

Guest Editorial

Jean-Michel Durr

National Institute of Statistics and Economic Studies (INSEE), France

E-mail: jean-michel.durr@insee.fr

I am very honoured to present this special edition of the Journal of the International Association of Official Statistics on “Population and Housing Censuses”.

Population and housing census is the oldest statistical operation, and still the cornerstone for national statistical systems. The characteristics of the population and housing units it provides at small area, regional and national levels are of paramount value for evidence-based policy making, in-depth academic research, as well as for other stakeholders.

National statistical offices (NSOs) are constantly improving their methodology, the technology they use and the organization they set up to conduct the census in a more efficient way. The traditional canvassing of the country, with enumerators going door to door to interview the inhabitants is still in effect in the majority of countries, but paper questionnaires are more and more replaced by e-questionnaires on tablets or smartphones. Many countries, not just developed countries, are increasingly using administrative sources to supplement or even replace field enumeration. The words “combined census” and “register-based census” are now items in the census dictionary. In the middle of the 2020 census round, which covers the period 2014–2025, as defined by the United Nations, this special issue of the Statistical Journal of the IAOS is particularly timely to take stock of the latest developments in census taking and present examples of these various situations.

1. International recommendations and support to population and housing censuses

Census has a long history, which Alphonse Mac Donald recalls the main milestones and the efforts to draw up international recommendations for conducting comparable censuses across countries. The first recom-

mendations were issued by the first International Statistical Conference held in 1853 in Brussels and the most recent ones were endorsed by the UN Statistical Commission in 2015. Along the time, they formalised the census principles and methodology, and recently incorporated alternative ways such as rolling census or register-based census. Mac Donald also discusses if the methodology of the census complies with the principle of scientific activities. In particular, whereas recognizing that the traditional census suffers from some flaws, he regrets the lack of theoretical basis of the register-based census and wonders to what extent this approach fully complies with the principles of universality and simultaneity.

From a recent survey conducted by the United Nations Statistics Division (UNSD), Srdjan Mrkic reveals that 69 percent of the countries will conduct a traditional census during the 2020 census round, 20 percent a census combining registers and field enumeration or sample survey, and 11 percent a fully register-based census. Among the countries performing field enumeration, multimode data collection is increasingly used: 72 percent use computer assisted interviews (CAPI), 38 percent internet response (CAWI) and 12 percent telephone interview (CATI). However, implementing new technologies is identified as a major challenge by countries, together with the risk of lack of financial resources. Therefore, support from international organizations is essential. In that regard, UNSD and the United Nations Fund for Population (UNFPA) maintain a complementary division of labour: UNSD overseeing the global guidelines and UNFPA overseeing in country-technical support.

UNFPA strategy in support of the 2020 census round is presented by Rachel Snow and colleagues Sabrina Juran, Tapiwa Jhamba and Maureen Jones. UNFPA advocates for the use of new technologies, including a greater use of digital mapping, for example for model-

ing populations at risk of environmental disasters, and for the inclusion in the census questionnaire of specific questions on migration and disabilities. UNFPA provides support to countries through guidelines, regional workshops and technical advice. In particular, UNFPA is committed to increasing the provision of Chief Technical Advisors on census. Finally, UNFPA develops South-South support, such as exchange of expertise or commodity sharing.

The International Program of the US Census Bureau (USCB) also provides census capacity-strengthening technical assistance. Oliver Fisher leads a team of nearly 40 experts to develop public domain tools and conduct technical assistance and trainings to facilitate the implementation of censuses in low and middle-income countries. Their famous tool, CSPPro, offers an integrated suite for data collection, editing and tabulation of censuses and surveys. The Bureau released in 2016 CSPPro Android to provide NSOs a mobile data capture solution. The US Census Bureau is part of the International Committee on Census Coordination, together with the UNSD and UNFPA. One example of the country-specific support of the USCB is the ongoing technical assistance to the Statistical Office of Malawi for the 2018 census, as presented further in the next article of this edition.

