

**Economic and Social Council**Distr.: General
22 July 2024

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Twenty-sixth Meeting**

Geneva, 2–4 October 2024

Item 2 (1) of the provisional agenda

**Revising the Conference of European Statisticians Recommendations
for Population and Housing Censuses for the 2030 round:****Dissemination of disaggregated census data****Developing the Recommendations on Dissemination and
related topics: security, documentation, metadata and
archiving****Note by the Conference of European Statisticians Task Force on
Dissemination of Disaggregated Census Data****Summary*

This document includes the draft chapter on Dissemination and related topics: security, documentation, metadata and archiving for the Conference of European Statisticians (CES) Recommendations for the 2030 round of population and housing censuses, and a summary of the changes introduced in comparison to the Recommendations for the previous, 2020 round. The main purpose of the document is to elicit comments and suggestions from national census experts on the proposed text, to ensure that it reflects the needs and priorities of national statistical offices.

* The Conference of European Statisticians Task Force on Dissemination of Disaggregated Census Data consists of the following members: Fabian Bach (European Commission – Eurostat, chair), Glen Hohlmann (Canada), David Price (Canada), Bernie Casey (Ireland), Klara Shalev (Israel), Alejandra Silva (United Nations Economic Commission for Latin America and the Caribbean), Baktybek Kainazarov (United Nations Population Fund Kyrgyzstan), Andrew Smith (United Nations Statistics Division), Ala Negruta (United Nations Women Regional Office for Europe and Central Asia), Lara Cleveland (IPUMS). The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission, which chairs the task force that authored this work.

NOTE: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.



I. Introduction

1. Every ten years the Conference of European Statisticians (CES) issues Recommendations to guide countries in conducting their population and housing censuses. The Recommendations are developed by expert task forces overseen by the CES Steering Group on Population and Housing Censuses.
2. The work of this Task Force is organized along three thematic strands (general dissemination aspects, statistical disclosure control, geographically referenced products). After the UNECE Survey on 2020 Census Round was concluded, the Task Force held three calls and various written exchanges. The Task Force analyzed the survey replies and developed proposals for updating the text parts of the CES Recommendations within its topical scope.
3. High-level topics that are addressed with the proposed updates include: modern dissemination technology and tools including table builders or data visualizations, new developments in statistical disclosure control including noise-based methods and implications from grid outputs or flexible dissemination tools, and recent developments on functional output geographies including grids.
4. Section II of this document summarizes the changes introduced in the content pertaining to dissemination, in comparison with the Recommendations for the previous, 2020 round. Section III gives a high-level overview of the draft chapter.
5. Section IV presents the draft chapter on dissemination (proposed as a stand-alone chapter entitled ‘Dissemination and related topics: security, documentation, metadata and archiving’) for the CES Recommendations for the 2030 round of population and housing censuses.
6. The main purpose of the document is to elicit comments and suggestions from national census experts on the proposed draft text, to ensure that it reflects the needs and priorities of national statistical offices.

II. Summary of changes

7. This Task Force (TF) revised the following Sections of the 2020 CES Recommendations (CESR):
 - (a) ‘Confidentiality and security’, paragraphs 140–148 in Chapter I (Methodology);
 - (b) ‘Dissemination, documentation, metadata and archiving’, paragraphs 292–318 in Chapter III (Field and other operational activities).
8. Based on its analysis of the 2023 survey replies, the TF focused on three high-level topical aspects that the CESR revision would need to address, as follows:
 - (a) *Modern dissemination technology and tools*: The survey confirmed that digital online formats are now by far the most popular dissemination methods among United Nations Economic Commission for Europe (UNECE) countries. Interactive online tools such as table builders and Geographic Information System (GIS) mapping tools are relatively new among the most popular formats, while user-friendly and accessible data visualization functions are considered a key element by many countries. This is a key area where UNECE countries keep innovating. The TF revised the CESR text in several places to make this general trend more visible and highlight its various benefits such as user satisfaction and more efficient dissemination;
 - (b) *New developments in statistical disclosure control (SDC)*: The survey showed that many UNECE countries have implemented significant updates to their SDC systems, most notably due to more efficient noise-based methods becoming available that allow for example to avoid cell suppression but also to address new risks that can arise with the use of more detailed and versatile dissemination products. Notably, additional risks from interactive dissemination tools interacting directly with the microdata and geographic differencing risks

