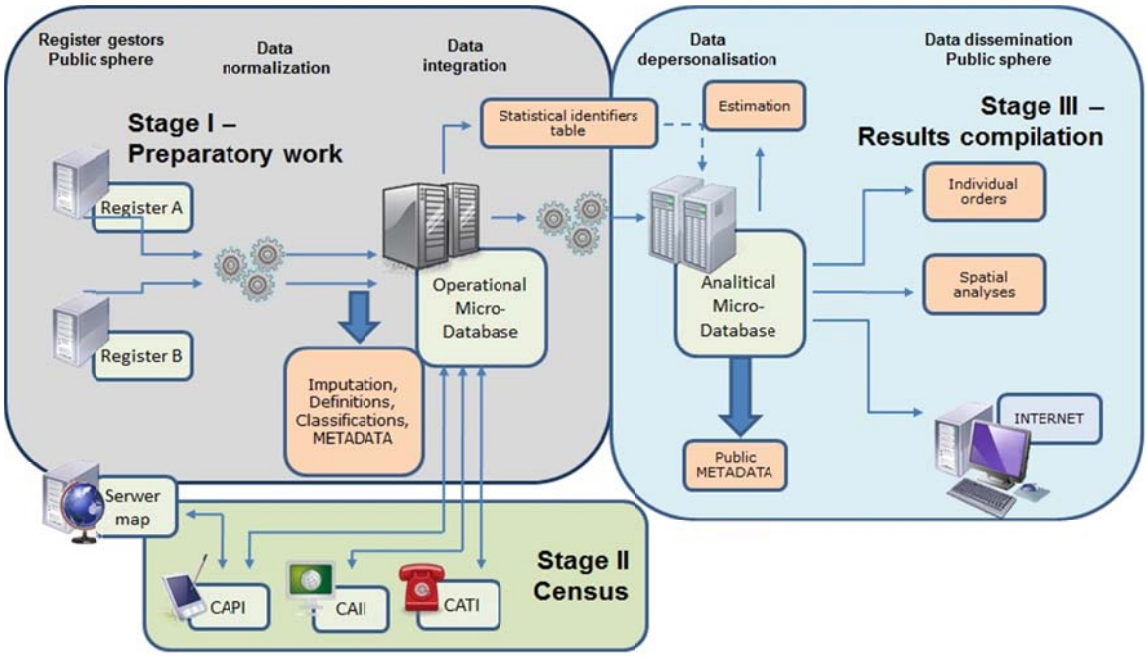
- CAII/CAWI (Computer Assisted Internet Interview/Computer Assisted Web Interview) – an online self-administered questionnaire, which verified the respondent data obtained from administrative sources, within a specified time frame, and, if need be, correcting the same and providing information missing from the registers.
- CATI (Computer Assisted Telephone Interview) – a computer assisted telephone interview, conducted by a statistical interviewer, to supplement data that were incomplete or missing in the sample survey.
- CAPI (Computer Assisted Personal Interview) – an interview conducted by a census enumerator in the field, and where the data were recorded on a hand-held device.

13. All three channels were based exclusively on an adaptive electronic questionnaire, ensuring high quality of data at the collection stage. The electronic questionnaire was designed specifically for implementation in each of the three CAxI data collection modes. An appropriate questionnaire application (available at a mobile terminal or Internet browser)

verified if the questionnaire had been filled in accurately, among other things, through logical and accounting control.

Census architecture: the IT Census System

14. To enable the optimal application of advanced IT and telecommunications technologies in the census, an appropriate census architecture had to be constructed (illustrated below) For the purposes of the census design and execution, the CSO implemented an IT Census System (ISS), consisting of more than 10 components - supplied by different contractors - that provided IT assistance for all operations within census. The ISS integrated various technologies (ranging from applications installed on mobile terminals, through those managing and assisting in telephone interviews, to specialist bases, data warehouses and analytical and reporting tools).



15. The ISS employed various solutions ensuring a high level of security for processed census data. Appropriate measures and controls were implemented, obliging all participants to observe statistical confidentiality and to guarantee the protection of personal data.

16. The CSO followed a number of stages as part of the work involved in the processing of census data, including data from administrative registers. These included:

- the preparation of normalisation, control, edit and imputation rules for datasets from the administrative sources;
- the preparation of rules for the synchronisation of data from administrative systems – the harmonisation of reference periods – tables of transition from the state in which data from administrative systems were acquired to the desired state;
- the preparation of rules for supplementing missing field data – imputation and calibration;
- the preparation of rules and methods for the precise and clear linkage of data from various administrative systems;
- the determination of the values of variables included in censuses (data-source rules);
- the preparation of the rules for calculating the values of census variables;

- the preparation of rules for creating derived objects – creating new objects (households, families);
- the preparation of a data-estimation model/method using data from administrative systems and statistical surveys; and
- the preparation of data anonymization rules.

17. Pursuant to the National Census Act of 2011, the Operational Microdata Base (OMB) was developed, prepared, and implemented at the Central Statistical Office. The system included hardware-system-tool infrastructure (computer hardware, system software, tool software) and application software (computer programs that are the result of programming work). This base enabled the inclusion of the census data transmitted in electronic form through four informational channels by those persons prescribed to do so under the terms of the Act – (that is, those responsible for maintaining administrative registers, enumerators, telephone interviewers or respondents), and facilitated further data processing (checking, editing, imputation and anonymization). In the next step, the anonymized data were transferred to the Analytical Microdata Base (AMB).

18. The meta-information repository sub-system collated indispensable metadata describing the data and census processes, including those necessary for preparing quality reports. The task of the meta-information sub-system was to ensure the coherent definition of statistical units for the OMB and AMB. The system was also used to store operational metadata of the OMB and AMB systems. This sub-system constitutes the Central Metadata Repository (CMR).

19. The function of the AMB is to store the anonymized census data in their final form. In this dataset every type of statistical analyses is carried out to produce the results required for publication - the census products. The AMB allows all the recipients of statistical information to readily access the data in the form of aggregates. The AMB system constitutes an analytical and reporting platform that currently enables the statistical preparation of the output data. The results of analyses in the form of reports, tabulations, maps and other graphical output are available to both internal and external users.

GIS Technology

20. For the first time in Poland, GIS (geographic information systems) have been used in both carrying out and reporting on the census.

21. The use of various reference materials and registers containing spatial information, has enabled the CSO to create spatial data for statistical address points and the areas of statistical divisions within the country. Digital maps were an indispensable reference source used by: census enumerators to enable them to navigate and verify dwelling locations in the field; gmina (local area equivalent to LAU2 level) leaders, for census monitoring within their area; and voivodship (regional area equivalent to NUTS2 level) and central supervisors, for census monitoring at the regional or national level. Maps were used to monitor the progress of the enumeration in a defined area or for a specific enumerator; an on-demand location or daily route could be visualised on the map.

The Geostatistics Portal

22. The Geostatistics Portal is a tool for interactive cartographic presentation and the publication of data collected in the census. It serves to store and present data, and to enable the sharing of information for a wide range of users.

23. The interface of the Geostatistics Portal allows its users quick and easy access to the published census statistical information. Data are presented using cartographic methods such as choropleth map and various diagram maps. It is also possible for users to define their own parameters for the visualisations of thematic phenomena for a given choropleth map. In addition to using ready-made spatial analyses, in the Geostatistics Portal, users can perform microdata querying by drawing a freehand polygon on the map and/or using sketching tools, which include linear / distance analyses and object buffering. Microdata queries can be performed on selected variables from the censuses

Summary

24. The 2011 Census in Poland turned out to be an innovative project not only nationally but also on a global basis on the grounds of the following criteria:

- For the first time in Europe, simultaneous data collection from four different channels (administrative registers, Internet self-enumeration, direct interviews conducted by census enumerators using electronic questionnaires on handheld devices, and telephone interviews conducted by statistical interviewers) was implemented on a national scale.
- Paper questionnaires were completely eliminated and were replaced by ICT solutions.
- Data from 28 administrative registers and three non-administrative sources were effectively integrated.
- The use of GIS technology was used in the preparatory work, to monitor the progress of the field enumeration, and offered the possibility to compile and present census results based on multi-dimensional spatial analyses.
- IT Census System comprised a number of solutions ensuring a high level of security and confidentiality of the processed data.
- Modern statistical data processing technologies were developed that will also have a considerable influence on the methodology of future statistical surveys.
- A comprehensive tele-information structure was established, considerably increasing the automation of statistical data processing.

25. A fully comprehensive review of the whole census operation has allowed the CSO to draw several conclusions, and the lessons learned provides the opportunity to assess the possibility for further improvement in future censuses. The new technology applied in the 2011 Census has proved that it can also be implemented in other questionnaire-based surveys. It is cheaper, employs up-to-date control mechanisms, enhancing the quality of the data collected, and reduces the burden of respondents.

26. The planned time frame was achieved, and none of the deadlines set for the data collection stage (specified in the Census Act) had to be extended, and neither was the census budget. The detailed schedule for the implementation of the census, comprising over 250 separate activities was regularly updated. The framework schedule, and the detailed schedules for the sub-tasks included therein (such as, the preparation and procedures for field control, and IT support for the census systems) were kept in separate files. The schedule comprised a total of several thousands of such tasks.

27. It should be acknowledged that the effectiveness of the census implementation was owed both to the methodological as well as the detailed organisational and logistic preparations.

28. There is another round of censuses ahead. Thanks to current experiences, in the 2020 round Poland is considering using even newer technology, with the aim of making the next census

even more effective and more innovative. During the inter-censal period, census implementation methods are being developed based on the experience gathered in 2011 as the starting point. However, considerable effort is still needed with a view to developing a new census strategy, so as to guarantee progressive solutions. Attempts will be made at further:

- reducing costs;
- using administrative sources in a more effective way;
- reducing public burden in data collection;
- improving the security and confidentiality of the transferred data; and
- improving the coherence and reliability of statistical outputs.

ANNEX D

AUSTRIA CASE STUDY

1. In the 2011 Population and Housing Census for the first time in Austria's census history information on persons, buildings and dwellings was derived exclusively from administrative and statistical registers. Ten years earlier, the census was conducted as a traditional enumeration, with questionnaires and enumerators. The transition from the traditional census to a register-based census was implemented in a relatively short time. This Annex describes the general conditions and the legal framework of the transition, and presents a brief description of the register-based census model.

The Government's decision

2. In 2000, the Austrian Government announced its plans for the move to a new census methodology after the 2001 Census. The questionnaire-based enumeration was being perceived as out of date; respondent burden was too high, results appeared too late, and costs were too high. Moreover, much of the information that was asked from respondents was already available in administrative registers. In the long run, register-based censuses were seen as a much cheaper alternative to the traditional approach.

3. In the late 1990s, Statistics Austria evaluated the possibilities for a replacement for the traditional census, of which the next was planned for the year 2001, by using data from administrative registers. However, due to the lack of a centralized population register (more than 2,300 separate municipal registers on population registration existed at this time), the lack of important basic registers (such as housing and educational attainment) and the lack of a unique identifier either for persons or addresses, the conclusion of this evaluation was that the quality of a register-based census in 2001 would be very poor. Linking personal data from different sources could only have been done by using the name and the characteristics from a significant number of topics that were, in any case, not available.

4. The Government followed Statistics Austria's recommendation to conduct a traditional census in 2001 and that a set of measures would have to be implemented in order to successfully replace the traditional census by a register-based census. Therefore, the Government's decision in 2000 was accompanied by the announcement that a central population register would be set up by the Ministry of the Interior, that legal and technical requirements for the anonymized linkage of administrative registers would be established, and that the quality of administrative registers would have to be improved.

Creating the Necessary Conditions

The Central Population Register (CPR)

5. In Austria, population registration is obligatory, but until 2002 each municipality had its own register. The central population register (CPR) became operative on 1 March 2002. The initial population stock was compiled from the 2,300 or more municipal population registers during the 2001 Census[30].

[30] As part of the 2001 Census, municipalities had to upload data from their population registration systems into a central data base (GSG **G**emeinde**S**oftware**G**roßzählung 2001) which was provided by Statistics Austria. The initial data for the Central Population Register were extracted from this data base. The central data base also supported the enumeration in the municipalities (delineating enumeration areas, assignment of enumeration areas to enumerators, reporting, etc.). It was pre-filled with addresses of buildings from the address register of Statistics Austria. Persons from the local population registers had to be assigned to these addresses, and the results of the enumeration with regard to the place of residence had to be entered.

6. The CPR contains variables such as sex, age, country of citizenship, place and country of birth, type of residence, address of the main residence and of other places of residence in Austria, country of former place of residence and country of destination in cases of emigration from Austria. In 2006, the variable 'legal marital status' was added as a new characteristic. The CPR does not, however, provide the information of family relationships that, for example, can be found in population registers in Nordic countries.

The Register of Educational Attainment, Register of Pupils and Students

7. One of the first activities in preparing the basic requirements for a register-based census in Austria was the implementation of a statistical register of educational attainment. Together with this new register, the basis for statistics on education in general was renewed (through the Educational Documentation Act of 2002). From 2003 onwards, schools have had to submit individual data on pupils and students once a year (reference date 30 September) providing information on current school enrolment (type of school, grade, etc.) and on performance at the end of the preceding school year including data on graduations.

8. According to the 2002 Act, the register of educational attainment was set up by using statistical information on highest educational level attained and other variables such as date of birth, sex and address code of the place of residence from the 2001 Population Census. The register is regularly updated with information on graduations from schools, universities and vocational training (apprenticeship). The unemployment register and those authorities responsible for the recognition of school or university diplomas from abroad provide further sources of information on highest educational attainment.

The Buildings and Dwellings Register (BDR)

9. The Buildings and Dwellings Register was set up subsequent to the Buildings and Dwellings Register Act of 2004; the register holder is Statistics Austria. It became operative in November 2004 and serves as an administrative register for municipalities who are required to update information if a building or dwelling is rebuilt or new buildings are under construction. The register contains information on the addresses and characteristic of buildings and dwellings and other housing units, and on construction activities (building permits, completion of buildings and dwellings).

10. The basic data of the register is made up of statistical information on buildings and dwellings from the 2001 Housing Census and from statistics of construction activities in order to fill the gap between 2001 and 2004. Initial data also came from the digital cadastral map and the land registry database of the Federal Office of Metrology and Surveying.

11. Addresses, buildings and dwellings have unique identification numbers that are also used in the central population register (CPR). The local authority cannot record persons in the CPR if the address is missing in the BDR.

The Branch-Specific Personal Identification Number for Official Statistics (bPIN OS)

12. Although the social security number is used in many administrative registers in Austria, its use to link data from different sources was not possible in the register-based census, mainly because of national data protection reasons. The social security number contains the date of birth of a person and does not guarantee anonymity.

