



Economic and Social Council

Distr.:
28 June 2017
Draft

Original: English
English and Russian only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Nineteenth Meeting

Geneva, 4–6 October 2017

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

Note by the Task Force on Register-Based and Combined Censuses¹

Summary

As a follow-up to the in-depth review of diversification of population census methodology and sources in October 2015, the Bureau asked the Secretariat to set up a Task Force on Register-Based and Combined Censuses. In this note the work of this Task Force is described: the new guidelines on the use of registers for population and housing censuses. **The terms of reference for the Task Force on register-based and combined censuses can be found in Annex A.**

¹ Prepared by Eric Schulte Nordholt, Chair of the Task Force on Register-Based and Combined Censuses.

I. Introduction

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*. Nine Task Forces were working together with the Steering Group on the Recommendations.

2. The CES adopted the Recommendations for the 2020 census round in June 2015. The Recommendations are available in electronic format on the UNECE website² and were printed in English, French and Russian.

3. In October 2015, the CES Bureau conducted an in-depth review of diversification of population census methodology 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).

4. This note presents the proposed new guidelines produced by the Task Force on Register-Based and Combined Censuses. To prevent confusion in the further work of the Task Force some of the basic concepts have been worked out first. After the presentation of the preliminary report in Geneva in September 2016, a start has been made with the work on the complete draft of the new guidelines on the use of registers and administrative data for population and housing censuses. Before the guidelines are presented, first information is given in the box below on census methods and their evolution over time in the UNECE region.

Census methods and their evolution over time in the UNECE region

There are many different ways to conduct a population and housing census. For the sake of simplicity, this document refers to three main categories of census methods: the traditional census, the register-based census, and the combined census.

The **traditional census** is intended here as a census based on the direct count of all individuals and the collection of their characteristics through the completion of census questionnaires, in paper form or electronically. The information is collected in the field across the whole country in a relatively short period of time, normally lasting a few weeks. When paper forms are used, they can be filled in directly by the households (with delivery and collection assured by enumerators, the post or other methods), or by the enumerators during an interview of the household.

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 vast temporary work force for the field data collection (enumerators, supervisors, etc.), and to print, distribute, and process a huge number of forms. Moreover, in most countries there are increasing difficulties to enumerate certain population groups, particularly those characterized by high mobility and multiple residences, and an increasing reluctance of the respondents to be enumerated, for

² <http://www.unece.org/publications/2020recomm.html>

various reasons. Finally, the traditional census is normally conducted every 10 years (for its cost and complexity) and the results are often available a relatively long time after data collection, while many users would like to have timelier and frequently updated information.

Some countries addressed the disadvantages of the traditional census using sampling (most households fill only a short form with basic information, while a sample fills a more detailed long form, reducing the amount of information collected and processed) or promoting the internet response, which may result in 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” developed in France.

A totally different approach from the traditional census is the **register-based census** that was developed by the Nordic countries in the 1970s. 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 coming from various registers (population register, building/address register, social security, etc.) through a matching process, normally making use of personal identification numbers. This approach permits the production of census data at a limited cost and with relatively limited work, once a good quality system of statistical registers has been set up.

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

In the 2000 census round only few countries in the UNECE region³ conducted a register-based or combined census (3 and 5 countries respectively⁴) and the traditional census was still by far the most popular approach in the region⁴ (40 countries). In the 2010 round, there was a significant increase in the number of countries conducting a register-based census (from 3 to 9) or a combined census (from 5 to 10), and a corresponding decrease of the traditional census (from 40 to 34 countries).

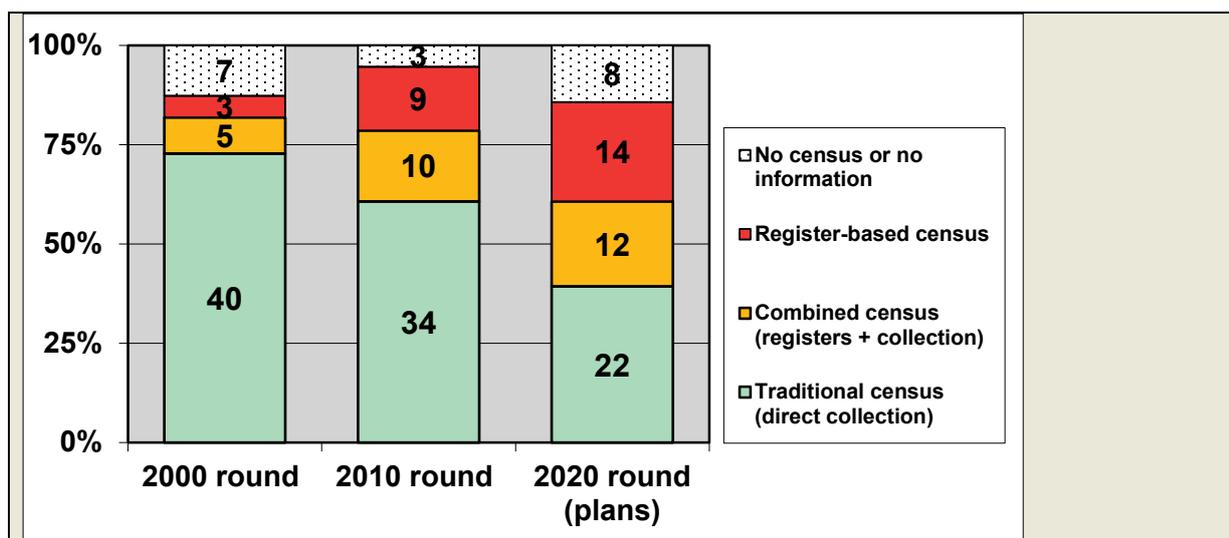
Figure 1
Number of UNECE countries by census method in the 2000-2020 Census rounds⁵

³ The UNECE region includes countries in Europe, North America, Central Asia, plus Turkey and Israel.

⁴ 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

⁵ In 2006 Montenegro became independent and the number of UNECE countries increased from 55 to 56.



Source: Valente, 2015.

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%), 12 countries a combined census (25%) and 22 countries a traditional census (46%). 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%), 9 countries a combined census (28%), and 10 countries a traditional census (31%).

II. Scope of the new UNECE Guidelines and definitions of register-based and combined censuses

5. 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 the Recommendations for the 2020 census round. In these new UNECE Guidelines different kinds of registers (on persons and buildings) are mentioned with a focus on those used in censuses.

6. For some of the definitions we can refer to those presented in the publication “Register based statistics in the Nordic countries”⁶. On page 15 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, i.e. keeping track of any changes in the data describing the units and their attributes. *Administrative data* sources 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.

7. In some countries the term ‘administrative data’ is used as a synonym of register data. In other countries a distinction is made. Administrative data then also include administrative sources other than registers.

8. A register-based population and housing 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 page 26 of the 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 by data from already existing microdata from sample surveys or full enumeration. All register-based census countries have in common that no census questionnaires are used to collect information about the population. Therefore, register-based censuses are in general much cheaper than combined censuses and especially than traditional censuses.

9. In a combined census statistics are created by using registers and other administrative sources, together with information from either sample field data for selected variables or full field enumeration for selected variables (see pages 12 and 24 of the Recommendations for the 2020 census round).

10. The remainder of the guidelines are organised as follows. Section III describes the essential features of a census and how these may be met by register-based or combined censuses. Section IV describes a number of elements that need to be taken into account when planning to transition from a traditional census to a register-based or combined census. In Section V we outline a common framework that describes the process of conducting these non-traditional censuses. Sections VI to IX provide more details on the processes and methods associated to each aspect of the framework and the role that quality assurance plays at each stage. The final section X presents case studies from a number of countries that have

⁶ http://www.unecce.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

transitioned or plan to transition from a traditional full-enumeration census to a register-based or combined census.

III. Essential features of population and housing census

11. The essential features of a population and housing census have already been defined by the International Conference of Statisticians in 1853 in Brussels and following these conditions warrant the comparability and high quality of census data across countries and time. Nowadays, these five essential features have been redefined and highlighted by the CES⁷ and the aim of them is to guarantee the consonance of census data gathered in different countries having different levels of technical development and also different cultures and grasps of life. Following all these features – regardless the concrete methodology of data collection – it is necessary to get the population data from the whole world having the quality that allows making decisions and forecasts on the population development of the world.

12. The five essential features of a census are:

- Individual enumeration;
- Simultaneity;
- Universality (within a precisely defined territory of a country);
- Small area data; and
- Defined periodicity.

13. We now describe each of these features and discuss how register-based and combined population and housing censuses can be designed to meet them.

III.1. Individual enumeration

14. Individual enumeration is an important feature for censuses. Usually, it has been realised by using individual questionnaires for each person. In the case of register-based censuses a different approach is adopted. If the data are taken from a register, it is important that each census unit has a special, uniquely identified, record in the register used. Then the register is a useful source for the census. In the case of a combined census only some of the variables are taken from administrative data sources and then the same approach as for a register-based census is used for those variables taken from a register.

15. Instead of one identifying variable there sometimes exists a group of variables that identify the units uniquely. In this case it is necessary to define a new identifier for statistical purposes using this group of identifying variables and carefully check its quality (errors and uniqueness).

16. 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 case 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.

⁷ Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing (United Nations, 2015), see <http://www.unecce.org/publications/2020recomm.html>, paragraphs 23-28.

17. The units of population and housing census are not only persons, but also (occupied) dwellings, households and families. All of those units need identification, but there is no need to use five different identification variables. The minimal necessary identification variables are the one for persons (person ID) and the one for dwellings (dwelling ID). These IDs must be linked with each other (for each person his dwelling ID has been fixed and for each occupied dwelling the list of person IDs of people living in it must be given. The dwelling ID is making use of the address code, which may contain also spatial coordinates.

18. Information about households is usually collected on the basis of the housekeeping concept⁸ 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 household. Whilst 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 relations.

19. In some countries (e.g. Slovenia, see Annex F) a household register exists. The existence of such a register eases the organisation of a register-based census, especially in case household IDs are included in this register. Then we have 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 household definition in use for statistical outputs i.e. a housekeeping definition, can be continued (see Annex B). As part of the work to understand the impact of transitioning to an Administrative Data Census, the Office for National Statistics (ONS) in the United Kingdom is currently exploring the potential impact of changing from a housekeeping concept to a household-dwelling concept with users of census data.

20. Sometimes it is useful to use identification codes for other units such as enterprises and organisations. If they are linked with person IDs and dwelling IDs they form a helpful tool for making statistics (e.g. on commuting between place of residence and place of work) and also for producing output of a register-based or combined census.

III.2. Simultaneity

21. The fixed census moment (day and hour) is the condition assuring the simultaneity of the census data. Traditionally, to ensure this condition, the census has been carried out during a very short time, often during one day only. This essential feature should be respected also in the case of a register-based or combined census.

22. If registers in use are permanently in the process of updating, then it is necessary to fix the census moment and to take the data from all registers referring to this moment. Sometimes the registers are updated regularly for some special date, e.g. the beginning of a year, and then it is plausible to use this date as census moment. Then the simultaneity of the census is guaranteed.

⁸ See page 162 of <http://www.unecce.org/publications/2020recomm.html>

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

24. When several registers are used in the census, it is important that all data taken from these registers have the same reference moment. Usually, census variables derived via special algorithms take some calculation time; hence those census variables are only ready for publishing sometime after the census moment. For some specific variables in combined or register-based censuses different reference dates are used on purpose to give a better picture of society. Demographic data can normally be taken from population registers at the beginning of the year. However, labour data could better be taken somewhat earlier or later as around Christmas and New Year's Eve less people have labour contracts. For some registers it is impossible to have Census Day as reference day. For example education registers often have relevant education data as reference days that may differ from a chosen Census Day. Then choosing a reference day close to Census Day for the education data to be used in the census may be a good compromise.

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

25. To warrant universality of the traditional census the questionnaires used in the enumeration process are the same for all members of the population. If there are questionnaires in different languages, it is important to check if their content and also meaning of all questions is exactly the same.

26. If the administrative registers used in the census are common for the whole country and all population groups, the condition is fulfilled automatically for both the register-based census and the combined census.

27. If there are different administrative data in different areas or for different population groups (e.g. an urban population register and a rural population register or different administrative data in different cities), then it is necessary to analyse the possible discrepancies between the different administrative sources and find a way to define the common unique census variables using these different administrative sources. In this case plausible results can be derived from these newly defined (via a special algorithm) variables in a statistical register.

III.4. Small area data

28. Getting correct information from small areas (either in geographical or social sense) is a very important task for censuses of all kinds, as there are no other ways how to investigate such data from small areas.

29. Small area data are also important in the case of register-based or combined censuses. Existence of small area census data follows from the high coverage of the administrative data, preferably covering the whole population accurately. If there are some small areas that are poorly covered, then the administrative data lack information and it is necessary to improve the administrative data before they can be used as a source in a register-based or combined census. If poor coverage is also a problem for the every-day usage of the administrative data, it should be improved anyway. Improving the statistical register can sometimes be done by adding information from another source.

30. Sometimes there might be special administrative data for some small areas (small groups of people). Then it will be necessary to combine these different administrative sources (see

III.2.). If the result is satisfactory, this combination is useable. In the case of a combined census it is also possible to complement the information from administrative sources by a survey that uses different methodologies in different areas, e.g. conduct telephone interviews, use mail or internet, etc. In case a census sample survey is used, problems can arise with small area data. Preferably, user needs regarding the required level of detail of the census publications are taken into account before a decision is made on the sample size of the census survey.

III.5. Defined periodicity

31. Nowadays the censuses are organised worldwide having a ten years cycle. The United Nations request that countries conduct at least one census every ten years (between 2015 and 2024 for the 2020 census round). For European Union requirements, member countries had to conduct a census in 2011 and will have to conduct the next one in 2021. However, some countries (e.g. Australia, Canada, Ireland, New Zealand, and Slovenia) have shorter periods between censuses. The same cycle should be kept in all censuses regardless the methodology used, even in the case of register-based censuses where other census cycle lengths could easily be introduced. In the case of a register-based census it is relatively easy to organise censuses more often. If a 5-year period has been used, then one of the two census years of the country coincides with the census year fixed by Eurostat.

32. An advantage of a register-based census is the opportunity to conduct the censuses more often than the usual 10-years cycle, as the register data are permanently available and it is advisable to prepare also census software in such a way that it is permanently ready for using at any moment. Then the time difference between censuses might be 10 years, 5 years, 2 years or censuses could be conducted annually. Yearly updates are a serious object of investigation in European countries applying different census methodologies.

33. It is also possible to conduct some census updates with a shortened list of variables. This could also save resources. From here it follows that also in countries where a regular ten yearly census will be done using combined methodology, some updates with a shortened list of variables can easily be done more often if the variables on the shortened list can be taken from registers.

III.6. Conclusions of features defined by the CES

34. From the above-mentioned it can be concluded that if in the country a good system of administrative registers exists that is consistent, well-useable and of high quality (that means, all units are uniquely identified using a common identifier), then it is possible to organise the population and housing census meeting all required CES features.

35. If the list of administrative registers of needed quality does not cover the whole range of census variables, it is still possible and useful to use those variables in a census using combined methodology or in a register-based census with a shortened list of variables in between regular censuses.

IV. Considerations when transitioning from a traditional census to a register-based or combined census

36. The decision to move to a combined or register-based census is motivated because of the many 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. The transition may give rise to some difficulties and disadvantages. These can be significant obstacles for some countries.

IV.1. Advantages

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

IV.1.1. Smaller costs per inhabitant

38. Traditional censuses are very expensive. In many countries conducting a traditional census it is common that the census costs the equivalent of about two yearly budgets of the statistical institute. It is understandable that governments put pressure on the institutes to cut these expenses, especially when other sources are available.

39. If a combined census is conducted with full field enumeration for some of the variables the costs saved by shortening the census questionnaires are partly lost again in combining the information from the field with register information. While the savings may be modest, the approach may still be preferable, e.g. in the case of a transformation process when more data from administrative sources are planned to be used in future censuses.

40. 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% compared to a traditional census⁹.

41. 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 98% of the costs of a traditional census can be saved this way¹⁰.

42. Some countries have switched in one census cycle from a traditional census to a register-based census. Then all costs for the change have been taken in 10 years. Normally, the move is done in several steps (often via an intermediate stage of a combined census). The costs for the change are then being spread out over 2 or 3 census cycles.

43. Moving towards combined censuses and, especially, towards register-based censuses contributes to working more efficiently while using lower budgets. It is clear that budgetary considerations of governments promote such moves, even if registers are lacking or of insufficient quality to use as sources for the census. In such cases, it should be made clear from the beginning that a country cannot move and has to continue conducting traditional censuses. However, even when countries continue conducting traditional censuses, innovations making greater use of administrative data could help the National Statistical Institute (NSI) work more efficiently. It helps if the relevant public authorities make administration data sources available for the statistical institutes to produce register-based statistics. The government can help a great deal both by removing legal barriers and by subsidising the transition.

IV.1.2. Quicker to conduct

⁹ 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

¹⁰ 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

44. 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. This is due to the fact that no field enumeration has to be conducted. The shorter production time does not hold for combined censuses as they also include fieldwork. Without fieldwork the census work may maintain the delivery time but start much later and that has of course advantages. However, one has to realise that the first time a register-based census is conducted it may cost more time than in later census rounds as census planning has to be set up anew.