(related to introducing grids) are becoming relevant for a notable number of UNECE countries. The TF revised the CESR text in several places and proposes some new dedicated paragraphs to cover these new developments and highlight the links between modern dissemination technology and new disclosure risks;

(c) *Recent developments on functional output geographies:* Collecting precise georeferenced information (coordinates) on the population and on buildings is becoming a standard in the UNECE region that brings various benefits for dissemination, such as free/custom choice of geographic output units. In this context, grids have become a key geographical output level in the UNECE region. Users benefit for instance through highly flexible interactive mapping tools. The TF revised the CESR text in several places to cover these recent developments and stress the main benefits in terms of comparability across countries and user satisfaction.

9. Finally, as part of a general effort by several Task Forces to rearrange the various Sections of Chapter III of the 2020 CESR into a more coherent structure, Task Force 12 proposes to merge the above-mentioned sections into a new dedicated Chapter entitled ‘Dissemination and related topics: security, documentation, metadata and archiving.’ Notable side benefits are that confidentiality and security become more visible, and that security and archiving aspects of access to closed census records are bundled.

III. Overview of chapter

10. Dissemination is one of the crucial aspects of a successful census operation, as it defines the interface between the comprehensive microdata records compiled by census authorities and the data products that users get to work with. As such, dissemination aspects do not only cover the end products but also how these products are obtained from the complete microdata. As outlined in Section II, recent developments in dissemination technology have led to more intricate interdependencies between personal data security and modern product features (interactivity, flexibility, geographical detail). This chapter therefore focuses on dissemination and various important related topics, namely security including SDC, documentation of the census production process and its outputs, metadata accompanying census data products, and longer-term archiving of census data including access to historic census records. Its individual sections are briefly outlined as follows.

11. ‘Confidentiality and security’ consists of two subsections:

(a) ‘Confidentiality principles’ sets out the key principles governing the measures to ensure the physical and logical security of personal data processed in the context of the census operation. Among them are putting the privacy of the individual at the centre and assuring the public that personal information processed by the census authorities is protected at all steps and only used for the specific statistical census purposes. Internally, sufficient internal measures need to be taken to implement this assurance, e.g. around Information Technology (IT) infrastructure and management of staff access;

(b) ‘Statistical disclosure control’ describes the process to protect statistical data in such a way that they can be released without giving away confidential information of any specific individual or entity. The key objective of any SDC methodology should be to ensure a sufficient protection level with minimum information loss and therefore maximum data utility retained. Several SDC measures are outlined to protect the dissemination of tabular and microdata. Links between new dissemination technologies and additional disclosure risks are highlighted.

12. ‘Dissemination’ covers the core aspects and recommendations around the dissemination strategy and ensuing products, most notably:

(a) Simultaneous dissemination to all users according to a timetable as a principle;

(b) Outline of the various ways of making the results of a census available to the user, including e.g. as static published reports or tables, as interactive data products optionally equipped with data visualization tools or mapping features; as commissioned or customized output on demand; or as microdata;

(c) Cost effectiveness of the dissemination strategy, highlighting potential efficiency gains by embracing new technological solutions (e.g. automation of SDC methods, interactive tools that may partly replace on-demand services);

(d) Easy and free access to products, including technical standards to interface online data products and serving modern information channels such as social media;

(e) Interlinkages between technological advancements in census production and more powerful dissemination tools, such as Geographic Information System (GIS) technology facilitating mapping and geographical visualization tools.

13. 'Documentation and metadata' covers, most notably, recommendations on:

(a) A comprehensive portfolio of supporting documentation and metadata to help explain, clarify, and enhance the value of the statistical outputs;

(b) A clearly structured metadata system based on international standards while corresponding to the specifics of national requirements;

(c) Standard elements of a metadata system;

(d) Methodological reports where the census methodology has changed since the previous round;

(e) Consulting stakeholders from diverse backgrounds to improve the metadata.

14. 'Archiving and access to closed census records' covers, most notably, recommendations on:

(a) The census as a unique source of data that puts a big responsibility on the census authority to keep this special historical picture of society for the future;

(b) The relevant purposes, contexts and data security responsibilities for census authorities to consider making historical census records accessible to users after a defined period (recommended to be at least 100 years);

(c) Where census records are archived, the importance of ensuring the preservation of, and easy access to, all the metadata and procedural/operational material.