13. In 2004, the eGovernment Act introduced the so-called branch-specific personal identification number (bPIN) for communication between public authorities within e-government. Each 'branch' such as for 'health', 'social security', 'taxes' or 'official statistics' uses its own PIN. The bPINs are derived from a source PIN that each person has in the central

population register. The authority responsible for this procedure is the "Stammzahlenregisterbehörde" of the Austrian Data Protection Authority (DPA).

14. In the register-based census, the use of the bPIN for Official Statistics (bPIN OS) is implemented in the following stages: Before data from a register are sent to Statistics Austria, the register owner submits identity data of the persons to the DPA. In a first step, these data are matched with the CPR data by using these identity data. If there is a match, the bPIN OS and the bPIN of the register owner can be derived from the source PIN by using a somewhat complex algorithm. Both bPINs are encrypted and their size is 172 digits. Back to the register holding authority, the data extraction for Statistics Austria is enriched by both encrypted bPINs and finally submitted. Only Statistics Austria has the key to decrypt the bPIN OS, which then has a size of 28 digits. The encrypted bPIN of the register holder serves to identify the respective record in case of inquiries by Statistics Austria. By using the decrypted bPIN OS, data from different administrative registers can be linked on individual level.

Census legislation

15. The existing census law was not a sufficient legal base for conducting a register-based census in 2011, and so a new census law had to be drafted based on the detailed concepts and requirements for such census provided by Statistics Austria in consultations with experts. This was ready in 2004 and was approved by an inter-ministerial working group in June 2005. The resulting register-based census legislation came into force in March 2006.

Main principles of the register-based census

Scope of the Census

16. The register-based census only comprises those core topics as suggested by the CES Recommendations and prescribed by the EU Census Regulation 763/2008. Topics for which information is available in registers but had not been part of the list of topics in previous censuses (such as 'income') are not included. Moreover, as some variables are not available in any register (such as language spoken or the means of transport) it is not possible to collect them solely using a register-based approach. However, it should be noted that the Census Act also regulates the census of local units of employment, which has been a part of the census in Austria since 1981.

17. Census day was fixed as 31 October 2011. A census day at the end/beginning of a year is not a good option for Austria, since in some municipalities the population numbers would be distorted because of seasonal employment in the winter tourism trade.

Linking of registers

Personal Identification Number

18. As noted in section II.4, register data are linked by using the bPIN OS. In practice, the assignment of a bPIN OS is sometimes not possible if identity data are inaccurate. Nevertheless, the register owner is required to submit these records to Statistics Austria without any PIN and without name. These data records are subject to statistical matching using characteristics such as date of birth, sex and postal code.

Address identifiers

19. The buildings and dwellings register is linked with the central population register by unique address codes, building and dwelling numbers. The same applies to the business register.

Redundancy

20. The various registers that are the basis of the census could contain different values for a characteristic of the same person. Therefore, it is not deemed reliable to trust data taken from just a single register. To ensure acceptable quality, the principle of redundancy is applied: information on sex, date of birth, citizenship, country of birth and legal marital status is collected from as many registers as possible. These are the so-called 'multiple attributes'. Each variable is assigned to a basic register and to 'comparison' registers which are used as a confirmation of the value in the basic register. If values are different, then an algorithm is applied to determine the best value of the variable.

21. Topics such as 'current activity status' have to be collected from different registers, in particular those holding information on economic activity and on school enrolment.

The Registers

22. The new Register-based Census Act defines eight base registers. A base register also has the function of a register which is used to assure the quality of a variable, if that variable is assigned to another base register. Consequently, these eight registers serve as both base registers and comparison registers, for different variables. The base registers (shown in Figure 1) and their owners are:

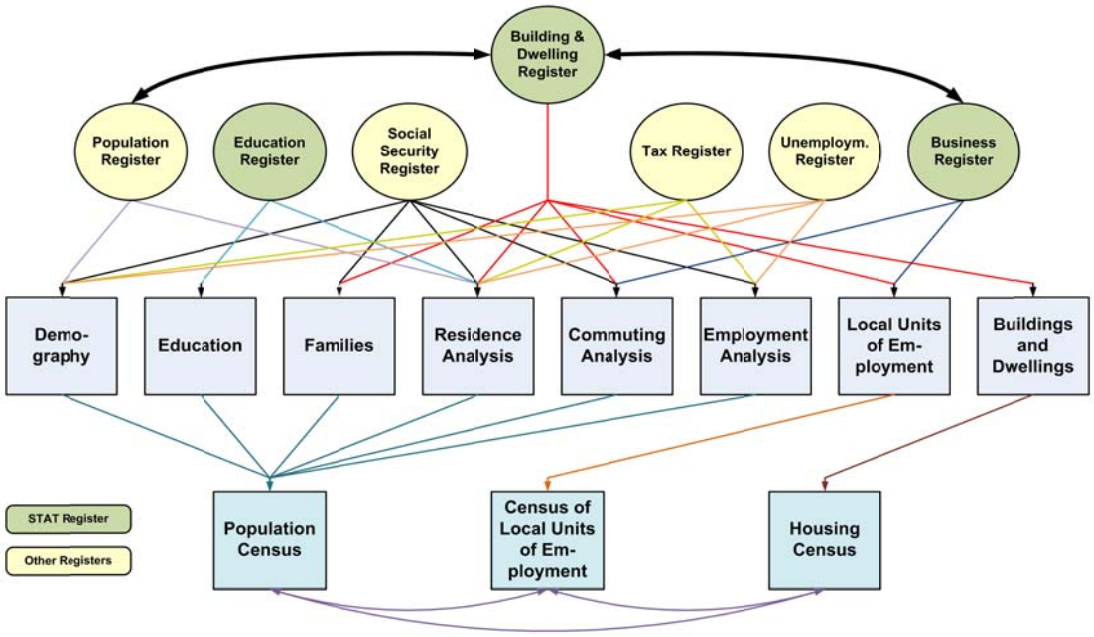
- Central Population Register (CPR), Ministry of the Interior
- Central Social Security Register (CSSR), Association of Social Security Institutions
- Tax Register (TR), Ministry of Finance
- Unemployment Register (UR), Public Employment Service Austria
- Register of Educational Attainment (EAR), Statistics Austria
- Register of Enrolled Pupils and Students (PSR), Statistics Austria
- Buildings and Dwellings Register (BDR), Statistics Austria
- Business Register of Enterprises and their Local Units (BR), Statistics Austria

23. These registers contain information on all variables required for the register-based census. Other registers that are used to assure the quality of the results ('comparison' registers) are:

- Family allowance register (FAR)
- Central register of foreigners (CRF; including asylum seeking persons)
- Registers of public servants of the federal state and the Länder (RPS)
- Register of car owners (COR)
- Register of social welfare recipients (SWR)
- Conscription register (CR)
- Register of alternative civilian service (ACSR)

24. The 'comparison-only' registers contain primarily basic demographic data but also additional information on employment of public servants (for example whether they are working full-time or part-time, place of work and branch of economic activity), or information on military or alternative civilian service as a supplement of the CSSR and TR.

Figure 1 The register model for the census, base registers* and the principle of redundancy



*The Education Register represents both the registers of educational attainment and enrolled pupils and students listed at paragraph 22.

Residence analysis – over-coverage in the Population Register

25. The Register-based Census Act prescribes that every person with a main residence as recorded on the CPR has to be counted if:

- she or he has lived in Austria for at least three months around Census day; and
- the residence is confirmed by a procedure referred to as the 'analysis of residence'.

26. The residence analysis is undertaken in order to avoid the over-coverage that is expected to exist in the CPR[31]. In Austria, the evidence of being registered is required for various necessary activities, so under-coverage in the CPR is expected to be of far less significance.

27. In a first step all register data are linked using the bPIN OS. Then, records without a bPIN OS are matched statistically to records of the CPR. At the end of this process there are inevitably records (persons) with no matches between the CPR and the other administrative data. To have a record in an administrative register other than the CPR means that a person shows signs of life on reference day (such as being employed or unemployed, receiving social welfare benefits, attending a university) and thus confirms the CPR.

28. Persons who are registered in the CPR but nowhere else do not show satisfactory signs of life. Their residence has then to be clarified by responding to an official letter[32], which asks about the place of the main residence on the reference day. In 2011, about 96.000 persons were queried and required to return a completed and signed form. A confirmation of a residence in Austria by a respondent has to be accepted, the procedure of the residence analysis is closed for them. This is not the case if a residence was not confirmed (such as when a person states, that he or she is living abroad; letter cannot be delivered because person is unknown; letter delivered but no answer at all). Instead of simply not counting all these people, Statistics

[31] Persons, who emigrate, often do not notify their local registration authority, or deaths may be registered with a time lag.
[32] Only for these persons, the Register-based Census Act allows a re-identification by the data protection authority.

Austria is obliged to inform the municipalities of the results of the queries. The local authority has the right to refute this but must prove that the person still has his/her main residence in the municipality, and such evidences has to be submitted to Statistics Austria.

29. On 31 October 2011 the CPR covered a population of about 8,466,000 (not including persons who had been registered for less than three months), whereas the population of the register-based census was reported as 8,401,940. About 62,800 persons were deleted from the population count due to the results of the query of persons with insignificant signs of life.

30. Residence analysis included other procedures such as checking for deaths by using data from the central social security register (about 3,800 persons were found to be dead but still registered), eliminating double counts and deleting delayed de-registrations. To a certain extent, also under-coverage from delayed registration is considered, for example in case of new born children or move to a new home by using information from CPR data extracted 6 months after reference day. The balance of retrospective corrections of the CPR was around 3,100 persons. So over-coverage in the CPR was around 0.8 per cent, under-coverage due to delayed registrations much smaller.

The 2006 Test Census

31. A major element of the preparation of the 2011 register-based census was to conduct a test census (on 31 October 2006) in order to evaluate the administrative data sources which were designated to be used in 2011, the residence analysis and the data processing procedures. It contained all the elements of the register-based census planned for 2011 (full enumeration, and full range of topics). The key difference was that the population count was not meant to have any legal consequences for the municipalities (regarding the fiscal transfer from the federal state to the municipalities).

32. A sample survey was conducted in order to evaluate the quality of the results. The sample was 25,000 people (representing 0.3 per cent of the total population) in 100 selected areas, and the data was collected in a field enumeration by interviewers using paper questionnaires.

33. The results of the Test (Statistik Austria, 2009) showed that information from registers was predominantly very good, and that in the expected problem areas such as household and family statistics was generally satisfactory. The conclusion was that the concept of the register-based census had been proved to work. Statistics Austria was then commissioned to continue with the preparation of the first register-based census in 2011.

Yearly Population Count on 31 October and the Register-based Labour Market Statistics

34. As a result of the quality of the data obtained in the 2006 Test Census, the Government decided to change the modalities of the fiscal transfer from the federal state to the municipalities. In Austria, the key factor for the calculation of the amount of tax revenues to be transferred was the population of the municipality as determined in the decennial population census. But from now on it was decided that the distribution should be re-calculated yearly on the basis of an annual population count as at 31 October, starting at 2008. The new law stipulated that the procedures for determining the population count as designated for the register-based census should be applied for these annual counts. There is one exception: the residence analysis between the census years does not include asking people with no signs of life as was done in the 2011 Census (and in the 2006 Test). In order, therefore, to reduce over-count an estimation model is used which is based on the results of the 2006 residence analysis (up to 2010) and since 2012 on the results of the 2011 analysis.

35. Based on the yearly provision of administrative register data, Statistics Austria decided to produce annual census-type statistics called Register-based Labour Market Statistics. In the first year, 2008, only a small range of topics was published. In 2009, topics such as highest education completed and place of work were added. In 2011, the Register-based Census was conducted including residence analysis as described in section III.5. From 2012 on, the yearly register-based statistics has comprised all census topics except housing.

Quality assessment

36. The method of collecting data from administrative registers required a completely new quality assessment concept. To this end, a quality framework[33] was developed that independently assesses the quality of administrative registers, the quality of the results and the quality of the processes at the individual variable level. Results of the quality assessment have been published for the 2011 Register-based Census and are available for the yearly Register-based Labour Market Statistics, additionally offering the possibility to compare results over time.

Improvements expected for 2021

37. The 2021 register-based census will be conducted based on the 2006 Census Act with no substantial changes in the methodology and processing procedures. From November 2014 onwards, data on the registration of births, deaths, marriages, registered partnerships and divorces are entered into a central database, which is kept by the Ministry of the Interior. At the same time a central citizenship register was established. Statistics Austria is allowed to access information from both administrative registers. It is expected that the quality of the data on demographic variables, especially the topics 'legal marital status' and 'family relationship' which, though already very good, will improve.

38. Currently, Statistics Austria is testing the use of information on the occupation of a person from the tax files and from the social security register. In order to use this information meaningfully, automatic coding will be required.

Summary

39. In Austria, the necessary conditions for the transition from a traditional to a register-based census were:

- a comprehensive register system developed for administrative needs (taxes, social security, registration of residences, etc.), readiness to set up new (statistical) registers;
- legal basis;
- ability to link data on individual level across sources while maintaining data protection;
- government and stakeholder approval, public approval;
- cooperation between the statistical office and other authorities and register owners;
- building up knowledge of the administrative data sources (by continuously working with the data and assessing their quality);

40. The use of administrative and statistical registers has become an integral part of today's statistical production process at Statistics Austria.

Reference

Statistik Austria (2009). *Bericht über die Probezählung 2006: Ergebnisse und Evaluierung*. Wien. Available only in German. Downloadable at:
http:///www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&dDocName=036181

[33] See references in Chapter V.

ANNEX E

SLOVENIA CASE STUDY

Census history and development in Slovenia from 1971 to 2002

1. The use of the administrative sources in statistical and census production has a long tradition in Slovenia. Following the long-term strategy to implement the Nordic model of statistics, the development had already started in the early 1970s. At that time the Statistical Office of the Republic of Slovenia (SURS) was an initiator and also the developer of the proposal for the legislation on national infrastructural registers. In addition, the introduction of a unique personnel identifier (PIN) into both administrative databases and statistical surveys was crucial for the development of register-based census statistics. As there were no similar initiatives in the other governmental bodies, SURS itself, as a producer, established in the 1980s four basic registers in close cooperation with the corresponding authorities:

- the Central Population Register (CPR),
- the Register of Spatial Units (addresses),
- the Statistical Register of Employment, and
- the Business Register.

2. As the administrative function of a register must differ from the statistical function, three of these (with the exception of the Statistical Register of Employment) have been transferred to relevant ministries after the adoption of the National Statistics Act in 1995. The National Statistics Act also requests all public authorities to use the general classifications and, where possible and feasible, align their administrative data concepts and variables to the statistical concepts and definitions.

