United Nations ECE/CES/GE.41/2024/4

Economic and Social Council

Distr.: General
12 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 (d) of the provisional agenda
Revising the Conference of European Statisticians Recommendations for Population and Housing Censuses for the 2030 round: Technology

Developing the Recommendations on Census Technology

Note by the Conference of European Statisticians Task Force on Census Technology*

Summary

This document includes the draft chapter on technology 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 of national statistical offices and the latest technological and methodological developments.

* The Conference of European Statisticians Task Force on Census Technology consists of the following members: Andres Vikat (UNECE, chair), Régis Plamondon (Canada), John Conway (Ireland), Ana Binstok (Israel), Johan Åmberg (Norway), Espen Andersen (Norway), Galina Sheverdova (Russian Federation), Hannes Reuter (Eurostat) and Seiffe Tadesse (United Nations Statistics Division).

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. In developing the Recommendations for the 2030 round of censuses, the Task Force on Census Technology has addressed technology-related matters in census-taking, with the exception of geographic information systems (GIS), for which a separate task force was established.

2. Section II of this document summarizes the changes introduced in comparison with the Recommendations for the previous, 2020 round.

3. Section III presents the draft chapter on technology for the CES Recommendations for the 2030 round of population and housing censuses.

4. 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 of national statistical offices and the latest technological and methodological developments.

II. Summary of changes from the 2020 Recommendations

5. The starting point for drafting the text in section III of this document was chapter II, Technology, in the CES Recommendations for the 2020 round of censuses[1]. Reflecting the development of technology since then and the practices in the 2020 round, seven new parts were added for the 2030 census round:

(a) The recommended components of testing (section III.B.3);

(b) Design of the electronic questionnaire (III.C.4);

(c) Technology to support the enumeration of the impaired and digitally disconnected (III.C.5);

(d) Interactive online platforms for dissemination (III.D.3);

(e) Storage and archiving (III.D.4);

(f) Cloud (III.D.5);

(g) Technologies for the use of administrative data (III.E).

6. The greatest attention is paid to technological solutions that have become widespread in the practice of the 2020 round of population censuses – electronic data collection tools, cloud technologies and integrated solutions for automatization all stages of census work from census planning to dissemination of its results.

7. Technologies related to geographic information systems (GIS) are proposed as a separate chapter of the Recommendations, the draft of which is presented in a separate document by a dedicated task force.

8. Data security issues receive more attention throughout the chapter. They are highlighted in relation to the need for testing of the integrated census system and its individual components, certification for cybersecurity, separation of critical infrastructures, the Internet response option, the use of portable devices and data storage. A section on security and confidentiality is included in the new part on technologies for the use of administrative data.

9. Well-established technologies, such as telephone interviewing, key-entry and optical reading systems are not covered in detail because these systems are already well understood

[1] UNECE, 2015. Conference of European Statisticians recommendations for the 2020 censuses of population and housing. Geneva: United Nations

and documented. Based on the experience in the 2020 round, automated telephone interviewing is not recommended.

III. Draft text for the chapter on census technology for the Recommendations for the 2030 round of population and housing censuses

A. Introduction

10. Technology has been used to assist in all phases of population and housing censuses. This chapter addresses general considerations for the use and testing of the technologies for the census and provides recommendations on technologies for data collection and processing. Technologies using the data from administrative sources for census purposes are also discussed.

11. Well-established technologies, such as key-entry systems are not covered in detail because these systems are already well understood and documented. The greatest attention is paid to technological solutions that have become widespread in the practice of the 2020 round of population censuses – electronic data collection tools, cloud technologies and integrated solutions for automatization all stages of census work from census planning to dissemination of its results. It is understood that new avenues for the use of technology, such as machine-learning and artificial intelligence, may still unfold in the years leading to the 2030 census round.

B. General considerations

1. Drivers for technological innovation

12. Technology has the *potential* to improve the census coverage and data quality, reduce costs and disseminate results timelier. However, in the short term, the introduction of new technologies may increase costs. Census agencies need to consider how the new opportunities provided by technological innovation may contribute to improving the relevance, accuracy and timeliness of the census data.[2] For example, with reductions in processing cost, it may be possible to expand the content of the census or increase sampling rate for sampled questions to improve the quality of data for thematically for small population groups, or spatially for small geographic areas and even 1km grids. Any content expansion, however, needs to be balanced against the impact on respondents.

13. The demand for evidence-based policy and planning generates a demand for census data from an increasingly *wide range of users*. Output systems therefore need to reach diverse users, ranging from those who would look for quick access to basic headline figures to those expecting to conduct advanced analyses.

2. Evaluation

14. *Adoption of new technologies or methodological approaches should only be considered where there is a sound understanding of their benefits and where their developments can be managed.*[3]

[2] For a detailed discussion on the drivers for technological innovation in direct enumeration, see section B.3 in United Nations Department of Economic and Social Affairs (2019), Guidelines on the use of electronic data collection technologies in population and housing censuses. New York: United Nations.

[3] For a detailed discussion on the decision-making process for using a particular technology, see section B.4 in United Nations Department of Economic and Social Affairs (2019), Guidelines on the use of electronic data collection technologies in population and housing censuses. New York: United Nations.

15. The feasibility of adopting any technology that is previously untested in a census environment should be carefully evaluated in advance, taking into consideration the national context, the relative costs compared with traditional solutions, the work needed for development and testing, the potential implications for the overall organization of the census operations, the potential effects on the quality of census results and the impact on the general population.

