United Nations ECE/CES/2012/25

Economic and Social Council

Distr.: General
21 March 2012

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixtieth plenary session
Paris, 6-8 June 2012
Item 3(b) of the provisional agenda
The future of censuses and their role for national statistical systems

A census is a census is a census ?

Note by Eurostat

Summary

This paper analysis the concepts of what a census is and what it should be in the future. It briefly describes the background of census taking and focuses on the new challenges. The changes in today's society, administration and technology are overwhelming and the census needs to adapt. The paper elaborates on the challenges to help countries find common ground for going forward. Statistical offices need to step up the integration of censuses into social statistics, continue to focus on coverage, make the geo-referencing of population data precise, make comparative quality assessments of multi-source/multi-mode processes to develop the methods of census taking and develop and use common tools in order to improve efficiency of work.

The title of the paper is inspired by the works of the author Gertrude Stein (1874-1946).

I. The 2010/2011 census round and beyond

1. 2010 and 2011 were years of hard work for the Statistical Institutes: following the recommendation of the United Nations they took stock of their population and housing. The United Nations Statistics Division's 'Census Clock' estimates that 87 % of the world population had been enumerated by 1 February 2012, with most of the remaining population to be enumerated soon. The work and resources invested in this gigantic, worldwide exercise will undoubtedly pay off. Censuses are still the backbone of demographic and social statistics.

2. The end of one census round is just the beginning of the next. It was a wise decision to organise the conference 'Challenges for future population and housing censuses' soon after the end of the reference year 2011. Now is the time to carefully re-visit our concepts of what a census is — and should be in the future.

3. So what is a census ?

II. From where we come

A. The 'white gold' of official statistics

4. Let us go for a trip to Meissen, a wonderful town in Saxony, Germany. Meissen is renowned for its historic porcelain manufacture, formerly located in the 'Albrechtsburg' castle, which offers a splendid view over the Elbe river.[1]

5. The story goes that in 1701 a young alchemist, Johann Friedrich Böttger, pretended he knew how to make gold. The high and mighty of the time, always in need of the stuff, took an interest. 'August the Strong', the ruling Prince of Saxony, managed to get hold of Böttger, demanded that he make gold and locked him up for that purpose for twelve years, most of the time in the 'Albrechtsburg'. The poor man did his best. But he never could make gold.

6. Instead he discovered something very useful: how to make porcelain ! Once the process was sufficiently understood, August founded a porcelain manufacture which provided a source of income to himself, and many useful items of crockery for those who bought them. The region prospered with the new industry, and porcelain was dubbed 'the white gold'.

7. The poor alchemist Böttger was released in 1714, but died five years later from the consequences of his experiments with poisonous substances.

8. We agree: this is a dangerous analogy to a census operation! Although conventional census operations bring an extremely high workload for the Statistical Institutes, hopefully no statistician has ever been locked up for doing a census, or has even died from it. And though the time lags accepted for censuses are way beyond what we consider acceptable in any other survey, certainly no census operation will drag on for twelve years.

9. However, there are similarities. Though no-one will ask a statistician to make gold, far too often the expectations people have of censuses are sky-high. Let us be honest —

[1] The statements made in this paper do not refer to any existing enterprise or to any marketable goods or services.

even statisticians are tempted to raise them when they are trying to convince stakeholders to contribute money, work and goodwill to the census. The risk is that, just like the poor alchemist, we may get trapped in the illusion that statisticians can achieve the unachievable. There is this 'golden feeling': 'For once, statisticians have got it 100 % right in the census!'

10. It is, as we all know, a simplified and exaggerated expectation. It does more harm than good. Statistical institutes reflect carefully on what can be expected from a census — and where the limits are. They invest a lot to assess their users' needs before they start the census operation, and carefully assess the quality of the data once it has been done. We are, after all, not alchemists but professional statisticians.

11. Striving for the perfect, we have done the useful. Just like the dishes produced in August's porcelain manufacture were not made of gold but porcelain, our census outputs are not 'the golden truth'. The term 'census' comes from the Latin word 'censere', which could be translated as 'to estimate'. Despite our best efforts, and despite the undoubted quality of census data: like any statistical output, census data are subject to errors. Just like the dishes produced in August's porcelain manufacture, census outputs are still man-made products. They are not perfect. Indeed, perhaps better still: they are made to be useful.

