

**UNITED NATIONS SECRETARIAT
ESA/STAT/AC.162**

**Department of Economic and Social Affairs
Statistics Division**

**July 2008
English only**

 **2010 World Population and Housing Census Programme**

**Report of the UNSD-ESCWA Regional Workshop on Census Data Processing
in the ESCWA Region: Contemporary Technologies for Data Capture,
Methodology and Practice of Data Editing**

Doha, Qatar, 18-22 May 2008

Table of Contents

Table of Contents	2
INTRODUCTION	3
Objectives of the Workshop.....	3
Attendance	3
Session 1: Opening	3
PRESENTATIONS AND DISCUSSIONS FROM THE VARIOUS SESSIONS	4
Session 2: 2010 World Population and Housing Census Programme	4
Session 3: Census Planning and Management	5
Session 4: Introduction to Data Capture	6
Session 5: Outsourcing versus in-house Processing.....	7
Session 6: Data Capture: Optical Mark Recognition (OMR)	7
Session 7: Data Capture: OCR / ICR / IR.....	8
Session 8: Data Collection: PDA-Handheld-computers/Internet	8
Sessions 9 and 10: Data Capture: Process Stages	9
Session 11: Data Capture: Overview of Major Distributors/ Commercial Suppliers.....	10
Session 12: Data Coding	10
Session 13: Introduction to Data Editing.....	11
Session 14: Concepts and Methods in Data Editing	11
Session 15: Data Editing (Practical Exercises).....	11
Session 16: Country Presentations on Data Processing	12
Session 17: Country Presentations on data Processing (continued).....	12
RECOMMENDATIONS & CONCLUSIONS	13
EVALUATION OF THE WORKSHOP	14
ANNEXES	15

INTRODUCTION

Objectives of the Workshop

1. The purpose of the Workshop was to present international standards for processing population and housing censuses and to highlight the significant additional capabilities of contemporary technologies and their use for census data capture and data editing. More specifically, the workshop aimed to: (1) Present an overview of the UNSD 2010 World Programme on Population and Housing Censuses; (2) Discuss methods of improving the management and planning of the census, including outsourcing issues; (3) Present and discuss contemporary technologies in census data capture, including the use of Optical Mark Recognition (OMR), Optical Character Recognition/Intelligent Character Recognition (OCR/ICR), Internet data collection, use of handheld devices (PDAs) for data collection; (4) Discuss the process stages for data capture; (5) Present an overview of major commercial suppliers for data capture; and (6) Present the principles and practices for census data coding and data editing. The workshop also offered the possibility to the participants to present the experience of their countries in census data processing.

Attendance

2. The seminar was attended by 44 participants from 12 countries (Bahrain, Egypt, Jordan, Iraq, Kuwait, Lebanon, Morocco, Palestine, Qatar, Saudi Arabia, Syria, UAE), by two International /regional organizations (ESCWA and UNSD), and by four commercial providers (DRS, Lockheed Martin, Betasystems, and Intergraph).

Session 1: Opening

3. After welcoming the participants to the workshop, His Excellency the Sheikh Hamad Bin Jabor Bin Jassim Al-Thani, Acting President of Qatar Statistics Authority (QSA) expressed his pleasure in co-organizing the workshop together with the UNSD and ESCWA. He stressed the importance of conducting the workshop as a part of preparation for the population and housing censuses in the region, as it aspired to deal with the state-of-the-art technology for capturing and processing data, that would help in production of better quality census results. He also mentioned the significant scope the workshop presented for exchange of experience between the countries of the ESCWA region and North Africa in conducting earlier censuses, particularly in respect of effective use of available technology. He wished the participants success in accomplishing the goals of producing accurate and timely census results with the knowledge of advanced technologies acquired in the workshop.