IV. Dissemination and related topics: security, documentation, metadata and archiving¹

A. Confidentiality and security

1. Confidentiality principles

15. Censuses collect information on each person and household in the country. Census collection may be through direct enumeration, through a compilation of information from administrative registers, or through a combination of register- and survey-based data collection. Its focus is not on specific details about individuals but rather on providing statistics about the community, and groups within the community, as a whole. The public, therefore, has a right to expect, and needs to be assured that, personal information provided in confidence will be respected. Names, addresses and personal identification numbers (PINs) should be separated from other data as soon as possible in the census process, and not released, so that the output data contains no personal identifiers. The confidentiality requirement, typically backed by specific statistical and personal data protection laws, encompasses the whole census operation, ranging from the security of any personal information process (such as completed census questionnaires or individual information from

¹ Cross-references internal to this chapter are given according to the numbering used in the present paper. For cross-references pointing outside the chapter, paragraph numbers in the counting of the 2020 Recommendations are given in square brackets. In both cases, it is understood that the cross-referenced paragraph numbers will differ in the final, published version of the complete 2030 Recommendations.

population registers or administrative data sources) to the protection of the information contained in the outputs and made available publicly.

16. Assurances should be given to the public that all the personal information (such as names, detailed address information, government identification numbers, or internal linking keys) will be treated in strict confidence by the census authorities and any person who is employed by, or provides a service to, the census authority for the purposes of carrying out the census. Many countries will have domestic legislation that protects such information in the form either of specific census legislation or of more general legislation relating to statistical confidentiality or personal data protection and freedom of information.

17. The following additional principles should govern the treatment of the information obtained from any part of the census processes:

(a) Only persons under the management of the census authorities, or agents acting on their behalf, should have access to any census-relevant personal information obtained;

(b) Privacy of the individual should be key, allowing their census-relevant personal information to be processed in such a way that it will not reveal any personal information to the public. Additionally, when surveys are conducted, individual household members should, if they wish, be able to give personal information on a separate questionnaire in a way that will not reveal it to others in their household, or to the enumerator. In censuses compiled using administrative register information, personal information should be stored separately from government linking keys, kept confidential, and not revealed in the dissemination of results;

(c) All members of the census organization and outside agents providing services to the census authority in connection with the census should be given strict instructions, and be required to sign legal undertakings, about confidentiality. They should be liable to prosecution for any breaches of the law;

(d) The physical security of any documents or digital data stored for census production and containing personal information held by the census authorities, by field staff or by authorized agents should be strictly enforced and, if felt necessary, independently reviewed;

(e) Any mode of accessing or processing personal census data, including IT systems, telephone applications, etc., should have strict safeguards to prevent unauthorized access;

(f) In releasing statistics from the census, sufficient steps should be taken to prevent the inadvertent disclosure of information about identifiable individuals and households. Special precautions may apply to statistical outputs for small areas.

2. Statistical disclosure control

18. Statistical disclosure control (SDC) is the process to protect statistical data in such a way that they can be released without giving away confidential information that can be linked to specific individuals or entities. The key objective of any SDC methodology should be to ensure a sufficient protection level with minimum information loss and therefore maximum data utility retained. SDC literature provides for a variety of standard risk and utility indicators to assess risk versus utility systematically and inform the choice of the most appropriate SDC setup for the country's data release programme.

19. Measures to prevent the disclosure of tabular data may include some, or all, of the following procedures:

(a) Restricting the number of output categories into which a variable may be classified, such as aggregated age groups rather than single years of age, particularly for the older ages ('global recoding');

(b) Where the number of people or households in an area falls below a minimum threshold, suppressing statistical output ('local suppression') – except, perhaps, for basic headcounts – or merging it with that for a sufficiently large neighbouring area;

(c) Adding ‘noise’ to the microdata records before producing tables (‘pre-tabular noise’), for example swapping some unit record characteristics of the most risky records by finding a match in the microdata based on a set of predetermined matching variables, and swapping all or some of the other variables between the matched records (‘targeted record swapping’);

(d) Adding ‘noise’ to the tables produced (‘post-tabular noise’), for example rounding cell values up or down to the nearest multiple of the predefined rounding base (conventional rounding) or adding noise of limited magnitude in a controlled and consistent way across tables (controlled noise injection, for instance with the ‘cell key method’).