12. The quest for perfection has yielded valuable results. For some purposes, our census outputs are of outstanding quality. They serve users who need population data with a high regional resolution or who want to focus on small groups of the population. Being the most reliable data on the regional distribution of the population, census results often serve to meet legal requirements, e.g. to delineate electoral districts, to determine the voting power of regions and states, or to allocate public funding. There is definitely a need for the quality that censuses provide.

13. Nevertheless, we accept that we do the useful, that we make dishes but not gold. That being so, we have to decide which products we want to fabricate. Just as the porcelain manufacture was run by managers (not alchemists) who were mindful of the market for tableware, so we have to have the overall market for social statistics in mind, with the whole range of products we offer. Each product must make sense and find a customer.

14. We also have to understand the interdependencies and potential synergies between censuses and other data collections. Far too often censuses are looked at as if they were singular achievements, separated from the rest. Though not ensconced behind the thick walls of a castle, census operations are often very distant from the rest of the statistical production. In the past, this might have been because of the sheer size of the operation, which required a separate organisation, financing and legislation. However, these reasons are not good reasons. Censuses are an integral element of our statistical output.

15. To draw a last parallel to our porcelain from Meissen. In the past, census outputs had to be handled with some care. They might not match well with the other dishes on the table. Comparing census data with data from other surveys has not always been self-explanatory, e.g. with regard to the definitions, specifications and breakdowns used, or the aspects of reference time. Though census outputs are detailed in regard to geography and small groups of the population, dedicated sample surveys might have taken more care to report precisely on particular aspects of topics. A dedicated survey on the labour force might take more care to measure employment and unemployment accurately and in accordance with international standards. Dedicated household surveys might pay more attention to the household relationships of persons. Users of census data must be aware that, just like the delicate tableware from Meissen, past census outputs do not always mingle easily with other data and might prefer exquisite usage and proper appreciation.

16. To preserve the value of the 'white gold' of official statistics, we have to rethink what aspects of our censuses are still desirable today (and why), and which of our beliefs have become relics of a past age. We have to be clear about what the *raison d'être* of a

census is and where integrating the census operation better into the statistical production system might offer better solutions for less money.

B. High quality comes at a high price — too high ?

17. The fundamental need behind censuses has never lost its force: to take stock of the population and its housing situation, and to have data of the precision and detail needed for administrative planning, governance and policy-making. There has been consensus on the census: it has been the straightforward answer to the simple request for 'authentic data you can rely on'.

18. Over the centuries, societies have accepted paying a high price to achieve this. Arguably, censuses are the most expensive operation statistical offices undertake. At times of tightening budgets, statisticians are having to explain whether there are no better opportunities to invest the money than on ten-yearly 'one-off' exercises — opportunities inside or outside statistics. Statisticians have to show that they are using the resources given to them to the best possible effect.

19. Moreover, like any data collection, censuses bring a burden to third parties. People accept some intrusion and open the door to enumerators who ask a wide range of questions. Municipal and other authorities accept cooperation and open their registers for statistical purposes.

20. Not always have censuses met with sympathy. The spectrum ranges from concerns about confidentiality, scepticism, refusals to respond, to outright public resistance. Today, governments and statisticians are attentive to these concerns, take them seriously and spend time informing citizens thoughtfully about why the census is 'their' census and how its results serve them.

21. The environment for doing censuses has not become easier.

C. Censuses need to adapt to changing environments

22. Despite their costs in terms of money and response burden, censuses have survived over the centuries. Apart from the constant quest for precision and detail, the main reason for this is that censuses have been adapted to changing environments.

23. *Societies change.* A characteristic feature of modern societies is the high mobility of people, within regions, between regions, crossing national borders. We all know that reporting internal and international migration is becoming increasingly difficult. Despite our best efforts even small reporting mistakes tend to accumulate to significant levels faster than in the past.

24. People live together in increasingly diverse arrangements. Today, a correct identification of the composition of households and families is more complex than ever.