4. On behalf of Dr. Paul Cheung, Director of the United Nations Statistics Division (UNSD), Ms. Diane Stukel welcomed the participants to the workshop. She explained that this workshop was part of the 2010 World Programme for Population and Housing Censuses, initiated by the United Nations Statistical Commission in March 2005 for the period 2005 to 2014. The three essential goals of the programme are: (i) to agree on a set of acceptable international principles

and recommendations governing the conduct of a census; (ii) to facilitate countries in conducting censuses during the period 2005-2014; and (iii) to assist countries in their efforts to disseminate census results in a timely manner. Ms. Stukel informed the participants that as part of the 2010 World Programme on censuses, the United Nations Statistics Division has conducted a series of regional workshops over the last two years, having had the two themes of the Principles and Recommendations for Population and Housing Censuses (2006) and Geographic Information Systems and Digital Mapping (2007), respectively. For this year (2008), the theme of the set of regional workshops is to be Data Processing, including the topics of data capture and data editing. The present workshop in Qatar is the inaugural one in a series of 5 regional workshops that are envisioned over the remainder of the calendar year. Ms. Stukel reiterated the expected outcomes of the workshop and she outlined how the workshop would be structured and reviewed the agenda for the five days.

5. On behalf of ESCWA, Mr Alope Kar thanked UNSD for co-sponsoring the workshop, and in particular the Qatar Statistical Authority for hosting and logistical support. He apologized on behalf of Mr Giovanni Savio, who could not attend the workshop due to the special circumstances in Lebanon. Mr Kar highlighted the importance of census-taking in general - as a source of information on social, demographic and economic characteristics for small geographic areas or sub-populations, and also as a basis for developing sampling frames for household surveys. He also emphasized the importance of new technologies used in OMR and OCR/ICR as a means of improving the collection, processing and dissemination of more accurate and timely data for census. He closed by saying that he hoped this important initiative would be an occasion for participants to be better informed about the new technologies and thus to make decisions about the use of such technologies in their own national population censuses.

PRESENTATIONS AND DISCUSSIONS FROM THE VARIOUS SESSIONS

Session 2: 2010 World Population and Housing Census Programme

6. An overview of the World Population and Housing Census Programme for 2010 was presented by UNSD. The three essential goals set for the 2010 programme were reiterated and the specific role of the UNSD in respect of these was outlined. UNSD has recently published the second revision of the *Principles and Recommendations for Population and Housing Censuses* and released it this year. UNSD, in partnership with the UNICEF and UNFPA, are developing dissemination software called *CENSUSINFO*, based on the original *DEVINFO*, but with some improved functionalities considered more appropriate for census data.

7. In the discussions that followed, it was clarified that the software would be provided free-of-charge, and training on the software would be available. It was reiterated that the dissemination and analysis of census results, which have been weak in the past, would be given more focused attention in this round.

Session 3: Census Planning and Management

8. This session, presented by UNSD, dealt with three main issues relating to planning and management of population censuses, that is: i) planning of census operations, ii) quality assurance and iii) evaluation.

Planning of Census Operations

9. It was noted that the census is usually the biggest statistical operation carried out in a country. Thus, it is essential to carefully plan each aspect of census operations from data capture to data dissemination through to proper evaluation. A full census cycle consists of the following phases:

- Preparation
- Field operations
- Processing of census data
- Dissemination of census results and preparation of report
- Evaluation

10. It was emphasized that each of these phases has to be properly resourced and organized, so that its output is of adequate quality. Each phase of the census cycle is dependent on a preceding phase. It is important to identify the dependencies between different phases. Due to long duration of census cycle, planning should not remain static but should be dynamic and continuous. It should be flexible enough to take into account changes that occur along the way.

Quality Assurance

11. In the second part of the presentation, an overview of the notion of quality assurance was given. It was mentioned that due to the size and complexity of census operations, it is likely that errors may arise at any stage of the census. Census data are usually subject to variety of errors including those related to coverage and content. To minimize and control errors, it is good practice to devote a part of the budget to quality assurance and control programs. Quality Assurance (Continuous Quality Improvement) has an emphasis on improving the processes rather than just fixing the errors (which is more the focus of quality control). Quality assurance systems recognize that there will be errors in the processes and aims to improve them as it proceeds.