2. Use of new technologies in traditional censuses

Implementing new technologies is not straightforward and bears risks. Therefore, technical assistance is essential and thorough cost-benefit analysis and risk enquiry for embarking on the technology is required in the planing of a census project. The 2018 census of Malawi was the first census of the country using electronic questionnaires. Mercy Kanyuka, Jameson Ndawala, Isaac Chirwa, Medson Makwemba, Richard A.P Phiri, Chisomo Singano and Grevazio Kapaswiche from the NSO-Malawi stress the advantage of using the full package CSPPro compared to tailor made developments, because of the on-site as well as remote technical assistance provided by the US Census Bureau. They also explain that they greatly benefited from experiences and skills from countries in the region that had pioneered in the use of CAPI technology and use of satellite imagery, and in return, they opened up for study visits and provision of support for other countries. This is a good example of the relevance of the UNFPA strategy to develop South-South support.

International technical assistance is also provided by NSOs as part of their cooperation activities. The Ital-

ian National Institute of Statistics (Istat) supported the Ethiopia Central Statistics Agency in designing and implementing a census monitoring system, fully integrated with the electronic data collection process managed by CSPPro. New technologies not only allow for implementing CAPI data collection, but also to computerize the whole data collection process, as described by Mauro Bruno, Filomena Grassia, Joshua Handley, Asres Abate, Deriba Mamo, and Atreshiwal Girma. The system developed with the assistance of Istat generates reports to monitor the progress of the fieldwork, providing coverage information at different levels. The system also provides a set of reports on questionnaire variables to monitor the quality of the data collected, e.g. the average household size or the sex distribution. In addition, the integration of digital maps allows users to instantly visualize the progress of enumeration activities and identify which areas have not been covered.

Censuses are large operations that require peaceful situation and well-defined boundaries, conditions that do not really apply to the situation in Palestine. However, the Palestinian Central Bureau of Statistics (PCBS) managed to face the challenge and successfully conducted the 2017 Population and Housing Census. Ola Awad and Sufian Abu Harb explain that the difficulties related to the transfer of paper questionnaires between the West Bank and the Gaza strip as well as the risks of confiscation of these questionnaires at checkpoints led PCBS to move to electronic data collection. Thus, census data were collected using tablets supported by GIS technology. The Palestinian experience shows that electronic data collection does not mean that network coverage should be universal in the country: the lack of full broad band coverage in some areas of Palestine was overcome using offline data collection. In addition, the cost of tablets, often considered as a barrier, can be reduced or even annulled by sharing equipment among countries. This is how PCBS could borrow more than 7,000 tablets from the Jordanian Department of Statistics.

Beyond the CAPI mode of interview, countries tend to develop multi-mode data collection, including self-enumeration using internet response. For the first time in Mexico, the 2020 Population and Housing Census data collection will be carried out through a CAPI scheme as the main enumeration method, but will also include the CATI and the CAWI methods. Other innovations, such as the use of a QR code for identifying the dwelling and serve as identifier for responding via the Internet, and GPS coordinates to indicate the location of the staff in the field in order to facilitate the

work of supervision, verification and post-enumeration will be implemented by INEGI, the National Statistical and Geographical Institute of Mexico. Edgar Vielma Orozco, Sabino Navarro Campos, and Isaac Salcedo Campos highlight that such a large use of IT technology requires to conduct a greater number of tests, both in controlled conditions and in the field.

In developed countries, census is collected through self-enumeration, and questionnaires are either dropped off by an enumerator or sent by mail. The only way to replace paper questionnaire is then to offer the possibility to respond via internet. The US Census Bureau is implementing the internet response for the first time for the 2020 Census with two kinds of Internet self-response: Identified or Non-Identified. The latter option will allow respondents to complete their questionnaires anytime and anywhere without requiring their census ID. Other areas of innovation include automatic comparison of satellite images to identify geographic areas that have changed since the last census in order to optimize field staff updating; and use of administrative and third-party data to identify vacant, non-existent, and occupied living quarters and thus reduce expensive in-person visits for households who do not self-respond. Robert Colosi presents these innovations, and underlines that the 2020 Census promises to be innovative, and improve over past censuses that relied heavily on paper questionnaires, while maintaining a high level of quality.