IV.1.3. Fewer problems with nonresponse and reduced response burden on the population

45. Fewer problems with nonresponse can be expected in the case where only, or mostly, administrative data sources are used. With growing nonresponse rates in censuses and surveys this is going to be a more and more important aspect.

46. If a country moves from a traditional census to a combined census the response burden on the population will be lower. This is especially evident if only sample field information is collected. A transition to a register-based census implies that the response burden reduces to zero and that there are no longer any survey nonresponse problems.

IV.1.4. Possibility of a permanent census

47. 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 new results could be published every day. Such real time censuses may be something for the future, but more regular census results are critical to keep the attention of the stakeholders. As information from other sources becomes real-time information, censuses will be expected to keep pace. As a by-effect it is easier to keep the knowledge and IT-infrastructure up-to-date if annual census updates are conducted.

IV.1.5. Better cooperation of units within the NSI

48. In some NSIs the different units look like isolated silos without too many contacts with other silos. If a NSI moves to a register-based statistical system that cannot hold. By moving from a survey-based statistical system organisation towards a register-based NSI the traditional 1 to 1 relationship between sources and statistics is replaced by a m to n relationship. A better cooperation of units within the NSI thus becomes vital. Moreover, by better integrating statistics the coherence in the statistical framework is improved. The statistical registers become the key driver of coherence in national statistical systems.

IV.1.6. More time and money remains for innovations

49. Innovations are crucial for the long term perspectives of a NSI. As introducing registers in the process of producing official statistics saves time and money, it is easier to innovate. The saved time and money could be used to stimulate innovations so that the work processes remain up-to-date.

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

50. If all data are stored appropriately, not only can the regular statistics be produced on time, new statistics also become easily possible. A NSI can then become more flexible and responsive to new information needs and enlarge its value for society.

IV.1.8. New output, data and services can be made available

51. By combining the saved time and money and larger flexibility it is clear that new output, data and services can be made available. Although that may not be an aim in itself, it can lead to a larger user satisfaction.

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

52. If a country wants to move to a combined or register-based census a number of conditions are relevant before information from administrative registers (and other sources) can successfully be integrated to create the underlying statistical register. A number of these necessary conditions are discussed below.

IV.2.1. Legal base

53. To be able to conduct a combined or register-based census there must be legal provisions that support the access and protection of the administrative data. The legal base is normally found in a Statistics or Census Act. Whatever census a NSI conducts, it should be within the legal framework.

54. The NSI must have legal authority to access the relevant administrative data sources, preferably including personal identifiers. Ideally, that access should be free of charge for data from other public authorities. To avoid legislative battles, 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.

55. The NSI must have a legal obligation to protect the confidentiality of the administrative data it obtains, i.e. adhere to the “one-way traffic”, except under specific circumstances mentioned in the legislation.

56. In certain countries, legal requirements may constrain of how a census can be conducted¹¹. In some countries the NSI has first started exploring administrative data sources and, thereafter, found a legal base to make register-based statistics possible. In other countries first the legal base has been arranged and, thereafter, register-based statistics were produced and published. To gain experience, it is often simpler to start with register-based statistics covering fewer topics than the census.

57. It always holds that moving to a census methodology where administrative data sources play a role needs careful preparation including pilot studies. One has to realise that when (part of) the fieldwork for the census stops, reinitiating this fieldwork is rather difficult. After some time the knowledge how to conduct a traditional census disappears, especially in the case of a register-based census where no census fieldwork is left.

58. 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

¹¹ Particularly where representation in the national legislature depends on census results.

59. While the law might give a legal licence to a combined or register-based census, public approval is also required. This might be more difficult to obtain than the legal base. In some countries people may get the impression that with a traditional census ‘big brother is watching you’. In some other countries using and combining administrative data sources may be seen as even more intrusive.

60. In a traditional census privacy concerns may lead to lower response rates or giving wrong answers on purpose. It is getting more and more difficult to correct for such unit and item nonresponse. So, on the one hand the population could prefer a situation where fewer questions are asked as the information is already available. On the other hand, part of the population may prefer to answer census questionnaires instead of seeing that their information is taken and combined from several administrative sources.

61. In a register-based or combined census people may feel uneasy or even object that 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 automatically be less aware of the fact that a census is conducted. However, this should not be misinterpreted as public approval.

62. It is desirable 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 in the census questionnaires and which variables are derived from administrative data sources.

IV.2.3. Stakeholder approval

63. Stakeholders typically want to have at least the same kind and detail of information as in the last census. This is not always possible when the census methodology changes.

64. It is important to inform and consult stakeholders beforehand. Stakeholders can become critical if expectations are not met. Even well-informed stakeholders could become very critical when they get the feeling that they are going to lose information that they had got in the previous census. However, disappointed stakeholders cannot always be prevented and disappointment among some of the stakeholders is sometimes unavoidable when adopting a new census methodology.

65. It is important to have a communication strategy for stakeholder engagement. A communication strategy can have 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 them 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.

66. Openness and identifying new opportunities and benefits for the stakeholders will help to keep their approval.

67. Change of census methodology that will provide statistics having financial consequences (e.g. transfers of money towards municipalities) should be communicated with the utmost care.

IV.2.4. Cooperation between the NSI and other authorities

68. Good cooperation between the NSI and other (mainly governmental) authorities is vital in using registers in the census. The NSI needs to know when microdata are available and needs accompanying metadata before register-based statistics can be produced. In a combined census and even more in a register-based census the NSI is heavily dependent on register holders and their willingness and capability to provide good quality data on time. If register holders fail to deliver, the NSI is held responsible for the failure to publish statistics on time.

69. It is vital to inform register holders how important their data are for the NSI and how their data are used. Additionally to a legal base (see 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, data sources from non-governmental authorities could also be used in the census, but it brings about other privacy and commercial considerations.

IV.2.5. Comprehensive and reliable register system

70. A comprehensive and reliable statistical register system is essential to conduct a combined or register-based census. Population registers and administrative data sources are not normally set up for the purpose of a census. A transformation process is undertaken to result in a reliable statistical register system (see section VIII).

71. To safeguard register-based statistics it is important to have good relations with administrative data holders. Conditional on the legal base, and if the administrative bodies are open to it, in some countries there is also a potential to improve the relations between the public bodies and the NSI by introducing new or extending existing register-based statistics, such as longitudinal studies to evaluate policy implementation. Of course there are contacts between the NSI and the administrative data holder when an administrative data source is introduced as a new source. However, permanent contact, e.g. facilitated via account managers, is vital to keep the administrative data holders aware of the importance their data plays and to keep the NSI informed about changes in the microdata and metadata they receive. Only with enduring contacts between the NSI and the administrative data holders can register-based statistics become a prolonged success.

72. The more users the 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 to base reliable census tables on?

IV.2.6. Unified identification system

73. A unified identification system across different administrative data sources greatly facilitates register-based censuses. It is preferable to have unique ID-numbers. 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

74. When a country wants to move from a traditional census towards a combined or register-based census, building up a wide-ranging knowledge on the administrative data sources available is important to prevent disappointments. Although building up knowledge can be done en route, the efforts needed to make this process successful should not be underestimated. Lots of lessons about failures and successes can be learnt from countries conducting combined or register-based censuses, but the national context should never be forgotten. It is never possible to simply copy the methodology of one country to another country in 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

75. If there are planned moves to a different census methodology, it is good practice to be transparent and share information on plans and tests as much as possible. As discussed in IV.2.3, it is important to inform users of moves towards register-based censuses as such moves may have an impact on output availability. Transparency and openness facilitates external review and feed-back for the new processes.

IV.3. Difficulties that may arise

76. 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

77. After having moved to a combined or register-based census a statistical office becomes heavily dependent on the relevant public authorities. One has to realise that statistics is not their priority. For the statistical office failures in administrative registers may cause failures in their statistics. However, the statistical office remains responsible for the quality of the official statistics.

IV.3.2. Differences in concepts and definitions

78. Registers and other administrative data often have different concepts and definitions than traditional censuses. One has to realise that such differences exist and decide whether these differences are acceptable when moving from a traditional to a combined or register-based census. It is an open question with different answers in the context of different countries whether one is 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 by derivations from different sources or by editing information from newly achieved census sources. However, this is certainly not always the case and, even if this is the case, it is normally only an approximation (artificial construction).

79. In practice acceptability of differences has to be weighed against costs. In case differences are accepted, breaks in time series have to be accepted. In case differences are not accepted, higher costs have to be accepted. Finally, one has to realise that full coverage of population may not be available on all administrative sources.

IV.3.3. Timeliness of administrative registers

80. The public authorities may have other priorities that delay the delivery of the relevant microdata and metadata. This can cause issues for the statistical office regarding the timeliness of their register-based statistics. The timeliness of delivery of different data source varies.

IV.3.4. Different reference periods in different sources

81. A problem that one encounters when moving towards combined and register-based censuses is that different sources often have different reference dates. Sometimes a source gives the option to distinguish clearly between reference dates and dates of events, but this good practice does not always apply. If these problems cannot be handled properly, the risk is that not all sources can be harmonised towards the same reference day. Then the question arises what is an acceptable difference in reference dates. The answer to this question is dependent on the variable concerned. Some variables are rather stable over time and then a small difference in reference date is not a problem. Large differences in reference dates are always unwanted. Finally, one has to acknowledge that discrepancies in reference dates also arise in census questionnaires as respondents do not always give the answers specific to the census reference date.

IV.3.5. Privacy and security concerns

82. Using registers leads to privacy and security concerns. These concerns often relate to linking different sources. In some countries the legal framework has been adapted and it seems there is public approval to use administrative data in official statistics. For other countries that is not (yet) the case or even looks like a Utopia.

IV.3.6. Difficulty of publishing small subpopulations

83. In a census it is important to ensure universality. Nevertheless, in case sample surveys are used for some of the census variables small subpopulations are sometimes difficult or even impossible to publish because of limited information. The range and detail of outputs may be even more limited in a register-based census than from a traditional census or a combined census with full field enumeration for variables that cannot be derived from administrative registers.

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

84. In cases where a country has an infrequent census timetable with large gaps between censuses, it is getting difficult to keep the knowledge and IT infrastructure up-to-date. However, when yearly census updates are introduced this difficulty disappears.

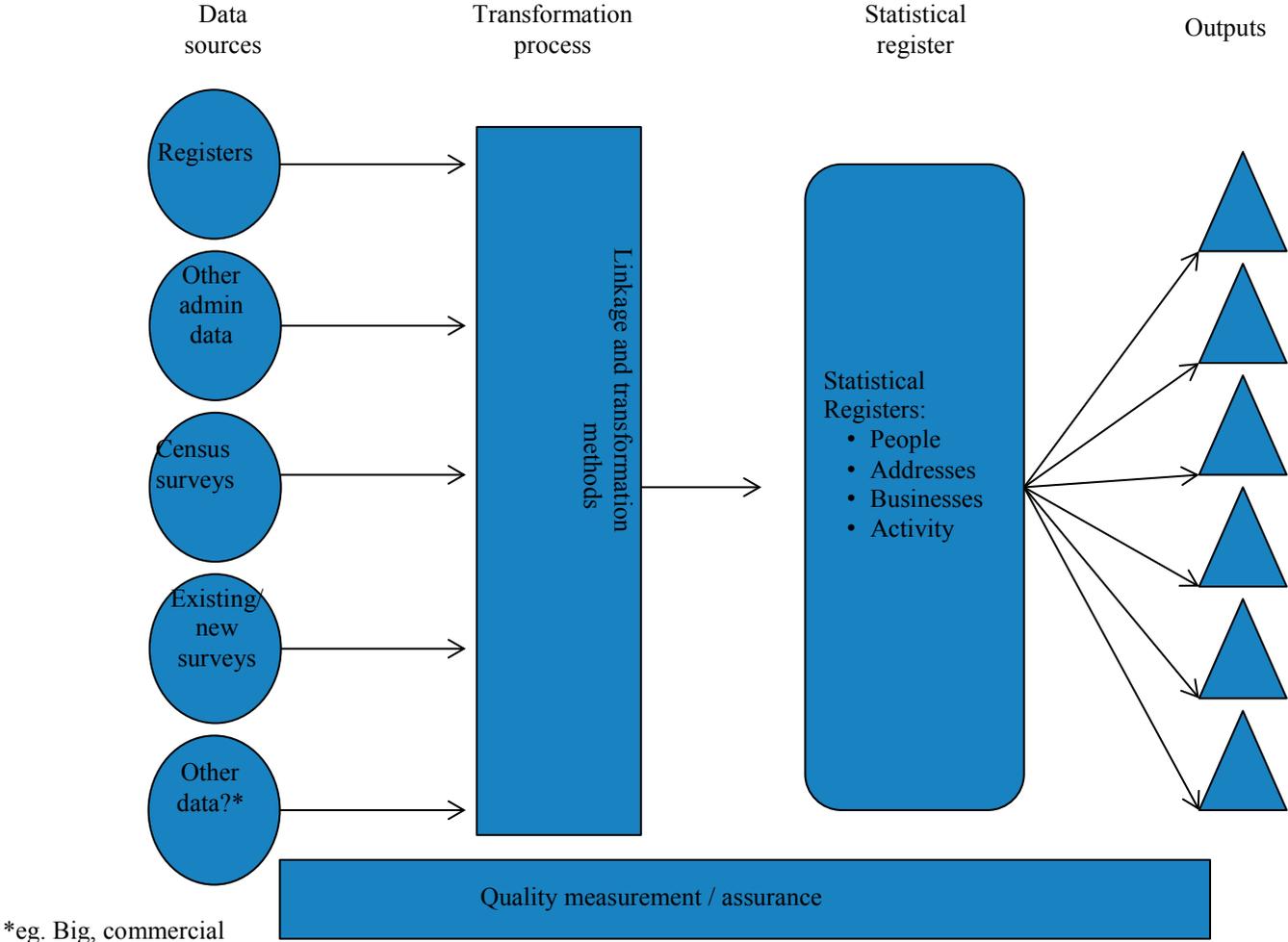
IV.3.8. Diminishing attention

85. In countries with census questionnaires people want to know the results of the census. In register-based countries, where people no longer fill in census forms, one sees a general decline in interest in census results. Many people are not aware that a census was conducted and therefore the automatic attention for the census no longer exists. In register-based census countries there is still a large interest in statistical information, but it becomes less relevant which sources were used to produce that information. The statistical office then has the task to find the best sources to produce the information required by society.

V. Common framework for register-based and combined censuses

86. As described in the introductory section 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 1.

Figure 1: common framework for register-based and combined censuses



87. This framework is divided into five key stages:
- a. Data sources;
 - b. Transformation process;
 - c. Constructing statistical registers;
 - d. Disseminating outputs;
 - e. Quality measurement/assurance.

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

V.1. Data sources

88. So far register-based censuses have been conducted based on population registers, but more recently countries that do not have national population registers 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.

89. For example, in its work to explore the potential to move to an Administrative Data Census (see Annex H), the Office for National Statistics (ONS) in the United Kingdom has created a Statistical Population Dataset (SPD)¹² from which to 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).

90. 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 possibility to integrate sources. The easiest way for linking different data sources is using unique identifiers for persons and for addresses. In some countries there are also other unique identifiers, e.g. for enterprises 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 constitute a barrier leading to the elimination 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 information occurring in the combined sources.

91. The range of sources that might be included is likely to differ for each country to meet the essential features outlined in section III. 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

92. As population registers and administrative data 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.

93. The first step is to link the different data sources that are being used. In countries where a unified identification system is available (for example, 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 relatively small.

94. This is more challenging for countries that do not have a unified identification system, or who are using administrative data sources without a unique identifier. However, sometimes this can be achieved by linking through a set of identifiers (e.g. name, sex, date of birth, numerical address) resulting in matches that can then be used to produce outputs similar to those produced through linking registers that do contain a unique identifier. Some methods are discussed in section 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 (i.e. links that are of good quality and thus not burdened by too many errors) are not possible, it is technically infeasible for a country to move to a combined or

¹² For the latest methodology, see <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20>

register-based census, but the administrative data may then be used for benchmarking purposes (see Annex B for an example in Ireland).

95. Once the data have been linked, further transformation may be required to create or improve the quality of, the statistical register. In section 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 one is data cleaning and editing. The second one is forming the census variables using special algorithms. Sometimes for 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’). This process is described in Wallgren and Wallgren (2014), and some of these processes are described in more detail in section VIII.

96. The ability to transform multiple administrative data sources into a statistical register is often a key challenge for NSIs without population registers. Statistical registers are useful for all NSIs conducting register-based censuses. 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. Statistical registers

97. A comprehensive and reliable system of statistical registers is essential 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 (e.g. 31 December), c) all changes during a specified period (e.g. calendar year) and d) the population that is present for a specified period (e.g. 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).

98. In the Estonian registers’ system there are also four registers, but the list of them is somewhat different. All the registers are linked with different identifiers (person ID, address ID, enterprise ID and farm ID):

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

99. Polish experience shows that it is strongly recommended that the address/dwelling register should include a geometry component for spatial location of each dwelling (building) with highest possible precision (x-y-coordinates). That will namely allow use of GIS technology supporting the main stages of census field operations (Poland conducts a combined census) and spatial analysis.

V.4. Disseminating outputs

100. An essential part of transitioning to an administrative data census is demonstrating the ability to produce a range of census-type statistics with associated measures of quality before the traditional census is suspended. Having an effective and compatible system of registers (including statistical registers) it is possible to disseminate census data periodically with shorter time intervals than 10 years. Yearly census like tables exist in a number of countries including e.g. the Netherlands.