Evaluation

12. In the third part of the presentation, evaluation of the census processes and outputs was highlighted. It was mentioned that all aspects and operations of the census program should be evaluated. Evaluation of census operations should include every aspect of the work from initial planning and user consultation through to final dissemination and analysis

13. In contrast, the limited exercises of evaluating the coverage and content of the census data can be undertaken through the following means:

- Post-enumeration surveys - mainly to measure extent of under or over-coverage of the population
- Comparing the census results with similar data from other sources such as surveys and previous census, and by using analytical methods, mainly to assess the degree of errors in the content and to analyze the basic distributional properties

14. The session concluded with a brief discussion on importance of dissemination and analysis of census results. It was stressed that this aspect should be recognized as an integral part of planning for census; otherwise the agency responsible for carrying out census would run out of resources to complete this part, since it takes place near the end of the cycle.

Session 4: Introduction to Data Capture

15. This session was devoted to a discussion on the methods of data capture, the relative advantages and disadvantages of the various methods, and issues relating to choice of appropriate method. The presentation made by the UNSD began by defining “data capture” as a process of converting collected data to a computer interpretable format. It described five main methods of data capture: i) keyboard data entry, ii) optical mark recognition/reading (OMR), iii) optical character recognition/intelligent character recognition (OCR/ICR), iv) personal digital assistant (PDA), v) Internet, and revealed the limitations and relative advantages of each method.

16. As for the choice of method of data capture, it was stressed that no single method is most appropriate uniformly over diverse national circumstances. Choice of method should be part of the overall strategic objective of the census in terms of timeliness, accuracy and cost. Choice of processing system and technology to use ought to be decided early in census cycle so that there is enough time to test different methods and adopt the most appropriate one. Successful use of imaging technology requires extensive testing in advance. Often a combination of more than one method is most suitable for a country.

17. The UNSD presentation was followed by a brief presentation by Lockheed Martin, from USA, on forms processing, i.e., collection and extraction of data in paper forms. It covered briefly the process flow and issues relating to management and planning of different phases of a census programme.

18. The session concluded with a lively discussion on use of PDA. The participants were in agreement that the PDA can be effectively used only when the questionnaire size is small. The participants observed that the introduction of selected validation rules at the data collection stage can save time and resources at the later phases of data processing but noted that the greater the number of checks, the more unwieldy the collection of data.

19. In this context, the representative of Lockheed Martin noted that introduction of validation checks during the data collection stage, in principle, restricts the permissible set of responses. As a result, the use of too many checks at the data collection stage may prevent the investigator from truthfully recording the data actually reported by the respondents.

20. In view of these comments, it was recommended that PDA may be used for data collection only when the size of the questionnaire is small; and validation checks at the data collection stage should be restricted to only those which detect logically unacceptable responses.

21. A discussion ensued on the possibility of current technology's capacity to recognize cursive or hand written text (IR or intelligent recognition, a technology still in the early stages of widespread use). The representative from Lockheed and Martin stated that, in their experience, machines perform better than humans in this regard.

Session 5: Outsourcing versus in-house Processing

22. In this session, the UNSD made a presentation on outsourcing of specific tasks at different stages of census operations. Most of the National Statistical Agencies responsible for conducting census are not capable of carrying out all the tasks involved in conduct of census. The reasons include i) lack of necessary technological expertise or equipment at NSO; ii) the need for improving timeliness and accuracy of the data, iii) a recognition of the complexity of job; and iv) the added advantage that the NSO gains access to external expertise and knowledge. Proper choice of tasks to be outsourced enables the NSO to concentrate on their core substantive work.