3. Increasing use of administrative sources and rolling census

Due to the high cost of a traditional census, countries explore other ways of collecting census information, in particular using administrative sources. Countries that do not have a population register face the challenge of building statistical datasets that represent all individuals of the population and all housing units. Nevertheless, the development of e-administration in many countries allow for the availability of more and more administrative sources of better quality.

The UK does not have a population register or a set of coherent identifiers across administrative datasets held by government, unlike many European countries. However, UK is exploring ways of replacing the decennial census, which is expensive and no more able to satisfy the demand for more timely statistics. Owen Abbott, Becky Tinsley, Steve Milner, Andrew Taylor and Rosalind Archer outline the current plans of the

Office for National Statistics (ONS), and some of the methodological challenges. A common challenge for countries attempting to use administrative sources as the UK is that the accuracy varies and is poor at small geographical levels, and that some sources cover only a sub-population. Another problem faced by the ONS is the linkage of various sources, in the absence of a common identifier. The usual way is to use deterministic or probabilistic algorithms to match records using characteristics such as names, sex and birth date. In the UK, for privacy protection, administrative sources must be anonymised before linkage. However, ONS was authorized to create match-keys by concatenating different parts of the identifiers in a string that can be used for deterministic matching. Despite these challenges, ONS considers that there are opportunities to produce statistics with improved timeliness and reduced cost and burden, even if replacing the census may not be achievable in the short term.

The examples of New Zealand and Australia show how administrative data can be used to enhance the quality of a traditional census. Statistics New Zealand (Stats NZ) is in a similar situation to that of the UK, and has engaged a Census Transformation Programme to explore the feasibility of a census based largely on administrative data, supported by sample surveys. Stats NZ recently released an experimental data series of administrative population estimates including for subnational levels and including ethnic groups. Christine Bycroft details how this work was very useful to supplement the high level of non-responses faced by the census 2018, by imputing administrative records. The final Census dataset consists of 89 percent census responses and 11 percent admin enumeration.

Imputation of missing values in a census is usually performed using hot-deck or nearest neighbour methods. In Australia, the 2016 census response rate was lower than in 2011 and 2006, resulting in higher rates of imputation. However, the Post enumeration survey (PES) revealed that the resulting age distribution was skewed and could be improved by enhanced selection information. James Farnell and Peta Darby¹ from the Australian Bureau of Statistics (ABS), proposed a method to incorporate administrative data to

¹James Farnell and Peta Darby are winners of the IAOS Young Statisticians Price 2019. The publication of their manuscript is part of the price. The manuscripts of the other three YSP 2019 price winners: Vianney Costemalle (INSEE, France), Nancy Wang (Statistics New Zealand) and Marlene Weinauer (Statistics Austria) were published in Vol. 35/4.

inform the choice of donors in donor imputation. The method consists in five steps: (i) preparation of administrative data; (ii) data linkage; (iii) selection of match variables; (iv) selection of donors; and (v) imputation of variables by duplicating selected census data from the donor unit to the recipient unit. The results of the method are closer to the PES estimates than the results of the initial imputation method.

Another way of improving the timeliness of the census while spreading its cost annually is the rolling census implemented in France since 2004. Every year, a sample of 8 percent of the population of all municipalities of 10,000 inhabitants and over is enumerated, and a full enumeration is conducted in one fifth of the municipalities under this threshold. Data collected over the last five years are compiled every year, in combination with information from administrative sources, to produce updated population counts for each municipality and detailed characteristics of the population and housing units at all geographical levels, down to the district in large cities. Valérie Roux shows that one of the advantages of the rolling census is the opportunity provided by the annual surveys to progressively introduce innovations, in comparison with the high-risk of introducing one-shot innovations in a decennial census, with only a few small scale tests. For example, the internet response was progressively offered in an increasing number of municipalities, allowing for full scale testing and setting up a secure and high-performance IT platform. Since the method is based on a five-year cycle, results of changes in the questionnaire are available at all geographical levels only after five years, but conversely annual surveys allow a more precise assessment of the effects of a change in methodology or questionnaire.