V.5. Quality

101. Quality measurement and assurance can occur during all stages of this framework, and the extent and methods for conducting quality measurement and assurance processes will vary between countries.

102. In a broad sense, this 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 on the resulting statistics). Some countries may add extra quality measurement processes, for example the case study by England and Wales (see Annex H) describes an extension of this framework to include a Population Coverage Survey for coverage adjustment). Countries who 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 England and Wales case study).

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

VI. Data sources and their quality

104. A population and housing census allows for the use of various data sources, including administrative registers. The quality of data used in the process of compiling statistics strongly affects the quality of output statistical information, i.e. products. Therefore, it is necessary to prepare and implement a standard method of assessing the quality of administrative registers as potential sources. The assessment of administrative registers is also important as far as the integration of their resources is concerned. The quality of registers is usually difficult to measure due to its complexity and multidimensionality. Many factors adding up to quality are non-measurable. The methodology of assessing the quality of registers is based on numerous aspects, criteria and indicators. It may take the form of a checklist or a survey (in case there is no other standard to test against) and should cover the following areas:

- general information about the register;
- information about the quality of the register;
- information about the quality of data from the register;
- general assessment of the administrative register.

As regards the quality assessment of the register, the key aspects include: availability, clarity, usefulness, comparability, relevance, consistency and the cost of using data in the register. The

key aspects used to assess the quality of sets and data contained therein include accuracy and comparability.

105. New statistical methods based increasingly on the use of administrative register data and other administrative data sources also require new methods for quality assurance. 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.

106. First of all 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 has to be developed which are used to analyse each source. On the basis of these indicators the NSI can decide whether to use or not to use a specific source. Secondly, the quality of administrative data has effects on the quality of census results. Indicators on the data sources may be integrated in a quality framework that assesses the input (administrative sources), the process quality and the product quality of a register-based census. The ESS.VIP ADMIN¹³ project aims to provide appropriate methods for quality assurance and facilitate access.

107. A NSI should gather information about the administrative data source:

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

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

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

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

- Quality of sources (supplier, relevance, privacy and security, delivery, procedures)

Supplier: confidence and reliability of the data holder (e.g. punctuality of data delivery), characterization of contact and communication with the supplier (e.g. good/bad, periodic/sporadic)

¹³ 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.

Relevance: administrative purpose of the attribute in question (is there only a legal basis or additively self-interest in recording and maintaining?)

Privacy and security: kind of data transmission, data decryption

Delivery: supply agreement, legal obligation, interval of periodic data delivery (monthly/quarterly/annually/sporadically), date of delivery, costs

Procedures: data collection and maintenance

- Quality of metadata (clarity, data treatment)

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 (technical checks, accuracy, completeness, time-related dimension, clarity, integrability, comparability, unique keys)

Technical checks: technical input checks, data format checks (e.g. readability of the data file), compliance of the data to the metadata definition

Accuracy: degree of precision, degree of certainty for faithful data records (presence of implausible values, presence of untrustworthy objects)

Completeness: definition of register population (e.g. under- and overcoverage), missing values for variables

Time-related dimension: updating process carried out by the supplier, changes comprehensible (i.e. historicised data), cut-off date possible

Clarity: clear structure and content of data

Integrability: extent to which the data source is capable of undergoing integration or of being integrated in the statistical system (on the level of objects and variables)

Comparability: data definitions compatible to those of the NSI, aggregation level sufficient for the statistical purpose

Unique keys: existence of a unique key on micro level, if there are multiple sources the linkability should be guaranteed (general key)

111. A quality framework can typically be used to study different potential sources and to decide which sources are going to be used in the census. Tools for a systematic evaluation of the quality of administrative data sources have been developed for example in the Netherlands and in Austria.¹⁴

Quality framework in 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 developed a quality framework. This framework 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 the Census 2011 project.

Austria

¹⁴ 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

The Austrian framework for assessing the quality of administrative data was developed for the 2011 register-based census. In every stage of the data processing a quality-indicator is 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 quality-framework can be applied individually. The approach for the assessment of administrative data was inspired by work from other National Statistical Institutes (P. Daas, Ossen, Vis-Visschers, & Arends-Tóth, 2009) and relies on four quality-related hyperdimensions (Documentation, Pre-processing, External Sources and Imputations) which aim to measure the quality at three production stages (raw data, i.e. administrative data; the combined dataset, i.e. integration of registers; and the final dataset, i.e. after data editing and imputations).

At the **raw data level**, three hyperdimensions are 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 is monitored by the use of a questionnaire containing several open and scored questions. The open questions gather information of general interest, like the timeliness of data delivery. Scored questions measure for example “Data History” (Are changes over time stored? Is information for the census day available?), “Definitions” (Are data definitions comparable to those of the NSI?), “Administrative Purpose” (Is the topic relevant for the data source keeper? Is there a legal basis for the topic in the administrative data source?) and “Data Treatment” (How fast changes of a topic are recorded in the administrative data source? Are data verified by the data source keeper, for example by requesting documents or identity cards? Are there technical checks and consistency checks between attributes by the data source keeper?) **Pre-processing** describes the share of useless data records. Records which cannot be used are those without a unique key, with item no-information (item non-response) or records with values 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 is calculated (for example for citizenship from the Central Population Register, for citizenship from the Unemployment Register, etc.).

112. 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 administrative sources. Those tasks are covered by the ESSnet on quality of multisource statistics¹⁵, 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 the quality of output based on multiple sources and a methodology for reporting on the quality of output; and to produce recommendations for updating the ESS Standard and the ESS Handbook for Quality Reports. 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

¹⁵ 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.

candidates for its input. Further work in the ADMIN project will include the production of quality guidelines when administrative sources are integrated in 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, transformation

113. There is an important difference between the quality of linkage of different administrative data sources with and without unique keys. If there is a unique identifier in all or most of the records, the linkage is easy and the quality of the linkage is normally good (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.

114. Due to the lack of a unique identifier across administrative data sources, the ONS has spent a number of years developing and refining methods to overcome this challenge and progress 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')¹⁶.

115. Further work has been done to refine these methods, including the development of a statistical spine to help resolve multiple matches across three or more datasets¹⁷. The links to reports provided in this and the previous paragraph also describe what has been done to quality assure these developing linkage methods. The methods show promise, but further refinements are needed to obtain the high-quality linkage that is required to produce robust estimates of the population.

116. Process quality is not straightforward in its definition. It consists of:

- **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.
- **Low response burden**: The reporting burden is proportionate to the needs of the users and is not excessive for the respondents. The NSI monitors the response burden and sets targets for its reduction over time. Whenever possible, administrative data sources should be used.

¹⁶ 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>.

¹⁷ More detail about these methods can be found in <https://www.ons.gov.uk/census/censustransformation/programme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20>.

VIII. Statistical registers

117. Using integrated data from various sources in a population and housing census is an important matter. In countries with registers using unique identification numbers, it is possible to identify basic registers and supplementary registers. The key registers enabling a statistical description of observation units in the census include: population register, housing register, business register and activity register. Combining their informational resources makes it possible to describe the population according to accessible data. Integrating data in censuses may include not only data from administrative registers but also data from current research referring to a specific population and its features, results from the previous census and information accumulated in a statistical sampling frame. The integrated dataset from the aforementioned sources assumes the form of a statistical register, a database used in the further process of collecting and compiling data within a census.

118. As mentioned in paragraph 88, the ONS and the CSO Poland have produced a Statistical Population Dataset by linking administrative data sources. The ONS linked four administrative data sources: one containing a list of patients registered with a GP, one containing a list of people who have a National Insurance Number, one containing a list of students who are registered on a Higher Education course and one containing a list of 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 are part of 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 sources¹⁸. The SPD is then used to produce estimates about the size of the population. To understand the quality of these admin-based population estimates, comparisons are made with census and official population estimates. These comparisons show promising results¹⁹.

IX. Outputs

119. Census outputs are produced for a range of users and, as discussed in Chapter IV, stakeholder approval and public approval 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 facilitate longitudinal studies. However, there may be some changes to variable definitions. The impact of these changes on the statistical outputs must be investigated and explained.

120. Assessing the quality of the output of census data has always been a serious and important task. Different ways and methods exist 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 generates relevant information on the reliability of the new census results. Here some different approaches of quality assessing will be described. It is always useful if several methods are being used in quality assessing.

121. The quality of the output of traditional, register-based or combined censuses can be assessed in different ways, e.g. by the assessment of product quality, by coverage studies, in quality reports and through quality committees. In quality and confidentiality issues among

¹⁸ More details about the ONS methods for constructing a SPD can be found in <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/methodologyofstatisticalpopulationdatasetv20>.

¹⁹ See <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation>.

other things the question is whether the quality of the output can be measured. In case sample surveys are used minimal cell frequencies (or minimal CVs) have to be defined to guarantee that table estimates are accurate enough to publish. In case a country only relies on registers or complete enumeration confidentiality rules have to be defined to prevent disclosure of individual entities. In the census context quality reports (in which results can be compared with e.g. LFS output) one typically finds that differences regarding demographic variables are relatively small and differences regarding economic variables (e.g. current activity status) are relatively large. Given the currently common ten years cycle, sometimes quality reports on future censuses are outmoded when a new census is set up.

122. As mentioned in the previous section, the ONS has compared their admin-data based population estimates with census and official population estimates to understand more about their quality. These comparisons use a set of quality standards to help understand the quality compared to that seen in census population estimates and other official population estimates²⁰. In 2017, work is being developed to produce 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 a Census Coverage Survey)²¹. This would be supported by other quality assurance processes, for example demographic analysis and comparisons with other data sources.

IX.1. Product quality

123. The quality in official statistics can be described on the basis of product quality and process quality previously described. Based on product and process quality, a wide range of quality indicators has been developed at the level of the European Statistical System and the United Nations Statistical Commission and implemented in quality reporting. The next paragraphs describe product quality in this context.

124. Product quality consists of:

- **Relevance** refers to the degree to which statistics meet current and potential needs of the users. It refers to whether all statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.) reflect user needs. Approaches to assess relevance is to ask users directly in a survey (user needs, user satisfaction surveys), to install stakeholders and client feedback mechanisms or by analysing the uses of data.
- **Accuracy** refers to the closeness of estimates to the unknown true values. The major components are: i) assessment of coverage, ii) unit and item non response and dealing with missing values (imputations) and iii) sampling errors. Usually, handbooks on quality assessment provide a set of indicators measuring accuracy.
- **Timeliness** refers to the period between the availability of the information and the event or phenomenon it describes. A common indicator is the time lag between the end of the reference period and the date of first release respectively final results.

²⁰ 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>.

²¹ 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>.

Sometimes there is a trade-off between accuracy and timeliness with a preference for results as early as possible at the expense of accuracy.

- **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 regulations or previously agreed among partners). A common indicator is the punctuality of time schedule of effective publication.
- **Accessibility** and **clarity** refer to the conditions and modalities by which users can obtain, use and interpret data (interpretability). Which kind of publications are available, is there an access to databases or even microdata? Are appropriate metadata including information on the quality of data available, and the extent to which additional information is provided by the NSI (National Statistical Institute).
- **Comparability** refers to the measurement of the impact of differences in applied statistical concepts, measurement tools and procedures where statistics are compared between geographical areas, sectoral domains or over time. Statistics should be consistent internally, over time and comparable between regions and countries.
- **Coherence** refers to the adequacy of the data to be reliably combined in different ways and for various uses. 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 approaches, 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, and between survey statistics and national accounts.

125. The Austrian framework introduced earlier also describes how the quality of all attributes of interest for their register-based census is computed.

Austria

In the processing stage, the entire information from the registers is combined in 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 than 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 conflicting information on date of birth in two or more administrative data sources) then a new 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.

126. The quality of a register-based or combined census is judged by the quality of the output. The common demands of census output are fixed in quality guidelines that are common for all types of censuses.

127. Regarding the quality of census output, three levels have to be considered:

- Quality of single census variables, measured by all quality criteria;
- Quality of all hypercubes²², assessed using the quality measures of census variables (being its components);
- Quality of census population (in the sense of coverage).

IX.1.1. Quality of a single census variable

128. The quality of a single census variable depends

- on the coverage of the register variable used for defining this census variable;
- on the correctness of the register variable used for defining this census variable;
- on the adequacy and correctness of the software used to derive this census variable;
- on technical or human errors.

129. If contextual checks of the variable have been made via comparisons, then the quality characteristics to be measured are:

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

130. As usually the outliers can be considered in a similar way as 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.

IX.1.2. Quality of a census hypercube

131. The quality of a census hypercube depends

- on the quality of all census variables included into the hypercube;
- on the adequacy of cube-generating software.

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

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

That means, the number of missing values in a hypercube is, in general, larger than in single variables. 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 $m(i)$ are the quality marks of census variables included in the hypercube, k is the number of dimensions of the hypercube 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 be good if at least one of the $m(i)$ has a very low mark.

²² Hypercubes are high-dimensional census tables.

133. Defining quality standards it is reasonable to define the maximal ratio of missing values allowed for excellent, good and satisfactory quality of a census hypercube. The levels of single variables' quality and hypercubes' quality must be consistent, that is, if all variables have excellent quality, then also the hypercube has excellent quality. And in the opposite situation, a hypercube having some marginals of poor quality cannot have good or excellent quality.

134. The quality report must include also the **confidentiality report** that includes the confidentiality rules defined in the country and the methodology of data protection transformations with the aim of warranting the confidentiality. Possible problems (e.g. loss of accuracy due to confidentiality transformations) must be added.

IX.2. Coverage

135. Coverage has been one of the most important quality problems of censuses. In the case of a traditional census the most common problem is under-coverage. A reason for under-coverage is that nowadays people are mobile, they might have different living places and it is difficult to enumerate them. Another problem is that some people appreciate their privacy very highly and do not like to give their data via enumerators to an administration. Also there might be some reasons to hide from authorities. Hiding from authorities, for example in case of illegal immigration, might be the reason why people do not have records in administrative registers. Illegal immigration is also a reason for under-coverage of register-based and combined censuses.

136. In the case of a register-based census, as well as under-coverage, there also exists the problem of over-coverage of registers. If people have not officially declared their emigration, their records may be kept in registers and so cause over-coverage of the register and also in the register-based census. If registers are not of enough coverage quality, a possibility to avoid coverage errors is to create a residency index on the basis of multiple registers (signs of life) that should be recalculated yearly to define the population size more accurately. To check the coverage errors in the case of combined censuses there are several possibilities. One option is to use the methodology common for traditional censuses, e.g. organising a post-enumeration survey and estimating both over- and under-coverage statistically. Another way is using several registers (and so-called signs of life) to estimate the real amount of residents that also give information on the adequate census population and the under- and overcoverage of the census.

137. If there are several registers in the country, then it is possible to use them for assessing the census coverage. Define for all possible residents the sign of life as a binary variable (with values 0 and 1) connected with each register existing in the country. Using these signs of life as explanatory variables it is possible to build a model forecasting the size of under- and overcoverage²³.

138. As previously described, ONS plan 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 a Census Coverage Survey). ONS is planning to begin testing such a survey in 2017, with the aim to run a full-scale test in 2020²⁴. This would be supported by other quality assurance processes, for example demographic analysis and comparisons with other data sources.

²³ 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.

²⁴ 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>.

139. The Dual System Estimation (DSE) is used in CSO, 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 SOL register, based on eleven administrative data sources and vital statistics, that aims to capture activity across all ages. Further information on this method can be found in the presentation ‘Compiling population estimates from administrative data’²⁵.

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 e.g. 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).

140. Finally, it should be mentioned that the method of assessing coverage does not depend on the census method and fits for all kinds of censuses.

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 over-coverage, which means that some people have been enumerated who do not belong to the factual population. Here the reason might be that some people have been enumerated repeatedly. When the ratio c/n is less than 1, then the census has under-coverage, which means that part of the factual population has not been enumerated.

Nowadays the **under-coverage** is quite a serious problem for several reasons: people are very mobile and it may happen that some of them are not at home at the census moment. Another reason is that some people assess their privacy so highly, that they do not want to share the 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 development is without any big changes in natural or migration increase. It is also necessary that the registration culture of population events is on a high level or the obligation of registration is otherwise very strict. This method has been used in Estonia several times and very good results were achieved in 1934 and in 1970.

²⁵ See

<http://www.cso.ie/en/csolatestnews/eventconferenceseminars/administrativedataseminars/5thadministrativedataseminar/>

Another method relies on **post-enumeration**. When using this methodology very shortly after the census a sample survey is organised with some census questions. Now the amounts of people $n(1)$ – enumerated by the census and by the survey, $n(2)$ – enumerated by the census and not enumerated by the survey, $n(3)$ – enumerated by the survey and not by the census and $n(4)$ – not enumerated by the census, but 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 to avoid the census or the survey, then the method might not give the adequate answer.

New methodology on the basis of existing administrative data sets

If a country has a set (system) of reliable registers or 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 we regard 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 we find their **signs of life** (SOL) 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**. Different possibilities exist: in Estonia 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 signs of life demonstrating their activity in Estonia during the last year). Not enumerated people who had according to the model a high probability to be 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% and the assessed model error was less than 5% of the set of people in question.