23. The presentation concluded with the following guidelines:

- Decision to outsource should be based on proper identification of technical needs, in terms of expected output.
- The requirements for the delivery of the output in terms of timeliness, quality assurance, accuracy, confidentiality, etc., should be precisely specified. Specifications of the contract should be laid down accordingly.
- Feasibility of outsourcing should be objectively assessed.
- Testing capability of the agency in delivering the expected output should be thoroughly tested.
- NSO should actively monitor the progress and quality of the job outsourced according to the laid down specifications.

Session 6: Data Capture: Optical Mark Recognition (OMR)

24. This session consisted of two presentations; one by the UNSD and the other by the representative of the DRS, from UK. The presentation of the UNSD mainly dealt with definitions and concepts of the method, while that of DRS provided the technical details of the two available OMR technologies, that is, OMR from image and OMR dedicated scanners. It was noted that OMR is a data capture technology that does not require a recognition engine. Thus, it is fast and relatively cheap. But, it requires a well-structured design and good quality printing of the forms to achieve high accuracy. However, OMR cannot recognize hand-printed or machine-printed characters. Thus, the use of OMR is often supplemented by other capture methods.

25. During the discussion that followed the presentations, a number of issues were raised by the participants. It was clarified that

- All data can not be captured through OMR;
- OMR is more suitable for questions that have limited number of responses;
- Usually some questions are suitable for tick-boxes while others are not. Thus a combination of both mark recognition and image capturing technologies are often required;

- Names can be captured as images and images can be stored for every form;
- This technology has been successfully used in Sudan for capturing data from multi-language forms;
- Error rate depends on the number of variables and thus varies from project to project.

Session 7: Data Capture: OCR / ICR / IR

26. This session consisted of a presentation by UNSD and one by DRS. The presentation by UNSD dealt with the definitions and the main difference between the OCR, ICR and IR technologies. The relative advantages and disadvantages were also highlighted in this presentation. The presentation by the DRS was focused mainly on form design, hardware/software requirements, workflow, accuracy, and relative advantages and disadvantages of the three methods.

27. To the questions raised during the discussion, the following clarifications were given by UNSD and DRS:

- Use of ICR technology helps in reducing production time.
- Editing of data is possible after scanning – changing the errors electronically.
- The performance of the ICR in recognizing Arabic characters is not yet known. Recognizing numeric characters would be easier than the Arabic alpha characters.
- Costs of applying ICR technology are scalable only to a certain extent.
- A combination of technologies, that is, OMR from image and OMR dedicated scanners, are mostly used, since that is found to be appropriate in most cases.
- There has been significant improvement since 2000 in the ICR, OCR and IR technologies. Usually the vendors tailor a package according to the clients' needs.

Session 8: Data Collection: PDA-Handheld-computers/Internet

28. This session consisted of two presentations – one by UNSD and the other by Lockheed Martin. While the UNSD presentation dealt mainly with different aspects of use of Personal Digital Assistant (PDA), the Lockheed Martin presentation was concerned with the procedure of multi-channel system of census data collection.

29. UNSD discussed the types of PDAs, as well as key specification features currently available on the market, relative advantages and disadvantages, and the criteria for making choice of the PDAs. It was noted that having extensive training prior to the deployment of PDAs is essential. It is critical that the vendor provide post implementation support – for both technical and hardware aspects.

30. The Lockheed Martin presentation stressed that application of multi-channel census data collection requires addressing issues such as completeness, scalability, data quality and consistency, security and authentication, channel data integration and management information.

31. The presentations were followed by a discussion on experiences of using the technology for data capture. The participant from the UAE mentioned that PDAs were used for the Labour Force Survey in UAE; and that the geographical divisions were collected, that technical support was provided; and that the use of internet for data transfer worked well. It was observed that PDA is the technology that is likely to be increasingly integrated in census data capture in the future, but that in the short run, paper forms for data collection might be best suited for many countries.

Sessions 9 and 10: Data Capture: Process Stages

32. Sessions 9 and 10 consisted of three presentations, made by Beta Systems/Intergraph, DRS and Lockheed Martin, as well as a demonstration from Morocco. All of them discussed and provided illustration of the different stages involved in a data capture process.

