**Census statistics – ensuring quality in a time of major change**

**David Thorogood**

**Eurostat**

**Abstract**

Censuses have long been a key part of official social statistics at both national and European level. However, the production and dissemination of census-type population statistics are now undergoing major changes in many countries, with related changes to the associated European statistics. Methodological developments have focussed on a more extensive use of population registers and other administrative data sources – either alone or in conjunction with sample surveys.

Amid the many opportunities that are emerging, there are technical and organisational challenges to ensure the effective management of statistical quality as part of these changing statistical processes. Often, the administrative sources used are outside of the direct control of the national statistical institute, with the accuracy of the source data managed primarily to meet the needs of its core administrative function. This implies that, compared to now, statistical quality will be assessed and managed in different ways and at different stages in the overall process.

These statistics must remain relevant to users. At European level, future requirements for harmonised census-type statistics must be developed to take into account user needs for flexibly-defined and geographically detailed outputs, as well as the ongoing changes and likely continuing diversity in national data sources and methods.

**1. Introduction**

1.1 This paper examines some of the quality considerations related to the recent trend seen in many countries in Europe and beyond for an increased use of data from administrative sources as part of the population and housing census. The use of administrative data sources offers a number of potential advantages over the traditional door-to-door census enumeration. There are though, legal, organisational and methodological issues that need to be addressed in order to maximise the value of the administrative data and to maintain the value of the census for its main users. Quality issues related to administrative data may be harder for the National Statistical Institute (NSI) to detect, quantify and manage. The challenges may be different to those previously faced by statisticians with experience of a traditional census.

1.2 In addition to the production of census data for national purposes, countries in the EU and EFTA also need to provide census data to Eurostat. The ways in which the decennial European census programme will need to evolve to facilitate the use of administrative data in censuses are also examined.

2. **The central importance of the census**

2.1 The census is frequently seen as the core of official social statistics. It provides a hugely detailed, albeit infrequent, snapshot of the demographic and social structure of a country. In addition, the census provides a basis for other - more regular - exercises, for example, through the rebasing of the annual demographic estimates of the population and as an input to the sampling and weighting procedures for major social surveys.

2.2 The census is a high profile event. Both the collection and the release of census data often attract considerable media attention. There can be controversy over aspects such as the topics and categories covered, the legal obligation to respond, privacy issues, and the interpretation of the results. In many countries, the census results are used in the allocation of central government money to regions, or as an input to decisions on electoral mapping and the number of representatives that an area may elect. This means that there is a particular pressure on NSIs to ensure that the census exercise is undertaken correctly, and that the results can be shown to be reliable.

2.3 Although there have been some variations and improvements between countries and over time, the overall census methodology has changed little for many decades. The established characteristics of the traditional or conventional census model with a 100% door-to-door enumeration have remained largely unaltered and could be recognised across many countries. Where there were changes in procedures and methods, this was usually gradual. In many countries, a census carried out in 2000 would often be conceptually similar to one in 1950 or earlier.

2.4 In recent years though, the rate of change in census methods has greatly accelerated. New approaches to the production of census-type data have been investigated and put in place in many countries in Europe and beyond. A factor often seen in these new methods is the increased use of data from administrative sources, sometimes supplemented by data from existing or ad hoc sample surveys. In 2001, and more particularly in 2011, the number of countries making use of administrative data, either as the only data source or as a key input to the census process, has greatly increased. By 2011, approximately half of the 40 countries in the United Nations Economic Commission for Europe region chose to move away from a traditional census [1], with most of these countries making use of administrative data. Among the EU and EFTA countries, the proportion using administrative data was higher.

2.5 This trend is continuing, with an even larger proportion of countries either considering or having already adopted administrative data sourced censuses for 2021 and beyond. Even where little or no use of administrative data is planned, significant changes to the census procedures are often being considered, including rolling censuses and greatly simplified 100% enumerations combined with large-scale sample surveys. The traditional census itself in the future may be less a door-to-door enumeration but instead a largely internet response-based exercise.

2.6 Five essential features have been set out that distinguish the census from other official statistics exercises: individual enumeration; simultaneity; universality; availability of data for small areas; and defined periodicity [2]. As countries consider moving away from the traditional census enumeration, these essential features can act as a useful measure against which to assess the appropriateness of the alternative methods and data sources selected. These essential features are also adopted within Regulation (EC) 763/2008 that provides the framework for the supply of census data to Eurostat [3].

**3. Pressures for a change in census methods**

3.1 Although the reasons for these changes to the census vary between countries, a number of common themes can be seen. A traditional door-to-door enumeration is an expensive and human resource intensive operation, often requiring the recruitment, training and management of a very large number of temporary staff. The costs of a traditional census are also concentrated in one or two years around the census date, which can cause particular budget problems.