Similarly, it is possible to find in the census population the persons who do not belong to the factual population, hence they form the over-coverage.

IX.3. Quality and confidentiality measuring

141. Quality and confidentiality are in some sense contradicting concepts, but both of them are essential aspects of census output. It is important to understand that it is impossible to always publish accurate output (especially in small regions) following at the same time the rules of confidentiality. This situation is different in the case of registers. As registers need not publish

their data (in small areas), they do not face this problem. From here it follows that in the case of a register-based or combined census transforming the register input to all output data should be done using the confidentiality rules.

IX.4. Comparison of census output with surveys

142. The well-known way to assess the accuracy is comparison of output results from different surveys and censuses. The same approach is also useful in the case when a register-based census has been conducted. It is important that the topic of comparison is census output (formed using the information retrieved from the register), not any specific part of the register itself.

143. The more reliable the survey is, the more reliable will also be the result of the check. In this sense the LFS is a good choice for assessing the census quality. However, there are also several problems – the definitions might be different, also the survey's reference moment might not coincide with the census reference moment, etc. Also, no surveys cover the whole range of census variables, that means, only several variables can be assessed in this way. And last but not least: if there are differences, then it is difficult to say, where the problem is – in the survey data or in the census (or even in both).

144. In spite of all problems, comparison of census results with survey results is always interesting and should be done. After the comparisons the results must be analysed thoroughly to try to explain all differences discovered.

IX.5. Quality reports

145. 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 explanations of differences between census and survey data. If explanations exist why the output of a register-based census (or the part of the combined census where variables come from register data) differs from the results of surveys, then there might be the following future steps.

- The quality of the register will be checked using alternative methods;
- The quality of the survey will be checked via a comparison with other surveys;
- The software (including algorithms) used for creating the census variables from register data will be checked;
- A check if the confidentiality measures applied caused the differences;
- The possibility of errors in all procedures will be checked.

All results must be reflected in the quality report that should be used as an input for the following census.

IX.6. Quality committees and final remark

146. The quality check of a register-based census should be made on several levels. These levels might include the following:

- **Internal quality committees** formed by the national census bureau and consisting of census team members or people who worked together with the census team. The committee checks the quality of all variables and hypercubes and writes also the quality report.
- **Audits based on quality guidelines** that might be carried out by experts outside the census team.
- **External audits (to check privacy or peer reviews)** performed by experts outside of the national statistical bureau, possibly from another country.

Finally, the ESS.VIP ADMIN project 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, e.g. estimation when combining 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 administrative sources.

X. Approaches and case studies from different countries

X.1. Technical approaches

147. A general overview of the approaches taken by countries to overcome some of the difficulties is given. A list of case studies that share specific country experiences is provided in the Annexes.

148. Although there is a high-level common framework country must develop its own path for implementation. It is not possible to simply copy the practice of another country and ignore the national context. Allowing that, it is also beneficial for national statistical institutes to share their experiences. Some of the technical approaches that have been used in different countries are presented in this section.

149. A number of countries have moved to a register-based census. Early adoption was generally facilitated by customary 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 administrative sources in 2011. Estonia and Poland both conducted a combined census in 2011 including on-line self-enumeration. Other countries, like the UK, Canada, Ireland and New Zealand, are working towards a register-based census system and have a mix of challenges to overcome.

150. Countries can put a safeguard system in place to monitor register quality, methodology and technological capacity during the transition phase. For example, the UK has a traffic light system and Estonia has the quality requirements for register holders.

151. 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 national statistical institutes.

152. Where statistical registers have been used the number of administrative sources and base registers available/used varies per country. Austria used eight base registers for the 2011 census but used other registers to test quality. Estonia integrated twenty-two sources to develop their statistical registers. Poland integrated twenty-eight administrative data sources and three non-governmental sources (out of 300 potential data sources originally assessed) into a statistical register called the Master Record.

153. Integration of the registers means linking the information for people and dwellings to obtain proper coverage of the population. A key difference between countries is whether unique identifiers are available which allow deterministic matching on the identity field. The main challenge in some countries is the linking of data records from different registers without

unique identifiers. In this case, probability-based or fuzzy matching over several fields is attempted.

154. Generally records go through an anonymisation process to protect the identity of individuals where direct and indirect identifiers like the personal identifier are removed from the register and replaced with a proxy. An example of a person unique identifier is the PPSN in place for Ireland. 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 work around 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. In addition to identifying individuals a ‘sign of life’ filter must be applied to indicate whether the person is a usual resident or a resident at a particular time.

155. While small area statistics may be derived from knowing the geographic area where a person 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 UPRN in UK. Some other countries have established new building and dwelling registers. Without a unique building identifier linkage is achieved via automated address matching. Specific commercial software is available. Some countries develop their own in-house system e.g. Canada. Address matching is a complex rules-based process that needs frequent up-dating. The address needs to be matched to addresses in other registers, to persons and be geo-coded to a location. An ‘Address Standard’ was introduced in Estonia to facilitate address matching. Rules must be put in place for when conflicting information is encountered such as a person linked to more than one address or for multiple household dwelling.

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

X.2. Specific country experiences

157. To illustrate how transitions can take place and what kind of problems one has to overcome in practice some specific country experiences are presented in detail in the annexes. These examples are based on the experiences of countries that have moved recently, or plan to do so in the future. Other countries can benefit from these experiences.

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

ANNEX A

TERMS OF REFERENCE FOR THE TASK FORCE ON REGISTER-BASED AND COMBINED CENSUSES

I. BACKGROUND

1. In October 2015, the CES Bureau conducted an in-depth review of diversification of population census methodology and sources. As an outcome of the review, the Bureau supported the preparation of new guidelines on the use of registers and administrative data for population and housing censuses, and decided to set up a Task Force for this purpose (Report of the Bureau meeting: ECE/CES/BUR/2015/OCT/21).

2. This note presents the terms of reference for the Task Force on Register-Based and Combined Censuses.

II. MANDATE

3. The Task Force reports to the UNECE Steering Group on Population and Housing Censuses, which in turn reports to the Conference of European Statisticians and its Bureau. The Task Force is created for a period from 2016 to 2018.

III. OBJECTIVE

4. The objective of the Task Force is to prepare new *UNECE Guidelines on the use of registers and administrative data for population and housing censuses*, covering operational, practical, technical and legal aspects. The new guidelines should cover the use of multiple sources (including the so-called combined censuses) and include methods for measuring coverage and quality of the various sources, without specific focus on the comparability of the various census methods.

IV. PLANNED ACTIVITIES AND OUTPUTS

5. The Task Force is expected to prepare a set of Guidelines on the use of registers and administrative data for population and housing censuses. The planned output is a publication in electronic and printed format, in English and possibly in Russian (subject to availability of resources), to be finalized by April 2018.

6. The Task Force will start its work by collecting available material on the use of registers and administrative data for population and housing censuses. In particular, the Task Force should consider the publication “Register-based statistics in the Nordic countries”²⁶, the relevant text in the census recommendations for the 2020 round, and the papers prepared in the last years for UNECE expert meetings on censuses, describing the preparation and conduction of register-based and combined censuses²⁷.

7. The Task Force should also establish cooperation with other relevant international activities, including in particular those conducted under the ESS.VIP ADMIN project (part of ESS Vision 2020) on the assessment of the quality of administrative sources.

²⁶ http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

²⁷ <http://www1.unece.org/stat/platform/display/censuses/Collection+of+papers+on+transitions+to+new+census+methods>

8. The available material should be reviewed, expanded and integrated with new text, in order to produce a comprehensive document covering the following topics (the list is tentative and not exhaustive):

- Definitions of register-based and combined censuses, and description of the methods more commonly adopted in the 2010 and previous rounds;
- How to assess the coverage and quality of registers and administrative data, and how to decide if and how they can be used for the population and housing census;
- Advantages, disadvantages, and necessary conditions for register-based and combined censuses;
- Transition from traditional census to combined or register-based censuses;
- Methods for assessing the quality of register-based and combined censuses.

9. The Task Force should use as much as possible the information available on national census practices, availability and use of data from registers and administrative sources, and plans for the 2020 round, including in particular the information collected by UNECE in the 2013 and 2015 surveys on population and housing censuses. If necessary, the Task Force may collect, through the UNECE, additional information from member countries, minimizing the burden on national statistical offices.

10. The Guidelines should present good practices identified by the Task Force that could be used as examples by other countries. The Guidelines may also include additional recommendations to complement those already provided in the CES Recommendations for the 2020 censuses.

V. TIMETABLE

Feb 2016	Establish the Task Force and draft detailed work plan with activities, timing and division of work
Mar-Apr 2016	Collect and review available material
Apr-July 2016	Prepare outline of the Guidelines, and first draft chapters
Sep 2016	Present progress report at UNECE-Eurostat Expert Meeting on Censuses
Oct 2016-Jun 2017	Revise draft and prepare complete draft Guidelines
Sep 2017	Present draft Guidelines at UNECE-Eurostat Expert Meeting on Censuses
Sep-Dec 2017	Revise draft Guidelines based on discussion at Expert Meeting
Dec 2017	Submit final draft Guidelines for review by CES Bureau in February 2018
Feb 2018	Review of draft Guidelines by CES Bureau
Feb-Mar 2018	Electronic consultation among all CES members
Apr 2018	Expected endorsement by the CES plenary session

VI. METHODS OF WORK

11. The Task Force is expected to work in English only, mainly by e-mail and telephone conferences. The UNECE Census Wiki platform can be used by the Task Force to share work material, develop documents, and possibly organize discussion forums.

12. Face-to-face meetings of the Task Force could be organized in connection with Joint UNECE-Eurostat Expert Meetings on Censuses, or other events attended by a significant number of members of the Task Force.

VII. MEMBERSHIP

13. The following countries and organizations participate in the work of the Task Force: Austria, Canada, Estonia, Germany, Ireland, Israel, Italy, Netherlands (Chair), New Zealand, Norway, Poland, Republic of Korea, Slovenia, United Kingdom, United States, Eurostat, FAO and UNECE. The secretariat is provided by UNECE.

ANNEX B

This note is based on a paper "The Irish Statistical System and The Emerging Census Opportunity" presented at NTTS 2015.

I. Environment

1. CSO, Ireland has made significant progress with developing administrative data sources for statistical purposes: the first phase of the framework introduced in Chapter V. This work is underpinned by the developing 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 for the use of common official identifiers for persons, business and property across all official systems (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 register. This register is maintained and updated by the department of social protection (DSP - government ministry that oversees social welfare administration); however there is no requirement for 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 linking 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:

- (a) to enable longitudinal analysis of population cohorts across different administration systems;
- (b) to explore population structures over time;
- (c) 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 a register the PAR is not a register and is simply a population frame.

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 employee tax returns. This file captures every paid employment in the state and facilitates linking of business and person based pillars of 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% 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 census operations (2011 and 2016).

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 linking.

II. Census background

9. Ireland typically has high population movements and migration flows that necessitate it conducting a census every 5 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 in the state to be filled in by the head of household with respect to every person present on census night. The forms are then picked up by the enumerators to be compiled and collated centrally by the CSO. In 2011 the census cost €55m or €12 per person living in the state. In 2016, experience found that it is becoming harder to contact each household.

III. Census 2021 intentions

10. The intention in 2021 is to include an internet response option as part of the traditional census. This could happen as follows: census forms or letters with a url, unique id and pin will be delivered to each property (by hand or post) and the head of household will be asked in the first place to complete the form online (in the first instance), return the form by post or give it back to the enumerator when they call. There is an intention to use administrative data sources to assist with person/household non response.

IV. Census 2026 intentions

11. It is not known how feasible a census based on administrative data sources and existing surveys would be in 2026. But if Ireland were to conduct such a census it might be undertaken along the following lines.

12. Step 0. Ensure an up to date master address file. This master address file 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.

13. Step 1. Identify all persons interacting with public administration systems for the census year and summarise this on a register against a person identification number. Let's call this register the person activity register (PAR). The PAR will contain a list of all persons that are active in a given year i.e., '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.

14. Step 2. 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.

15. Step 3. 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 (e.g., children's benefit) may also feed into this decision tree algorithm.

16. Step 4. Form household relationships using relationships identified in administrative data sources and persons identified as living at the same house.

17. Step 5. Estimate and include attributes for each person on the PAR using existing surveys, administrative data sources and appropriate methodologies.

18. Step 6. All census outputs are now compiled from the PAR as updated by the above steps.

V. References

19. MacFeely and Dunne, "Joining Up Public Service Information: The Rationale for A National Data Infrastructure" Administration, Vol. 61, No. 4 (2014), pp. 93–107 http://ipa.ie/pdf/forum_vol_61_4.pdf.

20. Dunne, "The Irish Statistical System and The Emerging Census Opportunity" NTTS/SJIAOS 2015.

ANNEX C

Estonian case study

1. During the 2000s, Estonia worked very productively on systematic development of registers, incl. identifying the data in all personal registers on the basis of personal identification codes and linking them with the X-Road system to facilitate exchange of information. However, an initial analysis indicated that Estonia's registers were not yet ready for successfully conducting a Population and Housing Census in 2011. The main reasons were as follows:

(a) According to the quality survey data from 2007 up to 2015 at least 20% 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 have been 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 different registers and information technological compatibility of registers had not been analysed.

2. So in 2009 combined census methodology was approved for the Census 2011.

3. Meaning for the combined census methodology:

(a) Coordination of data sources. Previously created data sources, i.e. registers are used in censuses alongside with internet and face-to-face interviews. In the Population and Housing Census 2011 (PHC 2011), 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 case of missing data. The information on the 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 as a method alongside interviews. This required compilation of extensive instructions and provision of comprehensive guidance and training to the persons enumerated, as well as the enumerators.

(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 writing instrument, e.g. a special machine-readable pencil.

4. Two new technologies were introduced in the PHC 2011: self-completion of questionnaires on the internet and entry of answers directly in laptop computers during census interviews. The option of using paper questionnaires was kept as a backup for emergency situations. The possibility of telephone interviews was also foreseen for particularly exceptional circumstances, especially in cases when there was a need to specify the answers to a questionnaire or a census location was extremely difficult to access (e.g. on a small island). In practice both these emergency forms of interviewing were used in less than 1% of the cases.

5. In general, the combined census went well in Estonia. About 67% of the persons to be enumerated used self-enumeration via internet and the remaining 33% were interviewed face-to-face. All materials were published and delivered in time during 2012-2013.

I. The background of register-based censuses in Estonia

6. Already in 2010 preparations began for a register-based census. Here, the experience of those countries that had already conducted a register-based census was taken into account. So a project (the REGREL project) was started to develop a methodology for a register-based population and housing census.

II. Preparatory works during the period of 2010-2016

7. In 2010 Statistics Estonia started preparations for the transition to register-based population and housing censuses (the REGREL project). The first stage of the preparations was extensive analysis, which began in autumn 2010 and was completed in September 2013.

8. The REGREL methodology project (with about 80% of the project 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.

9. The analysis was carried out in two parts:

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

10. In addition to the meta-analysis and the detailed analysis of characteristics, the team also made other preparations for register-based censuses:

- (c) legal analysis;
- (d) preparation of methodological guidelines for the creation of a census glossary;
- (e) analysis of international experience and practice.

11. All in all, the project team analysed nearly 20 registers and the data collected by these 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 obligatory characteristics).

12. 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, the most important work is to be done outside of Statistics Estonia – this includes work on the registers on the one hand, and work on the information on respondents, residents and enterprises, on the other hand.

13. In 2010-2013 the REGREL methodology project was developed. At the same time in Statistics Estonia the REGREL team worked intensively. A very important part of the preparatory work was connected with the software solutions for transporting the necessary data from registers, creating census characteristics and saving them in Statistics Estonia.

14. Also necessary proposals to make changes in laws were made with the aim to get administrative information for creating census variables.

15. The permanent collaboration with administrative registers and continued, the quality of register's data was checked and suggestions to improve the quality were made.

16. The big task in the census methodology was preparation and realisation of algorithms for creating census variables from administrative data.

17. An important task was also the creation of an environment for data transportation, checking and saving.

III. Census pilot in 2014

18. The main goal of the REGREL pilot census was testing of the production system with selected mandatory census characteristics (EU Regulation no 763/2008). The list of selected census characteristics was the following: Place of usual residence; Sex; Age; Legal marital status; Country of birth; Place of birth; Citizenship; Relationships between household members; Level of education. After the pilot census its results were analysed and the problematic points were forwarded to the plans of the following years.

IV. Action plan during the period of 2016-2020

19. The main preparatory work for the register-based census is related to the following tasks:

- (a) Data acquisition from registers (contracts, description of the data set, checks on data quality and the acquisition procedure);
- (b) Formation of census characteristics, programming of the necessary rules;
- (c) Testing the statistical system as a whole;
- (d) Testing statistical registers system and filling it with data 2015-2018;
- (e) Analysing the needs and expectations of potential users.

20. Statistics Estonia has developed a set of legal and organisational measures to improve the quality, timeliness and coverage of the data sets necessary for the census. The set of measures are submitted to the ministries.

21. Currently, the registers do not cover all the census characteristics in the entire population, and it is also unclear whether all register data are updated fast enough.

V. 2016 Census trial

22. The period of trial census activities was 02.01 – 08.12.16.

23. The trial census population consists of:

- (a) The usual resident population of Estonia;
- (b) The dwellings regardless of occupancy and occupied non-conventional dwellings located in Estonia.