16. Census agencies should undertake an evaluation well in advance of the census *to determine what systems and processes are appropriate* for their own situation. Issues to be considered include:

(a) The relative costs of staff and clerical-based processes compared with costs of possible computer systems and associated infrastructure;

(b) The technological capability and infrastructure within both the national statistical office (NSO) and the country as a whole;

(c) The capacity of the NSO to manage complex and sophisticated systems development processes;

(d) The availability of funding and time for developing and testing the technological solution.

17. *Design thinking and behavioural science principles* should be considered throughout the census process. Design thinking will allow for critical reflection of the current processes and evolve to meet future needs of customers and citizens. This should lead to reduced failure demand benefits as citizens are signposted in their journey of completing the census form.

18. Before the adoption of new technologies or methodological approaches, NSO should have a clear understanding of the associated *risks*. How will the new technology perform? What will be the reactions of respondent, or census staff? Because of the long intervals between census cycles (5 to 10 years), opportunities to learn first-hand from new approaches may be limited.

19. In considering the introduction of innovations, the NSO should strongly consider *learning from the experience of other census agencies internationally*. Consideration may also be given to collaboration with other organizations in jointly testing new approaches or technologies before their introduction.

20. The complexity of much of the new software and the infrastructure required for many of the new and emerging technologies may go beyond the current technical capabilities of a NSO to manage. It is likely, therefore, that some countries will want to consider whether significant components of technical solutions to the census operation could be *outsourced*.

3. Testing

21. The NSO should implement a strong *testing* strategy and consider testing activities as a priority for the census. The aspects to be tested include functional correctness, non-functional requirements, interfaces, error handling, performance, reliability, usability and security. Beside the testing cycle, the deployed solutions must be certified for cybersecurity.

22. *It is recommended to include the following components in the testing cycle*:

(a) Domain-specific or content-based tests:

- Qualitative or cognitive tests;
- Tests for new online services;
- Quantitative content test or behaviour test;
- A/B testing (split testing);

(b) Tests specific to information technology (IT):

- Unit tests;
- Field operation tests;

- Interface testing;
- Systems and operational readiness tests;
- Tests for backup and recovery of data;
- Tests against ransomware attacks;
- Volume and performance testing;
- Load stability measure test;
- Security testing such as penetration tests;
- Non-functional testing (validating the performance, reliability, utility, and other non-functional aspects of a software application);

(c) Functional tests:

- System-integrated testing: many iterations during the census cycle;
- Integrated end-to-end tests;
- Acceptance testing.

23. Testing of the coherence of the entire census system is especially important where respondents are provided with multi-mode options for completing the census returns, such as a mobile phone, tablet or desktop computer. Each of these modes need to be tested for accessibility, user experience, availability, performance and security.

24. In the context of the rapid development of IT and the resulting significant changes in many technological elements of the census after the 2020 round, *mandatory testing of the entire technological solution in a pilot census becomes especially important in preparing for the 2030 round census*.

4. Project management

25. Census operations involve a range of administrative processes that are common to other *large-scale projects*. For example, the planning of a complex operation such as the census would benefit from the use of appropriate project planning software. Also, many countries may require systems and processes to recruit and pay large numbers of temporarily employed census enumerators. NSO should consider how technology might assist in improving the efficiency and effectiveness of these operations. This can contribute both to containing the cost of the census and improving the overall quality of the census by allowing resources to be focused on the primary tasks of enumeration, processing and dissemination rather than on administrative processes such as paying staff.

C. Data collection in direct enumeration

26. The following three data collection technologies are considered: the Internet response option; telephone interviewing; and portable devices, such as laptop computers, tablet computers and smartphones.[4]

1. Internet

27. For direct enumeration, it is recommended to offer the option of responding over the Internet as the first or preferred option.

28. Responses to electronic questionnaires reduce overall collection **cost** and achieve data of better *quality*. The data collected over the Internet will be formatted and can benefit from the respondent's self-correction and edits. In comparison, paper data capture requires many processing steps such as scanning, data capture and more importantly keying by operators.

[4] For a comprehensive reading on these and some other technologies, reference is made to United Nations Department of Economic and Social Affairs (2019), Guidelines on the use of electronic data collection technologies in population and housing censuses. New York: United Nations.

The fact that the system needs to interpret different handwriting contributes to the complexity of the process and reduces the accuracy of the data. For all those reasons, the Internet response is clearly preferable.

29. Using the Internet as the medium means that the data is collected though *self-enumeration* rather than an interview. The Internet option can be incorporated into any of the traditional methods of delivering and collecting census forms such as drop-off/pick-up, mail-out, mail back.

30. The key factor to be considered is managing the collection *control* operations – that is, ensuring that every household and individual is counted once and once only. This requires the ability to provide linkage between each household and any individual within the household to its geographic location. An added complication for those countries where forms are also collected by census enumerators (rather than mailed back) is to have suitable and timely feedback to enumerators for them to update their own collection control information so that they do not visit households that have already returned forms.

31. The potential level of *take-up* of an Internet option should be considered by assessing the proportion of the population who can access the Internet from home, the proportion who use broadband services, or the general use of the Internet for other purposes such as banking, filing tax returns or shopping.

32. *Systems and processes* that allow for Internet return of census forms will also need to be developed. These are proven to save costs by reducing enumerator workloads, data capture, printing and postage.

33. Data security is a very important issue and should be a key consideration in designing the infrastructure. *Physically separate infrastructures should be set up to collect and to process the census information. Completed individual census forms, after their collection and capture, should be moved into a secure data processing infrastructure that is separate from the collection infrastructure.*

34. A standard census questionnaire that is downloadable from the Internet requires much less infrastructure than a form that is completed online. However, downloadable forms generally require a greater level of computer literacy than online forms. They will not necessarily work on different computer configurations and there will be an expectation that the NSO will be able to deal with each individual problem. From the respondents' point of view, recent experience has shown that they are much more likely to prefer completing the form online. *For these reasons, forms for online completion are recommended*.