3. In fact, the first (but unsuccessful) attempt to establish the CPR as a backbone for the population statistics and censuses was prior to the 1971 Population Census. But the second attempt in 1980 was the real starting point for register-based statistics as the PIN (still in use in the same format) was delivered prior to the 1981 Population Census to all permanent residents of Slovenia, and was also collected as a variable in the field resulting in 80 per cent coverage of PINs in the final 1981 Census database. In addition to the PIN, in the 1981 Census, for the first time, some administrative data (on educational attainment, occupation and industry) provided by employers, but only on paper forms, were used for rationalizing the data collecting stage and for improving quality. Data from the 1981 Census were also used as the base for regular daily updating of CPR data based on statistical demographic surveys and some administrative records, resulting in the dissemination of the stock population directly from the CPR using permanent residence definition from 1986 to 1994.

4. The subsequent (1991) Census - conducted only three months before the independence of Slovenia - can be described as the first transitional census. For the first time, the pre-printed questionnaires, using data from the CPR (PIN, name, surname, address) and the Register of Spatial Units (territorial codes), were used for collecting field data. In addition, data from the Statistical Register of Employment were used in the processing stage. This was the first real data integration process used in statistical production in Slovenia. Conversely, in the 1991 Census, the classic statistical processing (mostly manual editing) was performed for the last time. The population count derived from the CPR data was 1.8 per cent higher than 1991 Census count; the coverage of PINs in the final census database was 99 per cent.

5. The organization of the 2002 Census field enumeration and the statistical processing system were important steps to adopting the full register-based approach in the 2011 Census. Indeed,

the 2002 Census was the one and only combined census in Slovenia, since in addition to complete field face-to-face enumeration using paper questionnaires, data for some census topics (place/country of birth, last migration, citizenship, marital status, occupation, industry, place of work) were taken entirely from the registers and not collected in the field, while for some other topics (sex, date of birth, activity status) data were only collected in the field if they were not available in the pre-census database – a data set derived from various administrative and other statistical sources including the 1991 Census.

6. The main innovations in the 2002 Census were:

- composition of two databases complied from nine different administrative and statistical sources;
- uniform identifications and barcodes pre-printed on the questionnaires;
- an optical archive of images of all questionnaires;
- simultaneous verification and automatic coding based on images;
- online statistical editing (consistency check) supported by images of questionnaires.

7. The pre-census database was created six months before the reference date (31 March), and was used for pre-printing particular data (for example, the census area codes, address, name, sex and PIN) onto the questionnaires, and for planning several field enumeration and data processing activities. The final database used for data processing was established five months after the reference date. The coverage of PINs was complete. The difference between 2002 Census data and CPR data was slightly smaller than 11 years previously (1.6 per cent).

Creating the necessary conditions

General prerequisites

8. The decision to move to a completely register-based census was adopted in 2007 by the management of the SURS based on three prerequisites that SURS eventually fulfilled:

- legislation enabling free access to administrative data sources and linkage of data from different sources;
- availability of appropriate administrative or statistical sources with unique identifiers to link data on persons, households, and dwellings; and
- appropriate variables in the sources covering most of the demands of national users and corresponding to the (then draft) EU Regulation on population and housing censuses that was subsequently adopted in 2008.

Census legislation

9. There was no need for a law to specifically prescribe a register-based census as the legal basis already existed. The acquisition and integration of data is allowed by Articles 32 and 33 of the National Statistics Act (Official Journal of the Republic of Slovenia, No. 45/95 and 9/2001). Slovenia's decision on the register-based census as a method of collecting and processing data was adopted with the Medium-Term Programme of Statistical Surveys 2008-2012 (Official Journal of the RS, No. 119/2007) and the Annual Programme of Statistical Surveys for any particular year for which a complete register-based census is going to be conducted.

New census date

10. The coherence of census results with other statistics is an important step forward compared to the previous field census results. The new census reference date (1 January) instead of 31 March was selected for the following reasons:

- many administrative sources are linked to the calendar year;
- easier comparability of census data with annual demographic surveys; and
- greater consistency of administrative sources at the end of the calendar year.

New development after 2002 Census

11. From the content point of view, the decision to adopt a register-based census was viable because the only missing register - on dwellings (Real Estate Register) - was established in 2007 on the basis of both a special field real estate census conducted by the Surveying and Mapping Authority of Slovenia and already available sources (geodetic cadastre, court land register). According to the National Statistics Act it is not allowed to establish administrative registers based on statistical data; consequently, the 2002 Census data on housing could not be used for this purpose. At the same time, the numbers of dwellings in multi-dwelling buildings were determined and addresses of people in the CPR were supplemented accordingly.

12. The dwelling number is now also a part of the official address record in the CPR, and was the last missing link connecting people to their dwellings. In addition, as part of the computerization of administrative internal affairs, the Ministry of the Interior set up an electronic Household Register that used to be manually kept in the form of card files. The Household Register is a Slovene particularity, since other register-based countries do not have such a high-quality data source on the household structure. The most important advantage of the Household Register is the ability to implement the housekeeping concept and the availability to derive data on relation to the reference person of the household which is necessary to determine family composition.

Pre-census evaluation of quality of input data

13. The first step after the evaluation of methodological solutions based on available administrative and statistical data sources and approval at the appropriate SURS body, was to conduct a test census with the primary aim of analysing and evaluating the quality of the input data in terms of coverage, relevance, reliability, timeliness, accessibility and comparability. Three important obstacles to achieving acceptable quality were recognized at this early stage:

- inconsistencies in household composition (minor problem solved in the processing stage);
- excessive under-coverage of dwelling numbers (particularly in relation to multi-dwelling buildings) in the CPR;
- general poor quality of housing data as the main problem which was not adequately solved by the 2011 Census reference date.

Improving quality of data on dwelling number in CPR

14. That the completeness of updating dwelling numbers in the CPR was far below expectations created a challenge because this variable is crucial for matching dwellings with persons and households. These data were missing from the trial census of approximately 400,000 persons (more than half of the population living in multi-dwelling buildings). To rectify this, the Ministry of the Interior and SURS undertook two main activities in close cooperation. Firstly, methodological solutions for automated determination of missing dwelling

numbers were devised by linking data on ownership of dwellings with the registered residence of owners and their households (with the presumption being that most owners lived in their own dwelling). Then an official letter was sent to the reference person of the household living in a multi-dwelling building for which a dwelling number was not known to report this information; some 49,000 letters were sent out, and the response rate was 75 per cent.

Main principles of the register-based census

Linkage of data on persons, households, and dwellings

15. The linkage of data on persons, households, and dwellings using unique identifiers is one of the most important tasks in producing multivariate census data using field enumeration or register-based data. In the case of the register-based census, the direct linkage of all data sources for persons using a PIN (as shown in Figure 1) is the basic statistical operation. The PIN is transformed into a statistical identifier (SID) to protect confidentiality and privacy before the statistical processing of census data, and the household identifier from the Household Register is used. The household identifier is the serial number of the household running from 1 to NNNN at the same address. The dwelling number is an identifier linking persons and dwellings and is also connected to the address (via the serial number).

Figure 1 Identifiers used in the register-based census, Slovenia

Register / Database	PIN	Address	Dwelling ID	Household ID	Business ID
Central Population Register	X	X	X		
Real Estate Register	X	X	X		X
Household Register	X	X	X	X	
Statistical Register on Employment	X				X
Other population sources	X				

Quality of basic identifiers

16. The PIN is the most important identifier with (as has been noted) complete coverage in the CPR, but could be missing in some other administrative or statistical sources used for census purposes. The main quality obstacle of the household identifier is the fact that the household ID (and also the relationship to the household reference person which is also considered to be a key identifier in the Slovenian register-based system) is available only for permanent residences. Despite efforts to improve the coverage of dwelling numbers for persons living in multi-dwelling buildings in the CPR, many dwelling identifiers were still missing before the first stage of data integration. The distribution of input and output data for key identifiers in the whole statistical process for the 2011 and 2015 register-based censuses is presented in Table 1.

Table 1 Quality indicators for key identifiers, register-based census, Slovenia

Identifier	Number of records	Unchanged	Imputation	Correction	Number of records	Unchanged	Imputation	Correction
		Share in %				Share in %		
	2011				2015			
Dwelling ID[1)]	724,479	75.3	12.3	12.4	712,989	94.0	4.6	1.4
Household ID[2)]	2,016,423	94.9	2.1	3.0	2,024,604	93.9	1.5	4.6
Relation to the reference person[2)]	2,016,423	91.6	4.2	4.2	2,024,604	91.6	1.5	6.9

Source: SURS, Statistical Office of the Republic of Slovenia

[1)] Multi-dwelling buildings. [2)] Private households.

17. The main difference between the two censuses was the quality of the input data. Far fewer dwelling numbers were missing in 2015 than four years previously (30,000 in comparison to 89,000 in 2011). Even better quality was noticed in the Real Estate Register because of the (later invalidated) Mass Real Estate Valuation Act, which stipulated that dwellings would be taxed based *inter alia* on floor area and year of construction, with different rates for residential and unoccupied housing units. The share of correct records on dwelling ID consequently increased significantly, from 75 per cent in 2011 to 94 per cent in 2015.

Administrative and statistical sources

18. Three administrative registers form the backbone of the register-based census system:

- Central Population Register (CPR) maintained by the Ministry of the Interior;
- Household Register (HR) as a part of Central Population Register.
- Real Estate Register (RER) kept by the Surveying and Mapping Authority of Slovenia;

Most of data for the variables to be covered in Population and Housing Censuses (as prescribed in the EU Regulation 763/2008) have been extracted from one of these sources, for example:

- place of usual residence, sex, age, legal marital status, country of citizenship, place of usual residence one year prior to the census from the CPR;
- relationship between household members from the HR; and
- All housing topics including tenure status from the RER.

19. Data for the other personal variables have only been produced by using and combining input data from several data sources. The basic methodological principle for production of statistics in such cases is the hierarchy of the sources, which means that, in the case of availability of data from several sources for each person (identified by the PIN), priority is given to the source with the higher priority (indicated by the lower number in Tables 2 and 3) allocated after the quality evaluation of all sources has been evaluated. For educational attainment data, nine sources have been used. Data on educational attainment are now updated annually using the data sources displayed in Table 2 (with the exception of the 2002 Census).

Table 2 Data sources on educational attainment, 2011 Register-based Census, Slovenia

Priority	Source keeper	Source content	Period	Share (%)[34]
1	SURS	Tertiary education graduates	1989 - 2010	11.1
2	National Examination Centre	Graduates of general and vocational Matura	2002 - 2010	9.1
3	Chambers	Vocational upper secondary education	2002 - 2010	0.2
4	SURS	Student enrolment in tertiary education – education at enrolment	2002/03 - 2010/11	2.6
5	National Examination Centre	National examinations at the end of elementary education	2006 - 2010	4.6
6	SURS	Recipients of scholarships	2006 - 2010	0.5
7	SURS	Data from the Statistical Register of Employment on educational attainment	1986 - 2010	56.0
8	Employment Service of Slovenia	Registered unemployed persons	1. 1. 2011	0.8
9	SURS	2002 Population Census – highest level of education	31. 3. 2002	13.6

Source: SURS, Statistical Office of the Republic of Slovenia

[34] Population aged 15 or more years.

Table 3 Data sources on economic characteristics, 2011 Register-based Census, Slovenia

Priority	Source keeper	Source content	Period	Share (%)[1]
1	SURS - Statistical Register of Employment	Persons in paid employment Self-employed persons and farmers	Last week	45.7
2	Employment Service of Slovenia	Registered unemployed persons	1. 1. 2011	5.9
3	SURS	Full- and part-time students in vocational and professional higher education	Academic year	4.4
4	SURS	Recipients of scholarships in upper secondary and tertiary education	1. 1. 2011	1.0
5	Pension and Disability Insurance Institute	Recipients of old-age, disability, survivor's, and national pensions	1. 1. 2011	29.1
6	Health Insurance Institute	Family members of insured persons and other inactive persons with health insurance	1. 1. 2011	10.6
7	Ministry of Labour, Family and Social Affairs	Recipients of social and other assistance and benefits	2010	0.8
8	Tax Administration	Income tax payers	2010	0.6

Source: SURS, Statistical Office of the Republic of Slovenia

20. Data on economic characteristics (current activity status, occupation, industry, status in employment, location of place of work), derived from eight sources shown in Table 3, generally refer to the census reference date.

21. Data for migration characteristics (country/place of birth, ever resided abroad, previous place of usual residence) are produced from statistical surveys only which are based on CPR data:

- annual statistical survey on migration (data from 2002 to 2010);
- annual statistical survey on birth (data from 2002 to 2010);
- quarterly statistical survey on population, as of 1 January 2010; and
- 2002 Population Census.

Statistical process

22. Data availability determined the timing of the four-phase production and dissemination of the results of the 2011 register-based census:

- integration of input data for population, households, and housing (first release of some final population census data at the end of April 2011);
- processing of household and family data (first release June 2011);
- all other population census topics (economic and educational characteristics, migration, fertility), and preliminary data on occupied dwellings released at the end of 2011;
- occupied and unoccupied dwellings processed last because the housing characteristics from the RER as of 1 January 2012 were updated again (first release on 21 June 2012).

23. Data that have already been disseminated are not subsequently revised or updated at any later stage of the process, so special metadata tables are prepared to ensure that any changes are tracked. The goal was for the last status of an individual record to be retained in the final census database. Two other goals were also attained: traceability and repeatability. In other

words, all changes in data made during the statistical process were recorded transparently and clearly.

24. A special website was established for the dissemination of the 2011 register-based census (http://www.stat.si/popis2011/eng/Default.aspx?lang=eng) that included basic methodological explanations and information. In addition, data are also available online from SI-Stat Data Portal (http://pxweb.stat.si/pxweb/Database/Demographics/Demographics.asp) under the Population and Level of Living section There was only one printed report entitled *People, Families, Dwellings* (Statistical Office of the Republic of Slovenia, 2013).

Quality of the register-based census outputs

25. The overall quality of the statistical data depends, to a large part, on the quality of the underlying administrative data, as a register-based census becomes no more than a statistical operation transferring administrative records into statistical outputs. The requisite conditions for achieving quality outputs from this perspective are:

- the well-established and consistent use of administrative data in the statistical process;
- very close cooperation with the keeper of the administrative source; and

feedback from the statistical evaluation of the administrative source.

Regular quality monitoring

26. Over-counting is the most common problem of any register-based statistical system. Two main methods are used for the quality assessment of the coverage of the population in the Slovenian census:

- Imputation rates for data on educational attainment and labour force status could be an indicator of over-registration as data are available only in the CPR. The quality of input CPR data has improved as shown by the fact that the imputation rate for labour force status in 2011 was 1.50 while in 2016 it was only 1.14;
- The residence status of the selected respondent person in social sample surveys which could be:
    - living at the sampled address (correctness);
    - already died but not registered (administrative survivors) – only a few cases;
    - living elsewhere in Slovenia (differences between de facto and registered residence are the most relevant for territorial distribution of population data inside Slovenia) ranging from 3 per cent to 6 per cent;
    - living abroad – over-registration was in the order of 1 to 2 per cent;
    - no answer about residence status (non-response) - less than 1 per cent.