33. Beta systems, a data processing provider from Germany, presented its approach to the census/surveys data capture process stages, which consists of scanning, recognition (OMR, OCR, ICR), and verifying processes, with emphasis on the census data flow and quality assurance. The presentation gave an example of implementation of this approach for the Nigeria census conducted in 2006.

34. Intergraph, a GIS global provider from USA, working in association with Beta Systems, made a presentation on how GIS can assist in the process of census/surveys data capture. Census data are spatial data in its nature and GIS allows one to link a data value (e.g. census datasets) to a geographical feature (e.g. enumeration area) with improved capacities for collection, storage, management, analysis and reproduction of spatial data. Intergraph gave information on the development of Nigeria Census Geoportal.

35. DRS presented its approach to Census Data capture process, including the pre-census planning. The process stages include forms receiving, scanning, recognition, verification, quality assurance/management and logistics issues. All these process stages were presented in details with illustrative examples from the Sudanese census on the forms receiving in Arabic, the Ethiopian census on scanning process, the Sudanese and Malawi censuses on verification process, and the Tanzanian and Sudanese censuses on logistics. DRS stressed the fact that the largest issue for time and quality is how well the census forms are filled out by the enumerators.

36. Lockheed Martin provided a comprehensive review of the multi-channel approach: telephone, Internet, field hand-held devices, and outlined the characteristics of each. The presentation made a comparison of these contemporary methods with the paper-based approach in terms of the following factors: inventory control, questionnaire integrity, image quality, processing integrity, and capture accuracy.

37. It was explained that inventory control consists of ensuring that questionnaires are accurately accounted for and managed, and encompasses the following steps: delivery, system entry,

storage, processing, and disposition. Questionnaire integrity consists of ensuring that the questionnaire and its component parts are kept together for the complete process, meaning that the linkage between a questionnaire and its components is maintained, and that the mixing of data between questionnaires is prevented. On the other hand, image quality is about the qualitative value of the image representing the original document for automatic recognition and keying, while processing integrity consists of ensuring that all responses are processed in the proper sequence, priority and completely through all appropriate steps. Finally, capture accuracy is about the ability to assess and manage the accuracy of capture, regardless of the source.

38. Morocco made a presentation entitled: “Data Capture processes of large scale survey questionnaires: Case study of Census of Population and Housing 2004 of Morocco”. Data processing steps include (1) Reception of questionnaires, (2) Questionnaires preparation, (3) Scanning, (4) Image processing and OCR, (5) Normal video coding, (6) Inter-questionnaires control and correction, (7) Quality control, (8) Logical errors video coding, and (9) Data export. The presentation illustrated these stages for Morocco census which was conducted in 2004 and discussed in detail the different implementation issues. What is note-worthy for Morocco is the set-up of a National Centre for Automatic Reading of Documents, dedicated to data processing for census as well as for other government departments.

Session 11: Data Capture: Overview of Major Distributors/ Commercial Suppliers

39. During this session, three presentations were made by the data processing providers DRS, Lockheed Martin and Beta Systems/Intergraph. Each provider gave an overview on its specific solution to census data processing and some concrete examples illustrating its implementation. In addition, DRS gave a live demonstration of its scanner and its capabilities.

Session 12: Data Coding

40. UNSD gave a presentation on coding of data for censuses, which covered the basic concepts and definitions – including simple, structured and bounded coding as well as the concept of coding indexes. The appropriateness of closed-ended versus open-ended questions was explored. The relative merits of manual versus computer-assisted versus automatic coding was covered. Finally a few of the common international classification systems (ISIC, ISCO and ISCED) were discussed

41. Lockheed Martin gave a short presentation on their experiences with coding-related issues. It was mentioned that in the case of the UK, only 10% of occupational coding was done, due to the complexity of the exercise. For example, “engineer” is a very broad category that needs more information in order to do a reasonable job of coding the specific type of engineer. In general, a higher skill level is required for coding occupation, industry, educational level, and other more complex items. It was noted that higher levels of accuracy with coding could be achieved on items such as place of birth whereas often lower levels of accuracy could be achieved with items such as occupation.