24. The objective is to practice conducting the register-based population and housing census (REGREL) following Eurostat's quality requirements and rules on censuses. For this, data must be acquired from 24 national databases, followed by processing and analysis.

25. On the basis of census characteristics compiled from register data using the created algorithms a series of hypercubes will be created (part of the hypercubes demanded by EU).

26. The quality check will consist of three parts:

- (a) The quality of all census characteristics will be checked in the sense of five quality criteria;
- (b) 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.
- (c) The households with all members being residents of Estonia where the household dwelling principle has been used.

27. The quality report will be created where all quality problems will be listed. The registers vouching for these quality problems will be informed and the possible solutions will be discussed.

28. During the next census trial in 2019 the preparations for the register-based census are focused on datasets (improving the problems found in earlier stages) and data flow. Also the confidentiality check will be analysed in more detail.

VI. Summary

29. In Estonia preparations have to be made for the register-based census in the course of which data will be captured from various databases following the rules of data protection and statistical security. The requirements developed for databases are sufficient for ensuring the interoperability of state information systems if:

- (a) All data presented are submitted with metadata, including classification codes;

- (b) Capture and data updates take place via the X-Road service;
 - (c) Data are presented in XML format and the description of data will be submitted by the creator of the X-Road service as XSD and updates include the time of presentation.
30. The primary data in the databases are required to meet the quality requirements in order to guarantee that census objectives are fulfilled. Quality is indicated by the following:
- (a) 1. The coverage of the registers needs to be at least 97% for the population and 95% for single characteristics;
 - (b) 2. 95% of the data need to be linked with the classifications registered in RIHA (Administration System for the State Information System).
 - (c) 3. All legal and private persons and citizens of foreign states need to be assigned a code.

VII. References

31. REGRELi metoodika väljatöötamine. Projekti lõpparuanne (2013). A. Puur, L. Sakkeus, S. Aben, <http://www.stat.ee/regrel> (Final report of the methodology project, available only in Estonian)
32. Annual report 2013.(2014). Statistics Estonia, <http://www.stat.ee/annual-report>
33. Roadmap to a register-based census. D. Matteus. Quarterly Bulletin of Statistics Estonia 4/13. Direct link: <http://www.stat.ee/65376>
34. (Re)designing administrative data – towardsregister based census. Estonian experience. D. Matteus, K. Lehto. (Paper presented at Q2014 in Vienna in 2014) Direct link: http://www.q2014.at/fileadmin/user_upload/Q2014_paper_Matteus_Lehto_20140515.pdf
35. Opportunities and challenges of register-based census: observations based on methodology development for the 2020 round census in Estonia. A. Puur, L. Sakkeus (Presentation at 25th Conference of the Estonian Statistical Society in 2013). Direct link: http://www.stat.ee/public/yritused/ess-25/1/12.Puur_Sakkeus.pdf
36. The project for developing the methodology of register-based censuses in Estonia (2014). K. Lehto. (Paper presented at UNECE meeting - Group of Experts on Population and Housing Censuses - in Geneva in 2014) Direct link: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2014/mtg1/WP16_s3_Estonia.pdf

ANNEX D**Introduction to the Polish Population and Housing Census 2011**

1. The National Census of Population and Housing (NSP 2011) conducted in Poland in 2011 was designed and implemented with the application of a mixed model, i.e. employing data from administrative registers (full scope survey covering base demographic variables) and data obtained from respondents (20% sample survey), exclusively with the use of electronic questionnaires. The Act on the 2011 national population and housing census stipulated that the information systems of public administration shall be used as widely as possible as both the sources of data for the census purposes (which as a consequence meant that information to be gathered during the census was mostly obtained from available administrative sources and then 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 sampling survey) as well as direct source of census data. Data not included in the information system of public administration or data ineligible in terms of the statistical data quality were collected from persons covered by the census. As a result, paper questionnaires were eliminated altogether. It is a significant achievement on a global scale. We claimed that this method was 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 in administrative sources.

I. The use of administrative sources

2. The starting point was the use of administrative sources already existing within the State administration structures. In accordance with the National Census Act, all entities maintaining IT systems of public administration and non-administrative in the selected scope shall deliver data in the framework of census operations in the scope and time as specified therein.

3. The necessity to use data from administrative systems in Polish statistics resulted from:

- economic reasons – demand for effectiveness: minimisation of the costs of statistics production, including administrative burden on respondents,
- the risk of an increased non-response in statistical surveys, including censuses,
- an intensive development of IT systems of public administration, based on advanced technologies.

4. Census implementation based on administrative and non-administrative systems has brought numerous benefits, including:

- an effective use of administrative and non-administrative systems,
- reduced census costs,
- reduced social burden connected with data transfer,
- an improvement in data safety,
- a guarantee of surveys harmonisation,
- the availability information from future annual census based on registers,
- the availability of data from administrative registers for any level of territorial disaggregation,

- the possibility to identify double entry errors (overcounting),
- the creation of a micro-database supporting indirect estimation – modelling at the unit level,
- an improvement in estimation for small areas,
- an improvement in the coherence and reliability of statistical data.

5. The issue of using data from administrative sources required an in-depth recognition of information resources which were found in these sources. An analysis of all the sources and variables potentially useful for the censuses were carried out. The necessary metadata on approximately 300 administrative registers were collected, of which the 30 most useful ones were selected. For each of these registers separate records were opened and all variables from these sources were subjected to the 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. The knowledge concerning the quality and utility of variables from different registers was a basis for the rules of merging data, and their estimation and imputation in the operational base of microdata created. The result of this work was invaluable knowledge concerning the utility and possibility of integrating different registers of public administration which the statistical service had at its disposal.

6. Finally, in the census in Poland we applied 28 sources from Government and Local-Government administration, and from administrators outside public administration such as real estate administrators, housing co-operatives, power distribution plants and telecommunication operators. All the administrators of databases approached the need for statistics related to censuses with understanding and provided access to their information resources for the purposes of the population and housing census in 2011.

7. Data from administrative systems was used in the census:

- as a direct source of census data (personalisation of questionnaires).

And to create:

- compilations of buildings, dwellings and persons,
- an address-residence register,
- a sampling frame.

8. To enable the administrators to transfer data from dispersed systems via tele-transmission, the Central Statistical Office of Poland (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. These solutions were also applied when collecting from over 2500 local governments.

9. The unit data obtained from registers were converted into statistical registers, simultaneously being subject to the process of cleaning, de-duplication and standardisation of data. The process was carried out in the DQS SAS environment. At the same time, metadata were collected on quality of input data obtained from registers, the applied cleaning procedures and the final quality obtained after applying DQS procedures. Data from administrative sources converted to statistical data were used to derive the Master Record.

10. Master Record was the set of variables derived from the registers containing information that was introduced to the census forms in order to verify (confirm or update) by the respondents.

II. Quality evaluation of data from administrative sources

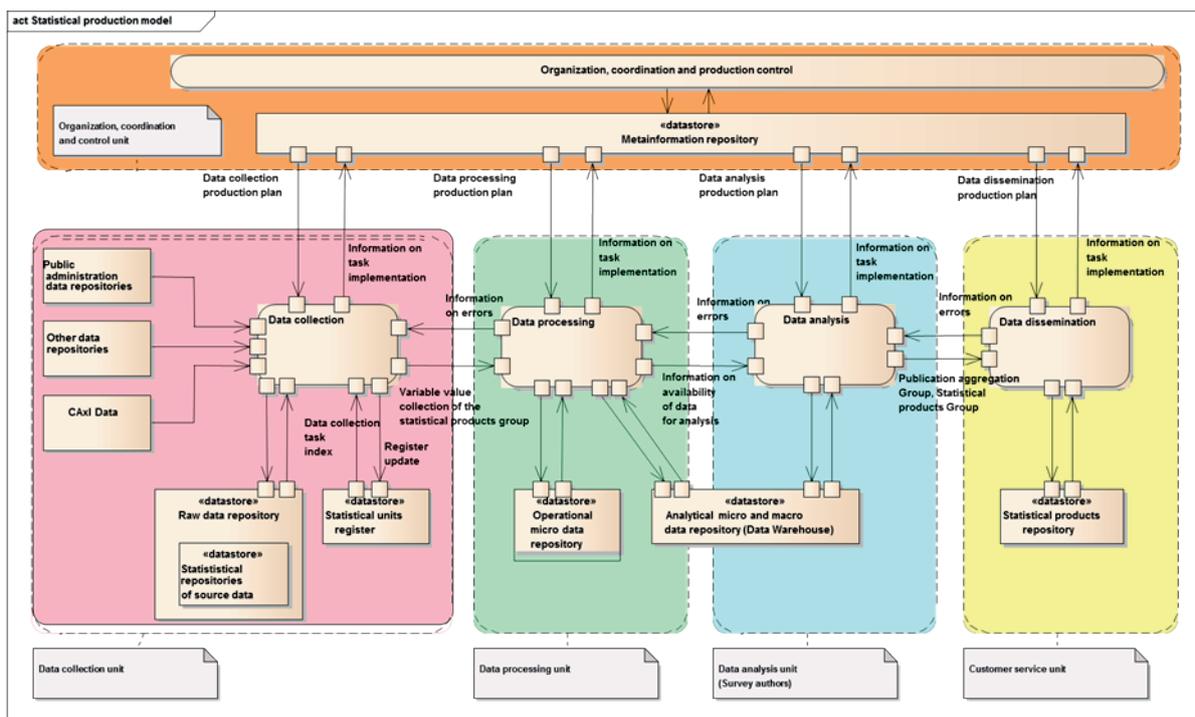
11. The quality of public administration systems was difficult to measure due to the complexity and multifaceted issues. Many factors cause that the quality was immeasurable, e.g. difficulties in assessing the quality result of the inapplicability of the synthetic instrument, hence assessing the quality of systems based on multiple indicators.

12. Before assessment of the quality systems there are defined criteria and indicators to measure quality. For purpose of census there were three stages of quality assessment: evaluation of data sources - systems, data sets, statistical products - resulting data.

13. With regard to the sets from administrative data sources, quality assessment of raw data sets provided by the administrators and sets after a transformation, i.e. after adjusting them for use in the census, was carried out. Quality measurement was performed also in all stages, in all processes of development of administrative data and the data combined with data from other sources.

14. The scheme below shows a graphical representation of the processes in the census, including quality assessment.

15. "Metainformation repository" was created to collect methodological, technical and operational metainformation. It ensured process control of data processing as well as monitoring of the course of processes, including measurement and collection of metainformation concerning quality at all stages of the process of data development, i.e. on stages: data collection, data processing, data analysis, data dissemination.



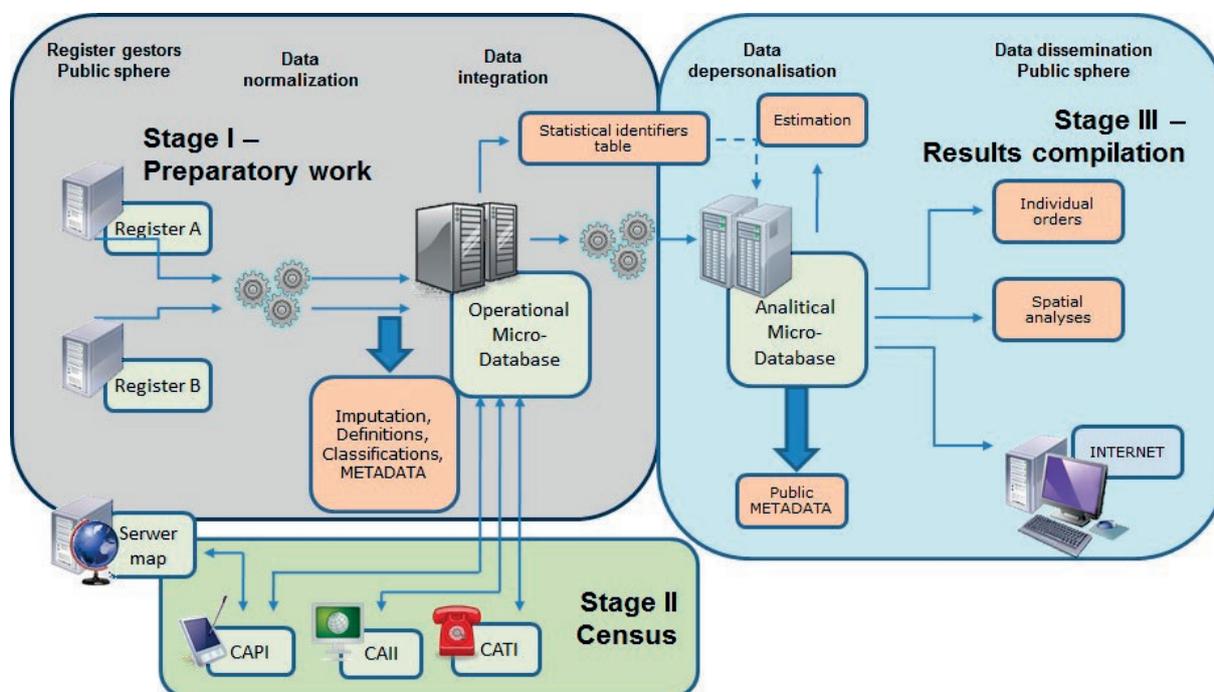
III. Others data acquisition methods used in the census

16. Poland was one of the first countries in the world which prepared a totally innovative method consisting of using several of the most modern techniques for collecting census data simultaneously. Apart from the use of IT systems of public administration, various data collection methods were applied, based on functioning of three channels simultaneously (known under the common name of CAxI):

- CAII/CAWI (Computer Assisted Internet Interview/Computer Assisted Web Interview) – an online self-administered questionnaire, which entails checking the respondent data obtained from administrative sources, within a specified time frame, and, if need be, correcting the same and providing missing information.
- CATI (Computer Assisted Telephone Interview) – a computer assisted telephone interview, conducted by a statistical interviewer.
- CAPI (Computer Assisted Personal Interview) – an interview conducted by a census enumerator, registered on a hand-held device.

17. All three channels were based exclusively on an adaptive electronic questionnaire, ensuring high quality of data at the collection stage. The electronic questionnaire was adjusted and implemented in accordance with the technology assisting particular modes of obtaining data based on CAxI. 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.

IV. Census architecture – The IT Census System



18. To enable the optimal application of advanced IT and telecommunications technologies in censuses, an appropriate census architecture had to be constructed. For the purposes of census design and conduction, the Central Statistical Office of Poland implemented the IT Census System (ISS). The system, consisting of more than 10 components implemented by

different contractors, provided IT assistance for all operations within census. The ISS integrated various technologies (from applications installed on mobile terminals, through applications managing and assisting in telephone interviews, to specialist bases, data warehouses and analytical and reporting tools).

19. The ISS employed various solutions ensuring a high level of security for processed census data. Certain organisational means were implemented, obliging census participants to observe statistical confidentiality and to guarantee the protection of personal data.

20. As part of the work involving the processing of census data, including data from administrative registers, numerous rules and solutions were established in the Central Statistical Office, which were later implemented in practice.

21. The solutions and rules included:

- the preparation of normalisation, control, and correction rules for datasets from administrative systems, including data imputation (in administrative datasets),
- the preparation of rules for the synchronisation of data from administrative systems – the harmonisation of base 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 census data – imputation and calibration,
- the preparation of rules for linking data from various administrative systems
 - the methods of precise/clear linking,
- 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,
- the preparation of data anonymisation rules.

22. Pursuant to the National Census Act of 2011, the Operational Microdata Base (OMB) was developed, prepared, and implemented at the Central Statistical Office. The created system included hardware-system-tool infrastructure (computer hardware, system software, tool software) and application software (computer programs that are the result of programming work).

23. This base enabled the inclusion of data transmitted in electronic form through four informational channels by entities obliged to do so by the Act, i.e. entities maintaining administrative registers, persons covered by the census (via the Internet, a telephone interview run by a statistical interviewer or a direct interview run by a census enumerator) and to conduct further data processing. In the OMB processes took place connected with the control, correction, and linking of data, up to their complete cleansing. Next, depersonalised data were transferred to the Analytical Microdata Base (AMB).

24. The Metainformation Subsystem gathered indispensable metainformation describing data and census processes, including the processes indispensable to drawing up quality reports. The task of the Metainformation Subsystem was to ensure the coherent definition of statistical objects for the OMB and AMB. The Metainformation Subsystem was also used to store depersonalised operational metadata of the OMB and AMB systems. This Subsystem constitutes the Central Metadata Repository (CMR).

25. The role of the Analytical Microdata Base is to store depersonalised census data in their final form. In this dataset every type of statistical analyses is carried out to acquire results for publication, i.e. the census products. The AMB allows all the recipients of statistical information to quickly acquire data in the form of aggregates. The AMB system constitutes an analytical and reporting platform that currently enables the statistical preparation of the outcome data from the National Population and Housing Census 2011. The results of analyses in the form of documents, reports and breakdowns are shared with internal and external users.

26. The AMB also allows the calculation of aggregates available in the Geostatistics Portal as maps (cartograms and cartodiagrams).

V. GIS Technology

27. For the first time in census history, GIS (geographic information systems) has been used in conducting and monitoring the survey.