Special survey on coverage

27. A special survey with a focus on plausible unregistered emigration was conducted in 2016. The sample frame consisted of persons for whom labour force status had to be imputed in the 2015 Census (15,500 persons); the second sample group consisted of persons marked in the administrative data (CPR) as non-residential group, but where data on their labour force status in Slovenia could be found in at least one out of the nine sources (2,700 persons).

28. Two methods were used: a postal self-response method for the whole sample, and face-to-face field inquiry (1,915 persons) for selected postal non-response (71 per cent). The final outcome of the survey was an estimate of over-registration of 0.5 per cent at the aggregated level (around 10,000 persons at most). In comparison with the last field census data in 2002, when almost 1 per cent of the population was counted twice (over-coverage) and slightly less

than 2 per cent of the population was not counted at all (under-coverage), the results of the survey are promising.

Conclusion

29. Basic demographic (census-type) data in Slovenia are produced quarterly using usual population definition. The statistical definition of population is completely harmonized with all existing EU Regulations defining usual residence. The main input administrative data are transmitted quarterly from the CPR approximately three months after the reference date. In addition to this, data on socio-economic characteristics using census statistical processes are produced annually. A complete register-based census as a regular statistical operation will have been conducted twice (2015, 2018) between obligatory years determined in EU Regulation (2011 and 2021).

30. The register-based approach in Slovenia achieves the key objectives set out in *Challenges for Future Population and Housing Censuses* prepared by Statistics Canada, CIS-STAT, and the UNECE secretariat for the 60th plenary session of the Conference of European Statisticians held in Paris in June 2012, namely:

- Increasing concerns over costs: there was no additional or special budget because the register-based census in Slovenia is now a regular statistical survey conducted under the Annual Programme of the Statistical Surveys.
- Improving data quality: a controlled methodological approach was used in all stages of the process, and there were no problems associated with field enumeration under-coverage or item non-response or difficulties with data entry and editing.
- Respondent burden and decreasing participation in the census: these are not a problem anymore in Slovenia.
- Privacy: far fewer persons are now handling information, in contrast to thousands of people having access to personal data in a field-based operation.

Reference

Statistical Office of the Republic of Slovenia (2013). *People, Families, Dwellings*. Population Census 2011. Issued and published by the Statistical Office of the Republic of Slovenia, Ljubljana. Downloadable at http://www.stat.si/StatWeb/File/DocSysFile/3712/people.pdf.

ANNEX F

PORTUGAL CASE STUDY

1. This annex summarises the work developed by Statistics Portugal (SP) between 2014 and 2016 to build a Statistical Population Dataset (SPD), which aims to replicate the country's resident population and characterise it through a set of demographic and socio-economic variables. Administrative data sources are presented, as well as the methodological approach, the quality indicators to assess its fitness for purpose and comparability with the 2011 Census, the annual Population Estimates (PEs) for 2015 and the 2016 Census Test (CT).

2. Unlike other countries that have already made the transition to a register-based or a combined census, in Portugal there is neither a central population register nor a unique Personal Identification Number (PIN). In addition, the country does not have a legal framework allowing access to the full name and address of persons in registers. Notwithstanding the limitations of the process, the results obtained in this short period of time are encouraging and pave the way for new measures that can lead to a paradigm change and the setting-up of medium to long-term strategies, to be implemented in 2021 and beyond.

Censuses in Portugal

3. Portugal has held censuses since 1864; every ten years since 1890. Throughout the census series, SP has introduced changes to the process, to make it more efficient. In 2011, online response was introduced quite successfully, with a response rate of 50 per cent.

4. Similar to some other countries, a feasibility study is being conducted in Portugal to analyse different methodological options for censuses. The contribution from registers is being assessed in order to improve the efficiency of census operations and to allow a more frequent and updated release of statistical data.

Construction of a statistical population dataset in Portugal

5. The main purpose of a census is to enumerate and characterise the resident population, particularly by releasing information for small geographical areas. So, with the aim of understanding whether or not the administrative information available in Portugal allows for the replication of a high-quality resident population enumeration, two exercises have been undertaken to create a Portuguese SPD. This database was built from a number of registers provided by different administrative data sources. The reference year for the first exercise was 2011, in order to use the 2011 Census as a benchmark for the comparison of results. The reference year for the second exercise was 2015, and the results were compared with the PEs for the same year.

6. The starting point for the SPD was the Civil Register (CR) file. This register contains the demographic characteristics of all Portuguese citizens. However, the CR is not a central population register, and relates citizens only to their legal, or registered, address in Portugal (which does not necessarily conform to the census concept of usual residence). The CR overestimates the country's resident population by 10 per cent (more than 1.1 million persons) when compared with the population enumerated in the 2011 Census and does not include most immigrants living in Portugal, who are registered separately in the Immigration Register (IR).

7. The SPD was built based on a 'Signs of life' methodology, given by the presence of a person in more than one register. In a very simplified way, a person is considered to be a resident in

the country if he/she is registered in the CR or in the IR and is, in addition, 'active' in at least one other register (for example, that persons studies, works, has used healthcare, pays taxes, etc.). The application of algorithms based on a person's presence in various registers made it possible to identify and distinguish those who actually live in Portugal from those who, though not living in Portugal, maintain their legal address in the Portuguese territory.

8. At a second stage, the relevant demographic and socio-economic administrative variables were assigned to the population in the SPD.

Administrative data sources and variables

9. In addition to the Civil Register and Immigration Register, eight administrative datasets were used to build the SPD:

- Social Protection for public servants,
- State Pension and Work Fund Register,
- Education Register,
- Private Employment Register,
- Unemployment Register,
- Social Security Register,
- Taxes Register, and
- the National Health Service Patient Register.

10. The administrative variables available in these different sources contribute towards providing data on 16 census topics, encompassing the 13 required at the geographic level prescribed by EU Regulation – sex, age, place of usual residence, place of residence one year before, marital status, citizenship, country/place of birth, labour status, occupation, branch of economic activity, status in employment, place of work, and educational attainment – and three others at the national level – number of hours worked, number of employees in the enterprise, and school attendance.

11. The coverage of each variable depends on the presence of individual records in the respective source files: only seven out of 16 variables (sex, age, place of usual residence, place of residence one year before, marital status, citizenship, country/place of birth) provide full coverage of information for the SPD. Coverage is only partial for the socio-economic variables. Moreover, for a number of important variables on the person and the household characteristics (specifically the household classification), there is either only partial information or no information at all currently available in the registers.

12. Moreover, there are some other sets of information – relevant to housing-related variables - that, although held in registers, cannot be used because of the particular legal restrictions noted at paragraph 15 below. Statistics Portugal has, consequently, created a Statistical Dwellings Dataset (SDD) that has evolved from the results of the 2011 Census and updated with inputs from several other sources, including the National Indicators System of Urban Operations (SIOU)[35], and other sample surveys of household and persons conducted by SP. But because there is no legal access to the full address (just local and postcode) it cannot be linked to the SPD.

[35] SIOU is based on administrative data from the 308 Portuguese municipalities relating to building permits and completed constructions.

13. The quality of the registers has been assessed according to a number of general criteria involving indicators such as the coverage rate for each variable, the accuracy and timeliness of the information, as well as a finer assessment at microdata level, comparing the information collected in the 2011 Census with information from 2011 reference dates registers.

Limitations in the Portuguese SPD construction process: privacy and protection

14. Access to registers is prescribed by the Law of the National Statistical System, (Law No. 22/2008 of 13 May 2008). The National Data Protection Commission has produced a set of recommendations to safeguard each person's confidentiality. Individualised records have been anonymized at the source by applying an algorithm that encrypts numerical identifiers, inhibits access to the person's full name (only the first three letters of the first name and the last three letters of the surname) as well as to the person's full address (only place of residence and postcode).

15. In Portugal there is no unique PIN. There are, instead, four numerical identifiers: the CR number (NIC) or the IR number (in the case of immigrants), the Taxes Register number (NIF), the Social Security number (NISS), and the National Health Service number. The registers have one, two, or three of these numerical identifiers, depending on the administrative source, but these do not always cover all of the records, for example: the CR only contains the NIC; the Taxes Register only contains the NIF; the Education Register contains both NIC and NISS, but in the first case for 90 per cent of the records and in the second case for less than 70 per cent.

16. In the absence of a unique PIN, which would no doubt increase the accuracy of linkage and improve matching rates, it was necessary to build linkage keys, essentially through deterministic methods, based on personal characteristics.

Results and quality assessment

17. For 2015, the population estimated from the SPD was 10,434,161 persons, with a deviation of 0.9 per cent (around 93 thousand persons) from the PEs for the same year (10,341,330 persons). The population's age structure and sex distribution given by the SPD were also consistent with those given by 2015 PEs.

18. Compared to the first exercise, relating to 2011, which estimated 8.6 million persons, this new edition of the SPD estimates the population size very accurately at a national level. Figure 1 illustrates the evolution of the results based on the introduction of new administrative sources. The improvement in the 2015 SPD when compared to the 2011 edition is due to two factors: the incorporation of the Taxes Register, with 9 million records, was relevant to determine the SOL, particularly for population segments that do not carry out any kind of business, do not receive any social benefit and are not studying, but being registered as part of the household in the Taxes Register. The second factor was the improvements in the record linkage process.

19. Also, at regional level (NUTS 2), deviations in the enumeration of the population were consistent with the 2015 PEs, ranging from -2 per cent to 0.7 per cent. But, when considering the population at the 'municipality' level, Figure 2 shows that there were minor differences between the 2015 SPD and the PEs for the same year for most areas. However, for 77 per cent of the country's municipalities, deviations between the population estimated by the SPD and the PEs were lower than 5 per cent.

Figure 1 – Resident population by age group, Portugal, 2011 and 2015

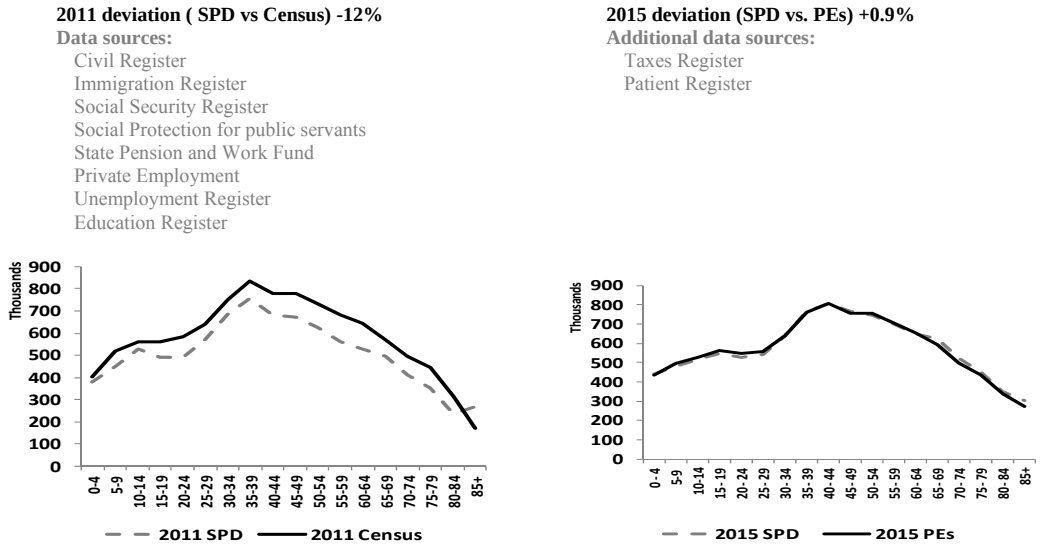
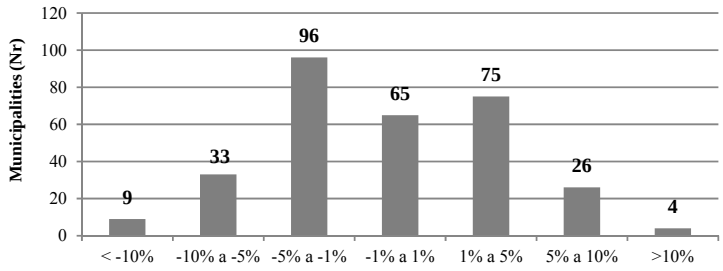


Figure 2 – Number of municipalities according to the deviations between the 2015 SPD and the 2015 PEs



20. The assessment of the deviation for small areas was conducted through a test survey (the 2016 CT), in five parishes (at the LAU2 level) one in each of five municipalities (at the LAU1 level). The sample size was 45 thousand dwellings - 1 per cent of housing units taken from the SDD. Around 70 thousand persons answered the survey. The comparison of the 2016 CT and the 2015 SPD results led to the following conclusions:

- Enumeration of the population, based on registers, still has some limitations. The 2015 SPD over-estimated the resident population in all 2016 CT parishes, with deviations ranging between -5.7 and -21.8 per cent;
- Although the size of the population at the parish level showed some differences, the structure and characterisation of the CT parishes' population given by the 2015 SPD was quite consistent with that enumerated in the 2016 CT;
- The level of agreement at the microdata level between the 2016 TC and the 2015 SPD (77 per cent) was not entirely satisfactory, as a consequence of the matching limitations relating to the availability of the persons' full name and address;
- Administrative data approximated to the high-quality of the information collected in the 2015 CT parishes, with a very high, or at least an acceptably high, level of agreement for the selected variables (citizenship, 99.3 per cent; place of birth, 96.8 per cent; marital status, 95.1 per cent; school attendance and education attainment, 94.1 per cent; and labour status, 79.4 per cent).

Conclusions and outlook

21. The construction of a SPD has made it possible, for the first time in Portugal, to conduct a qualitative and quantitative assessment of the potential of using administrative data for census purposes. The results obtained were encouraging, but not satisfactory enough to undertake a fully register-based census in 2021. This is primarily because:

- The enumeration of the resident population, based on registers, has limitations for small areas (this might be explained by non-updated addresses on registers and non-optimized data linkage methods); and
- The currently available administrative information is not sufficient to respond to all the variables provided by the census. Key domains such as housing, household and family characteristics or education attainment still cannot be derived from the available registers.

22. However, considering that only one methodology - that based solely on administrative information - will allow the availability of more frequent census-type information, the on-going studies on the use of registers should be further enhanced. Improving the methodology used for building the SPD is the first step for this strategy.

23. This on-going investigation benefits from the good institutional cooperation and the appropriate conditions within public administration, as a result of the country's modernisation processes in recent years that have led to the availability of new registers.

