



Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Sixteenth Meeting**

Geneva, 23-26 September 2014

Item 4 of the provisional agenda

Census technology**Census technology - Draft text for the Conference of European Statisticians Recommendations for the 2020 census round****Note by the UNECE Task Force on Census technology***Summary*

This document presents the draft text on census technology and outsourcing for the new Conference of European Statisticians (CES) Recommendations for the 2020 Round of Population and Housing Censuses. It was prepared by the UNECE Task Force on Census technology based on the first proposal discussed at the September 2013 meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, and on further discussion within the Task Force and with the UNECE Steering Group on Population and Housing Censuses.

I. Introduction

1. Technology has been used to assist in all phases of population censuses for many years. The focus of this chapter is on new technologies that might be used for direct census collection, processing and dissemination activities. Some of these technologies have been used in past census activity in some countries and are presented here to provide a review of possible options to other countries. There is no doubt that emerging, or yet to be developed, technologies will impact future census taking.

2. Well-established technologies such as key-entry-systems or Optical Character Recognition (OCR)/Optical Mark Recognition (OMR) technologies are not covered because these systems are already well understood and documented. Countries interested in traditional technologies should refer to *Principles and Recommendations for Population and Housing Censuses – Revision 3*, United Nations, New York 2015 and *Handbook on Census Management for Population and Housing Censuses*, United Nations, New York 2001 [NOTE: references to be updated when new Revision will be available]. These well-established approaches and technologies might continue to be the most viable option for many countries. Adoption of new technology or approaches should only be considered where there is a sound understanding of these benefits and where their developments can be managed. There should also be a clear understanding of associated risks.

3. The feasibility of adopting any technology that is untested in a census environment should be carefully evaluated in advance, taking into consideration the national context and in particular factors such as the size of the country, the relative costs of these technologies compared to traditional solutions, the work needed to develop and test the technologies, and the potential implications of the adoption of the technologies on the overall organisation of the census operations. Potential effects on the quality of census results as well as the impact on the general population need to be carefully considered. International cooperation and consultation would be a good practice in the decision to adopt a new technology.

4. Nor does this chapter also focus on ancillary systems that are required to conduct a census. Census operations involve a range of administrative processes that are common to other large-scale projects. For example, planning of a complex operation such as the census may be assisted by use of appropriate project planning software. Many countries may require systems and processes to recruit and pay large numbers of temporarily employed census enumerators. The National Statistics Office 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 as well as 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.

5. 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 many census agencies. It is likely therefore that some countries will want to consider whether or not significant components of any (technical) solutions to the census operation could be outsourced. This chapter presents some general considerations on the potential benefits of outsourcing specific components of the census, the necessary requirements, and the implications of outsourcing on the census process, in particular with regard to confidentiality and quality assurance.

II. Drivers for technological innovation

6. Census programs across countries have been fertile grounds for the introduction of innovations. Many of the new approaches that are used on survey programmes today had their origin on Censuses. The use of technology has most often been the principal enabler of innovation. In contemplating the numerous challenges facing the taking of a Census of Population or Housing, census organisations should consider how technological innovations can be applied to any steps of their process to increase efficiencies, quality, or even reduce costs. Such considerations should also be done in the larger context of the statistical organisation as in time of fiscal constraints, new approaches can benefit more than just the census.

7. The introduction of innovative technologies or approaches may present various risks to census operations. How will they perform? What will be the reactions of respondent, or census staff? Because of the long time spread between census cycles (5 to 10 years), opportunities to learn first-hand from new approaches may be limited. In considering the introduction of innovations, the census organisation should strongly consider learning from the experience of other census organisations internationally. Consideration may also be given to collaboration with other organisations in jointly testing new approaches or technologies before their introduction.

8. Technology has the potential to greatly reduce the cost and improve the quality of censuses. In the short term, however, introduction of new technology can actually increase costs. Seeking international cooperation and consultation would be a good practice in the decision to adopt a new technology. Census agencies need to consider how the new opportunities provided by technological innovation may contribute to improving the relevance, quality and timeliness of the census. 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 small population groups and small geographic areas. Any content expansion, however, needs to be balanced against the impact on respondents.