25. *Perceptions change.* Whereas in the past most people were prepared to comply with the obligation to provide their data, more and more people have become wary in recent years and refuse to answer. Whereas data protection was an issue reserved to experts not so long ago, today parents tell their 10-year-old children about it — for good reasons. Statisticians have to make it clear that they respect the confidentiality of personal data, and explain why it is important for people to help make the census a success.

26. Quite different sensitivities must be taken account of. A question on a person's ethnicity might be perfectly admissible in one society, but raise aggressive reactions in another. While people in some countries are quite relaxed about linking data from different

sources to create (and even publish) personal profiles, others might be very restrictive in this regard. The sensitivities in our societies are not static but evolve. Time and again we statisticians have to re-assess our understanding of how people perceive our data production activities.

27. *Technology changes.* A few decades ago, punch cards were welcomed as an innovation to make data processing less labour intensive. Today, the share of the data collection which happens via the Internet is growing; data processing and dissemination is done electronically everywhere.

28. Statisticians have continuously and successfully adapted the population and housing censuses to changing environments. In the course of time, the demand for 'authentic' census data has become less and less clear-cut, and the answers less and less straightforward. Today, the changes in society, administration and technology have become massive and the need to adapt is pressing. We must be up to the challenge.

III. Where we are

29. Let us come back to the fundamental question: what is a census?

30. Is a census a fixed operation using a fixed methodology whereby enumerators go from door to door presenting questionnaires to people? The answer is clearly 'No'. In the European Union, the 2011 census of population and housing clearly demonstrates the paradigm shift towards multi-mode/multi-source census operations. In 2001 seven countries of today's European Union (EU) used administrative registers as one of the data sources for their population and housing censuses; this has more than doubled to 16 in the 2011 round. In twelve EU Member States people could use the Internet to transmit their data for the 2011 census.

31. Committed to a course that leads to high quality, censuses have explored new territories. The 2010/2011 census round can be seen as a 'census laboratory'. Many new methods have been employed, new data sources have been tapped, innovative enumeration technology has been used.

32. The experiences made are an invaluable treasure for the experts who are advising governments with a view to changing their census methodology and improving the inter-play of censuses with other, more frequent data collections. All the experience gained should be reported in a transparent and accessible manner to support comparative studies.[2]

33. International and European organisations are well aware of the movement and support it. For example, the United Nations Economic Commission for Europe (UNECE) organises training courses, publications and seminars on the use of registers and administrative data in censuses. On the European level there is, for the first time, legislation on censuses which is adapted to modern multi-mode/multi-source operations.[3] Eurostat is

[2] The United Nations Economic Commission for Europe (UNECE) website is already a forum which is highly appreciated amongst census experts. Moreover, extensive metadata will be disseminated via the 'European Union (EU) Census Hub', including a description of the census methodology employed in the various countries.

[3] The EU legislation on the 2011 population and housing censuses is **output oriented**. It is open to conventional censuses, register-based censuses, rolling censuses, the use of sample surveys and any combination of the above. In turn, it provides harmonised definitions, technical specifications and breakdowns for the output variables, and defines a common dissemination programme.
The user will be able to access the results of the 2011 censuses via the '**EU Census Hub**'. Censuses provide exceptionally high volumes of data. Eurostat and the Statistical Offices in the European

committed to the excellent cooperation between itself, the National Statistical Institutes, and the UNECE. The saying ‘think globally, act locally’ applies unreservedly to statistical cooperation in the area of population and housing censuses. We work to make data for municipalities comparable on the international level!

IV. Where should we be heading?

34. As the 2011 censuses clearly show, the term ‘census’ no longer describes a fixed, conventional methodology. The shift of paradigm has already happened. International and EU organisations are fostering this process.

35. But are we clear enough in our minds about the role of census activities in a modern statistical environment? What essence can we distill from the many innovations going on in statistical offices already today?

36. Here are five proposals to help find common ground from where to go forward. We should:

(a) Step up the integration of censuses into social statistics;

(b) Focus on coverage;

(c) Make the geo-referencing of population data precise;

(d) Make comparative quality assessments of multi-source/multi-mode processes and outputs;

(e) Develop and use common tools.

A. Step up the integration of censuses into social statistics

37. The biggest challenge for censuses in the coming decade will be to strengthen their integration into social statistics.