24. Work is still in progress for the transition to a register-based census. The establishment of a more favourable legal framework (access to full name and address) is the key to overcoming problems related to data linkage, in particular to enable a liaison between the SPD and the SDD, and to the increase in accuracy of population estimates for small areas.

ANNEX G

UNITED KINGDOM CASE STUDY

Background

1. In May 2010, the UK's Office for National Statistics (ONS) began the 'Beyond 2011 Programme' to review the future provision of population statistics in England and Wales in order to inform the Government and Parliament about options for the next census. In particular the programme focused on the potential to replace the census with statistics based on administrative data already held by government, supplemented by household surveys.

2. On the basis of the research and evidence collected, the then National Statistician recommended, in March 2014 (ONS, 2014):

- *"An online census of all households and communal establishments in England and Wales in 2021 … [with] … special care taken to support those who are unable to complete the census online; and*
- *Increased use of administrative data and surveys in order to enhance the statistics from the 2021 Census and improve statistics between censuses."*

The National Statistician went on to note that:

> *"…[It] may offer a future Government and Parliament the possibility of moving further away from the traditional decennial census to annual population statistics provided by the use of administrative data and annual surveys."*

3. This approach was endorsed by the Government's formal response to the recommendation in July 2014 (Cabinet Office, 2014) which highlighted the ambition:

> "*that censuses after 2021 will be conducted using other sources of data and providing more timely statistical information.*…[subject to] *sufficiently validating the perceived feasibility of that approach."*

4. As a result, ONS is aiming to replicate the information collected through the census with administrative data already held by government, supplemented by surveys. The goal is to be able to compare outputs based on administrative data and targeted surveys against the 2021 Census to demonstrate that the alternative can produce high quality information at a lower cost, and can do so on a more regular basis. (Similar research is being carried out in parallel by National Records of Scotland and the Northern Ireland Statistics and Research Agency who are, respectively, responsible for the censuses in Scotland and Northern Ireland.)

5. However, such a change in approach is challenging in England and Wales (indeed, throughout the UK) because there is no population register and neither is there a unique identifier across administrative sources. Given the importance of producing accurate statistics, it would have been high risk to move straight to such a system without benchmarking new methods against the 2021 Census. This is in line with practice in other countries that have made the move more gradually.

6. This work addresses the Government's ambition, described above. It is also in line with ONS's strategy (ONS, 2013a) to be at the forefront of integrating and exploiting data from multiple sources, making greater use of administrative data across all statistics.

What is an Administrative Data Census?

7. It is ONS's ambition to produce the type of information that is collected by a ten-yearly census (on people, households and housing units) from an Administrative Data Census. Doing this will require a combination of:
• record-level administrative data held by government administration;
• a population coverage survey;
• a population characteristics survey.

Record-level administrative data held by Government

8. ONS will need access to a range of data held by government departments. Access will be required at the unit record-level to enable these sources to be linked together. High quality linking requires name, address, data of birth and sex - as described in more detail in the document *Matching Anonymous Data* (ONS, 2013b) - as combinations of these variables can be used to produce links that are made with a high-level of certainty. Linking together multiple sources will improve the quality and coverage of the outputs that can be produced, and will support the production of cross-tabulated outputs, such as employment by qualifications at small geographic levels within a local authority (LA).

Population Coverage Survey (PCS)

9. ONS anticipates the need to conduct a coverage survey, similar to a traditional Census Coverage Survey, to measure and adjust for under- or over-coverage in administrative data, and to enable the production of high-quality statistics on the size of the population. A PCS may cover approximately 350,000 households (1 per cent) on an annual basis, as described in the ONS Beyond 2011 information paper *Producing population estimates using administrative data: in theory* (ONS, 2013c). Further work is required to refine both the detail of the survey and the methods to subsequently produce estimates using the PCS and administrative data.

Survey to produce estimates about characteristics of the population and households

10. ONS currently anticipates the need for a separate survey to collect information on census characteristics. This may have two functions:

- For characteristics that are available on administrative data, the survey would need to measure, and adjust for, under- or over-coverage in administrative data, in a similar way to the PCS.
- The survey could also be used to provide direct survey estimates for topics that are not available from available administrative data (such as hours of unpaid caring). For such topics, it might be possible to produce estimates only at the LA level (where population size ranges from between 2,200 and 1,074,000).

11. The precise design (including size) of this survey will depend on the extent of ONS's access to administrative data and an understanding of its statistical quality.

Other sources of data

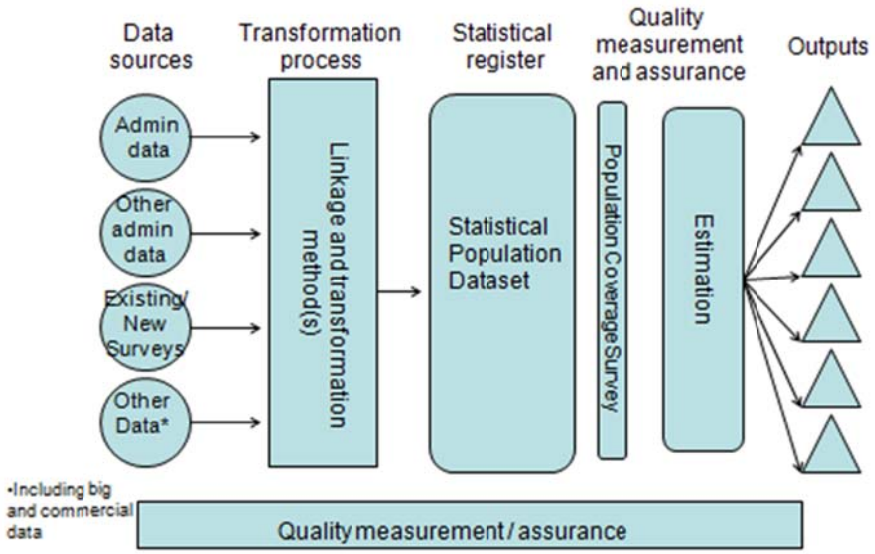
12. Other data (such as 'big data', private sector or commercial data) may also be considered for deriving some types of information traditionally produced by the census, such as commuting flows by using data from mobile phones, or information on tenure by using information from property websites.

13. These different sources of data may need to be linked together and used in combination with a range of methods and modelling techniques in order to produce the type of outputs that users require, and could also be used to quality assurance outputs. This approach may

additionally offer opportunities to provide users with the outputs that they need on a more frequent basis than provided by a ten-yearly census.

Adapting common framework for England and Wales approach

14. ONS has further developed and adopted the common framework presented in Chapter V of these Guidelines to take account of their specific challenges as follows:



What needs to be in place to move to an Administrative Data Census?

15. There are four key challenges to delivering an Administrative Data Census in England and Wales (and, indeed, throughout the UK):

- Accessing the range of data needed to produce outputs that are currently provided by the ten-yearly census;
- Linking together accurately lots of independently collected data whilst preserving personal privacy and the security of the data;
- Developing methods to transform the linked data into outputs that meet the needs of users;
- Making the linkage and use of such data acceptable to key stakeholders, for example by providing value for money, and providing reassurance that data will be kept safe and confidential throughout the whole approach.

To address these challenges, it would be necessary for a number of conditions to be in place.

Rapid access to existing and new data sources

16. To maximise the breadth and quality of statistics that could be provided by an Administrative Data Census, ONS would need to have rapid access to new and existing data sources from across government. This would also need to extend to other sources of existing data that would add value. ONS would also need to be consulted before changes are made to the administrative data that may affect the quality and stability of outputs from an Administrative Data Census over time. The Digital Economy Act, which passed into law in April 2017, offers a solution to at least some of these requirements.

The ability to link data efficiently and accurately

17. All countries that have moved away from conducting a five- or ten-yearly traditional census have adopted a combined or register-based census methodology that is underpinned by a population register and, usually, an ID card scheme. This usually means that administrative data can be linked to the register(s) through a unique ID number, resulting in highly accurate linkage. These registers also aim to provide complete coverage of the population, which administrative data does not always provide.

18. The UK has neither ID cards nor a population register. Instead, as described above, an Administrative Data Census would involve linking together by other means multiple administrative data sources and surveys to produce statistics on the range of topics that the census currently produces. This is, of course, not a simple task.

19. In the UK, individuals do not have a single unique reference number that is carried across all government-held data, making this linkage challenging. For example, data about tax and benefits from, respectively, Her Majesty's Revenue and Customs (HMRC) and the Department of Work and Pensions (DWP) use the National Insurance Number, while the General Practioner (GP) Register data uses the National Heath Service (NHS) number, and the School Census uses a unique pupil reference number. ONS, therefore, needs methods that can link together these independent data sources accurately to enable the production of high quality statistics, while, as already noted, preserving the confidentiality of personal information and the security of the data.

Methods to produce statistical outputs of sufficient quality that meet priority information needs of users

20. Accessing and linking data is only part of the puzzle. ONS needs to develop methods that can transform the linked administrative and survey data into statistical outputs that meet priority information needs of users. This means providing statistics on the topics that users need, at the right level of detail (for example, for small areas), and at the right quality. In response to a public consultation in 2013, users reported that any such statistical methodologies should provide:

- robust estimates about the size of the population and the number of households;
- estimates about population characteristics at a point in time to allow similar areas to be compared with one another;
- the granularity of information that users need to measure change over time (for example being able to spot changes over a decade in unemployment rates by ethnicity for small areas).

21. Another key area is developing the detail of the surveys that will be required and the methods to model from surveys and administrative data.

Acceptability to stakeholders (users, suppliers, public and Parliament)

22. In order to successfully move to an Administrative Data Census in the next decade, users of the data, data suppliers, the public and Parliament need to be convinced that this approach meets their needs and cost restraints. Acceptability to these four key stakeholders will be influenced by ensuring that:

- key information needs of users are met;
- data is held, processed and linked while providing privacy, confidentiality and security safeguards; and
- costs are significantly reduced.

Value for money

23. With particular regard to this third point, an Administrative Data Census will need to demonstrate that it provides value for money compared to a ten-yearly census. This means showing either that it can deliver the benefits that users get from a ten yearly Census at a lower cost, or that the cost saving is sufficient to justify a lower benefit. For example, the Administrative Data Census may not be able to deliver all the outputs that a ten-yearly census provides but it may include additional benefits such as more timely, frequent data and new outputs that are not currently provided by a ten-yearly census. This is the key trade-off that will need to be taken into account.

How will ONS know if an Administrative Data Census is possible?

24. For the government to make a decision after 2021 about the future of the census, ONS needs to provide evidence to show whether or not an Administrative Data Census is a viable approach to census-taking. In order to do that, ONS plans to do the following:

- Make progress in acquiring new administrative data sources, prioritising data sources that relate to, or may provide insight on, key topics that are currently produced by a ten-yearly census. For new data sources, record-level comparisons can be made with the 2011 Census, which provides a good benchmark of the statistical quality of the administrative data. For example, it can highlight whether an administrative source has coverage issues, or lags in updating address information. Comparisons with other data sources can also be useful to understand statistical quality. ONS will publish an update on the progress in acquiring data each year.

- Publish Administrative Data Census Research Outputs on an annual basis. Annual research outputs will demonstrate the type and quality of outputs that could be produced from an Administrative Data Census. To date, ONS has published such research outputs on:
    - The size of the population (ONS, 2017a);
    - Households and families (ONS, 2017b);
    - Population characteristics (ONS, 2017c).

  The range of topics will be expanded in future releases, depending on the availability of data and its statistical quality. A key aim of these outputs is to allow users the opportunity to provide feedback on the data and on the methods used to help focus future developments. ONS publishes a short summary of the feedback received from each set of outputs, demonstrating to users how their views and comments are being taken on board.

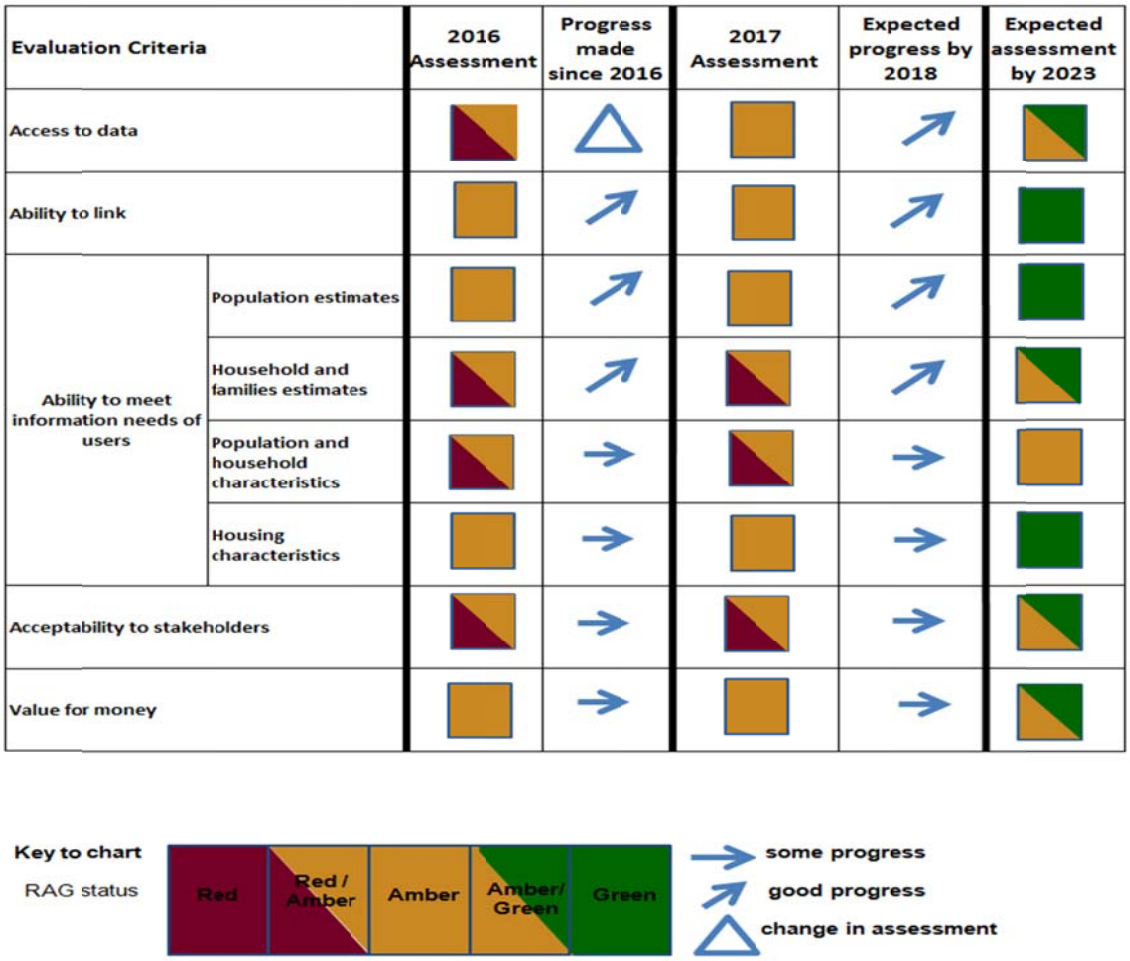
- Conduct an annual assessment of ONS's ability to move to an Administrative Data Census. This will ultimately conclude with a comparison of combined administrative data and survey based outputs with the 2021 Census outputs to benchmark this approach. This will culminate in a recommendation in 2023 on ONS's ability to switch to an Administrative Data Census.