9. The demand for evidence-based policy and planning generates a demand for census data from a wider range of users beyond the traditional government users. Output systems therefore need to be able to cope with a diverse range of users who may have limited knowledge of census data and who may no longer be content with the limited tabular output that may have been traditionally available.

10. In some countries there is a legislative requirement that governments should permit citizens to conduct government business electronically. Even without such provisions, growing use of the Internet in the community generates expectations that this is the way to do business. It should be noted that, in the census context, it is possible that the use of the Internet, at least in the initial stages, may increase the overall cost of the census.

III. Determining what systems are appropriate

11. Census agencies need to undertake an evaluation 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 census agency and the country as a whole; and
- (c) The capacity of the census agency to manage complex and sophisticated systems development processes.

IV. Outsourcing

12. 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 many census agencies. It is likely therefore that some countries will want to consider whether or not significant components of any technical solutions to the census operation (but also some non-technical activities like large scale recruitment of field staff) could be outsourced. The value of doing so is that external suppliers bring with them considerable technical experience and expertise which would otherwise be unavailable to

census takers, and allows NSIs to focus on their main task of carrying out the census rather than developing in-house procedures and skills that are not part of their core competencies. Furthermore, the 5- or 10-year cycle for the traditional census activities, the short processing timetable and extensive data systems required, mean that outsourcing can provide the opportunity for efficiencies and better value for money.

13. The appropriateness of contracting out should be determined following a methodological or step-by-step approach and after subdividing the overall census operation into separate stages. It is likely that a number of components will offer themselves as potential candidates for outsourcing. These might include (but are not limited to):

- (a) printing of questionnaires and other field documents/material
- (b) recruitment and training of field staff
- (c) publicity campaign
- (d) translation of field materials into other languages
- (e) delivery and/or collection/return of questionnaires/field documents
- (f) design and provision of questionnaire tracking system
- (g) provision of mapping services
- (h) primary data capture and coding
- (i) provision of call centre/telephone help line
- (j) design and provision of online response technology
- (k) online/web access design for outputs
- (l) data storage
- (m) data editing and/or imputation
- (n) post-processing questionnaire destruction
- (o) quality assurance
- (p) evaluation.

14. Many of these activities will, of course, be less relevant to those countries that carry out a full register-based census, in which the opportunities for effective outsourcing are clearly much reduced — if one considers that the creation and maintenance, by external agencies, of the registers from which the census information is extracted, is not “outsourcing” in the generally accepted meaning of the concept.

15. The decision to outsource will depend on the requirements of the census agency, whether the skills are available in-house and the ability of the census agency to manage complex system development projects. Total outsourcing might seem at first to be a simpler process to manage. However, it is unlikely that a census agency will choose to outsource the total solution to the census operation, but rather identify discrete components of the system that involve a combination of outsourced components, different external service providers working as contractors on specific projects, and in-house developments. Nevertheless, there will be advantages if several linked activities are provided by a single contractor; for example, the technical requirements for the printing and scanning of the questionnaires are so closely inter-related that a single contractor to provide both services would seem to be a prerequisite.

16. A clear understanding of requirements is needed before any contracts can be tendered so that these can be specified unambiguously to the contractor. These include

understanding the objectives of the project, the outputs to be achieved and the standards these outputs must meet (quality, timeliness, cost), and confidentiality and public sensitivity issues. Specifications must allow for the possibility of requirements changing over the lifetime of the project. How these changes are agreed and approved by the census agency and the provider needs to be determined and managed.

17. Timetabling, including milestones for key deliverables linked to payment schedules, needs to be agreed with the contractor. Regular monitoring on a routine basis needs to be undertaken at an operational level. In addition, processes should be established to allow senior staff to monitor progress and to deal with any major issues that cannot be resolved at the operational level.

18. A mixed approach to systems development is one in which the overall system may consist of outsourced systems, systems developed by external contractors working alongside census agency staff and systems developed in-house. This approach can have many advantages such as greater flexibility to adapt systems as more is learned through the systems development, systems testing program and actual census processing operations.