35. Adopting an Internet response option requires the provision of *credentials* to the respondents and methods of delivering of the logins and password needed to access the online form, which might include:

(a) Mailing the paper forms or letters;

(b) Delivery by enumerator directly to the respondent's address;

(c) Sending by email;

(d) Sending by SMS;

(e) Using the logins of online public service portals or other online services requiring a personal identification number.

36. An *online form* offers the possibility of interactive editing to improve response quality that is not possible on a paper form. People using electronic forms have a certain level of expectation that a certain amount of guidance will be offered – at the very minimum that they will be sequenced through the form and not asked questions that are not relevant to their situation. To ensure a high quality of data collected via the Internet, it is important to provide mechanisms to control response errors on the form. Such control should be conducted in real time, and the respondent should be immediately able to modify any incorrect data. If contradictions are found in the respondent's answers, the online form should report this and provide the respondent with the opportunity to correct one or more answers, or delete them, or confirm that reported situation exists in real life, although it is not provided for by the developers of the form.

37. Providing an Internet option may contribute to improving the quality of the census by making it easier for some *hard-to-enumerate groups* to respond. Most countries report difficulties in enumerating, for example, young adults and people living in secured accommodation where access is restricted. Some people with disabilities may also find it easier to complete an Internet form than a paper questionnaire. These groups are also more likely to be using the Internet for other purposes, and therefore, if available, this option should be promoted to these groups as a means of encouraging participation in the census.

38. Provision of sufficient *infrastructure* provides one of the major challenges for offering an Internet option. The census occurs over a relatively short period of time and involves the whole population of a country, and it is unlikely that the NSO will already have the needed infrastructure to cope with the peak demands of a census. It is therefore likely that this component, at least, of the Internet solution can justifiably be outsourced. It may be necessary for collection procedures to be modified to constrain demand. For example, staggering the delivery of census questionnaires or invitation letters or requiring people outside predetermined target populations/areas to contact the NSO before they can use the Internet form may be a means of restricting use of the Internet form.

39. Census agencies should, therefore, assess how they wish to *promote* the use of the Internet. Promotion of the Internet option should be determined by the capacity of the service to handle the expected load and should be coordinated with other data collection procedures. The public relations strategy should encompass assurances about the security and confidentiality of the information supplied via the Internet. Assuming that the Internet option is targeted to the whole population, the public relations strategy should also encompass managing public expectations about the ability to access the site during periods of peak demand. Simple messages of so-called 'graceful referrals' advising people to use the Internet option at 'off peak' times should be prepared and used, if necessary, on the census Internet site itself, through any census telephone inquiry service, and in any media promotion.

40. The take-up of the Internet response option can be expected to increase above the levels observed in the 2020 census round. During the data collection, census agencies should constantly monitor the levels of public response and make an *effort to increase* the level of online response if necessary.

2. Portable devices

41. The increasing sophistication and the reduction in unit costs of communication using laptop computers, tablet computers and smartphones means that these may be a *cost-effective solution* for some aspects of census data collection. Possible applications for such devices include the replacement of enumerator paper maps, address registers and lists, and as a means of data collection in the field. They have possible applications in the full range of census collection methodologies from drop-off/pick up through to the collection of the census questionnaires.

42. Portable devices have the advantage of being able to provide *real time two-way management information*. Census managers can be informed of the progress of the collection operations as the enumerators deliver census forms and collect completed returns. Likewise, census managers can provide the enumerator, via the portable device, with updates on forms received and which households need to be followed-up. Additionally, geospatial information for the collections (e.g. missing addresses, new developments) can be exchanged to allow efficient use of resources. Census managers can identify, in real time, areas where the enumeration is falling behind schedule or not meeting quality standards and instigate appropriate interventions.

43. Use of portable devices should allow greater opportunities for increased efficiency in data collection. However, several technical issues need to be considered in using such devices:

(a) Screen size may affect the ability of the enumerator to record and verify responses accurately. For the same reason, responding with mobile devices over the Internet risks fragmentation of data due to the small size of the screens;

(b) The compact and lightweight devices with sufficiently large storage capacity are most convenient for the field work of enumerators. The brightness and contrast of the screen should be adjustable to use the device both in bright and in dark light;

(c) To ensure the safety of data, completed information should be held in the devices for as short a time as possible – preferably no longer than 24 hours;

(d) Devices should be able to deal with being offline for periods of time. The length of battery life should be considered in relation to the daily workloads of field staff. It may be worth providing an additional power bank for the device;

(e) If system and software updates at the data collection stage, it is necessary to avoid the risks of loss of previously collected data or their inconsistency with the data collected after the update;

(f) The GPS accuracy (e.g. dense urban fabric) and the mobile signal reception (e.g. mountainous or forest areas) may not be satisfactory on some areas of the country. An assessment of mobile web connectivity should be done particularly if the portable device uses web-based collection.

44. Solutions based on portable devices should be extensively *tested* before the census phase, both on their own and in interaction with other elements of census technology that do not use portable devices.

45. There is also a range of **s***ecurity* issues associated with the use of portable devices:

(a) There is a greater risk of being stolen or lost compared with bundles of paper forms. However, regular uploading of the data from such devices should minimize the need to re-enumerate areas if the devices are lost;

(b) Measures are needed to protect the confidentiality of any data either on the device, in the event of loss of the device, or in transmission of the data. Data stored on the devices should be encrypted and only accessible through dedicated protection measures (e.g. passwords, fingerprints);

(c) Transmission of the data also needs to be secured through encryption and use of secure channels end to end;

(d) Security software should be loaded to the device and must be compatible with the other applications on the device. However, security software and passwords add an extra level of complication in use. These security measures will add to the support costs;

(e) Training and technical support for enumeration staff is an important issue. It should not be assumed that the people who are likely to be recruited for enumerator tasks are technically competent. These factors become increasingly complex and difficult to manage as the size of the enumerator work force, and the physical distances involved, increase. In larger countries, enumerators may be able to rely on training and technical support being delivered remotely via the Internet or phone.