3.2 In many countries, the level of public cooperation with the census enumeration process has declined, even though in most cases there is a legal obligation to provide information for the census. Even where people do not seek to avoid the census, it can be harder now for enumerators to make contact with people. Social changes such as the growing number of single person households and a more mobile population where people may divide their time between different addresses can man that the enumerator may have to return frequently to a dwelling to make contact with the occupants or indeed to find out if these dwelling is vacant.

3.3 As well as to creating specific data quality problems (in particular, biases in the results as certain demographic and social groups are most likely to be missed), these difficulties can greatly increase the costs of a traditional census operation, requiring repeated visits by enumerators. This is happening at a time when there is growing pressure on public finances, and a need to reduce the costs of the census. The use of administrative data has been seen by many NSIs as a way to avoid the costs and growing practical difficulties of the full census enumeration. However, it can be noted that adoption of administrative sources may require such extensive methodological development and testing, that the overall cost savings may be less than expected.

3.4 The scale and costs of the traditional census enumeration mean that it is generally undertaken only once every ten years (or in a few countries every five years). Such infrequent data do not fully meet the needs of some users, particularly where there the size and structure of the population may be changing more quickly. The use of alternative census methods based on administrative data, possibly combined with sample survey data, makes it feasible to consider more frequent, even annual, statistics.

**4. Planning a change to a greater use of administrative data**

4.1 Given the importance of the census, a decision to make large changes to the methodology and data sources requires extensive research, consultation and planning. The availability of appropriate administrative sources needs to be reviewed and the likely impact of the changes to the census in terms of outputs must be discussed with the main statistics users. An example of this can be seen in the major consultation exercise organised in 2013/14 by the UK Office for National Statistics regarding the 2021 census in England and Wales [4]. Depending on the national situation, it may be the cases that, despite the apparent advantages of moving to an administrative sources census, the availability and coverage of the administrative sources in that country may not be sufficient to ensure that the quality of census outputs can be maintained. As in the England and Wales example, the result of the research and consultation may be that a more traditional census enumeration, with some amendments such as the use of online enumeration, should be retained.

4.2 Any significant change to the census methodology and data sources implies a degree of risk. There is a danger that the new census procedures will fail to provide key users with the statistics that are required. Looking at those countries that have made such a change, or that are considering it, a number of common approaches and actions can be seen that are likely to minimise the risks associated with this change. Clearly, the specific national pressures and opportunities for change will vary. However, a number of countries have published very detailed accounts of the technical and user consultation that has been held, as well as the ways in which potential methods have been developed and tested [5].

4.3 Perhaps central to these actions is a comprehensive and detailed review of the available administrative sources in the country. This goes beyond simply knowing the population covered by the source and the data items held. There must be a detailed understanding of the underlying administrative process and the ways in which information is collected, coded and classified, quality checked, managed and updated. The impact of foreseeable future changes to the administrative systems on the availability of data also needs to be considered. In parallel, the holder of the administrative data must be helped to develop an understanding of the priorities and requirements of the NSI. Any changes need to be underpinned by agreements between the NSI and administrative data holders, appropriate legal frameworks and high level political approval.

4.4 Even without a major change in census methodology, a census takes years to develop and plan. Where a completely new methodology is being developed, it may even be necessary to begin preparations and initial testing during or immediately after the previous census exercise (when there are new conventional census data against which the output of a pilot administrative data project can be assessed). In some cases, the use of administrative data may be expanded gradually over several census rounds, moderating the potential risks and allowing the NSI to gain experience and refine procedures. In this context, it is interesting to note that several of the Nordic countries, often seen as being among the most advanced in the statistical use of administrative data sources, developed their fully register-based censuses over several decades [6].

**5. Specific quality considerations related to the use of administrative data**

*Types of administrative data*

5.1 A register may be seen as a dataset that is held (usually by an official body) to allow an administrative process or function to be undertaken. Most commonly associated with the census are population registers – in countries where these exist – that aim for a variety of administrative purposes to record all persons living in a country or area. However, there are other types of registers that can also provide important information for the census. For persons, these may include registers relating to social security, taxation, employment and education. Equally useful are registers of addresses, buildings or dwellings that can provide information about the housing conditions of the population. Similarly, registers of commercial enterprises or employers can provide information on the employment and economic situation of the population.

5.2 There are some other administrative data are not normally regarded as being from registers, but which may also contribute to an administrative census process. Whereas a register may be seen as a comprehensive list (of people, buildings, enterprises etc.), these other types of administrative data are usually produced as a by-product of administrative actions or processes, such as the granting of a permission to construct a building (from which the age of a building could be derived).