- Have methods and research reviewed by an external expert panel. These reviews are planned to take place in 2017, 2020 and 2022.

25. In June 2017, ONS published the second assessment of its progress towards an Administrative Data Census after 2021 (ONS, 2017d). Figure 1 shows the outcome of this

assessment as at mid-2017. The full assessment is supported by evidence and a description of what will be done in the future to improve the assessment.

Figure 1 – Current (2017) and future expected (by 2023) high-level assessment



References

Cabinet Office (2014). *Government's response to the National Statistician's recommendation*. Letter from the Rt Hon Francis Maude, MP, Minister for the Cabinet Office to Sir Andrew Dilnot, Chairman of the UK Statistics Authority. Downloadable at: https://www.statisticsauthority.gov.uk/archive/reports---correspondence/correspondence/letter-from-rt-hon-francis-maude-mp-to-sir-andrew-dilnot---180714.pdf

ONS (2013a). *ONS Strategy, 2013-2023. Downloadable at* https://www.ons.gov.uk/ons/dcp14298_323384.xml

ONS (2013b). *Beyond 2011: Matching Anonymous Data*. Beyond 2011 Information Paper. Downloadable at: http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-matching-anonymous-data--m9-.pdf

ONS (2013c). *Beyond 2011: Producing population estimates using Administrative Data: In Theory*. Beyond 2011 Information Paper. Downloadable at: http://www.ons.gov.uk/ons/about-

ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011--producing-population-estimates-using-administrative-data--in-theory--m8-.pdf

ONS (2014). *The Census and Future Provision of Population Statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority.* Downloadable at: http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations/national-statisticians-recommendation.pdf

ONS (2017a). *Size of Population: Statistical Dataset Population (SDP) estimates for England and Wales, for selected years.* ONS Administrative Data Census Research Outputs. Downloadable at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation

ONS (2017b). *Households and Families.* ONS Administrative Data Census Research Outputs. Downloadable at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/householdsandfamilies

ONS (2017c). *Population characteristics.* ONS Administrative Data Census Research Outputs. Downloadable at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics

ONS (2017d). *Annual assessment of ONS's progress towards an Administrative Data Census post-2021*. ONS Administrative Data Census annual assessments. Downloadable at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments

ANNEX H

ITALY CASE STUDY

Reference frame

1. Starting from 2018, the population Census in Italy will abandon the traditional decennial 'door-to-door' enumeration for a 'combined' approach which will integrate administrative data and sample surveys. In fact, in 2012, the so-called 'permanent' Census of Population and Housing (in Italian Censimento permanente della popolazione e delle abitazioni) was introduced in Italian legislation (Article 3 of Legislative Decree 179/2012, converted with amendments into Law 221/2012). The goal of the 'permanent' Census is to produce annual data - replacing the previous decennial cycle - using information from administrative sources integrated with sample surveys information. This will be done within the frame of Istat's (the Italian National Institute of Statistics) modernization program, whose focus is to integrate administrative data, create statistical registers and conduct supporting statistical surveys, in line with the new organizational, technological and methodological model aimed at fully exploiting data already available. The new Census strategy will allow a significant reduction of the cost of the census, of respondents' burden, and of the organizational impact on municipalities (that had traditionally been responsible for the census field-work).

2. The traditional Census in Italy typically reported a significant difference between the usually resident population and the individuals recorded in the local population registers. According to the Italian law, at the end of every census, the differences between the enumeration count and the population registers should be analysed by municipal officers and population registers should be updated and revised on the basis of the Census outcomes.

3. Despite the strong efforts made during the 2011 Census to increase the coverage, thanks also to the use of different modes of data collection, some sub-populations remained very difficult to count. Based on the 2011 post-enumeration survey, which covered more than 320,000 individuals, Istat estimated that about 650,000 usual residents in Italy had not been counted and that about 80 per cent of them had foreign citizenship (Istat, 2016). These results, together with the increasingly less sustainable cost of conducting the traditional enumeration, led Istat to resort to multiple data sources and yearly surveys to conduct the census (Crescenzi and Sindoni, 2015).

4. The Italian public administration is currently completing its digitalization process to realize a centralized population register, the so-called 'Anagrafe Nazionale della Popolazione Residente'. This centralized register will have a unique ID number which will provide a link to those other administrative sources that are of particular importance in the creation of a register-based census.

5. From the beginning of 2011, Istat has been collecting individual and household data from municipal population registers on a yearly basis. These data were used to distribute the 2011 Census forms to households (for the first time a mail-out system was used instead of enumerator delivery). Municipal population registers are also the sampling frame of social surveys (such as the Labour Force, EU-SILC and Household Budget surveys) and variables such as place of residence, age, gender and citizenship are used for stratification and for treatment of non-response.

6. Moreover, by using the 2011 population census microdata and adding vital events (births, deaths, internal and international migrations), Istat has been computing and managing, at

municipal level, a yearly statistical population, the so-called 'ANagrafe VIrtuale Statistica' (ANVIS). This statistical population ensures a higher quality level of data than the population registers microdata and represents a more solid frame to implement a register-based Census.

7. Since 2015, Istat has initiated several projects to explore the use of administrative sources for statistical purposes. To manage the increasing number of administrative data sets and to maximize the benefit, Istat has built an integrated system of available administrative sources, called ISM (Integrated System of Microdata). This system identifies each object (an individual person or economic unit) in the administrative data sources and gives it a permanent ID number. ISM manages and links social and economic variables of individuals, households, and economic units to the place where people perform their activities or spend their time (Di Bella and Ambroselli, 2014). By using a syntax to define the rules of extraction, the reference time, the statistical domains or the subsets of variables, the ISM IDs make it possible to construct data structures for statistical processes and to create a thematic database.

8. The administrative sources already stored and integrated in the ISM are the following:

- population data from the municipal registers
- ANVIS
- foreign population with permit to stay
- employees and self-employed workers
- compulsory education students and university students
- retired people
- non-pension benefits records
- individual data on income and taxation.

9. There is also a hierarchy of the sources, in that the signals coming from the 'activity registers', related to labour and education, could be considered as robust signals of presence in a territory[36]. So, the comparison at the micro level between ANVIS and individuals recorded in administrative sources with labour or study activities could contribute to the determination of the usually-resident population at the municipal level[37]. This last database may represent the second frame to support the register-based Census in Italy.

10. The final decision on which basic data structure should be used for the 'permanent' Census will depend on a quality assessment both of the sources and of the processes used for deriving all statistical outputs. For this reason, Istat implemented, in 2017, a pilot survey to evaluate the coverage level of the main data source that should be used for the new census methodology in Italy.

11. Other than population counts by sex, citizenship, age and place of births, the population Census should provide a set of hypercubes currently required by Eurostat on socio-economic variables (employment status, educational level, migrant status, etc.). For this purpose, and with the aim of introducing the changes required for the integration of European social surveys (EUROSTAT, 2013), Istat is implementing the integration of the 'permanent' Census with the Social Surveys Integrated System.

[36] These signals may be determined on a monthly basis and could be aligned to the concept and the definition of 'usual-residence' of the European Union Regulation No. 1260/2013.

[37] However, in these data, some hard-to-count sub-populations, such as the homeless and other people who do not have usual residence in the same place during the year, are not included.

The Census and the Social Surveys Integrated System

12. This section presents a possible scenario for the integration of social surveys whose purpose is to achieve a complete integration within the system of social surveys and ensure the maximum integration with the registers system available at Istat.

13. The Census and the Social Surveys Integrated System (CSSIS) is a complex statistical process exploiting and integrating the information derived from registers and collected in surveys on socio-economic variables. It is designed as a two-phase design based on a Master Sample (MS) and on a set of balanced and coordinated sample surveys. It is planned for supporting the Population Register (PR) in order to increase the amount of statistical information provided and to improve the level of coverage and quality.

14. The PR is the backbone of the system for the production of social statistics, with a row for each target unit i.e. a *usually resident person* (whether living in private or institutional households). For each target unit, the core information (taken from demographic sources) is extended to all the basic social variables (obtained from administrative sources and/or social surveys) among which employment status and health conditions.

15. For an optimal design of the CSSIS, it is useful to classify the variables included in the PR as *totally*, *partially* or *not replaceable*. The first category encompasses those variables for which the administrative sources provide the corresponding *proxy* information and which, at the end of the statistical process - including editing and imputation for partial non-response, are considered to be *complete* (because they are available for all units in the PR), and *accurate* (i.e. having a good level of coverage and quality). For instance, sex and age are variables which are known for all the individuals in the PR and, therefore, they are considered totally replaceable variables.

16. Administrative sources also provide the corresponding *proxy* information for *partially replaceable* variables, which are considered complete and accurate only for a sub-set of the target population. For the remaining population, these variables are either unknown or cannot be considered accurate because of the failure of the synthetic model of imputation. For instance, this is the case for the ‘employed’ variable, which is completely replaceable only with respect to the sub-population of the ‘regularly employed’.

17. Finally, for *not replaceable* variables the corresponding proxy information coming from administrative registers is not directly available. For these variables, target parameters can be estimated by means of sample surveys and exploiting the auxiliary information coming from the PR, that is the set of variables contained in the register which are supposed to be predictive for the non-replaceable variables under study. The set of estimates should meet the requirements of:

> (a) *reliability* obtained by means of an approximately design-unbiased estimator, or by a model-based method in which the model used is plausible in some sense. In both cases the coefficients of variation of the estimates should be kept lower than a chosen threshold; and
> (b) *consistency* in that the data obtained by combining estimates in different ways must produce the same results.

18. The main function of the CSSIS is filling the gaps in information of the PR for the estimation of those target parameters referred to above as partially replaceable and not

replaceable variables on socio-economic data. To this aim the MS is designed for exploiting together (*pooling*) in an efficient way all the common information (target and auxiliary variables) observed by the different sample surveys belonging to the system. Furthermore, the MS estimation strategy uses all the complete auxiliary information of the PR. This strategy should be able to produce more efficient direct estimates than the estimates produced by adopting a separate estimation strategy for each survey. Within this context, the harmonization of the common variables – the *core structural* variables (which are the target variables for all surveys) and the *harmonized* variables (which are target variables for more than one survey but not all) - and the harmonization of the statistical production processes are crucial issues.

19. With regard to the basic objectives of the 'permanent' census, the first phase of the MS design is based on two different component samples, namely A and L.

20. The component A sample - based on a sample of Enumeration Areas (EA) or of addresses selected from an Integrated Address File (IAF) - is designed to satisfy the needs of estimating under-coverage (S_U) and over-coverage (S_O) rates of the PR both at national and local level for different sub-population profiles such as several different combinations of sex, age and nationality. These rates should be applied to the PR for obtaining weighted population counts corrected for coverage errors. The estimated population counts are obtained using the Extended Dual System Estimator (EDSE), i.e. taking into account both under-coverage and over-coverage.

21. The component L sample - based on a list of households - is designed with the purpose of: (T_I) thematic integration that is estimating the hypercubes which cannot be obtained using the replaceable information coming from registers. Furthermore, in order to pool the information coming from the two components, component L could be planned to provide reliable information on spatial variability of over-coverage indicators (S_O) of the PR. On the other hand, the component A sample could be designed to also meet the thematic integration target (T_I). In turn, the component L sample could also be modified to improve the estimation process with the aim of estimating via indirect sampling some aspects of under-coverage (S_U).

22. More generally, the first phase survey (i.e. the two components - A and L) should be focused on the following aims:

(a1) obtaining sampled information on partially replaceable and not replaceable variables useful for the PR;

(b1) establishing a first contact with the sampled households, a sub-sample of which will be re-interviewed the following year in the second phase. The first contact could be managed in order to reduce the second-phase potential non-response;

(c1) obtaining updated contact information on telephone numbers and e-mail addresses, which is not available in the sampling frame, that may facilitate less expensive interview techniques (such as CAWI or CATI) in the second phase.

23. From the first phase sample a set of negatively coordinated samples of households can be selected for the second phase surveys, aiming:

(a2) to provide information on harmonized and specific socio-economic variables currently observed by Labour Force (LFS), EU-SILC and Household Budget (HBS) surveys;

(b2) to confirm the common structural variables already surveyed in the first phase interviews and to make consistent the variables which are common to the above mentioned social surveys. These surveys are currently based on stratified two-stage sampling designs (municipalities-households), and each survey is planned, selected and realized separately. For these reasons, it may be that significant differences are observed among the estimates related to

the same variables derived from the different social surveys even if the definitions and the wording of the related questions are the same. In this case, in order to be able to pool, in the future, the same information coming from the different surveys, a strategic issue will be to improve harmonization between the social surveys. As a matter of fact, one of the main purposes of the system described above is to reduce potential systematic differences among the surveys via harmonization of survey designs.

24. Furthermore, the first phase sample (MS) can be stratified or balanced, using variables in the PR, to identify those areal or structural sub-populations that pose coverage problems or are subjected to structural and characteristic changes in the short-term period. Similarly, the second phase sample can be balanced on the set of harmonized specific variables, and to observe directly variables correlated with target variables (such as the self-declared employment condition) to be used in the estimation process.

25. The component L of the first phase sample should be based on a yearly sample size of about 2,000 municipalities out of 8,100 and around 300,000 households. The first-phase sample size should be at least large enough to cover the 140,000 households sample size needed for the second phase.

26. An overview of CSSIS, based on the two-phase MS design is shown in Figure 1. The administrative records mainly support the development of the Census Population Frame (CPF) from which the component L is selected. This integrates the PR with other sources related to labour and educational archives, and tax returns. A further goal is the correction of individual addresses in order to obtain the correct geographic population. These corrections are made on individual records and, therefore, all sources of information have to be linked by using a unique identification code. In the CPF, records are needed to be associated to their dwelling unit, via the centroid of their building. The component L is selected from CPF and the Final Sampling Units (FSU) are households belonging to the CPF.