46. The training tools for the portable devices should be uploaded to the device for the convenience of their use by the enumerators for the training and during the field work, cover all the elements of the enumerator's work, be interactive, have easy navigation, contain illustrative examples of the enumerator's reaction in all possible situations of using this device.

47. Census agencies should think ahead about using the large number of devices after the census. It is impractical to store devices for the next census since they can become technologically outdated and unusable in 5 to 10 years without using and recharging. Census agencies may transfer some of these devices to other users (e.g. the government organizations) while keeping some of the devices.

3. Telephone

48. In the past, automated telephone interviewing has previously been suggested as a potentially cost-effective solution for countries that have a ‘short form’ census questionnaire

requiring only the capture of basic demographic information. However, no country applied it in the 2020 census round. *Automated telephone interviewing is not recommended.*

49. *Computer Assisted Telephone Interviewing* (CATI) method can be used to collect data via the census questionnaire and/or to verify and complete any missing data collected on a long-form questionnaire. Countries should be aware, however, that the user-friendliness of such systems decreases greatly as either the number and complexity of the questions increase or the number of people in the household increases.

4. Design of the electronic questionnaire

50. The design of the electronic questionnaire is a very important part of the technological solution when responses are provided online or collected using portable devices.

51. *The design of the electronic questionnaire should take into account the following requirements*:

(a) Contain a complete set and a clear sequence of the questions, which are divided into open and closed questions;

(b) Have answer options with the possibility to choose only one option or several answer options for closed questions, and if the "other" option is selected, allow typing in the respondent's own answer;

(c) Fit the entire one question and its answer on the device screen without scroll or skip to the next screen, if possible, because the hidden part of question or answer options may be missed when filling in;

(d) Make available the help option is for the questions such as a text hint or a jump to the appropriate element of the metadata or the training materials;

(e) Provide for easy navigation in between the questions to one respondent, between members of the same household and between different sections of the questionnaire (e.g. about the housing conditions, about the household, about the person);

(f) Use built-in controls for the validity of the entered data, taking into account previously entered information about this person and other members of this household;

(g) Display the progress bar for questionnaire completion as well as the general quantitative characteristics of the completed questionnaire, such as number of persons in the household and completeness of answers for each person.

52. In the case of using an electronic questionnaire for both online self-completion by respondent and for enumerator’s device, the design of the questionnaire may differ because the respondents have no knowledge of the census methodology, whereas the enumerator is pre-trained and familiar with census terminology and the metadata built in the questionnaire.

53. *The online form should additionally* contain *a summary of the basic requirements* for completing it by the respondent. These include:

(a) A description of the general structure of the questionnaire and the sequence of its completion;

(b) An estimated time of completing the questions per one respondent or one household;

(c) A description of ways to call up help information and respond to error messages;

(d) A description of the possibility to correct, delete or add some information to the previously completed questionnaire, if necessary;

(e) The signs of successful and unsuccessful completion of the census process and further actions of the respondent (e.g. obtaining confirmation of participation in the census);

(f) A way of feedback to NSO, such as a telephone number, email address of the census hotline or of the NSO, to evaluate the quality of online services or ask questions that could not be answered when filling out the form;

(g) The ability to get translation of the form into the most popular languages in the country if necessary;

(h) It may also be useful to provide answers to frequently asked questions with terminology accessible to respondents and provide a clickable link to the page of the NSO website where the census legal, methodological and organizational principles are described.

5. Technology to support the enumeration of the impaired and digitally disconnected

54. When introducing new technologies, it is necessary to keep in mind that the census must *cover the entire population*, regardless of the technical equipment used and computer use proficiency of the respondent.

55. Technology can support the enumeration of the impaired and digitally disconnected *in two main ways*: (a) in reaching respondents that do not have the necessary Internet connection; (b) to allow for as many people as possible to enter their responses electronically.

56. *The digitally disconnected*. In order to reach as many people as possible it will be required to think about locations that do not have proper internet connectivity. One aspect that we should think about in those cases is unfortunately have a paper option. However, other options should also be explored such as having people with satellite internet access being able to gather information with portable devices. Also, giving the respondent a phone number to call so they can fill out their form with an operator.

57. *Accessible internet response*. Another aspect that we should consider is how people with limitations could fill out their census form over the Internet. Internet response should follow accessibility standards as defined by Web Content Accessibility Guidelines (WCAG) 2.2. While adhering to those standards technically, it is important to consider accessibility already at the stage of developing the content. Before developing any new features that could impact accessibility, consultation with a centre of expertise in accessibility and a user experience group should be performed to ensure new functionalities are designed to be properly accessible. The following paragraphs describe examples of what could be done.

58. *Hidden text for auto-generated character mask fields*. This includes numerous masked input types that help ensure that the content conforms to expected formats, while auto-generating delimiting characters such as hyphens in the date field. Hidden instruction text that is only visible to screen readers is provided, to inform vision impaired users of the necessary inputs required to be typed by the user, and those that will be provided automatically via JavaScript.

59. *Required colour contrast for viewing of text content*. To ensure users can view text content, analysis is done on all text to ensure that their RGB hexadecimal colour differences are sufficient to meet the standards of WCAG 2.0 level AA for a contrast ratio of at least 4.5:1 for normal text and 3:1 for large text (14 point (typically 18.66px) and bold or larger, or 18 point (typically 24px) or larger).