28. With use of various reference materials and registers containing spatial information Polish official statistics has created spatial data for statistical address points and borders of statistical division of the country. Digital maps used by census enumerators were an indispensable data source (to navigate and verify dwelling locations in the field), by gmina leaders (for census monitoring within the gmina), and by voivodship and central supervisors (for census monitoring on voivodship or global level). Maps were used to monitor the census progress in a defined area or for specific enumerator (an on-demand location or daily route could be visualised on the map).

VI. The Geostatistics Portal

29. The Geostatistics Portal is a tool for interactive cartographic presentation and the publication of data acquired in censuses. It serves the following functions:

- storing,
- presenting,
- sharing information for a broad group of recipients.

30. The Portal functions on two levels: for internal (official statistics) and external users, and the scope of presented data is defined through the appropriate roles and authorisations. Internal users have access to both unit and aggregated data, whereas external users only to aggregate data, published taking into account statistical confidentiality.

31. The interface of the Geostatistics Portal allows its users quick and easy access to resulting statistical information. Data are presented using such cartographical presentation methods as cartograms (choropleth map) and various cartodiagrams. It is also possible to set one's own parameters for the visualisation of a thematic area for a given cartogram. These include measure, aggregation level (territorial division unit), the number of intervals, etc. Aside

from the possibility of using ready-made spatial analyses, in the Geostatistics Portal, internal users can draw up custom thematic maps based on a selected feature of the data model, using dynamic spatial analyses, i.e. linear or distance analyses, or object buffering.

VII. Summary

32. The census in Poland turned out to be an innovative project not only countrywide but also worldwide on grounds of the following facts and figures:

- simultaneous data collection, without paper, from four different channels (i.e. administrative registers, Internet self-enumeration (CAI), direct interviews conducted by census enumerators, using electronic questionnaires (CAPI), and telephone interviews conducted by statistical interviewers (CATI)) was used and implemented on such a large scale for the first time in Europe,
- data from 28 administrative registers and 3 non-administrative systems were effectively integrated,
- paper questionnaires were completely eliminated, and were replaced by ICT solutions,
- the use of GIS technology helped to conduct the census preparatory work and an on-going census process monitoring and give possibility to compile and present census results based on multi-dimensional spatial analyses,
- IT Census System comprised a number of solutions ensuring the high level of security of the processed data,
- the modern statistical data processing technologies have been developed – they will 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.

33. A comprehensive analysis of census conduction, accounting for all its participants, thus also the members of the field census frame, allows one to draw certain conclusions and to assess the possibility for their further implementation. The new technology applied in the census has proven that it can also be implemented in questionnaire-based surveys. It is cheaper, employs up-to-date control mechanisms, enhancing the quality of the material collected, and, in consequence, reduces the burden of respondents.

34. The time frame proved right, since the deadlines set for the data collection stage (specified in the Act on NSP) did not have to be extended, and neither did the census budget. The detailed schedule for the implementation of NSP 2011 was regularly updated. The framework schedule comprised over 250 items, and the detailed schedules for tasks included therein (for example the preparation and procedure of a control census, support of the census systems) were kept in separate files. The schedule comprised a total of several thousand tasks.

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

36. There is another round of censuses ahead. Thanks to current experiences, in the next round of censuses Poland is going to use newer technology making it more effective and more inventive, which is still delivered and modified by the IT world.

37. Until then, census implementation methods should be developed and implemented, treating the experience gathered in 2011 as the starting point. Considerable efforts need to be expended with a view to developing a new census strategy, so as to guarantee progressive solutions. Attempts should be made at:

- reducing census costs,
- using administrative sources in an effective way,
- reducing social burdens connected with data transfer,
- improving the safety of transferred data,
- improving the coherence and reliability of statistical data.

ANNEX E

TRANSITION FROM TRADITIONAL CENSUSES TO REGISTER-BASED CENSUSES THE REGISTER-BASED CENSUS IN AUSTRIA

1. In the 2011 Population and Housing Census for the first time in census history information on persons, buildings and dwellings was derived from administrative and statistical registers. Ten years before, the census was conducted as a traditional enumeration, with questionnaires and enumerators. The transition from traditional censuses to a register-based census was implemented in a relatively short time. The following paragraphs describe the general conditions and the legal framework of the transition and present in short the register-based census model.

I. Government Decision

2. In the year 2000, the 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 (respondents burden too high, results too late, costs too high). Much of the information that is asked from respondents is already available in administrative registers. In the long run, register-based censuses are much cheaper than traditional censuses.

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

4. The government followed the 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 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 anonymized linking of administrative registers would be established and that the quality of administrative registers would have to be improved.

II. Creating the Necessary Conditions

(a) The Central Population Register (CPR)

5. In Austria, population registration is obligatory, but up to 2002 each municipality had its own register. The central population register (CPR) became operative on March 1st, 2002. The initial population stock was compiled from more than 2 300 municipal population registers during the 2001 Census²⁸.

²⁸ As part of the 2001 census, municipalities had to upload data from their population registration systems into a central data base (GSG GemeindeSoftwareGroßzählung 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 topics 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 case of emigration from Austria. In 2006, the topic “legal marital status” was added as a new characteristic. The CPR does not provide information of family relationships which, for example, can be found in population registers in Nordic countries.

(b) The Register of Educational Attainment, Register of Pupils and Students

7. One of the first activities of 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 (Educational Documentation Act of 2002). From 2003 on, schools have 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 law, the register of educational attainment was set up by using statistical information on highest education 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 is a further source of information on highest education attained as well as authorities which are responsible for the recognition of school or university diplomas from abroad.

(c) The Buildings and Dwellings Register (BDR)

9. The Buildings and Dwellings Register was set up based on 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 which have to update information if a building or dwelling is rebuilt or new buildings are under construction. The register contains addresses of buildings and dwellings, characteristics of buildings as well as of dwellings and other housing units and information 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.

12. Addresses, buildings and dwellings have unique identification numbers which 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.

(d) The bPIN OS

13. 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 data protection reasons. The social security number contains the date of birth of a person and does not guarantee anonymity.

14. 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” uses its own PIN such as for “health”, “social security”, “taxes” or “official statistics”. The bPINs are derived from a source PIN each person has in the central

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

15. In the register-based census, the use of the bPIN OS is implemented as follows: 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 central population register data by using these identity data. If there is a match, the bPIN OS (official statistics) and the bPIN of the register owner can be derived from the source PIN by using a very 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.

(e) Census Legislation

16. The existing census law was not a sufficient legal base for conducting a register-based census in 2011. So a new census law had to be worked out based on a detailed concept of Statistics Austria for a register-based census which was ready in 2004 after many consultations with experts. It was approved by an inter-ministerial working group in June 2005. The register-based census act came into force in March 2006.

III. Main Principles of the Register-based Census

(a) Scope of the Census

17. The register-based census only comprises core topics according to the CES Recommendations and the EU Census Regulation. Information that is available in registers but had not been part of the list of topics in previous censuses, for example “income”, is not included.

18. As some variables are not available in any register, it is not possible to collect them using a register-based census, e.g. language spoken or characteristics of commuting such as the means of transport.

19. Census day was determined as to be October, 31st. A census day at the end or at the beginning of a year is not a good option for Austria. In some municipalities the population number would be distorted because of seasonal employment in winter tourism.

20. It should be mentioned 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.

(b) Linking of Registers

Personal Identification Number

21. As already mentioned (see 2.4), register data are linked by using the bPIN OS. In practice, the assignment of a bPIN OS sometimes is not possible if identity data are inaccurate. Nevertheless, the register owner has 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

22. 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.

(c) Redundancy

23. The various registers that are the basis of the census could contain different values for a characteristic of a person. Therefore, it is not feasible to trust data taken from only one register. To ensure satisfactory 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 topic is assigned to a basic register and to “comparison” registers which are used as confirmation of the values in the basic register. If values are different, then a set of rules is applied to determine the value of the variable.

24. Topics such as “current activity status” have to be collected from different registers which, for example, provide information on economic activity on the one hand, and on school enrolment the other hand.

(d) The Registers

25. The 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 topic, if the topic is assigned to another base register. So these eight registers are both, base registers and comparison registers, for different topics. The base registers and their owners are:

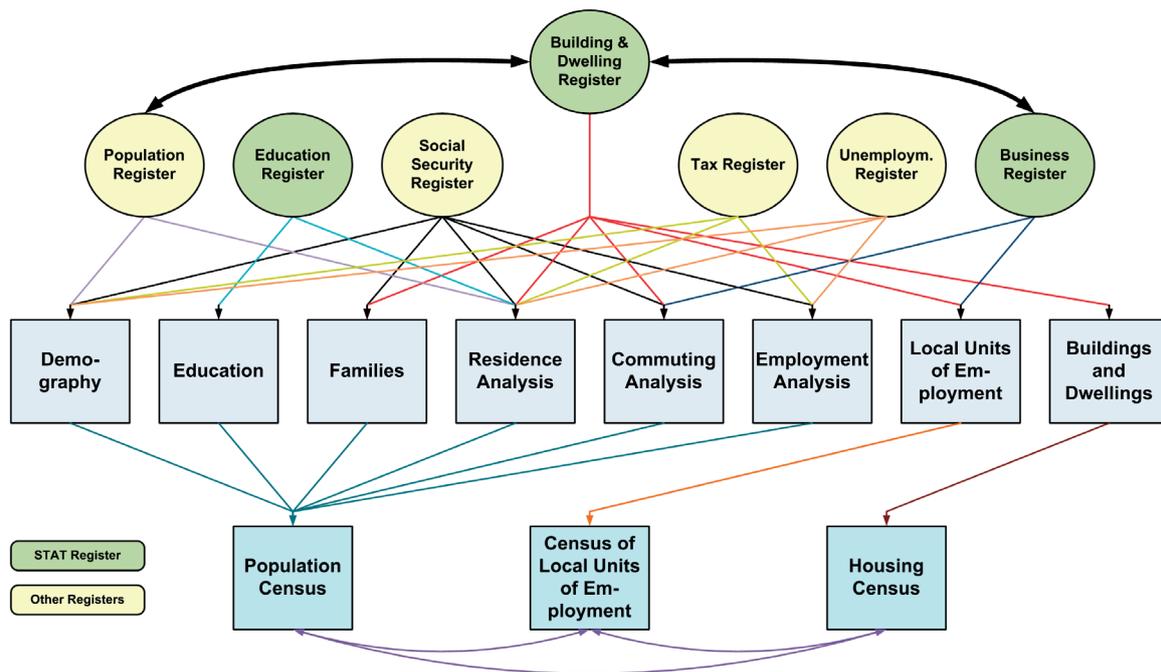
- Central Population Register (CPR), Ministry of the Interior
- Central Social Security Register (CSSR), Association of Social Security Institutions
- Tax Register (TR), Ministry of Finance
- Unemployment Register (UR), Public Employment Service Austria
- Register of Educational Attainment (EAR), Statistics 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

26. These registers contain information on all topics required for the register-based census. Other registers used to improve the quality of the results (“comparison” registers) are:

- Child allowance register (CAR)
- Register on non-citizens (including asylum seeking persons)
- Registers of public servants of the federal state and the Länder (RSP)
- Register of car owners (COR)
- Register of social welfare recipients (SWR)
- Conscription register (CR)
- Register of alternative civilian service (ACSR)

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

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



(e) *Residence Analysis – Over-Coverage in the Population Register*

28. The Register-based Census Act defines that every person with a main residence according to the central population register 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 called analysis of residence.

29. The residence analysis is part of the census in order to avoid over-coverage which was expected to exist in the central population register²⁹. In Austria, the evidence of being registered is required in various circumstances, so under-coverage is expected to be of low significance.

30. 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 records (persons) with no matches between the CPR and the other administrative data. To have a record in other administrative registers 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.

31. Persons who are registered in the CPR and nowhere else do not show satisfactory signs of life. Their residence has to be clarified by responding to an official letter³⁰, which asks about the place of the main residence on the reference day. In 2011, about 96.000 persons were queried and they should send back a filled in 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 (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 Austria is obliged to inform the municipalities of the results of the queries. The local authority has the right to oppose but must prove that the person still has his/her main residence in the municipality. The evidences have to be sent to Statistics Austria.

²⁹ Persons, who emigrate, often do not notify their local registration authority, or deaths may be registered with a time lag.

³⁰ Only for these persons, the Register-based Census Act allows a re-identification by the data protection authority.

32. On 31 October 2011 the CPR covered a population of about 8 466 000 (without persons who were registered less than 3 months) whereas the population of the register-based census was 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.

33. 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%, under-coverage due to delayed registrations much smaller.

IV. The 2006 Test Census

34. A major element of the preparation of the 2011 register-based census was to conduct a test census (census day 31-10-2006) 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, full scope of topics). The only difference was that the population count was not meant to have legal consequences for the municipalities (regarding the fiscal transfer from the federal state to the municipalities) and that a sample survey was conducted in order to evaluate the quality of the results.

35. The modalities of the sample survey:

- A sample of 25 000 people = three per thousand of the total population, in 100 sample areas;
- Field enumeration by interviewers, paper questionnaires;
- Participation of the population in the sampled areas was obligatory.

36. The results of the register-based test census (report available in 2008, published on the website) showed that information from registers is predominantly very good, in the expected problem areas such as household and family statistics mostly satisfactory. The conclusion was that the concept of the register-based census has proved to work. Statistics Austria was commissioned to continue with the preparation of the first register-based census in 2011.

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

37. Because of the good quality of 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 transferred was the population number of the municipality as determined in the decennial population census. But now, the distribution keys should be re-calculated yearly on the basis of a yearly population count, starting with October 31st, 2008. The new law stipulated that the procedures for determining the population count as designated for the register-based census should be applied. There is one exception: The residence analysis between the census years does not include asking people

with no signs of life as done in the 2011 (and 2006 test) census. In order 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.

38. Based on the yearly provision of administrative register data, Statistics Austria decided to produce yearly 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 3.5. From 2012 on, the yearly register-based statistics comprises all census topics, only housing is excluded.

VI. Quality assessment

39. The method of collecting data from administrative registers required a completely new quality assessment concept. To this end, a quality framework³¹ was developed that independently assesses the quality of administrative registers, the quality of the results and the quality of the processes at topic 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.

VII. Improvements Expected for 2021

40. 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 on, on the occasion of registration of births, deaths, marriages, registered partnerships and divorces data have to be 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 use information from both administrative registers. It is expected that the quality of demographic variables, especially of the topic “legal marital status” and family relationship information, which is already very high, will improve.

41. Currently, Statistics Austria is testing to use 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.

VIII. Summary

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

- A comprehensive register system developed for administrative needs (taxes, social security, registration of residences), readiness to set up new (statistical) registers;
- Legal basis;
- Possibility 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;

³¹ See references in chapter III.

- Building up knowledge on the administrative data sources (by constantly working with the data and assessing their quality);
43. The use of administrative and statistical registers has become an integral part of today's statistical production process at Statistics Austria.

IX. Reference

44. Berka C., Humer S., Lenk M., Moser M., Rechta H. and Schwerer E. (2010). A Quality Framework for Statistics based on Administrative Data Sources using the Example of the Austrian Census 2011. *Austrian Journal of Statistics* 39 (4), 299-308.
45. Lenk M. (2009). *Methods of the Register-based Census in Austria*. Seminar on Innovations in Official Statistics, United Nations, New York.
46. Rechta H., Schwerer E., Waldner C. (2014). Family Generation Process from Administrative Data Sources and the Austrian Register-based Census 2011. *Austrian Journal of Statistics* 43 (1), 77-88.
47. Statistics Austria, *Quality Assessment of Administrative Data. Documentation of Methods*. Vienna 2014
http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&RevisionSelectionMethod=LatestReleased&dDocName=079140
48. Statistics Austria, *Census 2011 Austria. Results of the Register-based Census*. Vienna 2014.
http://www.statistik.at/web_en/publications_services/Publicationsdetails/index.html?includePage=detailedView§ionName=Population&pubId=685

ANNEX F

Transition from a traditional census to the register-based census: Case of Slovenia

I. 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 started already in the early 1970s. At that time the Statistical Office of the Republic of Slovenia (hereinafter 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 (hereinafter PIN) to administrative databases and also to the statistical surveys was crucial for further 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 corresponding authorities: the Central Population Register (hereinafter the CPR), the Register of Spatial Units (addresses), the Statistical Register of Employment and the Business Register. As the administrative function of a register must differ from the statistical function, three of them (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 from all public authorities to use the general classifications and, where possible and feasible, bring the data concepts and variables close to statistical concepts.

2. The first 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% coverage of PINs in the final 1981 Census database. In addition to PIN, in the 1981 Census for the first time some administrative data (on educational attainment, occupation and industry) provided by employers but on paper forms were used for rationalizing of the data collecting stage and for improving quality. Data from the 1981 Census were also the basis for regular daily updating of CPR data based on statistical demographic surveys and some administrative records, resulting in the dissemination of stock population directly from the CPR using permanent residence definition from 1986 to 1994.

3. The next 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, also the 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. On the other hand, in the 1991 Census for the last time the classic statistical processing (mostly manual editing) was performed. The difference between 1991 Census data and CPR data was 1.8%; the coverage of PINs in the final census database was 99%.

4. The organization of the 2002 Census field enumeration and the statistical processing system were important steps to the full register-based approach in the 2011 Census. We can say that the 2002 Census was the first and the last combined census in Slovenia as in addition to complete field face-to-face enumeration using paper questionnaires data for some census topics were entirely taken over from the registers and not collected in the field (e.g. place/country of

birth, last migration, citizenship, marital status, occupation, industry, place of work) and for some topics only if data were not available in the pre-census database (sex, date of birth, activity status). Main novelties in the 2002 Census were:

- Composition of two databases compiled from 9 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.