5.3 The use of administrative sources means that certain very specific quality issues that must be addressed. In many cases, these will be different from the quality issues that NSI census experts have experience of managing. The adoption of an administrative data census means that different approaches to the management of the statistical process are necessary to limit, assess and report on quality problems. It is not suggested that an administrative sources census presents an increased risk of quality problems compared to a traditional census. It is clear however that the types of quality problems that must be addressed differ considerably.

5.4 The exact types of problems to be addressed will vary according to the types of administrative sources used, the particular strengths and weaknesses of these sources in a particular country, and the ways in which the administrative data are used. One common issue though is that, compared to the traditional enumeration, the administrative data sources will often be outside of the direct control of the NSI.

*Using multiple data sources*

5.5 One important aspect of the census is the wide range of information that it offers on different topics. Although countries differ as to the statistical topics covered, there are international and European standards that countries should follow. For example the European legislation on the census requires data to be supplied on a range of population and housing topics, covering for example, age, sex, marital status, citizenship and country of birth, the type and size of the household, occupation, activity status, educational attainment, previous residence abroad, type of dwelling, and availability of facilities such as toilets and heating. In a traditional census, the NSI can design the census questionnaires to include questions by which data on all necessary topics can be produced or derived. However, it is highly improbable that a single administrative data source would provide information on all of these topics. This means that data usually needs to be taken and combined from several administrative sources. Countries have adopted different approaches to combining data, but have often developed a composite register database for census purposes. The availability of a common administrative, personal or ID number for people, buildings and enterprises across different registers greatly facilitates this process.

5.6 Depending on the availability and content of administrative sources in that country, it may be impossible to find appropriate administrative sources to cover particular topics. For this reason, many countries have combined the use of administrative data with information from sample surveys.

*Concepts, classifications and accuracy*

5.7 Administrative sources are designed and managed primarily for the efficient undertaking of the associated administrative processes. The administrative information must be appropriate for the administrative process in relation to individual persons, buildings or enterprises. In contrast, the statistical use of the data is intended to provide information on the population. This impacts on the type of information that is collected and held, how the information is processed and coded, and on how the accuracy of this information is assessed and managed [7]. This can also result in important conceptual and definitional differences in the statistics. A traditionally enumerated census aims to provide statistics on the population present at a reference date and/or the usually resident population. In administrative data, people are normally counted at the place where they are registered (rather than necessarily where they really live). People who cannot register – perhaps because they are undocumented migrants – will be excluded.

5.8 As noted above, some key census topics or classifications will be missing from most or all of the available administrative data. This can particularly be the case for census topics that require information on several individuals to be linked – as required to produce statistics on household and family size and composition. To take one example, the available administrative sources may contain separate records showing two adults and two children living in the same dwelling. However, the administrative data may not have sufficient information to define whether the two adults live as a couple and whether they are both parents to one or both of the children. It may not be possible to distinguish whether this is a couple living with their two children, a couple living each with a child from a previous relationship, or indeed a father living with his adult daughter and her two children. In a traditional census, these problems are addressed by asking about each person's relationship to either the first person recorded on the census form or to all other persons living in that dwelling. Using available administrative information, it can be possible to infer the household and family structure but the overall result may be less reliable. Alternative definitions of a household may need to be applied to take into account the lack of complete information about the links between the different people living at that dwelling.

5.9 While the presence or absence of particular information in an administrative record can be obvious, the accuracy, definition and underlying meaning of the information that is available may be harder to assess. If the administrative process is managed effectively, it might be assumed that the information collected is accurate. A noted above though, the focus of the administrative source is primarily on the administrative purpose. Information will be recorded and coded in a way that is more useful for the administrative purpose. The concepts that are recoded may appear to match those addressed by the statistics, but there can be differences in definition. For example, a traditional census may record citizenship or nationality on the basis of a person's self-declared answer to a question. It is unlikely that a census respondent would be asked to prove their citizenship. In contrast, a register may well require the person to show evidence of citizenship when registering, or citizenship may be officially assigned to the person based on other information given. This may not be significant in many places, but may be important in places where territory, national identity or citizenship rights are controversial or disputed.