60. *Labels, tab-stops and aria-described by usage to provide context of the hierarchy of fields***.** As some question layouts in surveys can have multiple layers of text, following the WCAG standards for Meaningful Sequences has led to automatic generation of tab-stops (i.e., blue rectangles) that follow the page in a way someone would read it (left to right, top to bottom). The adherence to the WCAG requirement for Labels or Instructions has led to the creation of label text for all inputs (i.e., the Yes radio button, and checkboxes[5]), which is read by screen readers when focus is put on the input itself. Lastly, to comply with the WCAG success criterion of Info and Relationships, *aria-described by* attributes could be implemented on inputs, where for example, focusing on the checkbox in the green rectangle below, the application would read the text sequentially. This provides context to help the user decide whether to mark the checkbox.

[5] W3C 2023. Web content accessibility guidelines (WCAG) 2.2. W3C Recommendation 5 October 2023.

6. Census management software

61. At a basic level, multi-modal collection operations require that *timely information* be provided to census enumerators so that they do not visit households that have already submitted a census form. This is both an efficiency issue and a public relations issue. Modern technologies provide opportunities to improve the management of field operations and thus the quality of the census itself.

62. While the key issue is the flow of timely information to the census enumerator, the same systems should also provide for a close to real time *two-way flow of information between census managers and enumeration staff*. Such monitoring of enumerator work will allow for more timely interventions where the data collection process is falling behind schedule or there are some problems with the quality of the data collected.

63. The NSO may need to rely on external organizations for key parts of the solution. Regardless of if these systems are internal or external, they must adhere to internationally (e.g. ISO27000 family), regionally (e.g. NIS2[6] in the European Union) or nationally agreed *cyber security standards*. This is especially important as software for various external or internal operations (collection, processing, geographical information systems, imputation, dissemination) are provided by a variety of solutions (e.g. one-stop-shop, proprietary, custom developments).

64. An integrated field communication system can use, and build on, already existing IT infrastructure. This implies, for example, direct access to the national geocoding infrastructure providing unit record data of single addresses. *The geospatial information (addresses, buildings, cadastral parcels) should rely on using permanent identifiers (PID).*

65. *A census management infrastructure should contain the following* elements:

(a) a register of dwellings addresses with geospatial coordinates, in which all addresses are distributed by enumeration areas;

(b) a register of enumerators and their contact details with the possibility of linking to each of them a certain enumeration area (from the register of dwellings) and the including addresses – if this data collection method is used;

(c) a register of devices for data collection (for example, tablet computers or smartphones) and their unique serial numbers with the possibility of linking them to the enumerators (from the register of enumerators) – if this data collection method is used;

(d) a central census data storage using for the collecting, processing and accumulation of all data linking to the respondent's residential address (from the register of dwellings) that received by all data collection methods used in the census (online responses, enumerators data, data from administrative sources), and the data collecting method should be indicated in the storage;

(e) communication software to enable timely exchange between enumerators, supervisors and the census management team.

66. All elements of this infrastructure should be *interconnected and managed centrally* using software and technology tools specially developed for the census purposes. *It is recommended to perform the installation, configuration and performance check of the devices in the census management centre before delivering the devices to the field personnel.*

67. Centralized management of the portable devices for the census data collection includes *the automatization of the following functions* (if this data collection method is used):

(a) Installation on the devices of the following: special software for the census data collection by the enumerator (the electronic questionnaire); list of dwellings addresses and maps of the enumeration area; metadata and classifiers used in the electronic questionnaire; tools for the monitoring the operation of the device/enumerator; training materials for the enumerators/device users;

[6] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union.

(b) Single-valued linking between the device, the enumerator and the enumeration area in the registers of dwellings, enumerators and devices, update of the registers and links between their elements in case of the enumerators or devices replacing;

(c) Online management and monitoring of each device's operation after its initialization by the field staff;

(d) Clearing the device of all information and software used for the census purpose after the successful transfer of the collected data to the central data storage at the end of the census and preparing the devices for a long-term keeping (conservation) or use for other purposes (for example, transfer to another agency).

68. The *online management and monitoring of device's operation* includes:

(a) Obtaining information about the time when the device is turned on and off;

(b) Obtaining geo-coordinates for filling out an electronic questionnaire at each address and filling statistics, such as the duration of filling, and number of completed questions, errors and corrections in the forms;

(c) Transmission to the devices the information on the online responses addresses and the control data on these respondents to enable the enumerator to verify the census completeness at this address or make corrections to the data according to the respondent's request, if necessary;

(d) Remote installation of software updates on the devices in case of emergency. It is recommended to use this only when critical software errors are detected during the census, since it is important to ensure the consistency of data collected using different software versions;

(e) Locking the device in case a field worker reports the loss or theft of the device to prevent illegal use of the device or information leakage from it;

(f) Providing a means of remote consultation of the enumerator with the central census office.

69. The *central census data storage management tools* provide the following functions:

(a) Downloading of forms filled in by the online census respondents (if this data collection method is used);

(b) Verification of their suitability for processing (availability of answers necessary for the respondent's identification);

(c) Linking completed forms to the addresses of dwellings;

(d) Creation of a confirmation, such as a QR code, of successful participation in the census for the respondent's feedback (depending on the method of receiving the completed form – to the personal account of the online census respondent, by email or other means) and for the transmission to the enumerator – if this data collection method is used;

(e) Downloading completed electronic questionnaires from enumerator's devices, from administrative sources and other data collection channels used in the census, linking them to the addresses of dwellings;

(f) Consolidation of versions of completed census forms obtained from all data collection channels used in the census relating to the same dwelling address – selection of the reference version, enriching it with data from other versions, removing duplicates if necessary;

(g) Calculation of the census progress statistics and visualization of work monitoring to identify problematic census areas that need to be affected to correct the situation during the census.