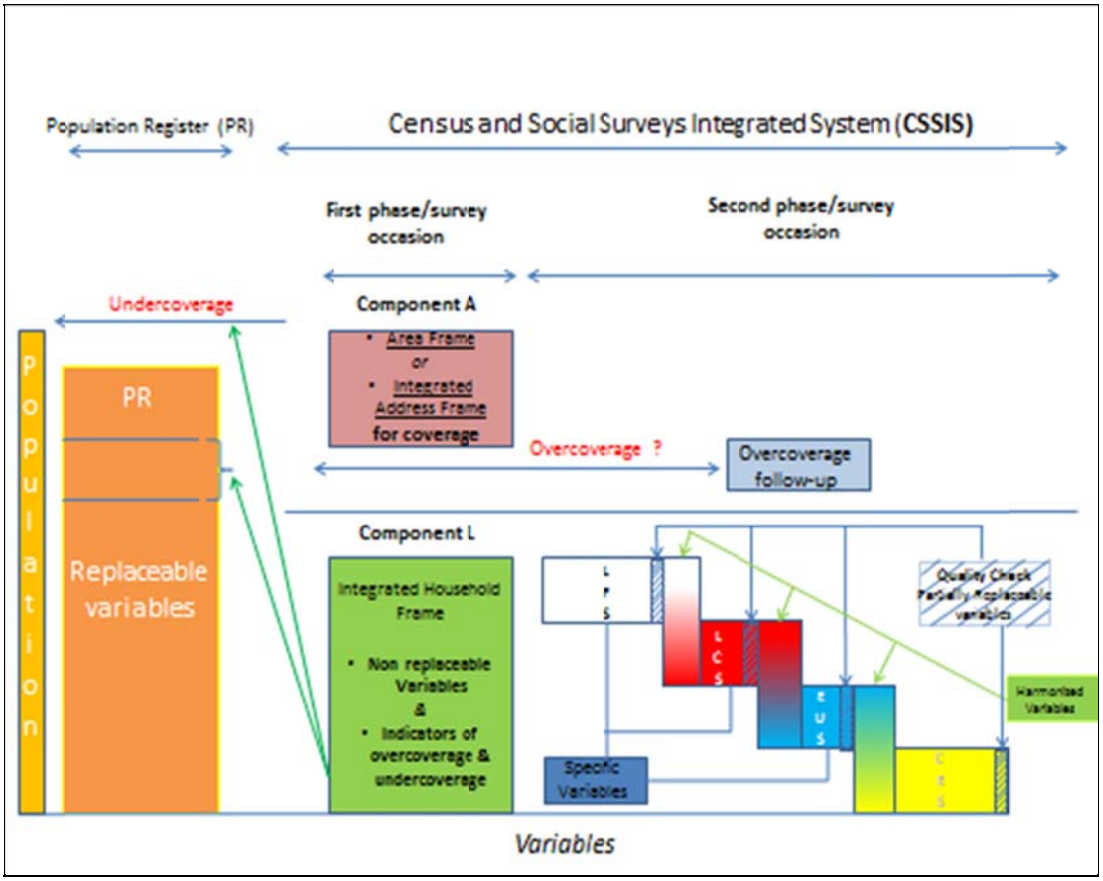
27. The component A is based on a sample design, in which the FSUs are census EAs or the addresses of an Integrated Addresses Frame (IAF). The IAF is obtained by integrating the addresses belonging to the CPF with addresses related to new buildings.

28. The main difference between the component L, using households as FSUs, and the component A, using addresses as FSUs, is that the latter must be ‘blind’ with respect to the information and the units belonging to the CPR. In this way the necessary conditions underlying the EDSE are completely satisfied.

29. Referring to similar international initiatives, designs analogous to the Italian model of the general Master Sample design for social surveys have been proposed by Eurostat when considering a modular approach for the design of integrated social surveys. Furthermore, the Australian Bureau of Statistics (ABS) is designing an integrated system of surveys very similar to the model that has been described here. In their case, however, this surveys’ system, called Australian Population Survey, does not replace the census.

30. Furthermore, the Italian model with two components supporting the register-based census is similar to the methodology that the UK’s Office for National Statistics (ONS) has been researching for its proposed Administrative Data Census to be introduced possibly in 2023 after the 2021 Census (ONS, 2017). In particular, in 2021 the ONS will conduct a traditional census and, at the same time, will carry out a parallel census based on the construction of an integrated

Figure 1 – An overview of the CSSIS.



population register using several administrative sources and two surveys with characteristics similar to those of the components L and A of the Italian Master Sample. It is worthwhile to mention that every year since 2015 and until 2023 the ONS will produce an assessment to evaluate how much they are away from the future model (see Annex G for more details).

31. The Israeli rolling integrated census represents another model that also shows similarities with what is planned in Italy. The Central Bureau of Statistics uses an integrated register which is adjusted based on weights computed by means of an EDSE (Pfeffermann, 2015).

32. In order to define the final survey design of the CSSIS, Istat conducted a pilot survey in 2017 in order to test the quality of the IAF, to evaluate the first and the second phase response rates and to fine tune the operational aspects of the survey.

References

Crescenzi, F and Sindoni, G (2015). *The combined use of multiple data sources in the population census*. Paper presented at the UNECE Group of Expert on Population and Housing Censuses. Geneva, 30 September-2 October 2017, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2015/mtg1/CES_GE.41_2015_7-Istat_rev.pdf.

Di Bella, G and Ambroselli, S (2014). *Towards a more efficient system of administrative data management and quality evaluation to support statistics production in ISTAT*. Paper presented at the European Conference on Quality in Official Statistics Q2014. Vienna, 2-5 June 2014.

EUROSTAT (2013). D12. (2013). Roadmap for the integration of European social surveys, http://ec.europa.eu/eurostat/cros/sites/crosportal/files/D12_Roadmap.pdf.

Istat (2016). *Atti del 15° Censimento generale della popolazione e delle abitazioni. 6 La valutazione della qualità,* http://www.istat.it/it/files/2016/07/AttiPOP_Fascicolo-6-web.pdf

ONS (2017). *Annual assessment of ONS's progress towards an Administrative Data Census post-2021*.
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments/annualassessmentofonssprogresstowardsanadministrativedatacensuspost2021

Pfeffermann, D (2015). Methodological Issues and Challenges in the Production of Official Statistics, *Journal of Survey Statistics and Methodology*, 3: 425–483.

ANNEX I

GERMANY CASE STUDY

Introduction

1. Introducing a combined census model in Germany was not straightforward for several reasons. Firstly, fewer suitable registers are available than in other countries that are adopting register-based approaches. Secondly, strict data protection regulations that were established in the context of the last traditional German census in 1987 make it challenging to find solutions to link registers from different areas. Due to these regulations neither a person ID nor a dwelling ID have been introduced so far, which makes any linkage between registers a burdensome undertaking.

2. At the same time, a traditional census based on interviewer-administered data collection is not popular among stakeholders due to the sheer size of the cost. This is even a more critical factor as, given the federal structure of official statistics in Germany, the costs have to be born jointly by the federal government and the governments of the 16 regions (Länder), which used to be an issue of long debate in the past.

3. Since 1983, the traditional census data collection has also enjoyed only limited popularity among the respondents. The last traditional census in Germany finally took place in 1987, but only after some protracted delay. As a consequence of this special context in Germany, a combined census model was developed, tested in a large-scale test in 2001 and finally implemented in 2011.

Legal and institutional background

4. The creation of the combined model for the German census can only be understood against the background of the difficult implementation of the last traditional census, which was stopped by the German constitutional court only few weeks prior to its implementation in 1983 (for a short history of censuses in Germany before the 1980s see Scholz and Kreyenfeld, 2016). It was implemented in modified form in 1987. After the controversial discussions of the census during the period 1983-1987, the Federal Government was reluctant to engage in a traditional census again. So, instead of carrying out a full census in the 2001 census round, a large-scale census test was conducted to assess the viability of a register-assisted approach, that combined data obtained from registers with a number of primary data collections.

5. The new model had to comply with the judgement of the German constitutional court that was delivered on the occasion of the planned census 1983. This judgement has, since then, had a major impact on data protection regulation in Germany. It stated that the right of informational self-determination directly follows from the fundamental right of personal freedom, guaranteed by article 2 of the constitution. Any data collection required from the public therefore is only considered constitutional if justified by a legal basis, which needs to be specific and clear as well as commensurate compared to the public interest at stake. While data for administrative purposes may only be collected for specific, well justified and commensurate purposes, collection for official statistics, given its specific role, is allowed for a certain stock of information that can be used for multiple purposes. Consequently, data collected for statistical purposes must be used for statistical purposes only and under no circumstances can be transferred to other public bodies ("Rückspielverbot"). More generally, any matching of registers that allows for an index of people was rejected as unconstitutional. Similarly, the constitutional court stressed that a universal personal ID number, which would facilitate the

matching of data held in different registers, was also regarded as unconstitutional (Bundesverfassungsgericht, 1983).

The combined census model and its implementation in 2011

6. The basic idea of the combined census model in Germany was to use the data in the fields of demography and employment from the available administrative registers (such as the population registers maintained by the municipalities, and the employment statistics register of the Federal Employment Agency). Together with a complete enumeration of buildings and dwellings (as no sufficient register information was available on such units) and a supplementary sample survey (for variables on persons not available from registers), a 'census-typical' data set was to be constructed. The 2001 census test revealed that it was also necessary to use the supplementary sample survey to correct for the errors detected in the registers (see Statistische Ämter des Bundes und der Länder, 2004).

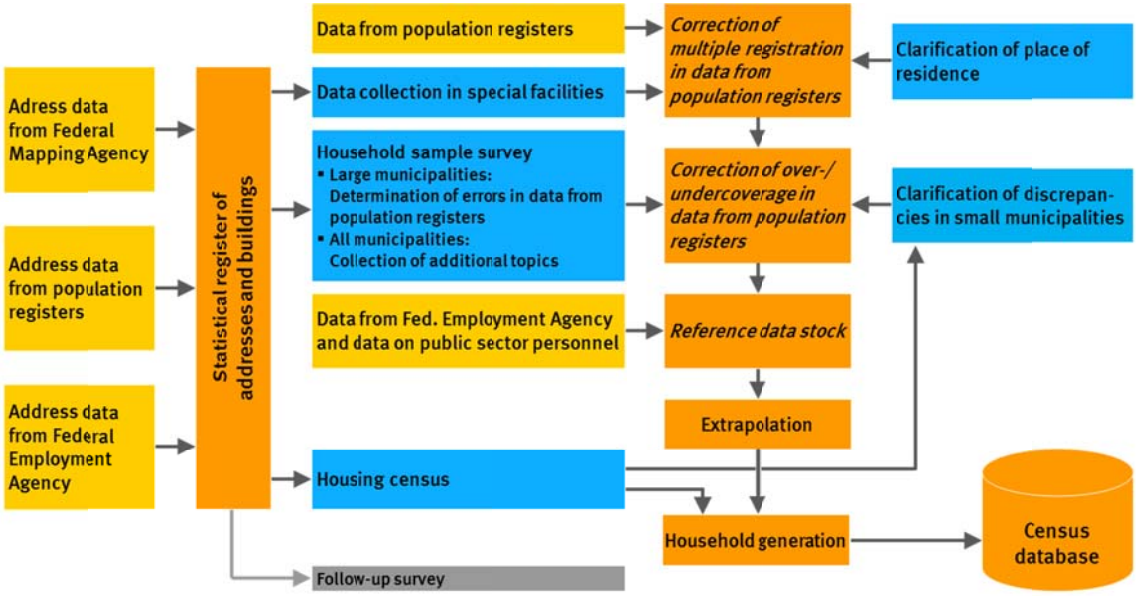
7. The primary aim of the 2011 census was to use the demographic information available from the decentralised population registers and to complete – and where necessary, correct – this data by merging it with information from other registers and mandatory primary surveys. (The following discussion is based on papers presented by Bechtold in 2013 and 2016; a more detailed presentation of the methodology is given by Statistische Ämter des Bundes und der Länder, 2015). By combining different data sources and methods of automatic data generation, a distinct data set containing all required census information could be created for each person, each household and each building with dwellings.

8. In order to merge the data of the different parts of the census data collection, first a basic register was established, containing a list of all addresses where buildings with residential space existed at the census reference day. This address and building register was the key link for all data collections during the census. It was also used as the statistical population for the sampling procedure of private households and for the housing census.

9. The main data sources used in the combined model (illustrated in Figure 1) were the following:

- The *population registers* provided the main demographic data as well as information on family relationships for all individuals that belong to the target population (about 86 million data records). The data from the municipal population registers were collected at the census reference day (9 May 2011) and were updated three months later in order to take into account delayed register entries and delayed de-registrations. The register data were merged in a nationwide data set, and it was subsequently tested to determine if people were registered at more than one sole or main place of residence on the census reference day. If such cases were identified in large municipalities (with at least 10,000 inhabitants), they were automatically corrected by using the most current information. Multiple residences in small municipalities (with less than 10,000 inhabitants) were investigated using a postal inquiry. The same applied to cases where a person was registered at a secondary place of residence only.

Figure 1: The German Census Model in 2011



- The *supplementary household sample survey*, covering almost 10 per cent of the population was used to adjust the register data in municipalities with 10,000 inhabitants or more, after the registers had been corrected for multiple residences. For the calculation of the population of large municipalities, the level of error of the population registers (over- and under-coverage) detected by the household survey was taken into account. The sample was designed to ensure that the population figures of large municipalities met a 1 per cent error margin target at a 95 per cent confidence level. The method applied to optimise the sampling process was dedicated individually to each municipality and the sample size ranged between 2.1 per cent and 45.6 per cent and differed significantly even for municipalities of a similar size. For municipalities with fewer than 10,000 inhabitants, a survey was carried out among those households that had been identified as needing clarification by combining and analysing register information and information of the housing survey.

  In addition to the objective to establish the population figures, the supplementary household survey was also used to cover further census variables required by virtue of EU regulation that were not available from registers (relating in particular, to labour market participation and educational attainment). The additional census topics were collected in all municipalities (not just those with 10,000 inhabitants or more). The sample size was designed to allow publication at the NUTS-3 level.

- For persons living in *special facilities* - such as a communal accommodation, care institutions, dormitory or similar types of housing - census information was collected using a complete enumeration, because fluctuation and missing registrations for this sub-population in the population registers lead to high rates of error. Addresses that were considered to be sensitive or potentially stigmatising – relating, for example, to psychiatric hospitals or prisons (and referred to as ‘confidential special facilities’) -

were distinguished from non-confidential special facilities, such as student dormitories. In confidential special facilities, the privacy of data collection was secured by a special procedure and only a reduced set of variables was collected.

- As there are no registers of buildings and dwellings covering the whole of Germany, the compulsory EU variables of the housing census needed to be obtained through a postal *survey of buildings and dwellings* that was conducted among all property owners (for the total of just under 20 million buildings with residential space, data were collected at approximately 19 million owners). In addition, the census of buildings and housing covered auxiliary variables (number of persons living in a dwelling and names of two persons) which were used in the household generation procedure (see below).