5. The pre-census database was created 6 months before the reference date, which was 31 March, and was used for printing data to the questionnaires and for planning of all organizational 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 ago (1.6%).

II. Creating the necessary conditions

(a) General prerequisites

6. 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;
- Appropriate topics in the sources covering most of the demands of national users and corresponding to the draft EU Regulation on population and housing censuses which was adopted in 2008.

(b) Census legislation

7. There was no need for special law on census as legal bases already existed. Acquisition and integration of data is allowed by Articles 32 and 33 of the National Statistics Act (Official Journal of the RS, 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 a particular year that a complete register-based census is going to be conducted.

(c) New census date

8. 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;
- Greater consistency of administrative sources at the end of the calendar year.

(d) New development after 2002 Census

9. From the content point of view the decision to go for a register-based census was possible because the only missing register - on dwellings (Real Estate Register) - was established in 2007 on the basis of 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 on the basis of statistical data, so 2002 Census data on housing could not be used for this purpose. At the same time dwelling numbers in multi-dwelling buildings were determined (exist also in physical form) and addresses of people in the CPR were supplemented with these numbers. The dwelling number is now also a part of our official address in the CPR and it was the last missing link connecting people and their dwellings. In addition, within the computerization of administrative internal affairs the Ministry of the Interior set up an electronic Household Register, which 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 of data on relation to the reference person of the household which are used for direct family formation.

(e) Pre-census evaluation of quality of input data

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

- Inconsistencies in household composition (minor problem solved in the processing stage);
- Under-coverage of dwelling numbers (appendix to the address in case of multi-dwelling building) in the CPR;
- General unsatisfactory quality of housing data as the main problem which was not adequately solved by the 2011 census reference date.

(f) Improving quality of data on dwelling number in CPR

11. The completeness of updating dwelling numbers in the CPR was far below expectations 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). The Ministry of the Interior and SURS undertook two main activities in close cooperation. First, methodological solutions for automated determination of missing dwelling numbers were prepared by linking data on ownership of dwellings and 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 that did not have a dwelling number to report the data (with financial, organizational, and processing support of SURS); 49,000 letters were sent out, and the response rate was 75%.

III. Main principles of the Register-based Census

(a) Linkage of data on persons, households, and dwellings

12. Linkage of data on persons, households, and dwellings using unique identifiers is one of the most important tasks in producing census data using field enumeration or register data. In the case of the Register-based Census, the direct linkage of all data sources for persons using a PIN is the basic statistical operation. The administrative personal identifier (EMŠO) is transformed into statistical one (SID) due to confidentiality and privacy reasons before the statistical processing of census data. Further, 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 (serial number).

Chart 1: Identifiers used in 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				

(b) Quality of basic identifiers

13. The PIN is the most important identifier with complete coverage in 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 relation to the reference person of the household which is also considered as key identifier in our register-based system) is available only for permanent residence. 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 Census 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.

14. The main difference between the two Register-based Censuses was the quality of the input data. Much fewer dwelling numbers were missing in 2015 than 4 years ago (30,000 in 66

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% in the 2011 to 94% in the 2015.

(c) Administrative and statistical sources

15. Three administrative registers already described before 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;

16. Most of data for topics to be covered in Population and Housing Censuses from the EU Regulation 763/2008 has been extracted from only one source:

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

17. Data for other topics on persons have only been produced by using and combining input data from several data sources. The basic methodological principle for production of statistics in that case is the hierarchy of the sources, which means that, in the case of availability of data from several sources for each person (PIN), priority is given to the source with the highest priority after the quality evaluation of all sources foreseen. For educational attainment data 9 sources have been used (Table 2). 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 (%) ³²
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	31. 3. 2002	13.6

³² Population aged 15 or more years.

		education		
--	--	-----------	--	--

Source: SURS, Statistical Office of the Republic of Slovenia

18. Data on economic characteristics (current activity status, occupation, industry, status in employment, location of place of work) prepared from 8 sources generally refer to the census reference date.

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

19. Data for migration characteristics (country/place of birth, ever resided abroad, previous place of usual residence) were 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;
- 2002 Population Census.

(d) Statistical process

20. Data availability was the basis for the four-phase production and dissemination plan for the 2011 Register-based Census data:

- 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 housing characteristics from the RER as of 1 January 2012 were updated again (first release on 21 June 2012).

21. Data that have already been disseminated cannot be altered in the later stages of the process, so special metadata tables are prepared to ensure that 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 must be recorded transparently and clearly.

IV. Quality of the Register-based Census outputs

22. The overall quality of the statistical data in great depends on the quality of the administrative data as a register-based census is in fact a statistical operation transferring administrative concepts into statistical ones. The preconditions for quality outputs from this perspective are:

- Permanent use of administrative data in the statistical process;
- Long-continued and long-lasting use of the administrative data in the statistical process;
- Very close cooperation with the keeper of the administrative source;
- Feedback from statistical evaluation implemented in the administrative source.

(a) Regular quality monitoring

23. The over-registration is the common problem of all register-based statistical systems. Two main methods are used for the quality assessment of population (census) data in Slovenia in terms of coverage:

- 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 is better and better as the imputation rate for labour force status in 2011 was 1.50 while in 2016 only 1.14;
- The residence status of the selected respondent person 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) - 3% to 6%;
 - living abroad – over-registration in terms of coverage – 1% to 2%;
 - no answer about residence status – less than 1% (non-response).

(b) Special survey on coverage

24. A special survey with emphasis 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 Register-based Census (15,500 persons); the second sample group consisted of persons marked in the administrative data (CPR) as non-residential group, but data on their labour force status in Slovenia could be found in at least one out of nine sources (2,700 persons). Two methods were used: postal method for the whole sample and face-to-face field inquiry (1,915 persons) for selected postal non-response (71%). The final outcome of the survey is estimation of over-registration, which was approximately 0.5% at the aggregated level ($\approx 10,000$ persons at most). In comparison with the last field census data in 2002, when almost 1% of the population was counted twice (over-coverage) and slightly less than 2% of the population was not counted at all even though they were living in Slovenia (under-coverage), the results of the survey are promising.

V. Conclusion

25. Basic demographic (census) 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. Our main input administrative data are transmitted quarterly from the CPR approximately 3 months after the reference date. Beside it data on socio-economic characteristics using census statistical processes are produced annually. A complete register-based census as a regular statistical survey is going to be conducted twice (2015, 2018) between obligatory years determined in EU Regulation (2011, 2021).

26. The register-based approach in Slovenia achieves 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:

- Increasing concerns over costs: there was no 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 with coverage or item nonresponse or difficulties with data entry and editing.
- Respondent burden and decreasing participation in the census: these are not a problem anymore in Slovenia.
- Privacy: much fewer persons are now handling information, in contrast to thousands of people having access to personal data in the field operation.

ANNEX G

Transition from a traditional census to a register-based census: construction of a Statistical Population Dataset in Portugal

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 socioeconomic variables. Administrative data sources are presented, as well as the methodological approach, the quality indicators to assess its fitness and comparability with the 2011 Census, the annual Population Estimates (PEs) for 2015 and the 2016 Census Test (TC). 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 personnel identifier number (PIN). In addition, the country does not have a legal framework allowing access to the full name and address of the 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 lead to a paradigm change and the setting-up of medium to long-term strategies, to be implemented in 2021 and beyond.

I. Censuses in Portugal

2. 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%.

3. Similarly to 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 allowing a more frequent and updated release of statistical data.

II. Construction of a statistical population dataset in Portugal

4. 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 purpose to understand whether the administrative information available in Portugal allows for high-quality resident population enumeration, two exercises have been held for creating a Portuguese SPD. This database was built from a number of registers provided by different administrative data sources. The reference period for the first exercise was 2011, in order to use the 2011 Census as a benchmark for the comparison of results. 2015 was the reference year for the second exercise, and the results are compared with the PEs for the same year.

5. 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, considering that it is a dataset of citizens with legal address in Portugal (which does not meet the census concept). The CR overestimates the country's resident population (more than 1.1 million persons, considering the population enumerated in the 2011 Census, which corresponds to 10% of the resident population) and does not include non-citizens living in Portugal, who are registered in the Immigration Register (IR).

6. 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 is registered in the CR or in the IR and is, in addition, ‘active’ in at least another register (e.g. he 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, not living in Portugal, maintain their legal address in the Portuguese territory.

7. Hence, the SPD is composed by the citizens in the CR and the non-citizens of the IR for which ‘signs of life’ are given by the presence in at least one more register. On a second stage, the demographic and socioeconomic administrative variables have been associated to the population in the SPD.

III. Administrative data sources and variables

8. Additionally 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 National Health Service – Patient Register.

9. The administrative variables available in the different sources converge towards 16 census topics, i.e. 13 at EU Regulation level – 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 at national level – number of hours worked, number of employees in the enterprise, and school attendance.

10. The coverage of each variable depends on the presence of individual records in the respective source files: only 7 out of 16 variables (sex, age, place of usual residence, place of residence one year before, marital status, citizenship, country/place of birth) contain information for all SPD. Coverage is partial for socioeconomic variables. For a number of important variables on the person and the household characteristics (specifically the household classification), there is no information available in the registers.

11. The quality of the registers has been assessed according to a series of general criteria involving indicators such as variables’ coverage rates, quality of the information, timeliness of the information, among others, and a finer assessment at microdata level, comparing the information collected in the 2011 Census with information from 2011 reference dates registers.

IV. Limitations in the Portuguese SPD construction process: privacy and protection

12. Access to registers is ruled by the Law of the National Statistical System (NSS), i.e. Law No 22/2008 of 13 May 2008. The National Data Protection Commission has produced a set of recommendations to safeguard persons’ anonymisation. 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 postal code).

13. In Portugal there is no unique PIN. There are four numerical identifiers: the CR number (NIC) or the IR number (in the case of non-citizens), the Taxes Register number (NIF), the Social Security number (NISS), and the National Health Service number. The registers have one, two, or three numerical identifiers, depending on the administrative source, and not always covers all of the records (e.g. the CR only contains NIC; the Taxes Register only contains NIF; Education Register contains both NIC and NISS, but in the first case for 90% of the records and in the second case for less than 70%).

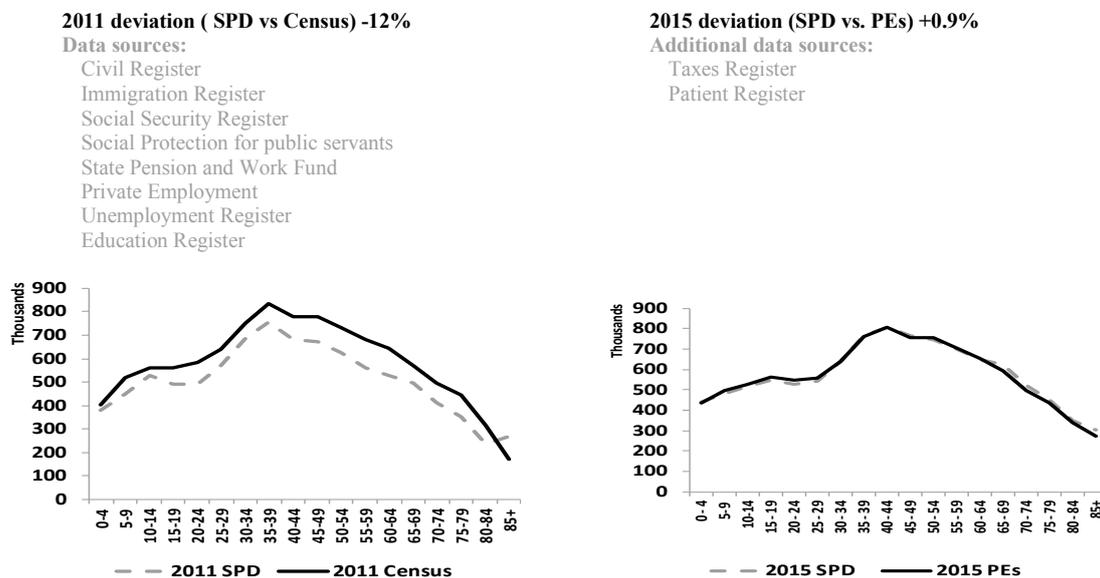
14. In the absence of a unique PIN, which would increase the accuracy of linkage and improve matching rates, it was necessary to build linkage keys, essentially through deterministic methods, based on personal characteristics.

V. Results and quality assessment

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

16. Compared to the first exercise, held for 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.

Figure 1 – Resident population in Portugal, by age group

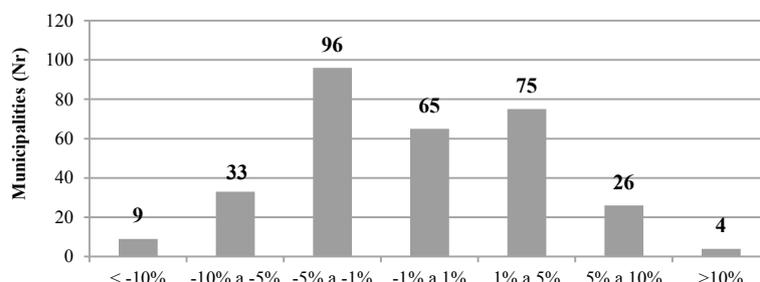


17. The improvement in the 2015 SPD when compared to the 2011 edition is due to two factors: the incorporation of the Taxes Register, with 9 million records, was relevant to determine the 'signs of life', particularly for population segments that do not carry out any kind of business, do not receive any social benefit and are not studying, being, however, registered as part of the household in the Taxes Register. The second factor was the improvements on the record linkage process.

18. Also at regional level (NUTS 2), deviations in the enumeration of the population are consistent with the 2015 PEs, from -2% to 0.7%.

19. Considering the geographical level ‘municipality’, there are minor differences between the 2015 SPD and the PEs for the same year for most municipalities (figure 2). For 77% of the country’s municipalities deviations between the population estimated by the SPD and the PEs are lower than 5%.

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



20. The assessment of deviations for small areas was conducted through a test survey, the 2016 TC, in 5 parishes (LAU2) of 5 municipalities (LAU1). The sample size was 45 thousand dwellings (1% of housing units of the National Dwellings Register), and around 70 thousand persons answered the survey. The comparison of the 2016 TC and the 2015 SPD results led to the following conclusions:

- Enumeration of the population, based on registers, still has some limitations. The 2015 SPD overestimates the resident population in all 2016 TC parishes, with deviations between -5.7% and -21.8%;
- Although the size of the population at the parish level shows some differences, the structure and characterisation of the TC parishes’ population given by the 2015 SPD is quite consistent with that enumerated in the 2016 TC;
- The microdata equality rate obtained between the 2016 TC and the 2015 SPD (77%) is not satisfactory, as a consequence of the limitations relating to the availability of the persons’ full name and address;
- Administrative data approximates with high-quality the information collected in the 2015 TC parishes, with a very high level of equality for the selected variables (marital status: 95.1%; place of birth: 96.8%; citizenship: 99.3%; school attendance and education attainment: 94.1%; labour status: 79.4%).

VI. Conclusions and outlook

21. The construction of a SPD allowed, 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 are encouraging, but not enough to not run a survey in 2021, because:

- 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);
- The currently available administrative information is not sufficient to respond to all the variables provided by the census. Key domains such as households and family characteristics or education attainment still cannot be characterised based on the available registers.

22. However, considering that only one methodology based on administrative information will allow for census information to be available more often, 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. The on-going investigation benefits from the good institutional cooperation and the appropriate conditions in public administration, as a result of the country's modernisation process in the past few years, which have led to the availability of new registers.

24. Work is still in progress for the transition to a register-based census. The setting-up of a more favourable legal framework is the key to overcoming problems related to data linkage, increasing the accuracy of population estimates to small areas, and liaising between the Statistical Population Dataset and the National Dwellings Register.

ANNEX H

Evaluating the potential for moving away from a traditional census – a case study for England and Wales

(Office for National Statistics, England and Wales)

I. Background

1. In May 2010, ONS began the ‘Beyond 2011 Programme’ to review the future provision of population statistics in England and Wales in order to inform 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 on-going household surveys.

2. On the basis of the research and evidence collected, the then National Statistician recommended, in March 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.*

...[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 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 to government and other users that the alternative can produce high quality information at a lower cost, and can do so on a more regular basis.

5. However, such a change in approach is challenging in England and Wales 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 ambition, described above. It is also in line with ONS’s Strategy to be at the forefront of integrating and exploiting data from multiple sources, making greater use of administrative data across all statistics.

II. 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 housing, households and people) from an Administrative Data Census. Doing this will require a combination of:

- *record-level administrative data held by Government;*
- *population coverage survey;*
- *population characteristics survey.*

a. Record-level administrative data held by Government

8. ONS will need access to a range of data held by Government. Access will be required at a 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 Matching Anonymous Data) 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, for example employment by qualifications at small geographic levels within a local authority.

b. Population Coverage Survey (PCS)

9. ONS anticipates the need to conduct a coverage survey, similar to a Census Coverage Survey, to measure and adjust for under- or over-coverage on administrative data and to enable the production of high-quality statistics about the size of the population. A PCS may cover approximately 350,000 (1%) households on an annual basis, as described in Paper M8. Further work is required to refine the detail of the survey and the methods to subsequently produce estimates using the PCS and administrative data.

c. Survey to produce estimates about characteristics of the population and households

10. ONS currently anticipates the need for a separate survey about 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 on administrative data, in a similar way to the PCS.
- A survey could also be used to provide direct survey estimates for topics that are not available on the administrative data (such as hours of unpaid caring). For such topics, it might be possible to produce estimates only at LA level³³.