20. In case of the release of census microdata (such as microdata under contract or public use files) it is important that all information from databases relating to name, address and any unique characteristics that might permit the identification of individuals is removed. Microdata sets for scientific use allow for complex analysis beyond that provided by the published tables, visualizations, and indicators. Constructing a microdata sample, thereby providing access to only a fraction of the total population, adds a layer of protection while preserving information about population level characteristics. In addition, applying global recodes and local suppressions to the microdata can be used to diminish the risk of disclosure. Also, perturbing the microdata or making use of synthetic methods in a targeted way may help to protect confidential information. Perturbation and synthetic approaches should be used within bounds and rigorously tested as they have the potential to decrease the accuracy and utility of the microdata.

21. A systematic risk versus utility assessment may indicate that a combination of several protection measures is most suitable for a given data release programme. Moreover, it is important to apply any SDC scheme consistently across the entire statistical output, as inconsistencies typically entail additional disclosure risks. Some of the noise-based methods mentioned in paragraph 19 were developed specifically to overcome known shortcomings of more traditional methods, for instance excessive information loss and loss of consistency of local suppression in large table sets.

22. Interactive publication tools are becoming increasingly popular (see paragraph 26). Where these are used, it is important that the SDC scheme can be automated and integrated into these tools. Very flexible tools that serve highly customizable user requests by direct queries to the underlying microdata entail specific additional disclosure risks, for instance from scripted massive and systematic querying attacks. Dedicated SDC schemes are typically needed in such cases.

23. High geographical detail is one of the unique features of census outputs in many countries, reflected in the increasing trend to release more data at the smallest geographical units, nowadays often complemented by grid products (see paragraphs [428–433]). However, releasing data on very small and non-overlapping geographical units (such as administrative versus grid boundaries) entails specific additional disclosure risks (‘geographic differencing’) that should be accounted for in the SDC scheme. For example, post-tabular noise methods have been found effective in this regard by several countries during their preparation of the past census round.

24. Irrespective of the specific SDC scheme adopted, it is essential to explain the broad properties of the scheme so that users are aware. This holds in particular for SDC methods where released data look like untreated data, or for methods that may lead to a limited loss of internal additivity of tables.

B. Dissemination

25. A census is not complete until the information collected is made available to users in a form, and to a timetable, suited to their ever-changing needs. Thus, in disseminating the results of the census, much emphasis should be put on responsiveness to users and on high standards of quality in the production of statistics. Census results should be disseminated simultaneously to all users, and the greatest care should be exercised to avoid the inadvertent

disclosure of information about identifiable individuals. To protect confidentiality, various statistical measures can be applied (see paragraphs 18–24).

26. There are several conventional ways of making the results of a census available to the user:

(a) As published reports (either in hard copy or, more commonly, in digital media) containing standard and pre-agreed tabulations, usually at the national, regional or local district area level, which may be obtained from government agencies or directly from other outlets;

(b) As ad-hoc unpublished reports/abstracts comprising the standard census data but disaggregated by other geographical boundaries and/or sub-groups not published previously (a contribution towards the cost of production may be paid by the requester);

(c) As data products available online through National Statistical Office (NSO) websites or other electronic media. These data products may range from aggregated to micro databases, available for online processing or free download, optionally equipped with dynamic or interactive data visualization tools including mapping features to enhance the value of the statistics;

(d) As commissioned or customized output produced from a database, automated table builder or statistical data service, comprising customized cross-tabulations of variables not otherwise available from standard reports or abstracts but which should conform to the same statistical disclosure controls applied to standard outputs; and

(e) As microdata (often referred to as public use samples), usually available only as a sampled fraction of the total population (often sampled at the household level to include individuals from the household) with applied SDC methods (see paragraphs 18–24). To balance the increased risk of providing microdata, a combination of additional controls may be applied: potential data users are often screened; data are made available in a restricted format only; and data are often supplied or accessed under secure and strictly controlled conditions where thorough steps have been taken to protect the confidentiality of the data.