- A large part of information on the employment of the population was taken from *registers of the Federal Employment Agency* (for about 36 million employees subject to social insurance contributions) and from the administrative files of the public service agencies with personnel (for about 3 million public officials, judges and soldiers). These registers were similarly used to supplement the demographic information obtained from the population registers, the household sample survey and the survey of addresses with non-confidential special facilities. Together with the register of addresses and buildings, this information constituted the reference data stock.

- To obtain information about household and family structures and their housing conditions (such information is not included in any register) data from the various census components had to be combined in a so-called *household generating procedure*. In this multistage procedure, information about persons from the population registers, the household sample survey and the survey conducted at special facilities was used to form households and to link them to dwellings collected in the housing census.

10. Merging data sets from different sources for individual persons was one of the great challenges of the 2011 census, because it had to be accomplished without an existing unique personal identification number available in the different registers. An already existing set of ID numbers for the purpose of the tax authorities was available in some of the registers, but could not be used due to legal restrictions. Therefore, individual and address-based information such as name, sex, date of birth, municipal code, post code, street name, and house number were used to link respective records of different data sets.

Lessons learned and modifications for the 2021 census

11. Retrospectively, the modular concept of the German 2011 census, combining the use of registers and primary statistical data collections, was successful. A qualitative evaluation came to the result that the model worked well, and (while identifying proposals for an improved implementation) should be the basis for the census in 2021. At the same time, it was concluded that the quality of the population registers was still not considered solely sufficient for the purposes of a census so that the supplementary household sample survey would still be necessary as an element to correct for over- and under-coverage in the registers (and to collect the data for variables not covered by the registers). Thus, in considering the results from 2011 census, the household sample survey remains an integral elemental part for the optimization of the census model. The model was scientifically accepted and achieved a high precision. At the same time, the model reduced considerably the burden on the respondents and the cost of data collection on the statistical offices compared to the former traditional census. The household generating procedure was another new element which turned out to be an acceptably accurate way to determine family and household data at the local area level.

12. Changes for the 2021 census therefore considered the 2011 census experiences in quality aspects and potential to reduce complexity and thus enable more timely results. Additionally, modifications aim to make the results easier to understand and raise general acceptance of the model (among both users and the public alike). The main changes proposed for the 2021 census design (as illustrated in Figure 2) are:

- The interaction of the different census components need to be designed early on to allow a comprehensive technical approach that integrates the individual parts of the model together. The results of the different surveys and census components will therefore be linked in a central data stock instead of storing them separately as in 2011. In doing so, data can be cross-checked and validated at an early stage of data processing. Inconsistencies and implausibilities can be removed by rules or even by manual checks. This will help both to improve data quality and to reduce efforts to link the data with each other consecutively.

- The use of paper questionnaires will be reduced by a rigorous 'online-first' strategy. This is an important component in the further effort to reduce costs, improve timeliness and minimize response burden by guiding respondents more easily through the questionnaire.

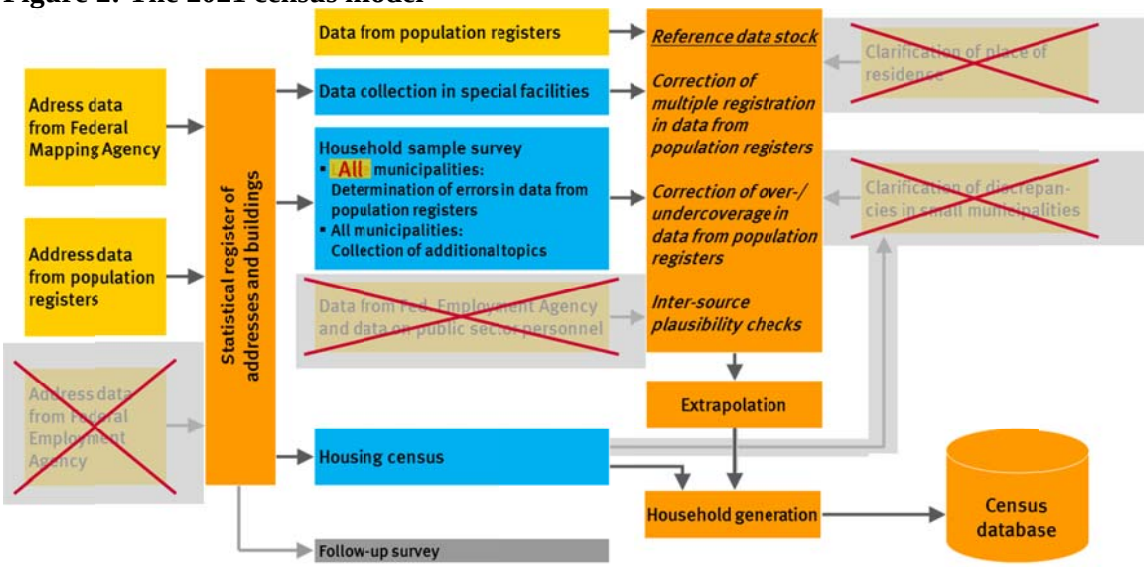
- Building up the address register has to start earlier, and one of the data sources will no longer be used. In 2011, three main sources were leveraged to collect addresses: data from the Population Registers, the Federal Mapping Agency, and the registers of the Federal Employment Agency. The latter will not be used in 2021 as there have been no further addresses added to this source since 2011, while many cross-checks were necessary due to different spellings of cities, streets and house numbers.

- In 2011, data of the Federal Employment Agency were, furthermore, also used to generate data on employment. These data were of high quality, but users complained about the complexity of analyses, since different employment figures were released depending on whether they were based on the combined model or the household survey only. Looking at employment in a broader sense, this source had to be analyzed in combination with the household survey to cover self-employed or unemployed as well. Since this additional complexity turned out to confuse many users, the use of the registers of the Federal Employment Agency will be discontinued in 2021.

- The different models of correcting the number of people registered in big cities (with 10,000 inhabitants or more) and small cities was criticized by many municipalities. Even though the different models produced results with the quality expected from the beginning, the use of such different models created a barrier to user acceptance. Furthermore, with the knowledge of hindsight, the model of correcting population registers in small municipalities has to be optimized.

- The weighting scheme of the supplementary household survey was targeted primarily at a highly precise number of inhabitants. The production of results for census variables that were not available from registers was only considered as a secondary priority in the development of the estimators. The weighting procedure needs to be optimized in order to minimize any risk of bias in case of the census variables not available from the registers.

Figure 2: The 2021 census model



References

Bechtold S. (2013): *Lessons learned from a mixed-mode census for the future of social statistics*. Paper presented at the 97th DGINS Conference on 'New conceptual design for household and social statistics', Wiesbaden, 27 September 2011. Available at https://www.destatis.de/EN/AboutUs/Events/DGINS/Document_Destatis_MixedModeCensus.pdf?__blob=publicationFile (accessed 14 July 2017)

Bechtold S. (2016): 'The 2011 Census Model in Germany'. In: *Comparative Population Studies* D1-D9 (date of release: 04.08.2016). Available at DOI: 10.12765/CPoS-2016-07en (accessed 14 July 2017)

Bundesverfassungsgericht (1983): Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983. (BVerfG 65, 1) Available at https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Volkszaehlungsurteil_1983.pdf?__blob=publicationFile&v=9 (accessed 4 July 2017)

Scholz, Rembrandt D and Kreyenfeld, Michaela (2016): 'The register-based census in Germany: historical context and relevance for population research'. In: *Comparative Population Studies - Zeitschrift für Bevölkerungswissenschaft* 41 (2016), 2, pp. 175-204. Available at: DOI: http://dx.doi.org/10.12765/CPoS-2016-08en (accessed 14 July 2017)

Statistische Ämter des Bundes und der Länder (2004): *Results of the census test*. Statistisches Bundesamt: Wiesbaden.

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011. Methoden und Verfahren. Statistisches Bundesamt: Wiesbaden. Available at: https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=6 (accessed 14 July 2017)

GLOSSARY OF TERMS, DEFINITIONS AND ACRONYMS

ABS: Australian Bureau of Statistics

Accessibility: A measure of data quality relating to the conditions and modalities by which users can obtain, use and interpret the data.

Accuracy: A measure of data quality relating to the closeness of estimates to the unknown 'true' values.

Activity register: A register that holds information about residents' different activities that indicate a presence in the country or area. Such activities can include information on, for example, employment or other economic status, receipt of benefits or pensions, or student status.

ACSR: Register of alternative civilian service (Austria)

Administrative data: Data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.

AMB: Analytical microdata base (Poland)

Anonymization: The process of protecting the confidentiality of personal information by removing all unique identifiers from the unit records.

ANVIS: ANagrafe VIrtuale Statistica - yearly statistical population (Italy)

BDR: Buildings and dwellings register (Austria)

bPIN: Branch-specific personal identification number (Austria)

bPIN OS: Branch-specific personal identification number for Official Statistics (Austria)

BR: Business Register of Enterprises and their Local Units (Austria)

CAII: Computer assisted Internet interview

CAPI: Computer assisted personal interview

CAR: Child allowance register (Austria)

CATI: Computer assisted telephone interview

CAWI: Computer assisted web interview

CAxI: Computer assisted multi-mode interview (Poland)

CBS: Central Bureau of Statistics (for example, in Israel)

CES: Conference of European Statisticians

CMR: Central metadata repository (Poland)

CPF: Census population frame (Italy)

Coherence: A measure of data quality relating to the degree to which the census data can be combined in different ways and for various purposes with statistical information from other sources.

Comparability: A measure of data quality relating to the degree to which statistics are comparable between geographic areas and over time.

Combined census: A census in which some information on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes, but where other information that is not available from such sources is collected directly from individual persons and households by means of full or partial field enumeration or from other sample surveys.

COR: Register of car owners (Austria)

CPR: Central population register (for example, in Austria and Slovenia)

CR: Conscription register (Austria); Civil register (Portugal)

CSO: Central Statistical Office (for example, in Ireland and Poland)

CSSIS: Census and Social Surveys Integrated System (Italy)

CSSR: Central Social Security Register (Austria)

CT: Census Test

De facto census: A census based on a count of persons at where they were present on the reference date.

De jure census: A census based on a count of persons at their place of usual residence on the reference date.

Deterministic method: A method without a random component that thus always leads to the same outcome

DPA: Data Protection Authority (Austria)

DSE: Duel System Estimation - a statistical method, based on a capture-recapture technique, applied to estimate the size of a population.

DSP: Department of Social Protection (Ireland)

DWP: Department of Work and Pensions (England and Wales)

EAR: Register of Educational Attainment (Austria)

EDSE: Extended Dual System Estimator (Italy)

EFTA: European Free Trade Association

EHIS: The Information System of Education (Estonia).

ESS: European Statistical System

EU: European Union

Eurostat: Statistical Office of the European Union

EU-SILC: European Union statistics of income and living conditions

FAO: Food and Agriculture Organization of the United Nations

FSU: Final sampling units (Italy)

GIS: Geographic information system

Gmina: One of 2,478 basic units of administrative division within Poland (equivalent to a municipality)

HMRC: Her Majesty's Revenue and Customs (England and Wales)

HR: Household Register (Slovenia)

Hypercube: A high-dimensional statistical tabulation of, typically, four or more dimensions

IAF: Integrated addresses frame (Italy)

ID: Identification

IR: Immigration Register (Portugal)

ISM: Integrated system of microdata (Italy)

ISS: IT census system (Poland)

ISTAT: Italian National Institute of Statistics

IT: Information technology

LA: Local authority (England and Wales)

LAU: Local administrative unit. The classification of local administrative areas used by Eurostat. Levels 1 and 2 equate to those areas that were previously classified, respectively at the NUTS 4 and 5 levels.

LFS: Labour Force Survey

Metadata: Information about the content, structure, quality and other relevant characteristics of a register

Microdata: As used in these Guidelines, information in a register relating to a single entity or entities

MS: Master Sample (Italy)

NDI: National data infrastructure (Ireland)

NHS: National Health Service (England and Wales)

NIC: Civil Register number (Portugal)

NIF: Taxes Register number (Portugal)

NISS: Social Security number (Portugal)

NSI: National Statistics Institute (or Office)

NSS: National Statistical System (Portugal)

Numerical address: Code linking to the address

OMB: Operational microdata base (Poland)

ONS: Office for National Statistics (England and Wales)

PAR: Person Activity Register (Ireland)

PAYE: Pay as you earn (England and Wales)

PCS: Population Coverage Survey (England and Wales)

PE: Population Estimate (Portugal)

PIK: Protected identifier key (Ireland)

PIN: Personal Identification Number

PHC: Population and Housing Census

Population register: A register of residents of the country

PPSN: Personal public service number (Ireland)

PR: Population register

Probabilistic method: A method with a random component that thus not always leads to the same outcome.

Process quality: The quality of a statistical process as evaluated by the methods used, cost effectiveness and response burden.

PSR: Register of Enrolled Pupils and Students (Austria)

Punctuality: A measure of data quality relating to the delay between the date of the release of the results and the target date (the date by which the data should have been delivered).

Register: A systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up-to-date, correcting, or extending, the register, that is, keeping track of any changes in the data describing the units and their attributes.

Register-based census: A census in which the data on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes. No information is collected directly from individual persons or households.

RER: Real Estate Register (Slovenia)

Relevance: A measure of data quality relating to the degree to which statistics meet current and potential needs of the users.

RSP: Registers of public servants of the federal state and the Länder (Austria)

SDD: Statistical dwellings dataset (Portugal)

SID: Statistical identifier (Slovenia)

SOL: Signs of life

SP: Statistics Portugal

SPD: Statistical population dataset

Statistical register: A register created for statistical purposes. They are typically created by transforming data from registers or other administrative data sources.

SURS: Statistical Office of the Republic of Slovenia

SWR: Register of social welfare recipients (Austria)

Synchronisation: The transmission of data recorded on handheld devices in the field to a central server.

Timeliness: A measure of data quality relating to the period between the availability of the information and the event or phenomenon it describes.

TR: Tax Register (Austria)

Traditional census: A census based on the direct count of all individuals and the collection of information on their characteristics through the completion of either a self-completion or interview-based questionnaires, either in a paper or electronic format.

UNECE: United Nations Economic Commission for Europe

Unique ID/key: A single alpha numeric identifier that relates a characteristic or variable to a particular entity (person, household or dwelling) across a range of different registers or administrative data sources.

UPRN: Unique Property Reference Number (Great Britain)

UR: Unemployment Register (Austria)

Voivodship: One of 16 administrative areas in Poland (equivalent to a province or region)

XML: Extensible markup language. In computing a language that that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

XSD: Extended Data Services (Estonia)

* * * * *