4. Moving to a register-based census

The following countries have gone a step further, planning for a register-based census for the current round. Through the examples of Abu Dhabi, Italy, Saudi Arabia, Estonia, Poland, Turkey and Spain, one can see the steps to move towards a register-based census. The path is not straightforward but is worth the efforts as it saves the huge cost of a decennial census while providing regularly updated census like results.

First of all, building a census based on a set of administrative registers and sources requires extensive coordination from the statistical office. Aisha Ali Turki underlines that the Statistics Centre of Abu Dhabi had

to establish coordination with the entities owning the administrative registers, resulting in the signing of service level agreements, to ensure the sustainability of the provision of data. But administrative data are rarely built following the statistical concepts, and Ali Turki describes the methodology to identify the resident population in the registers. She also stresses that the lack of accuracy of addresses of the resident population is one of the most important challenges faced to develop a register-based census in 2020.

To tackle this issue, Fabio Crescenzi and Fabio Lipizzi present how the Italian Institute of Statistics (Istat) has built a Base Register of Territorial and Geographical Entities (RSBL) as part of the integrated system of registers that replace the decennial census. The aim of the RSBL is to provide a unique source of accurate and standardized addresses to georeference statistical information from administrative registers and thus allow for releasing statistical information with a strong territorial detail.

Saudi Arabia has opted for an intermediate approach for the 2020 census. Fahad Altekhaifi, Fahad Alfahead, and Arwa Alshanghiti explain that the General Authority for Statistics (GASTAT) will use a combined methodology based on administrative data supported by fieldwork. To develop its coding system of buildings and housing units, GASTAT works on using a unified national coding system for addressing. This system includes more than 7 million national addresses managed by Saudi Post. A fieldwork operation was conducted to update all national addresses for all buildings and estimate the number of housing units for every building. Then, census data collection will be carried out by field enumerators who will check or complete the administrative data of the household with the household response.

Dana Beltadze recalls that moving towards a register-based census takes time to ensure at least the same level of quality of a traditional census. In Estonia, the possibility of a register-based census has been considered since 2002. In 2007, Statistics Estonia conducted a quality assessment study of available administrative sources that concluded they did not have sufficient information for census variables. The new Official Statistics Act in 2010 tasked Statistics Estonia to start preparations for a register-based census. A comparative analysis with the census 2011 concluded that it was possible to collect the most mandatory census variables by combining registers and census data, but that in many cases the quality of administrative data could still be insufficient. Intensive collaboration be-

tween statisticians and register holders was initiated, and by 2014 a new level had been reached in terms of the interoperability of state information systems, allowing for the preparation of a register-based census. A pilot was conducted in 2019 of a full-scale register-based census, where almost all census variables are obtained on the basis of register information.

The combined census can represent an intermediary step before a full register-based census, as testify Janusz Dygaszewicz (Statistics Poland) and Şebnem Beşe Canpolat and Muharrem Gürleyen Gök (Turkish Statistical Office – TurkStat). In Poland, the 2021 census will be based on the mixed model of the Census 2011, using data from administrative registers and data obtained from respondents (20% ad-hoc sample survey), with the use of electronic questionnaires. According to Dygaszewicz, censuses in Poland after 2021 will undoubtedly be based on the full use of administrative and non-administrative registers. Annual updates after 2024 will require full automation of data processing from administrative sources and other alternative sources such as Big Data combined with statistical data.

In Turkey, the Government established between 1997 and 1999 a population register, which involved transferring civil records from paper-based family ledgers to a digital format. By the end of 2000, all Turkish citizens were given a unique identification number, but the register still had no information available on place of usual residence. As such, TurkStat could not use the register for the 2000 Census. In 2006, the Address Based Population Registration System was created: TurkStat was given the responsibility for developing the system and the Ministry of Interior for maintaining the register. The first step was to establish the National Address Database. The 2011 Census used a combined method to produce census-type statistics that were absent in the register. The next step is to add some other individual-related registers to build an integrated system for a full register-based census in 2021.