27. Where customizable dissemination tools are not available or not sufficient to provide specific tabulations required by only a few users, such as certain government offices or specialized research organizations, these can be supplied on demand. Once produced, however, there should be no restriction on making them publicly available through the NSO client and user support teams.

28. Cost effectiveness of print is becoming an issue as the physical copies cannot reach as many users as a publication online. The role of traditional publications, especially in printed form, is near complete obsolescence. While print publications can provide coherent and consistent commentary on individual topics and therefore may suit particular users or markets, users nowadays generally expect interactive, dynamic, digital forms of dissemination.

29. Data publications in electronic format online should provide users with easy means of data retrieval in standard formats or through application programming interfaces (APIs), ideally complemented by user-friendly interactive functionalities to customize outputs. Immediate usability of the data tables online is the most important feature to ensure the searchability and relevance of the census information. International standards for metadata such as the Statistical Data and Metadata eXchange standard (SDMX) should be considered a priority for formatting the output data base. Dissemination strategies should also be harmonized with any national government policies on open data.

30. Online tools for ordering, specifying, customizing and receiving census data products and public use samples (microdata) should be developed wherever possible, ensuring that appropriate measures are in place to protect statistical confidentiality of the data and the security of transmission. In the design of census outputs, consideration should be given to all forms of developing technology used by users, such as smart phones and other portable devices. Also important to consider is the flexibility of the content that users will need. Trends change quickly and a comprehensive set of tables or products that can quickly ensure access to data is critical for the relevance of the census information. Online table builders

allow users to quickly customize their requests and query the data base directly for immediate results. This would reduce the need for a massive data table presence online for users to attempt to search through to find their data if it has been published.

31. Social media have become an increasingly popular and effective means of disseminating small amounts of output, particularly to the non-specialist user or for disseminating timely results relevant to events or important dates. Given the varied platforms that are currently available, text images and infographics can all be used through social media campaigns to highlight the usability of census outputs. The use of such media will often demonstrate an NSO's commitment to engage and establish a dialogue with users in order to respond more readily to their questions and concerns; see also paragraphs [297].

32. While online access or dissemination of such micro- and/or macro-data bases on computer media can greatly contribute to an enlargement of the user base and thus to a greater demand for census data, two cautionary notes are important to keep in mind:

(a) Certain cross-tabulations may have notable quality issues because of non-response, sampling or processing errors, or because of processing or imputation procedures. The census authorities should establish procedures for warning potential users about such problems to help safeguard the credibility of the entire census. Some NSOs suppress the release of certain cross-tabulations for reasons related to substantive quality, although such a policy may often alienate users. Other NSOs will release such cross-tabulations only where there is a clear policy that takes into account both substantive and technical considerations;

(b) Some detailed cross-tabulations and all files with individual records pose risks in respect of disclosing information about identifiable individuals in violation of the rules on census confidentiality. This issue is more fully discussed in paragraphs 15–24.

Both the substantive quality and confidentiality issues need to be addressed and appropriate safeguards established. On the other hand, neither issue should pose any problem with respect to the dissemination of a wide range of census products.

33. A range of products should be available to meet the constantly evolving requirements of users. There is likely to be a need for:

(a) National, regional and local area summaries;

(b) Reports on key findings on particular topics, supplemented by detailed results and analyses either in a standard form for areas down to the more local geographic levels, or more detailed disaggregated statistics on particular topics and populations of interest;

(c) Population profiles or key summary statistics for small areas and small population groups;

(d) Spatial and graphical analyses, including a census atlas, possibly complemented by interactive graphical analysis or mapping tools allowing for user-customized analysis;

(e) Value-added products such as area and/or household classifications; and

(f) Supplementary metadata covering definitions, classifications, and coverage and quality assessments. Metadata may differ for enumeration-based, combined, and register-based censuses, but should be equally thorough across all production modes.

34. The dissemination of census results must adhere to a predetermined structured plan including a pre-announced timetable. This entails releasing national figures promptly according to the schedule, followed by the subsequent release of disaggregated data after a comprehensive processing phase, all in accordance with the established plan. Nevertheless, timeliness is a significant concern in the release phase, and efforts to shorten the release period should be made while ensuring the integrity of the data.