19. This can lead to improved data quality and savings in processing costs as systems are optimised. However, census agencies will need to be aware that, for such an approach, management becomes much more complex. The census agency must be skilled in the management of complex projects, have a clear understanding of business processes and manage carefully the integration of both the technological and clerical processes. Team-based working, where external contractors work very closely with census agency staff is essential, if this method of systems development is to be successful

20. Throughout the overall process, activities should be conducted by a method that can best meet the requirements of users (with regard to the accuracy and timeliness of the results) and reassure the general public on matters relating to confidentiality and data protection. Indeed, it is important that no part of the outsourced operations should be done by a method that may result in loss of trust of the general public. So, in judging the propriety of contracting out, it is recommended that census agencies should carefully consider the following criteria:

- (a) strict protection of data confidentiality;
- (b) the method of confidentiality assurance that satisfies the general public;
- (c) relevant government policies and procedures, if applicable;
- (d) guaranteed measures of quality assurance;
- (e) ability to manage and monitor the outsourced census tasks/activities; and
- (f) control over the core competence of the national statistical office.

21. Confidentiality assurance is perhaps the first foremost issue that has to be considered, rising above cost and efficiency. NSIs are responsible for data confidentiality, in terms of both reality and perception. Consequently, contracting out of tasks that pose an actual or perceived risk to security of data confidentiality risk should be avoided. For example, for any the data collection or processing operations carried out by external suppliers safeguards should be put in place to ensure that there is strict protection of confidentiality. The contracting out of such services should be carefully considered so that public trust and confidence in the census is not eroded. Where, for example, temporary enumeration staff are engaged under contract, this should be done in such a way that they are subject to strict measures of monitoring and control by the census agency. These enumeration staff should be engaged in such a way that their activities are governed by the relevant statistical legislation to preserve the confidentiality of the data they collect.

22. Another key issue to be considered in outsourcing is the quality assurance that should be guaranteed. A key point is that the census agency must be satisfied that the goods or services paid for are provided to an agreed and acceptable standard. In this respect, cost should not be the first priority in considering and judging the successful bidder. Although it is desirable to engage in fair competition among several companies to reduce costs, the census agency must be aware that merely considering low price bidding as a determining factor may adversely affect the quality of the service provided by the successful bidder. Low quality work could cause as significant a loss of trust among the user and general public as would the risk to confidentiality.

23. To assess the quality of work, as part of any contract allocation process, potential contractors should be required to provide samples of their work or to list referees who could be contacted to verify their claims and/or sites at which previous work can be inspected. Once the contract has been awarded, continuous monitoring of the progress of work entrusted to the selected company is necessary and the census agency should ensure that a system for monitoring quality is built into the contract.

24. Further discussion of the issues to be considered in outsourcing, and in the evaluation of software and hardware can be found in the Principles and Recommendations for Population and Housing Censuses, Revision 3, United Nations, New York, 2015, and the Handbook on Census Management for Population and Housing Censuses, United Nations, New York 2001. [NOTE: references to be updated when new Revision will be available]

V. Data collection

25. The following three data collection technologies are considered: Internet response option, telephone interviewing, hand-held devices e.g. tablet computers, laptop computers, pocket computers, smart phones, mobile, or cellular phone. Technology now allows collection and processing for some steps to be done simultaneously. The opportunities that new technologies offer for better managing the collection operations are also discussed. The use of technology in developing the manner of connections with administrative registers for census purposes is also discussed.

A. Internet response option

26. Using the Internet as a collection method means that the census collection methodology will need to be self-enumeration rather than interview based. The Internet option can be incorporated into any of the traditional methods of delivering and collecting census forms (for example drop-off/pick-up, mail-out, mail back). The key factor is managing 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 individual to theirs a geographic location. An added complication for those countries where forms are collected by census enumerators (rather than mailed back) is to have adequate and timely feedback to enumerators so that they can update their own collection control information so that they do not visit households that have already returned forms.