70. *Behavioural science* can also assist through the understanding of the respondent's motivations, feelings and experiences. Examples to increase response rates include timely interventions, better communication, and lowering the cognitive level required to complete an online or paper survey return.

71. Another key factor in reducing risk is the relationship with key *technology partners*. Strong governance where the in-house census management team remains at the core, are critical to ensuring Census systems are designed, implemented, and delivered successfully.

72. Technology solutions are now available which can *combine multiple field management functions*. Customer Relationship Management (CRM) can provide solutions for field application management, front-end website, enhanced communication through chat-bot functionality, and a citizen helpdesk for knowledge management.

73. NSOs could consider *remote-access technology* to support flexible working arrangements of staff who process the census data while ensuring data security and confidentiality.

D. Data processing

74. Recent years have seen significant *improvements in processing technologies* that have reduced the cost of census processing, and improved data quality.

75. Based on the conclusions from a review of countries' practices in the 2020 census round, it can be assumed that most countries with direct enumeration in their next census will use the *Internet response option*. The use of paper forms and OCR/OMR technology could be assumed as limited.

76. Comprehensive planning for *integration between key census systems* will be required, to ensure a better seamless journey as survey return data moves from the field to being processed. Instilling strong governance on the different partners who may be involved will also be critical in the successful integration between the field, processing and dissemination systems.

1. Processing of paper questionnaires

77. Data collection with paper questionnaires may be necessary because of respondents' preference or lack of access to the Internet. **F***or processing the paper questionnaires, it is recommended to use automated processes such as Intelligent Character Recognition (ICR).*

78. Optical Mark Recognition (OMR) can be a cost-effective option where the census form contains only **t***ick-box responses*. Additional means of data capture or computer-assisted coding operation are required to handle write-in responses. However, OMR has largely been superseded by ICR technologies.

79. The most cost-effective option is likely to be *a combination of digital imaging, ICR, repair and automated coding*. An example of this process is briefly described below.

(a) The census forms are processed through scanners to produce an image. Recognition software is used to identify tick box responses and translate handwritten responses into textual values. Confidence levels are set to determine which responses are of acceptable quality and which responses require further repair or validation;

(b) Automated repair is designed to reduce the need for operator intervention and typically involves the use of dictionary look-up tables and contextual editing. The dictionaries are tailored according to the census question being processed. Thus, for example, the dictionary for country of birth question would only contain names of countries. Preparatory work on the construction of natural language dictionaries of terms will greatly increase the efficiency of coding;

(c) Operator repair can be undertaken on images not recognized. This is only cost-effective for those questions where there is a high probability that the repaired data can then be automatically coded;

(d) Automatic coding uses computerized algorithms to match captured responses against indexes. Those responses that cannot be matched are then passed to a computer-assisted coding process. For the responses that cannot be automatically coded, it is recommended to use a machine learning algorithm that could replace human coders with as good and even better data quality and highly reduce cost. Data from a previous census can

be used to train the machine learning algorithm. Data from the current census testing cycle could also be used, especially if new variables must be coded. To further constrain costs and improve quality, responses that are not coded should be analysed for common responses. These responses could either be added to the coding indexes and resubmitted through automated coding, or some other form of bulk coding could be undertaken.

80. Further considerations on the use of digital imaging, ICR, repair, automated coding, Optical Mark Recognition (OMR) and Optical Character Recognition (OCR) are presented in the CES census *recommendations for the 2020 round*.[7] Generative artificial intelligence can be expected to lead to new possibilities and replace keyers. However, it requires investment to build proper models to keep high quality of data.

2. Output production

81. Traditionally, census output comprises aggregated tables, statistics, illustrations and maps with appropriate metadata. (See [cross-reference to the relevant chapter].)

82. *Online dissemination* via the Internet additionally allows for the design of products to meet better the needs of different kinds of census data users, for the cost-effective dissemination of a much wider range of census data and for the improved usability of the data. Application programming interfaces (API) should be made available to increase usability of the provided data.

83. Functionality and data content can be targeted to satisfy *the different levels of users*. This functionality should be seamless – from the simple to the sophisticated – with the users being led by the nature of the query or analysis they are wishing to undertake using different products.

84. One of the main objectives of the census is to produce information for *small geographic areas and for small population groups* (both social and economic). Internet dissemination can support both types of use of the data. For small geographic areas, GIS technology can be used as means for both defining areas of interest in searching for data and for mapping of the outputs of the search. There is a range of software packages that can be used to zone in on populations of interest from large pre-defined matrix tables.

85. The Internet dissemination system should provide *flexibility* for users to export the results into a range of commonly available packages for statistical analysis, tabulation or mapping.

3. Interactive online platforms for dissemination

86. As the amount of data gathered increases at spatial, temporal or semantic levels (e.g. European 1km grid), *data visualizations* have been introduced to help people, business owners, academics, and management at all levels, to understand key information derived from the data. Data visualization can be used to communicate a message quickly, to simplify the presentation of large amounts of data, to see data patterns and relationships, and to monitor changes in variables over time.[8]

87. *It is recommended to disseminate the data on interactive online* platforms *where different user groups could directly make the tables and visualizations (e.g. graphs, maps) according to their needs.*

88. Focus should be on the application of the FAIR guiding principles for scientific data management and stewardship.[9] The principles emphasize the capacity of computational systems to **f**ind, **a**ccess, **i**nteroperate, and **r**euse data with none or minimal human intervention. The analytical solutions in the platforms may remain limited as this is not the main purpose of the census.