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

d. Other sources of data

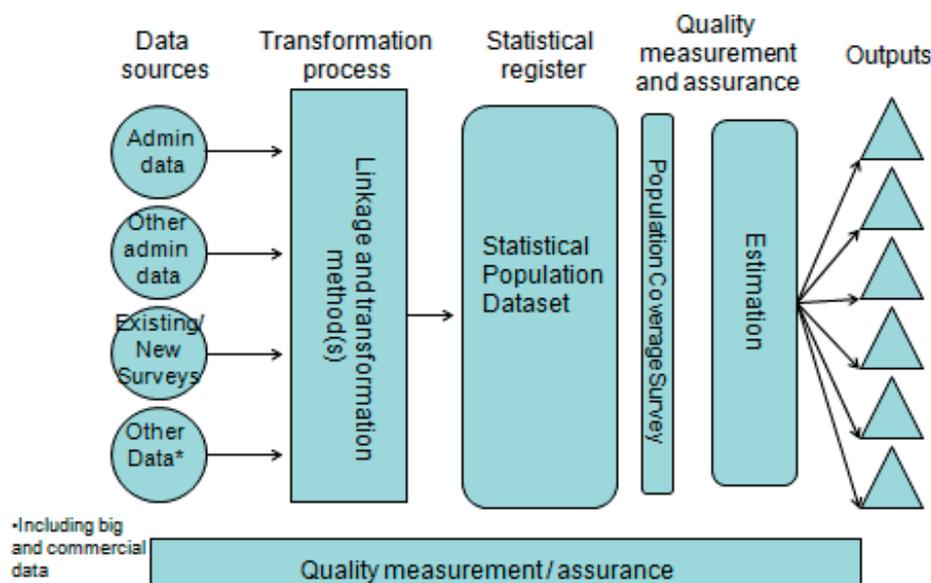
12. Other data (such as 'big data', private sector or commercial data) may also be necessary to produce some types of information traditionally produced by the census, such as commuting flows by using data from mobile phones, or information on tenure by using information from property websites.

13. These different sources of data will 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. 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.

³³ Geographical area with a total population of between 2,200 and 1,074,000 people.

e. Adapting common framework for England and Wales approach

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



III. 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:
- i. Accessing the range of data needed to produce outputs that are currently provided by the ten-yearly census;
 - ii. Linking together lots of independently collected data accurately whilst preserving the privacy and security of the data;
 - iii. Developing methods that can transform the linked data into outputs that meet the needs of users;
 - iv. Making an Administrative Data Census acceptable to key stakeholders, for example by providing value for money, and providing reassurance that data will be kept safe through this approach.
16. To address these challenges, the following would need to be in place.

a. Rapid access to existing and new data sources

17. 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 these requirements.

b. The ability to link data efficiently and accurately

18. All countries that have moved away from conducting a five- or ten-yearly census have moved to a combined or register-based census 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 linking. These registers

also aim to provide complete coverage of the population, which administrative data does not always provide.

19. In the UK we do not have ID cards or a population register. Instead, as described above, an Administrative Data Census would involve linking together 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.

20. 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 DWP and HMRC use the National Insurance Number, while GP Register data uses NHS number and School Census uses a unique pupil reference number. We therefore need methods that can link together these independent data sources accurately to enable the production of high quality statistics. An additional challenge is to do this while preserving the privacy and security of the data.

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

21. Accessing and linking data is only part of the puzzle. We need to deliver 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 told us that we need to develop statistical methodologies that:

- provide robust estimates about the size of the population and the number of households;
- provide estimates about population characteristics at a point in time to allow similar areas to be compared with one another;
- provide 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).

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

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

23. 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. 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.

e. Value for money

24. 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.

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

25. 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, we have published [Research Outputs](#) on:
 - [The size of the population](#);
 - [The number of households](#);
 - [Personal income using Pay As You Earn \(PAYE\) and benefits data](#).

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. We publish a [short summary](#) of the feedback received from each set of outputs, demonstrating to users how we are taking on board the feedback.

- 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 against 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 currently planned to take place in 2017, 2020 and 2022.

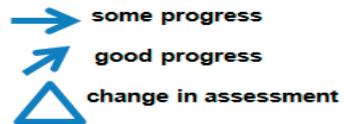
26. In June 2017, ONS published the second assessment of its progress towards an Administrative Data Census after 2021. Figure 1 shows the outcome of this assessment. The full assessment can be found [here](#), and is supported by evidence and a description of what will be done in the future to improve the assessment.

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

Evaluation Criteria		2016 Assessment	Progress made since 2016	2017 Assessment	Expected progress by 2018	Expected assessment by 2023
Access to data						
Ability to link						
Ability to meet information needs of users	Population estimates					
	Household and families estimates					
	Population and household characteristics					
	Housing characteristics					
Acceptability to stakeholders						
Value for money						

Key to chart

RAG status



ANNEX I

Permanent Census and Social Surveys Integrated System in Italy (Italian National Institute of Statistics: Stefano Falorsi and Gerardo Gallo)

I. Reference frame

1. Starting from 2018, the population Census in Italy will leave the decennial “door-to-door” enumeration for a “register-based” system which will combine administrative data and sampling surveys. The Italian National Institute of Statistics (ISTAT) is implementing a new Census plane which will involve a yearly sampling survey so called “Censimento permanente” (or permanent Census).
2. The traditional Census usually accounted a significant difference between the usually-resident population and the individuals recorded in the local registers. According to the Italian law, the differences between enumeration outcomes and population registers should be verified, at the end of every Census, and corrected by the municipal officers.
3. Despite the strong efforts made during the 2011 Census to increase the coverage and the use of different modes of data collection, some sub-populations remained very difficult to count. After the 2011 post-enumeration survey, which involved more than 320,000 individuals, the Italian national institute of statistics (Istat) estimated that about 650,000 usual residents in Italy had not been counted and that about 80% of them had foreign citizenship (Istat, 2014). These results, together with the huge cost of 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 so-called “Anagrafe Nazionale della Popolazione Residente”. This centralized register will have a unique ID number which will provide a link to other administrative sources being of great importance in the creation of a register-based Census.
5. From the beginning of the 2011, Istat has been collecting individual and households’ data from municipal population registers on a yearly basis. These data were used to send the 2011 Census form to households. Currently, variables collected from population registers (i.e. citizenship, age, gender and place of residence) are used to fit the sampling social surveys (i.e. Labour Force, Living Conditions, EU-SILC and Consumer Expenditure surveys).
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 so-called “ANagrafe Virtuale Statistica” (ANVIS). This statistical population ensures higher level of quality than population register microdata and represents a solid frame to implement a register-based Census.
7. Since 2015 Istat has organized 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 built an integrated system of available administrative sources, called ISM (Integrated System of Microdata). This system identifies each object (individual 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. The administrative sources already stored and integrated in the ISM are the following: population data coming from municipal registers, ANVIS, foreign population with permits to stay, employees and self-employed workers, compulsory education students, university students, retired people, non-pension benefits records, and individual data on income and taxation. 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 robust signals of presence on a territory³⁴. So, the comparison at micro level between ANVIS and individuals of administrative source with labour or study activities could provide the usually-resident population at municipal level³⁵. This last database may represent the second frame to support the register-based Census in Italy.

8. The managers at Istat favour a register-based Census, but the final decision on which basic data structure should be used for the permanent Census will depend on a quality assessment of the sources and also the processes used for deriving all statistical outputs. For this reason, Istat will implement in 2017 a pilot survey to evaluate the coverage level of the basic data structure that should be used for the permanent Census in Italy.

9. Other than population counts by gender, citizenship, age and place of births, the population Census should provide a set of hypercubes traditionally produced for Eurostat every ten years on socio-economic variables (employment status, educational level, commuters, etc.). For this purpose and with the aim to follow the changes of EU social statistics, Istat is implementing the permanent Census and Social Surveys Integrated System.

10. This contribution presents a possible scenario for the integration of social surveys which arises from a specific strategy associated with a specific sampling design. The whole purpose is to achieve a complete integration of the system of social surveys and ensure maximum integration with the registries system available in Istat. The performance of the scenario integrating the social surveys is compared with the scenario in which the social surveys are pooled together.

II. Pooled Sample and Master Sample

11. The Census and Social Surveys Integrated System (CSSIS) is a complex statistical process exploiting and integrating the information arising from registers and surveys on socio-economic variables. It is designed as a two phases Master Sample (MS) design based on a set of balanced and coordinated sampling 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.

12. The PR is the backbone of the system for the production of social statistics, with a row for each target unit referred to a usual resident persons (living in households or in institutional households). For each target unit, the core information, coming from demographic sources, is extended to all the basic social variables (coming from administrative sources and/or social surveys) among which employment status, economic and health conditions.

13. For an optimal design of the CSSIS to support the PR, it is useful to classify the variables included as *totally, partially or not replaceable* ones. The first class contains the variables for which the administrative sources provide the correspondent proxy information. At

³⁴ These signals may be handled according to a monthly criterion and could be managed even by considering the concept and the definition of “usual-residence” of the European Union Regulation n. 1260/2013.

³⁵ In these data some hard-to-count sub-populations, like homeless and other people who do not have usual residence in the same place during the year, are not included.

the end of the statistical process, including editing and imputation for partial non response, these variables are considered complete, because they are available for all units in PR, and accurate, having a good level of coverage and quality. Administrative sources provide the corresponding proxy information, also for partially replaceable variables, but these are considered complete and accurate only for a subset of the target population. For the remaining subset of the target population, this type of variables is unknown or cannot be considered accurate, because of the failure of the synthetic model of imputation. For instance, this is the case of the “Regular employed in the Labour Market” variable. Finally, for not replaceable variables the corresponding *proxy* information coming from administrative registers is not directly available. Then, for these variables, target parameters can be estimated by means of sample surveys and exploiting the auxiliary information coming from the PR. 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 CV of the estimates should be kept lower than a chosen threshold; (b) consistency that the data obtained by combining estimates in different ways must produce the same results.

14. The main scope of the CSSIS is filling the informative gap of the PR for the estimation of target parameters referred to partially replaceable and not replaceable variables on social and economic data. To this aim the MS design is planned for exploiting together (*pooling*) and in an efficient way all the common information (target and auxiliary variables) observed by the different sampling 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 separate estimation strategy for each survey. Within this context the harmonization of the common variables – i.e. *core structural* variables (which are target variables for all surveys) and *harmonized* variables (which are target variables for more than one survey) - and the harmonization statistical production process are crucial issues.

15. The permanent census process is integrated within this context. It is a register-based census using both the information produced by the PR, for replaceable variables, and the CSSIS for the remaining variables. More precisely, the permanent census is aimed to produce, starting from 2018, both annual data for a subset of the target parameters (hypercubes), and multi-annual data for the complete set of hypercubes produced traditionally every ten years by the population census, possibly pooling sampling data over a period of consecutive years. Of course, statistics for replaceable variables can be disseminated every year.

16. As regards the basic objectives of support to the permanent census, the first phase of MS design is based on two different component samples, namely A and L.

17. The component A - based on an area sample of Enumeration Areas (EA) or selected by an Integrated Address File (IAF) - is designed to satisfy the needs of estimating under-coverage (SU) and over-coverage (SO) rates of the PR at national and local level for different sub-population profiles like sex, age classes, 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), taking into account both under-coverage and over-coverage.

18. The component L - based on a list sample - is designed with the purpose of: (TI) 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 (SOI) of the PR. On the other hand, the component A could be designed to also meet the target TI. In turn, the component L could also be modified to improve the estimation process with the focus of estimating via indirect sampling some aspects of Undercoverage SU.

19. More in general, the first phase survey should be focused on the following aims:
 - (a1) obtaining sampling information on partial and not replaceable core structural variables useful for the PR;
 - (b1) establishing a first contact with the sample households, a subsample of which will be re-interviewed the following year in the second phase. The first contact could be managed in order to reduce potential second-phase non-response;
 - (c1) obtaining updated contact information on telephone numbers and e-mail addresses, which is not available on the sampling frame, may allow to carry out less expensive interview techniques (CAWI or CATI) in the second phase.

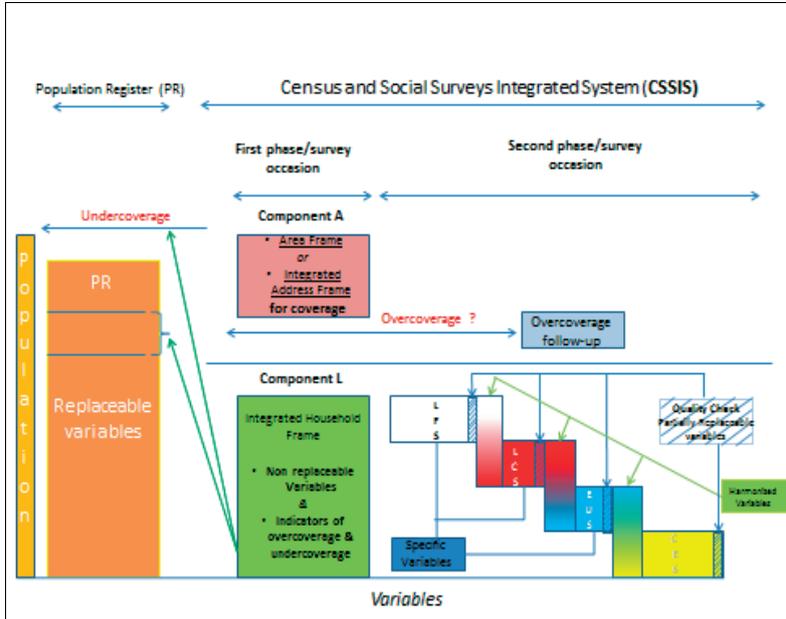
20. From the first phase sample a set of negatively coordinated samples of households can be selected for the second phase surveys, aiming:
 - (a2) to provide information on harmonized and specific socio-economic variables currently observed by Labour Force (LFS), Living Conditions (LCS); EU-SILC (EUS) and Consumer Expenditure (CES) surveys;
 - (b2) to confirm the common structural variables already surveyed in the first phase interview. These surveys are currently based on stratified two stage sampling designs (municipalities-households), and they are planned, selected and realized separately. For these reasons, it may happen that important differences are observed among the estimates related to the same variables observed from the different social surveys even if the definitions and the wording of the related questions are the same. Then, in order to be able in the future to pool the same information coming from the different two phase 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.

21. Furthermore, the first phase sample can be stratified or balanced, using variables in the PR, to identify areal or structural subpopulations supposed to be problematic for coverage or subjected to structure and characteristics changes in the short time 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 (for instance the self-declared employment condition) useful to be used in the estimation process.

22. The component L of the first phase sample should be based on a yearly sample size of about 2000 municipalities out of 8100 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.

23. For an overview of CSSIS, based on the two phase MS design, see figure 1 below.

Figure 1 – An overview of the CSSIS.



24. The administrative records support mainly 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, 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 or addresses belonging to the CPF.

25. 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 integrating the addresses belonging to CPF with addresses related to new buildings.

26. The main difference between the components L and A sampling schemes with 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 hypotheses below the DSE are completely satisfied.

27. Referring to similar international experiences, for the definition of a general master sample design for social surveys, analogous designs have been proposed by Eurostat considering a modular approach for the design of integrated social surveys. Furthermore, the ABS is designing an integrated system of investigations very similar to what has been described here. In that case, this survey system, called Australian Population Survey, does not replace the census.

28. The design with two components supporting the register census is similar to what ONS has been studying for the register-based census supposed to start in 2023 after the 2021 census run (ONS, 2016). In particular, in 2021 the ONS will conduct a traditional census and, at the same time, will carry on a parallel census run based on the construction of an integrated population registry using several administrative sources and two investigations with characteristics similar to those of the components L and A of the Italian strategy. 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.

29. Another international experience showing similarities with what is planned in Italy is the Israelian rolling integrated census. They use an integrated register which is adjusted by means of weights computed by means of an EDSE (Pfeffermann, 2015).

30. In order to define the final survey design of the CSSIS a pilot survey will be conducted in the next months. In particular to test the quality of the IAF for estimating register coverage weights and first and second phase response rates together with the need to get a fine tuning operational aspects of the survey.

III. References

31. Battese, G. E., Harter, R. M., Fuller, W. A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401): 28–36.

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

33. 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 to the European Conference on Quality in Official Statistics Q2014. Vienna: 2-5, June: 1-14.

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

35. Ioannidis, E., Merkouris, T., Zhang, L.C., Karlberg, M., Petrakos, M., Reis, F. and Stavropoulos, P. (2016). On a Modular Approach to the Design of Integrated Social Surveys, *Journal of Official Statistics*, 32(2): 259–286.

36. Kim, J.K., Rao, J.N.K. (2012). Combining data from two independent surveys: a model-assisted approach, *Biometrika*, Vol. 99(1): 85-100.

37. ONS (2016). Annual assessment of ONS's progress towards an Administrative Data Census post-2021, downloadable at <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments>.

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

* * * * *