27. 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 and the general use of the Internet for other business purposes (for example on-line banking, filing tax forms, shopping). It is worth mentioning that some target groups (e.g. younger age classes in urban areas) may prefer to

fill in the forms via the Internet, and therefore delivering the paper questionnaire to the respondent may be unnecessary. It is worth considering therefore only sending out access information to the internet response option. Systems and processes that allow for Internet return of census forms will also need to be developed. These may increase costs. On the other end there are potential savings in data capture costs, as well as printing and postage costs. However, scanning and Intelligent Character Recognition are in themselves cost efficient. Therefore, savings in data capture costs may be less than the costs of developing and implementing the internet system.

28. Security is an important consideration. Industry standard encryption (SSL128) offers two-way encryption (encrypts data flowing both from and to the user's computer) and has been accepted by nearly all countries as adequate to protect the census information. Security should be a key consideration in designing the infrastructure. A physically separate infrastructure should be set up to collect the census information. Completed individual census forms should be moved behind firewalls into infrastructure that is completely separate from the collection infrastructure.

29. A downloadable on-line form requires much less infrastructure than for forms that are completed on line. However, downloadable forms require a greater level of computer literacy than on-line forms. They will not necessarily work on different computer configurations and there will be an expectation that the census agency will be able to deal with each individual problem. From the respondents' point of view, they are much more likely to prefer completing the form on-line. For these reasons it is expected that most countries will adopt on-line completion of census forms.

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

- (a) posting the paper forms or letters ;
- (b) delivery by enumerator directly to the respondent's address;
- (c) defining the logins and passwords based on data coming from registers. The combination of data typically for the whole population should be used to authenticate the respondents identity (by using for example Personal Identification Number).

31. An electronic 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 a minimum that they will be sequenced through the form and not asked questions that are not relevant to their situation. In order to ensure high quality of data collected via the Internet, it is important to provide mechanisms to control errors in the form. Control should be conducted in real time. The respondent should be able to modify the incorrect data.

32. 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 young, adults and people living in secured accommodation where access is restricted. Some people with disabilities will also find it easier to complete an Internet form than a paper form. These groups are also more likely to be using the Internet and, if available, this option should be promoted to these groups as a means of encouraging participation in the census.

33. 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 affects the whole population of a country, and it is unlikely that the census agency will have adequate 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 census agency before they can use the Internet form may be a means of restricting use of the Internet form. Census agencies need to 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 the collection procedures. The public relations strategy will need to encompass assurance about security of information supplied via the Internet. Assuming that the Internet option is targeted to the whole population, the public relations strategy should encompass managing public expectations about the ability to access the site during periods of peak demand. Simple messages advising people to use the internet option at “off peak” times should be prepared and used if necessary on the census internet site itself and through the census telephone inquiry service, radio and print.

34. Taking into account experience from 2010 census round it is not possible to predict the percentage of shares response in CAWI method. It depends on promote the use of the Internet and attitude the society to modern technology. During the census, NSIs should constantly monitor public participation and make an effort to increase the per cent of response.

B. Telephone Interviewing

35. Automated telephone interviewing generally may be a cost-effective solution particularly for countries that have a “short form” census questionnaire requiring only the capture of basic family and demographic information.

36. As with the Internet solution, each household would need a unique code to enable proper collection control.

37. Voice recognition software can be used to lead the respondent through the census form with either voice recognition or the phone keypad used as the response mechanism. Confirmation that important census variables such as date of birth or age have been captured correctly would need to be provided.

38. 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.

39. 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.

C. Hand-held devices

40. The increasing sophistication and the reduction in unit costs for hand-held devices means that these may be a cost-effective solution for some aspects of census collection. Possible applications for such devices include the replacement of enumerator paper maps, address registers and lists as a means of data capture in the field. They have application in the full range of census collection methodologies from drop-off/pick up through to census form collection.

41. Hand held 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 and collect completed census forms. Likewise census managers can provide the enumerator, via the hand held device, with updates on forms received and which households need to be followed-up. Census managers

can identify, in real time, areas where the enumeration is falling behind or not meeting quality standards and undertake appropriate interventions.