[7] UNECE 2015. Conference of European Statisticians recommendations for the 2020 censuses of population and housing. Geneva: United Nations.

[8] For example, Eurostat Census Hub.

[9] GO FAIR Initiative.

89. Census data users have very different levels of statistical literacy or computer proficiency. *It is recommended to consider the following groups of users:*

(a) Light users who mainly look at data visualizations or report highlights;

(b) Intermediate users who customise visualizations or select specific data;

(c) Technically advanced users who would download data to perform their own analysis and use application programming interfaces (API).

90. *International standards* such as SDMX should be considered a priority for formatting the output data and metadata that is made available to users.

91. Whatever the means of access or dissemination, *protecting the statistical confidentiality* of the census data is a prime consideration in any such systems and statistical disclosure control procedures should be in place.

4. Storage and archiving

92. Data storage technology has evolved in line with new technology trends. Historically, data has been stored on-premises, using traditional media tapes sometimes complemented by disk storage. However, as the volume of data has continued to increase exponentially, the capabilities of storage solutions have had to adapt. Many organizations are now considering implementing a *hybrid storage* approach, utilizing a mix of Cloud and on-premises storage solutions based on individual needs.

93. Certain factors will need to be considered in *determining what storage policy is appropriate* for individual statistical organizations. These include:

(a) The risk for utilizing additional storage solutions that are not on-premises;

(b) New skillsets will be required to develop and maintain advanced storage solutions such as a hybrid approach. These can be difficult to obtain and costly to retain;

(c) Each statistical organization should ensure that they understand their data in terms of data classification. This will aid the decision making required when determining how to build their storage design capability;

(d) As organizations consider new storage solutions, cyber security safeguards are required to identify, detect, protect, respond and recover from cyber risks and threats. Their cyber defence strategy will need to ensure storage protections are available in a hybrid environment;

(e) There are many cloud-based storage approaches including glacier storage for long term retention of data that is not frequently used. Again organizations which historically have relied in backing up all data using full, incremental and/or differential methods, will need to understand to develop a deeper understanding of their data when to leverage the capabilities of these new storage approaches;

(f) In partnering with external storage service providers, organizations will need to ensure appropriate service level agreements are defined that meet their expectations in service delivery;

(g) Storage and archiving should follow legal texts and well-established rules. For example, in the European Union, any scientific manuscript data should be kept for 10 years;[10]

(h) Backup and recovery procedures must be implemented and tested to perform any steps needed to restart the systems and services, to verify their operations and data integrity;

(i) A business continuity plan is recommended to be in place and tested for the prevention and recovery of the systems and their related data for threats like natural disasters or cyber-attacks.

[10] Commission Decision (EU) 2021/2121 of 6 July 2020 on records management and archives.

5. Cloud

94. Cloud computing is growing very rapidly and is expected to become *the most common IT infrastructure in businesses* across the world. Cloud adoption strategies have ranged between cloud-only, cloud-first and, more recently, cloud-smart.[11]

95. Cloud adoption is also becoming prominent for government entities worldwide, as they recognize the transformative *potential* of cloud computing in enhancing operational efficiency, scalability, and service delivery. Embracing cloud technology allows them to optimize resource utilization, improve data accessibility, and foster innovation across various domains.

96. The development of the Cloud has evolved to include *multiple platforms and service models*. Service models include Infrastructure as a Service (IAAS), Platform as a Service (PAAS), and Software as a Service (SAAS). Platforms include public, private and community cloud. Each of these individual models can provide opportunities and challenges for statistical organizations in their adoption and deployment.

97. Many *considerations* to the adoption of Cloud by statistical organizations should be assessed. For example, global Cloud providers (hyperscalers) rely on global contracts that may not align to local laws and legislation. One area for consideration include jurisdiction in the event of a dispute. This will require localized agreements agreed at a national level. This will help reduce a significant barrier to Cloud adoption. Data sovereignty is also a major consideration and requires agreement.

98. The 2024 UNECE *guidelines for Cloud adoption*[12] explore many of the key themes, opportunities and challenges associated with the Cloud. They can assist statistical organizations to gain a deeper understanding of the issues that are relevant to them.

99. A *transitionary approach* to the adoption of the Cloud with incremental steps could be appropriate for many statistical organizations. Beginning with less sensitive applications and systems, and learning from the experiences, can provide a greater degree of comfort to the key stakeholders making the decision to adopt the Cloud. Ensuring that the risk is assessed and measured on an ongoing basis will also help with the transition.

100. Cloud infrastructure can bring important advantages *if the census has a 5 to 10 years cycle*. A benefit of using the Cloud is the possibility to scale up and down easily when required. This could end up in cost saving for the census programme but requires nonetheless to proceed with caution and consideration of broader implications for the agency. For example, on-premises infrastructure built for the census could be reused for other purposes of the agency. So, for the census programme, there will be cost-saving possibilities, but it should be evaluated at a higher level. For a shorter cycle such as an annual census, it may not be the best approach.

101. Further *advantages* are related to the investment to the new cloud technologies such as cloud native and Paas. Also, a multi-cloud architecture should also be considered in order of benefits from the strength of different cloud providers. For example, there could be one cloud infrastructure for the application and another one that will host the database.

102. Cloud infrastructure allows a lot of flexibility but should be *strictly managed* in order not to pay for infrastructure that is not used. Otherwise, using the Cloud may turn out to be the more costly option. Cloud servers should be scaled down when not used and scaled up when required.