35. The initial release of population counts is generally awaited with anticipation among users ranging from the general public to programme and policy administrators. Thus, some countries release provisional results very soon after enumeration is completed. Subject to change once the full data-processing and verification operations have been completed, these nevertheless provide a general picture of population trends. Data users should, however, be

made aware of implications of using provisional population counts which may differ substantially from the finally produced and validated figures.

36. The schedule and description of upcoming releases of final results and products should be made public early in the process in order to maintain public interest in the census. The releases can be staggered, from simple, short descriptive summaries covering a country's major geographical divisions initially, to more comprehensive cross-tabulations and descriptive thematic and analytical reports later on.

37. Data should in principle be free at the point of access or delivery, but charges, where they are necessary (for example in the case of customized/commissioned outputs), should be set to make access to the results affordable to all types of users. Flexible interactive publication tools that process customized data queries automatically (see paragraph 26) may help to address custom user needs efficiently and thus increase the value of for-free data releases. Online dissemination options have reduced reliance on extensive paper outputs. However, there may still be occasions for NSOs to provide a paid print-on-demand service to supply census material to users who are unable to access or receive digital copies. Such users should not be disadvantaged by the lack of paper-based output.

38. Products should be developed which will allow statistical and geographical information to be delivered together with GIS and/or the use of other APIs to meet as widespread an interest, and with as much flexibility and inter-connectivity, as possible commensurate with the necessary assurances on confidentiality (see paragraph 23). Some desirable product properties are highlighted as follows:

(a) Users should be able to find information quickly and simply and in an open or multiple format;

(b) NSOs will greatly increase the usefulness of their census data by having associated graphic and mapping capabilities embedded within products. Ideally, users should, themselves, be able to generate graphs and/or maps easily, and then to print or plot them or make the images available for other uses. Several countries now produce this kind of census product, sometimes in cooperation with commercial agencies;

(c) Open data formats, specifically those that facilitate interchange, should be prioritized, allowing the 'mashing-up' of census information with other databases, which will offer even further opportunity for the data to be more widely utilized.

39. Thematic mapping and data visualization has become an important element of dissemination of outputs and appeals to NSOs because of its ability to engage with users and increase the reach of census data. But data visualization is a broad field, with content and structures ranging from simple infographics through to sophisticated tools for exploring multi-dimensional data analysis. Moreover, data visualization may present difficulties for some census agencies due to high costs, infrastructure requirements and compliance with individual countries' technical requirements (e.g. accessibility, common look-and-feel). It may be the case that the skills required for effective visualization are in short supply, and that there are problems with dedicating sufficient resources to its development, especially given the budgetary constraints that are being faced by many agencies during the decennial period. However, users are now expecting web content to be visual, engaging and personal, so developing a data visualization capability should be a high priority on many NSOs' wish lists.

40. NSOs must attempt to achieve a balance of producing more traditional tabular data versus the higher production costs of creating data visualizations. This requires a thorough understanding of user segments and their specific needs. Research should be undertaken to better understand end users' data needs (for more on user segmentation see Chapter [x] on communication and outreach). There is no one-size-fits-all solution and many NSOs account for this by maintaining dedicated units or teams responsible for data visualization technology.

C. Documentation and metadata

41. An important component of any country's programme of disseminating the results of its census is a comprehensive portfolio of supporting documentation and metadata to help explain, clarify, and enhance the value of the statistical outputs, particularly with regard to making comparisons with previous censuses and other data sources.

42. A metadata system provides supplementary information on characteristics of surveyed and published data. Each NSO will use its own metadata system based on international standards while corresponding, at the same time, to the specifics of national requirements. Since a census and its results are often closely connected with other areas of statistical activities, it is recommended that the census metadata system in each country should use the same elements as the entire metadata system of the particular NSO. What is also usually necessary, however, is that the census metadata should contain some elements that are used only for that census. The metadata system should also ensure the widest possible comparability of data internationally.

43. The census for the 2030 round should also ensure comparability with data from the previous censuses, while at the same time including new elements relevant for any development that has taken place during the time since that previous census. Thus the metadata system should also be comparable with that of the previous census but updated in line with the needs arising from subsequent developments. The metadata systems of individual NSOs should also reflect the extent to which they use data from direct enumeration and/or administrative data sources.