42. Use of hand-held devices should allow greater opportunities for greater efficiency in data collection. Several technical issues need to be considered in using these devices. Storage capacity is related to the cost of the devices, but is of itself not now a limitation on their use. To ensure the safety of data, completed information should be kept in the hand-held devices for as short a time as possible, and preferably no longer than 24 hours. The length of battery life should be considered. If the census enumerators work all day, they will need an additional charging source for hand-held devices. Hand-held devices should be able to deal with being off line for periods of time. System and software updates should be avoided during the data collection phase since this may result in loss of data. An assessment of mobile web connectivity should be done particularly if the hand-held device uses web based collection. Solutions based on hand-held devices should be extensively tested before the census phase.

43. There is also a range of security issues associated with the use of hand held devices. These carry, for example, a greater risk of being stolen or lost compared with paper forms. Regular uploading of the data from the hand held devices should minimise the need to re-enumerate areas if the devices are lost. Measures are also needed to protect the confidentiality of any data either on the device in case of loss of the device or in transmission of the data. Data on hand held devices should be encrypted and only accessible through password protection. Transmission of the data also needs to be secured through encryption and use of secure channels. There are several commercial software packages that can be used for these purposes. These security measures will add to the support costs. Security software has to be loaded to the hand held device and must be compatible with the other applications on the hand held device. Security software and passwords add an extra level of complication in use.

44. Training and technical support for enumeration staff is an important issue. It cannot 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 increase. In larger countries, enumerators may be relying on training, and technical support delivered remotely via the Internet or phone.

VII. Census management software

45. 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. Also, the modern technologies provide opportunities to improve the management of field operations and thus the quality of the census itself.

46. While the key issue is the flow of timely information to the census enumerator, the same systems set up to ensure this can also provide for a close to real time two-way flow of information between census managers and enumeration staff. Timely monitoring of enumerator work will allow for more timely interventions where census collection is falling behind.

47. It is unlikely that the census agency will have the knowledge or capacity to develop and run these systems in-house and will need to rely on external organisations for key parts of the solution. An integrated field communication system can use and build on already existing infrastructure present in most developed countries. The following contains a brief description of a possible solution that uses a combination of census agency developed

systems and processes, call centres and mobile (cell) phone technology. It also presumes that the census agency has a central register of all enumerators, their enumeration areas and their mobile (cell) phone number:

(a) The census agency might want to establish an electronic central register of forms received either by mail or by the Internet, which could be compared against the census frame (address register or other source). Such a register would need to contain the unique identification number for each form and/or dwelling (e.g. geographic coordinates, internet access code, barcode information, etc.). This census reference number would allow the identification of the enumerator responsible for that enumeration area and ultimately a physical address (for doorstep collections the actual address may only be known at this stage by the census enumerator).

(b) Most census agencies establish some form of telephone inquiry service to handle queries from the public during the census collection period. Call centres use technologies that can be readily adapted to meet census requirements. The call centre systems record callers' "census reference number" or derive the census reference number from the address supplied by the caller. The call centre also records what action is required for the enumerator (for example census form needs to be picked up, assistance is required in completing the form, a form has not been received etc.)

(c) The information from the electronic register and the call centre could be sent to the enumerator as a Short Messaging Service (SMS) (or text) message. This message only need contain the census reference number and a code to indicate the action required on the part of the enumerator.

(d) The solution can be cost-effective as it relies on readily available commercial infrastructure (call centres) in a highly competitive industry (and most of this cost is not a new requirement for censuses) and the fact that most enumerators are likely to own their own mobile phones. The cost of SMS messaging is small. The census agency needs to develop the electronic register and manage the integration of the various systems.

VIII. GIS Technology

48. There is widespread recognition that it is important for national statistical agencies to develop a continuing capability to serve their specialized cartographic needs. It should be stressed that there is now a wide range of techniques and technologies available for use in a census mapping exercise. There is a strong and growing interest in the use of GIS as a tool to support the process of conducting the census (use of vector data and digital statistical division boundaries), and as a tool for friendly visualization of statistical results (GIS web-based mapping tools to disseminate census results). Maps, which are now commonly in the form of digital products, play an increasingly important role in the dissemination phase.