103. Cloud migration process should be well evaluated and well planned. Just migrating to the Cloud to use another data centre would not be a recommended approach. This was proved in some country that this approach is cost more than staying on premises. Different countries should look at being cloud-smart, that is moving to the Cloud by rebuilding the application into the Cloud or leveraging on cloud-native aspect (Containerized, Paas, continuous integration and continuous deployment).

[11] UNECE 2024. Cloud for official statistics. Geneva: United Nations.
[12] ibid.

104. *Using the Cloud can bring important advantages to the census but is not always the best approach. It is recommended to do a thorough evaluation before moving to the Cloud.*

E. Technologies for the use of administrative data

1. Scope

105. The technologies applied for the use of administrative data differ from those for data collection in the field. The development and increase in the availability of *new information and telecommunication technology* (ICT) allows administrative registers to be utilized more widely in population and housing censuses. Bearing in mind the development of state-of-the-art technologies and the commitment of agencies to implement innovative solutions in censuses in the 2030 census round, it will be necessary to create or modernize the software and hardware infrastructure for collecting, storing and linking data from administrative sources and storing metadata on processes and products.

106. The quality of the source data has a large impact on the quality of output products. Therefore, the methodology for *improving the quality of* data *from administrative sources*, for example, by adjusting them to statistical requirements, is of vital importance. State-of-the-art ICTs may prove very useful here and have a key impact on improving the efficiency and effectiveness of these operations. For assessing the quality of administrative sources for use in censuses, reference is made to the UNECE guidelines on this matter, published in 2021[13].

107. As part of the preparatory work for the census, particularly in the design phase, the necessary *technical requirements* related to the use of data from administrative registers, which may affect the need to modernize infrastructure, should be determined in the following *areas*:

(a) Data collection;

(b) Data storage;

(c) Data linking;

(d) Storage of metadata on processes and products.

108. The application of several techniques of collecting data from administrative registers and other sources for use in population and housing censuses will require a more comprehensive *organization and management processes* and more complex systems. Modern technologies provide opportunities for improvement in this case as well. The process of collecting data from administrative registers should include the preparation of a data-collection strategy using various data-collection modes.

2. Security and confidentiality

109. It is crucial to consider the *growing emphasis* on data security, privacy, and data protection in society. This is evident in the legal and statistical frameworks of many countries. Moreover, EU member states must comply to the General Data Protection Regulation (GDPR)[14]. Consequently, there are increased requirements on how the agencies conducting a register-based population and housing census to receive and process data. To address these demands, a strong emphasis on the need-to-know principle and processing anonymized data is recommended.

110. Secure IT infrastructure is the necessary condition for data collection from administrative registers. A crucial issue connected with the process of data collection is the protection of data. Regardless of the technology applied, the data collection strategy should

[13] UNECE 2021. Guidelines for assessing the quality of administrative sources for use in censuses. Geneva: United Nations.

[14] Regulation (Eu) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

ensure *information security*. This requirement should be addressed at the early stages of designing the process of obtaining and gathering data from administrative registers and designing the proper software and hardware infrastructure. The technical issues concerning the coding of data transmission should be considered in detail, together with the use of secure transmission channels.

111. Security must be ensured against loss of data implement *appropriate technical and organizational measures* to protect data against accidental or unlawful destruction or accidental loss (including backups), alteration, unauthorized disclosure or access.

112. Security must be implemented in *multiple layers*. Register owners, data-collecting NSOs, and other partners must establish secure transmission channels with corresponding certificates to control access granted to collectors.

113. By organizing data into different "states" — such as source data, prepared data, statistical data, and output data — *authorization management* can be effectively implemented for various roles and teams within the NSO. Data access should be enforced on a need-to-know basis, with anonymized identifiers and disclosure controls applied across all aggregation levels.

3. Data linking, transfer and storage

114. Modern technologies are certainly useful in the process of linking records and data. After identifying the administrative sources to be used in the census, it would be necessary to map them and create application programming interfaces (API) for the *automatic flow of data* to the central.

115. Administrative sources are different from each other, both within different subject areas within a country and between countries. If possible, it is advised to use standardized APIs. Regardless of the interface used or differences encountered, *good metadata and good understanding of data* are very important for their inclusion in the census.

116. Many countries are currently transitioning to cloud-based storage and computing. Compared to traditional on-premises solutions, this shift offers new possibilities, including the potential for a more flexible, variable-controlled universe of data, or *data lakes*, rather than traditionally thematically stored, structured, and updated data. These data lakes will place high demands on metadata and accurate periodization, made more feasible by the increased processing power of cloud-based computing. Centralizing the administrative data into a data lake would support the business process and can provide a strong and reliable infrastructure for storing, creating metadata and synchronizing the different sources of data.

4. Improving the administrative data

117. Various techniques are useful for converting administrative data into statistical data. With the procedure of *automatic data cleaning* in place, it is possible to eliminate errors in source data from administrative registers and edit the data efficiently.

118. *Machine learning models* could be used, for example, in determining the "census address" of the individual or type of the private household. Building a deep neural network (DNN) based on artificial neural networks (ANN) training model system that teaches the model.

119. *Artificial intelligence* (AI) could be used for determining the classification of employment of economic branch and professional occupation of the individual. An IT team with data engineers and AI specialists could be built to develop neuro-linguistic programming (NLP) and DNN models. These NLP classification models need constant updates and quality analysis to evaluate the results. Classification using AI can be more efficient, give quality results and save time with approximately 90 per cent of the cases classified in this way. It is understood that specific security and confidentiality requirements may not always allow working with such data and models in the Cloud. There are also known limitations of using classification models developed in English in the context of specific other languages.

IV. Conclusion

120. The draft recommendations on census technology for the 2030 round of population and housing censuses are presented for comments and discussion.