In Spain, the Population Register (Padrón) was created in 1996. For each person, Padrón contains the following information: sex, date of birth, place of birth, nationality, educational attainment and national identity number. To determine the population count, Jorge Vega Valle, Antonio Argüeso Jiménez, and Marina Pérez Julián explain the steps implemented by the Institute of Statistics (INE): verification that births and deaths are up-to-date as of the reference date; analysis of the expiration date of Padrón foreigners; detection

of “signs of life” of those people that appear in other files like Tax Agency or Social Security; and application of the 12-month criterion. Census variables that are not in the Padrón or other administrative sources will be collected through a 1 percent sample survey (200,000 households). However, the question remains on the precision of the survey for small geographical areas.

5. Specific challenges in conducting population and housing censuses

This special edition would not be complete if it did not shed light on some specific census situations. Philomen Harrison draws our attention to the specific challenges facing small island countries for the conduct of a census in the Caribbean region: vulnerability to natural disasters, scarcity of financial resources, lack of skilled and stable staff, poor IT infrastructure, hard to reach populations and security issues in some areas. Various strategies are used to mitigate these challenges, the most prominent one being strong coordination among countries under the CARICOM umbrella.

A census shall include everybody living in a country, but indigenous populations have been frequently ignored or underrepresented by censuses in the past. Michele Connolly, Sam Notzon and Bette Jacobs point out that much has changed since the first US Census, conducted in 1790, that explicitly excluded American Indians from being counted, in accordance with the US Constitution. Joint efforts by American Indian Tribes, the National Congress of American Indians and the U.S. Census Bureau have evolved to obtain accurate and complete counts. It is exemplary that the very first Americans to be counted in the 2020 Census reside in the remote Indigenous village of Toksook Bay, Alaska.

Among the challenges facing a census is the risk of political interference. Population counts at all administrative levels produced by the census are the basis for resource sharing, apportion of seats in the parliament among other. On the other hand, the large scale of census operations requires governmental support, to ensure funding and logistics. On the basis of various examples, both in developing and developed countries, Jean-Michel Durr discusses the balance between the necessary support from the Government to the census and influence avoided and stresses the importance of a solid legal framework to guarantee the independence of census operations.

Timeliness of the dissemination of census results is of paramount importance to preserve the high value of

the information and justify the high cost of a census. Vikas Kumar gives examples of censuses of India that were not released in due time, in particular regarding sensitive information such as religion or ethnicity.

To terminate this overview of census practices, it is noticeable that population and housing censuses share many common features with censuses of agriculture. Linking both censuses may present some advantages, for example reducing the cost by sharing equipment, or integrating agriculture and population statistics for more in-depth analysis of interactions between agriculture and the socio-economic environment to contribute to monitoring the Sustainable Development Goals. The technical guidelines, presented by Jairo Castano and Naman Keita, from FAO, issued by UNFPA and FAO provide detailed guidance on conceptual and other methodological issues for the linkage of the Population and Housing Census with the Census of Agriculture with several country examples.

6. Conclusion

I sincerely hope that this special edition will help census takers in their endeavour to achieve successful censuses, and generate intensive discussions on census methodologies, new technologies and other census issues among the vibrant community of passionate census practitioners.

This edition could not have happened without the assistance and inspiration of Dr. Pieter Everaers, the current editor of the *Statistical Journal of the IAOS*, who approached me with the idea of a special edition. I also wish to thank UNFPA and the French Institute of Statistics (INSEE), of which I have the honour to belong, for their financial contribution to make this edition freely available, thus ensuring a wide audience.

Finally, I would like to deeply thank the reviewers of this issue: Ali Ahmad, Marius Andersen, Olivia Blum, Christine Bycroft, Cynthia Clark, Roberto Bianchini, Janusz Dygaszewicz, Frank Eelens, Katerina Kostadinova Daskalovska, Richard Madden, Vincent Mule, Eric Schulte Nordholt, for their outstanding contribution to the quality of the papers presented.