44. A metadata system should encompass, as a minimum:

- (a) Definitions of terms and concepts used;
- (b) Data dictionary or glossary of terms;
- (c) Explanatory notes to the tables;
- (d) Classifications and nomenclatures;
- (e) The census questions (where the information is collected through a conventional field enumeration process);
- (f) The purposes for which the information is collected, particularly in the case of administrative data; and
- (g) The data sources used, most particularly where data is derived from administrative registers.

45. For indicators for which international standard classifications have been created, those international classifications should be used. For indicators that cannot be classified by such international classifications, new nomenclatures may need to be created. Supporting documentation might cover a wide range of specific issues such as: basic methodology, coverage, response, data sources, pilots and tests, derived variables, Internet responses, imputation, and post-enumeration surveys, as well as reports covering more general descriptions of the census operation as a whole and the quality of the data. The extent of documentation dealing specifically with coverage and quality issues is covered in Chapter [IV] but it is recommended here that, as a minimum, countries should include specific quality and coverage measurements such as levels of response (nationally and locally) and levels of data imputation (for the data source as a whole and for individual topics), as part of the census metadata.

46. Methodological reports are particularly important where the underlying methodology has changed since the previous census (such as moving from a traditional field enumeration to a wholly or partially register-based approach). Such changes are likely to affect the definitions and concepts used and hence the comparability between censuses.

47. In order to have full and comprehensive metadata, census authorities should consult with stakeholders from diverse backgrounds, including government agencies, researchers, policymakers, and the public using, for instance, structured surveys, focus groups, or other

consultations (see paragraphs [255–265]). Feedback may be collected to refine metadata descriptions, making them more intuitive and comprehensive.

D. Archiving and access to closed census records

48. The census is a special statistical data source where continuous and comparable information covering a period of up to 100–150 years may be available. This provides the opportunity to create a unique source of data but puts a big responsibility on the NSO (or appropriate national archive agency) to keep this special historical picture of society for the future. Census data is thus valuable not only for present decision makers and users but also for future generations. The NSO has the responsibility to handle, archive and store this ‘treasure’.

49. Many countries retain the census information relating to individual persons and households only for as long it is required for data processing and the production of the statistical results, or until the entire census operation is conducted. However, the scientific, socio-historical and genealogical value of the individual records should not be underestimated when considering the overall costs and benefits of the census. NSOs may therefore want to allow (public or restricted) access to the full set of census records after a period of closure. If countries do intend to retain the records for such research, they should ensure that there is a robust legal and physical framework in place to protect the security and confidentiality of the records until they become open to the public. National governments should recognize that the ability of NSOs to collect information from the general public may be seriously compromised if assurances given about the confidentiality of the information collected are not honoured. Public confidence in the security and confidentiality of the personal information processed for the census should be regarded, therefore, as paramount.

50. Closure of census records should extend to cover a period that is sufficient to protect the confidentiality of the information, particularly any sensitive information, about living people, or at least to minimize the risk of breaching such confidentiality. The period of closure in many countries is prescribed specifically by statute but may vary from country to country. Other countries may rely on more general provisions within data protection and/or freedom of information legislation to keep confidential records closed until the risk of disclosure of personal information about living individuals has expired. A period of 100 years is therefore recommended, although with life expectancy ever increasing, countries may wish to consider extending this threshold, depending on national circumstances.

51. In addition to requests for public access, NSOs may also receive requests from other government agencies for access to census records for the purpose of validating or corroborating existing information when historical records are sparse or non-existent. Access to these records should be considered when compelled by law or if it clearly serves the public good.

52. In addition to the archiving of the census records (for those countries that do so) it is equally important – more so in fact – for all countries to ensure the preservation of, and easy access to, all the metadata and procedural/operational material, including all project management documentation, created during the entire census process. Not only does this provide a valuable audit trail when coming to evaluate the success and effectiveness of the census, but it will also enable future census planners to learn from the successes achieved, and the challenges faced, by their predecessors.

53. In doing so, countries should ensure that, as technology develops rapidly, the media and systems on which this valuable information is archived are reviewed regularly in order to ensure that it can be retrieved readily whenever it may be required in future years, perhaps as much as 20–50 years hence.

V. Conclusion

54. The draft recommendations on Dissemination and related topics: security, documentation, metadata and archiving for the 2030 round of population and housing censuses are presented for comments and discussion.