49. GIS as far as it is possible should be used at all stages of the census (inventory preparation, progress monitoring, and dissemination of census results). Despite a significant majority of countries implementing GIS technology in their National Statistical Institute it is still the case that in many countries there are only a limited range of maps available and these often do not show sufficient detail to enable the boundaries of small areas to be clearly defined.

50. GIS technology should be considered only at a level appropriate to the skills and resources available, and constitute an integral part of the overall work of a national statistical organization. Collaborative arrangements with other agencies, such as national mapping and survey agencies should be pursued particularly with regard to the acquisition and maintenance of base maps and digital databases, which should not become a

responsibility of national statistical organizations. Therefore as many as possible administrative sources should be inspected for presence of spatial referenced data and tested for possibility of using it for the statistical purposes.

51. Wherever it is possible, data should be collected with reference to an address point- results can then be disseminated using any desired spatial divisions;

52. A very broad and comprehensive description of the use of maps and GIS technology at all stages of census has been developed in the document "*Principles and Recommendations for Population and Housing Censuses, Revision 3*" prepared by the United Nations Statistics Division [NOTE: references to be updated when new Revision will be available]. It is highly recommended to use those recommendations- particular attention should be paid to chapters:

- (a) III.A.8 Mapping;
- (b) III.F.3 Geographic information systems;
- (c) VIII.B Census mapping;
- (d) VIII.C Interactive digital outputs.[NOTE: references to the publication and the chapters will be updated when new Revision will be available].

IX. Processing

53. The last decade has seen significant improvements in intelligent character recognition, data repair, imaging and automated coding technologies that have reduced the cost of census processing, and improved data quality.

54. Based on the conclusions from the 2010 Census Round, it can be assumed that more countries will use an Internet response option. Many countries will, however, still choose to use paper forms and OCR/OMR technology, but a declining trend in the use of this technology has been observed.

A. Data capture methods

55. Several data capture technologies have traditionally been used such as key entry and optical mark recognition (OMR).

56. Key entry requires simple software and low-end computing hardware. However, it requires many more staff than other automated methods of data entry and is likely to take more elapsed time to complete. The cost-effectiveness of this method is dependent on the relativities between staff costs and hardware/system development costs required for other methods.

57. OMR can be a cost-effective option where the census form contains only tick-box responses. Additional means of data capture/computer assisted coding operation are required to handle write-in responses. However, OMR has largely been superseded by intelligent character recognition (ICR) technologies.

58. For most countries, 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 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 reduce the efficiency of coding.

(c) Operator repair can be undertaken on images not recognised. 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 computerised algorithms to match captured responses against indexes. Those responses that cannot be matched are then passed to a computer assisted coding process. To further contain costs and improve quality, responses that are not coded should be analysed for common responses. These responses could be either, added to the coding indexes and resubmitted through automated coding, or some other form of bulk coding be undertaken.

59. The combination of ICR/automatic coding/imaging technology described above is likely to prove the most cost-effective solution for most countries. Staffing can be reduced through automatic coding and use of these systems. Use of images greatly reduces the need to move paper forms and it has been shown that referring to images for follow up coding of responses that cannot be automatically captured is much more efficient than referring to paper forms.

60. Importantly, this methodology offers the opportunity to improve the quality of the data. The consistent treatment of identical responses can be guaranteed. However, the quality of the automated capture and coding need to be carefully monitored during processing to ensure that the system is functioning as specified. Character substitution rates should be monitored closely and for critical questions or parts of questions (such as the year of birth as compared with the day of birth) may require more stringent confidence rules that require higher level of inspection and quality assurance than other fields or values. Numeric values in particular may require extra scrutiny as there may be no contextual information that can be used to automatically check their validity. The design of the work flow for repair and coding can greatly increase the efficiency and accuracy of the process by routing the results of particular questions to specialised operators or coding in blocks of questions.

61. There should be ongoing quality assurance of the final outputs of the system – such as manually recoding from the images a sample of the responses and comparing them with the automatically captured and coded responses. This should allow a proper balancing of the quality/cost equation including the reducing the amount of manual repair and not therefore wasting resources for marginal improvements in quality.