38. We should move to see the term ‘census’ as referring to a set of requirements for data and quality to be met by our systems of demographic and social statistics. The strong points of censuses are known: a high level of accuracy, great regional detail, a great capacity to cross-tabulate data, the best possible coverage. The challenge is to achieve the desired quality level once and enable the whole system of demographic and social statistics to profit from it.

Economic Area (EEA) countries accepted the challenge to make these huge datasets available to users in a flexible and user-friendly manner. Innovative ways of dissemination were required in order to achieve this. This was the main motivation behind the ‘Census Hub’. This pioneer project creates an integrated ‘data at source’ transmission and dissemination of census data, providing users with easy access to detailed data that are comparable between Member States and structured in the same way. The EU’s census legislation was formulated with a view to making that development possible.
The EU legislation on the 2011 censuses describes **harmonised quality reporting for a multi-source/multi-mode data collection**. This was pioneer work. One main challenge was to develop a concept that can be based on information available in the national statistical offices’ census databases and which will make it possible to draw viable conclusions. The guiding principles are the quality dimensions established in the European Social Survey (ESS) (relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity), and the essential features of population and housing censuses adopted by the United Nations (individual enumeration, universality, availability of small-area data, simultaneity, and defined periodicity).

39. Whatever sources or processes population frames are based on, they will always be the starting point for any census. Assessing and improving the coverage of population frames on a regular basis boosts the quality not only of the census, but of any statistical survey using the same frame. In the 'Wiesbaden Memorandum on the Modernisation of European Social Statistics' the Directors General of the National Statistical Institutes of the EU agreed that '*social statistics should be based on reliable and up-to-date sampling frames of individuals or dwellings.*'[4] Often such frames are derived from registers of the population and of buildings or dwellings. A 'culture of unified frames' generates great synergies between sample surveys and the census, between different sample surveys, and between the usage of administrative data and sample surveys.

40. Census work might require the linkage of different data sources. Again, this work should not be a one-off exercise, but a sustainable achievement that will help to improve the production of other statistical data. The more 'stovepipes'[5] are torn down, the more the census-related investment will benefit the production of other statistical data.

41. Often, administrative data sources undergo a fitness check-up for the census. Consequently, the multiple use of the same data source (i.e. for censuses and for other statistics) carries the potential for synergies.

42. The UNECE and Eurostat have worked hard in recent years to harmonise the specifications and breakdowns for the census variables. The EU census legislation often defines several nested breakdowns, with the difference being the level of detail. This can help later to define the output flexibly according to user needs, but on the basis of breakdowns that are consistent for different data sets. Generalising and extending this approach to other statistical data collections would be an important milestone towards data integration in official statistics.

43. One major development within the modernisation of social statistics is the integration of the different surveys done in that area. In this context, the census should be seen as *primus inter pares*. Improved integration of the different data collections in social statistics — including the census — can bring economies of scale.

44. Integrated data collections should be based on integrated user consultation. For which data is high geographical detail needed, and for which data might less detail be sufficient? Which variables should be cross-tabulated in great detail with other variables, and where is more aggregate information sufficient? The National Statistical Institutes put a lot of work into learning about the needs of their users, and balance the desired output carefully against costs and response burden. Much of their knowledge has found its way into the Conference of European Statisticians (CES) Recommendations[6] and the EU legislation on the 2011 censuses. Nevertheless, nothing is cast in stone, particularly in view of the efficiency gains of a better integration of censuses and other surveys. Strong, ongoing cooperation is welcome if we are to enhance the responsiveness of statistical

[4] Wiesbaden Memorandum on a new conceptual design for household and social statistics (adopted by the Directors General of the National Statistical Institutes (DGINS) conference on 28 September 2011), point 5.a.

[5] According to the 'Communication (2009) 404 from the Commission on the production method of EU statistics: a vision for the next decade' a stovepipe model is a business model where the production processes are organised by distinct statistical products, i.e. in numerous parallel processes, country by country (even region by region) and domain by domain. In such a model a stovepipe corresponds to a specific domain of statistics and the corresponding production system.

[6] Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing.

outputs to user needs across the board, viewing the census as one element of a larger system.