5.10 There is also a question as to whether that information remains accurate – whether the information is kept up-to-date. This will depend on the requirements of the administrative process, as well as on any obligations, incentive or disincentives that individual persons may have to keep their information up-to-date. A particular problem seen with population registers in a number of countries is that of register inflation. There may be strong incentives for a person migrating to a country or region to make sure that they are included in the population register or another administrative register. Registration may be necessary to gain access to social security or health services. The incentives to de-register for emigrants may well be less, and indeed there may be benefits in trying to remain registered if the person expects to return to the country in the future. Particularly where registers are managed at local level and where a larger registered population may result in greater public finding for that area, there can also be little incentive for local officials to ensure that people no longer resident in the area are promptly de-registered. There may be similar problems associated with ensuring that the individual data items attached to a record remain up-to-date. Information that was correct at the time of collection may have changed at the point when data are extracted for the census. For example, in the case of housing or building registers, the number of rooms or type of water supply may well change over time. There will be variations in how well and how frequently this information is updated.

***6. European and wider international considerations***

6.1 Although censuses are principally undertaken to meet national needs, countries are encouraged to comply with relevant international recommendations. EU Member States must also comply with European statistical legislation on the census statistics that must be supplied to Eurostat [8]. For the 2011 census round, there was a comprehensive package of legislation involving a framework regulation of the European Parliament and the Council, as well as three separate Commission implementing regulations that defined aspects such as the definitions of census topics, breakdowns, the tabulation programme, metadata and quality reporting. The legislation is intended to result in census data that are comparable between countries. An output-orientated approach is taken, in that countries remain free to use any appropriate data sources, technologies and methods to produce these comparable census results, according to national circumstances and preferences.

6.2 For the 2021 census round, it has been agreed that the existing framework regulation from 2011 will still be used. However, revised implementing regulations that define the details of the European census outputs will be prepared. Eurostat will shortly begin consultations with Member States to begin the process of drafting these implementing regulations. A key aspect of these consultations will be to facilitate the use of administrative sources in those countries that wish to do so by trying to ensure that the detailed requirements of the European census programme can be met with an administrative-source based census.

6.3 Eurostat has also begun consultations on the types of population and housing data that countries might be able to supply to Eurostat beyond the 2021 census round. Given the ongoing and expected changes to data sources and methods in many countries, as well as evolving user needs, it is possible that this longer-term consultation and development may reflect moves in some countries away from a single decennial census exercise towards more frequent data production. Discussions on this between Eurostat and Member States are still at a very early stage, and will focus on reaching workable and practical solutions to the longer-term needs for European-level census-type data on population and housing.

6.4 The Conference of European Statisticians Recommendations for the 2010 round of censuses included an annex intended to provide methodological information and support to countries making use of administrative data sources and other ‘non-traditional’ census methods. Similarly, the United Nations Statistics Division's Principles and Recommendations [9] for censuses at global level provide information and practical considerations for countries wishing to use different census data sources. Updates for both of these sets of recommendations are currently being drafted. Reflecting the growing use of administrative sources in some countries, and increasing interest in their use in others, the updates will include expanded information relevant to the use of administrative data sources for censuses.

**7. Conclusions**

7.1 Many countries in Europe and beyond are reconsidering the data sources and methods used as a basis for the population and housing censuses. Some countries have already made changes; others are planning changes for 2021. Depending on different national circumstances, the use of registers and other administrative data sources can result in good quality results, either with a wholly register-based census or in combination with sample survey data. However, the adoption of administrative sources cannot be regarded as an easy option, presenting particular methodological challenges and risks. In the short- to medium-term, the use of administrative sources may not necessarily be a significantly cheaper option than the traditional census, given the extensive development and testing that will be necessary.

7.2 Such major changes in data sources take time, involving extensive investigation of potential sources, consultation with users and with the holders of administrative sources, changes to the legal basis forte census and pilot studies. A gradual introduction and increase in the role played by administrative sources in the overall census exercise may be a preferable option. In some countries, appropriate administrative data sources may not be available.

7.3 At European and wider international levels, it is important that NSIs should be supported as they develop new approaches to the census. In particular, it is important that European requirements for census data should take into account the ongoing changes at national level.

**References**

1. Valente, P. (2010). Population & Societies - no. 467 – Census taking in Europe: how are populations counted in 2010?

2. United Nations. (2006). Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing.

3. Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008

on population and housing censuses

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:218:0014:0020:EN:PDF>

4. The Census and Future Provision of Population Statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority

March 2014

<http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations/national-statisticians-recommendation.pdf>

5. See, for example, the documents presented at the series of specialist meetings organised by the UN Economic Commission for Europe <http://www.unece.org/index.php?id=3214>

6. United Nations. (2007). Register-based statistics in the Nordic countries.

7. Walgren, A. & Walgren, B. (2007). Register-based statistics: Administrative Data for Statistical Purposes

8. Eurostat, European Commission. (2011). EU legislation on the 2011 Population and Housing Censuses – Explanatory notes.

9.United Nations. (2008). Principles and Recommendations for Population and housing Censuses – Revision 2