62. For this reason, it is vital that even if these systems are outsourced, that census managers have a good understanding of quality/cost equation implicit in the confidence levels being implemented in the ICR/data repair software, their affect on the substitution rate and the ultimate quality of the census data. Contracts should allow ready adjustments of these parameters to meet the quality and the operational requirements of the census agency. This issue is considered in further detail in paragraphs below [NOTE: references to be updated when new Revision will be available]

63. Census agencies need to consider how the data are going to be held through the processing stream. Traditionally, census processing has been conducted using a flat file that gets progressively updated with the earlier version of the file retained for backup and

recovery processes. Typically this has been allied with batch processing where a discrete group of forms (typically for an enumeration area) are processed together. Thus the forms will be data entered, edited and coded as a group. This allows a high degree of workload control. Databases allow information to be held and processed at the individual field level. This provides a greater degree of flexibility as once census data is electronically captured it easily organised to maximise both processing efficiency and quality as similar responses can be readily grouped and coded together. However, holding the census data in a database requires more complex systems to manage and deliver work. Consideration also needs to be given to backup and recovery mechanisms.

64. Early examination of standard tables of processed results for a defined area should be done to ensure that results are internally consistent and compare well with other sources this can identify coding biases and sub optimal OMR recognition. Examination of coding profiles across individual operators should be done identify coding biases and overuse of bucket codes.

65. These systems typically require far more extensive systems development and testing than traditional census processing system. There are a number of factors that need to be considered through the systems design and integrated into the systems development such as the work organisation of the remaining clerical processes.

66. Adequate network capacity is critical because of the large number and size of the files associated with images –techniques such as form drop-out, where fixed background information is dropped from the final image can greatly reduce the size of the files. This need to be designed into the Census form and tested throughout the form printing process to ensure that print density is consistent with form drop out.

B. Output

[Note: It might be useful to cross reference with text on Dissemination in the Chapter on operational issues]

67. Traditionally, census output comprises aggregated tables, statistics, illustrations, maps, etc. with appropriate Metadata.

68. Internet dissemination allows for: the design of appropriate products to meet the needs of different types of census data users from novice to sophisticated users; the cost-effective dissemination of a much wider range of census data; and improved usability of the data.

69. Functionality and data content can be targeted to satisfy the different levels of users. This functionality should be seamless from simple to sophisticate with the clients being led by the nature of the query or analysis they are wishing to undertake through the different products.

70. 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 packages that can be used to zone in on populations of interest from large pre-defined matrix tables.

71. The Internet dissemination system should provide freedom for clients to specify the form of the output – whether as hard copy or as a data file that can be exported into a range of commonly available statistical analysis, tabulation or mapping packages.

72. Some countries may wish to consider providing access to clients to submit tabulations directly off-line or on-line to be run against the census unit record file.

73. Protecting the confidentiality of the census data is a prime consideration in such systems. As well as implementing statistical disclosure control procedures (see paragraph 61) [NOTE: Update reference to text in Chapter on Methodology], there may be a requirement to limit the size of tabulations that can be submitted through this method.

C. The use of technology in adopting new census methodologies

74. The use of technology in the census depends on the applied methodology. Technology, which is used in a register-based censuses is applied to a lesser extent than that used in the combined censuses, where apart from technology used for administrative data, following technologies are used additionally: GIS, CAWI, CATI, CAPI or OMR, OCR technologies, etc.

75. The development and increase in availability of new information and telecommunication technologies (ICT) for national statistical institutes allows administrative registers to be utilised more widely in population and housing censuses. The efforts of statistical institutes to make greater use of data from administrative registers in censuses are primarily dictated by the necessity to reduce data collection costs and the burden on respondents. The use of data from administrative registers is particularly effective in statistical institutes where data from registers is used not as part of a single action, but continuously and consistently in many data collection operations for the purposes of various surveys. This issue is important due to the necessity to incur increased expenditure in the initial period, which is needed to use data from administrative registers.