45. The 'multi-purpose' statistical infrastructure in social statistics has been outlined in a recent working paper for the High-Level Group 'Business Architecture in Statistics'.[7] Censuses certainly have a fundamental role to play in this context.

46. Much of the above could be summed up as 'enhancing the continuity of census work'. This is particularly true for the quality check-ups and updates of frames and administrative data sources.

47. More regular activities can have positive effects for *non-census data collections*, which could be based on more up-to-date and accurate frames and raw data.

48. *Censuses* would cause flatter peaks in the workload. Peaks are a well known enemy to efficiency. They are hard to finance and can lead to strained budgets. They are hard to plan, as work processes have to be programmed from scratch and assessed ex-ante from beginning to end, each time round. They are hard to sustain as peaks in the overall process translate into long working hours for individual employees. At the end of a census round they lead to the loss of staff who have acquired valuable skills — the kind of skills that will be sought after (and hard to find) in the next census round.

49. Who would deny that decennial censuses are at particular risk? Adopting a steadier pace of census activities over the years carries the potential for high efficiency gains.

B. Focus on coverage

50. Good coverage of the regional and national population is *the* key objective of censuses. Have all members of the population been counted once and once only?

51. Certainly, coverage is very important in any country. Data on regional and national population stocks is top of the agenda not only of regional and national authorities,[8] but also of the European Institutions. The system of Qualified Majority Voting in the Council of the EU[9] and the allocation of Structural Funds[10] both depend on an accurate population

[7] Excerpt from a working paper for the High-Level Group 'Business Architecture in Statistics — The business case for 21st century official statistics 'by Walter Radermacher and Martina Hahn (December 2011): 'With world-wide harmonised classifications and — to some extent — definitions, a wide range of statistical products can be put together in a modular way, generating additional outputs. Accounting systems provide a common umbrella for building blocks of data that can be combined in a coherent and consistent manner, yielding substantial value added to users. At the same time the 'combinability' of various data sources reduces response burden, as information from respondents will be collected only once and then used for various purposes. It also allows for observing multi-dimensional phenomena, such as globalisation. This approach of multiple-purpose statistical infrastructure has prospered particularly in the area of economic statistics and there is still large scope for extending statistics' modularity to be deepened and extended to cover further statistical domains, most notably in social statistics.'

[8] For example for allocating local and/or regional authority budgets, or for delineating electoral districts.

[9] Population data are used for European Union decision-making processes. Qualified majority voting in the Council of the EU has been extended to new policy areas to make decision-making faster and more efficient (under the Treaty of Lisbon). From 2014 on, the calculation of the qualified majority will be based on the 'double majority' of Member States and people, to represent the dual legitimacy of the EU. A double majority will be achieved when a decision is taken by 55 % of the Member States representing at least 65 % of the EU population. The population figures for qualified majority voting are collected and provided by Eurostat every year.

count. This is why the EU legislation on the 2011 population and housing censuses requires Member States to conduct a coverage assessment for their censuses.

52. One issue which will have increasing importance in the European Union is how to assess and correct cross-border coverage problems, in particular how to avoid the double enumeration of people with bonds in different countries. Just as regions work together to overcome national coverage problems, so must countries work together to overcome European coverage problems. We need to address this better in the medium term.

53. Getting a good assessment of the coverage of frames is complex and expensive. A whole host of problems tend to gang up against the statistician.[11] Clearly, good coverage is much too valuable an asset to be left to the census alone.

54. Frames that have been updated for a census should also be used — and maintained — for other statistical surveys. Like the foundation of a house must stand firm, frames used in social statistics must be of high quality. To achieve economies of scale, periodic reviews of statistical frames (and the registers on which they are based) and the periodicity of the census should be harmonised. One day perhaps 'doing a census' will mean updating the statistical frames used for many other data collections of social statistics!

C. Make the geo-referencing of population data precise

55. The geographical referencing of population data is at the heart of every census, and its importance is growing. Current global concerns like climate change, threats to ecosystems or limited resources call for integrated policies for sustainable development. Statistical data are mostly reported for sizeable, mostly administrative areas,[12] with all their inherent diversity. But the causes of social and environmental phenomena do not follow the delineations of administrative areas. One way of overcoming this gap is to use spatially referenced statistics.