76. The new approaches to statistics production based on multimodal data collection and integration, now increasingly adopted by national statistical institutes, trigger the necessity of creating or modernising a suitable software and hardware infrastructure. The dynamic development, in the 2010 census round, of new ICT and their increasingly broad application in statistics production, even in countries with a long-standing tradition of administrative data use in censuses, caused the necessity to modernise the infrastructure for collecting, storing and linking data from administrative registers and storing meta-information on data, processes and products. Bearing in mind the development of state-of-the-art technologies and the commitment of national statistical institutes to implement innovative solutions in censuses in the 2020 census round it will be inevitable to create or modernise the software and hardware infrastructure.

77. The quality of the source data including data from administrative registers, has a large impact on the quality of output data (products). Therefore, the methodology of improving the quality of data from administrative registers, i.e. adjusting them to statistical requirements, is of vital importance. The use of data from administrative registers as a rule requires a more comprehensive preparation than in the case of traditional censuses. State-of-the-art information and communication technologies (ICT) may prove very useful here and have a key impact on improving the efficiency and effectiveness of these operations.

78. Census results are the outcome of the implementation of a large number of processes and sub-processes. As the quality of census results largely depends on the quality of these processes, measures should be undertaken to improve the quality of statistical processes by determining and measuring the key variables in the process, i.e. the variables which have the greatest impact on product quality. Process-quality indicators should be used to evaluate the impact of process quality on product quality. The use of the Generic Statistical Business Process Model (GSBPM) to design, describe, and define the set of statistical processes in the census might contribute to improved census quality.

79. As part of the preparatory work for the census, in the process of designing, 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 or information on processes and products.

80. The application of several techniques of collecting data from administrative registers and data from other sources used in population and housing censuses will require a more comprehensive organisation 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.

81. In the 2010 round censuses National Statistical Institutes used various methods of obtaining data from administrative registers. The data obtaining process was dominated by electronic data collection:

- (a) transfer of data sets through a secure ICT channel, by administrative data keepers,
- (b) delivery/receipt of external electronic data carriers containing data files directly by/from administrative data keepers,
- (c) remote access and capturing of data in electronic form, allowing its processing.

82. The condition necessary for electronic data collection from administrative registers is the preparation of secure IT infrastructure. A crucial issue connected with the process of data collection is the protection of data. Regardless of the technology applied, the data collection strategy, often including the entire population, should ensure information security. This issue should be taken into consideration already at the stage of designing the process of obtaining and gathering data from administrative registers and designing the proper software and hardware infrastructure. A number of technical issues concerning the coding of data transmission should be considered in detail, together with the use of secure transmission channels.

83. In the 2010 census round many countries used various techniques of automatically converting administrative data into statistical data, which in the context of a dynamic development of ICT technologies and a growing quantity and availability of administrative registers for statistical purposes should be taken into consideration in the 2020 census round. During the 2010 census round the following techniques were used: validation, conversion, parsing, standardization and upcase (convert all letters to uppercase letters). With the procedure of automatic data cleaning in place, it is possible to eliminate errors in source data from administrative registers. The use of data cleaning tools makes data editing process efficient. Modern technologies might prove useful in the process of linking records and data.

84. In the processing of data from administrative registers, i.e. improving their applicability to censuses as data sources, without the negative impact of the applied sources on the quality of output data, the following elements should be taken into account: the collection transformation model, i.e. the applied set of rules and principles of data control and adjustment reflecting the concept of the required quality level of input data. Moreover,

it is recommended that national statistical institutes should prepare descriptions of specific procedures of transforming data from administrative registers and data quality measurements on the basis of the applied scope of validation checks and the agreed principles. It is also essential to ensure the documentation of the data transformation process; this will allow the monitoring of the progress and the evaluation of the data processing.

85. It is recommended as good practice to use metadata in steering and monitoring the processing of data from administrative registers, including their transformation. The use of metadata and principles should be used in the programming of the implementation of processes, collecting information on the course of the processes, and the outcomes of their implementation. Ensuring the high quality of census data is the overriding objective of these operations.