56. What matters primarily is precision of the referencing. Addresses should be linked to their geographical coordinates in order to geo-reference statistical information in a way that best serves spatial analysis across disciplines.[13] Once that investment has been made, the geographical referencing can be used for local, regional, national and European information requirements. The result will be greatly improved flexibility for any analyst.

[10] The EU has competency in the area of regional cohesion, which constitutes one of the largest budgetary items. The EU supports the development of its less prosperous regions. The 'Convergence' objective of the Structural Funds is a main priority within the EU's cohesion policy. Geographical areas lagging behind in their development are identified as those whose Gross Domestic Product per head is less than 75% of the EU average.

[11] According to the statistical system these might include challenges related to the organisation of post-enumeration surveys, population groups that are hard to enumerate and/or to register (e.g. frequent movers, fringe groups of the population), non-registration of individuals, non-deregistration, multiple registration, registration delays, the readiness/obligation of the register-owner to cooperate with the statistician, editing of data records by the register-owner, the linkage of registers to find unusual patterns, data protection legislation, etc.

[12] In the EU: Nomenclature of Territorial Units for Statistics (NUTS) areas.

[13] For example, Eurostat manages the GEOSTAT project with the aim of creating a standardised European Social Survey (ESS)-wide population grid map for use in spatial analysis.

D. Make comparative quality assessments of multi-source/multi-mode processes and outputs

57. Today, many censuses have become multi-source/multi-mode operations based on innovative methodology and technology. They are at the forefront of a process of modernising the way social statistics are produced.

58. This creates a new challenge for the quality assessment. The starting point is settled: a sound framework within which we can assess quality in official statistics,[14] the clear criteria to pin-point census qualities.[15]

59. However, there are new and different challenges when it comes to making quality criteria operational, and measuring them in a comparative manner for multi-source/multi-mode operations. With the old habit of 'one source/one methodology' becoming obsolete, a lot of methodological work is now in the pipeline. Tried and tested concepts require re-thinking.

60. How can the *accuracy* of data from a multi-source/multi-mode operation be measured best? How can we ensure the *comparability* of data extracted from administrative registers which might use definitions and specifications that deviate from statistical ones? How should we measure the extent to which different data sources uphold the principles of *universality* and *individual enumeration*?[16]

61. In a future marked by increasing complexity in the production of statistics, a comprehensive quality assessment has to evaluate outputs *and* processes. When processes are transparent, they boost efficiency and generate trust and acceptance in data users and other stakeholders.[17]

62. There are plenty of challenging questions to keep our methodologists and practitioners busy over the coming decades!

E. Develop and use common tools

63. Technology should go hand in hand with standardisation and an innovative business model. The reward will be huge synergies and efficiency gains. One example is the EU Census Hub. EU Member States describe the output they will make available in the Census Hub in a unified way in Statistical Data and Metadata eXchange (SDMX). This opens the door to shared software developments, ranging from data validation to dissemination. The European Statistical System is venturing ahead to enhance data integration and to explore the areas where the development and usage of common tools can help us do a better job at lower cost.

[14] In the ESS we refer to established quality dimensions, namely relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity.

[15] The United Nations has adopted essential features of population and housing censuses, namely individual enumeration, universality, availability of small area data, simultaneity, and defined periodicity.

[16] If the extent is low: how to measure the impact on the output data? What are the differences for different levels of geographical detail? For small groups of the population? For different estimation methodologies?

[17] The European Statistics Code of Practice identifies sound methodology, appropriate statistical procedures, non-excessive burden on respondents, and cost-effectiveness as principles for statistical processes. Apart from assessing process quality, comparative descriptions of processes are useful for defining and applying standards, exchanging expertise, granting certification, etc.

V. Epilogue

64. A census is a census is … a lot of work!

65. As we rise to the challenges, let us reflect about what the term ‘census’ means to us today.

66. It means transition in progress. The successful 2010/2011 census round is an impressive demonstration of the credibility of official statistics, yielding detailed and high-quality data and, at the same time, making responsible use of taxpayers’ money. But we must not rest on our laurels. Let us seek to integrate our census activities better into the production of social statistics. Let ‘census qualities’ spill over into other data collections.