42. In the discussion that followed, it was noted that if only a percentage or sample of occupation responses are coded, and estimates are achieved by using weights to inflate the coded responses,

that care should be taken to ensure that the sample that is coded should be representative of the entire population. One of the participants asked whether it was possible to automatically code from PDAs in the field and Lockheed Martin replied that complex items such as occupation should not be attempted since unwieldy computer algorithms would be needed and this would slow down the data collection process considerably. There is also an issue of adequate training for the enumerators as well as considerable bias being introduced by enumerators in this case.

Session 13: Introduction to Data Editing

43. UNSD gave a presentation on the introduction to data editing, commencing with a description of the types of errors typically encountered in the census process – including both content and coverage errors. An illustration was given showing why it is important to edit, especially in terms of overall trend-related distributions of data. Some basic principles of editing were given and the concepts of fatal versus query edits, micro- versus macro-editing, and manual versus automated editing were presented. Finally the pitfalls of over-editing were discussed.

44. One of the delegates remarked that it might be possible to do “real time” editing even with OMR or OCR/ICR combined with scanning, although it seemed clear from later discussions that the definition of “real time” needed elaboration before agreement could be reached on this issue. In other discussions, it was noted that the method of “eye editing”, although used in some countries for censuses, is very labour-intensive, prone to human error and is time consuming – and therefore, should be avoided.

Session 14: Concepts and Methods in Data Editing

45. UNSD gave a second presentation on data editing, this time going into more detail into the notions of within record editing and across record editing. The concepts of geography edits were covered as was the notion of the geographical hierarchy of records for census data, and the correspondence between housing and population records. The concepts of validity and consistency checks were discussed as were two basic approaches to editing: top-down and multiple variables. The group was led through two in-depth examples of sequential hot deck editing and then several issues and considerations in relation to the use of hot-deck were described.

Session 15: Data Editing (Practical Exercises)

46. UNSD gave a short introductory presentation on the basics of CSPRO, a software package freely available on-line, developed by US Census Bureau, and having 3 main functionalities: data entry, data editing and data tabulation. Each pair of participants was outfitted with a laptop and then the delegate from Morocco (an expert in CSPRO) led the participants through a hands-on exercise, using the data editing module. The exercise focused on trying to impute for missing sex data within a household.

Session 16: Country Presentations on Data Processing

47. The ESCWA representative gave a summary of the results of the Questionnaire on census data processing, which was sent to the participating countries prior to the Workshop. Nine countries replied: Egypt, Iraq, Jordan, Kuwait, Lebanon, Palestine, Qatar, Syria, and Morocco. The main findings are: (i) manual data entry is used by most of the countries; (ii) optical data capture is used by two countries; (iii) software, and good scanners are available in the countries that use optical data capture method; (iv) outsourcing of data capture and data editing is not applied in most of the countries; (v) optical data capture for the next census/survey has not been decided by most of the countries, but a few countries mentioned that they are planning to use OMR/OCR/ICR; (vi) three countries plan to use PDAs; (vii) concerns were pointed out by some countries regarding data capture with regards to accuracy, quality and timely dissemination of data; (viii) archiving methods/policies are used quite extensively, including storage of paper forms and electronic forms.

48. Egypt made a presentation on the use of OMR/ICR data capture for their 2006 Census. The presentation described the workflow through the process stages of data capture, coding and editing. A particularly emphasis was put on the first experience on the use of ICR engine for the Arabic language, with support from ESCWA. The presentation stressed some issues on archiving, by presenting the systems used for data warehouse and data mining. The presentation highlighted the fact that the work of data processing was carried out in the regions, in a decentralized approach, with the advantages and the challenges/lessons learned from this experience. It was requested that Egypt convey more detail on its experience to Iraq and other countries in the region.

49. Morocco presented a video (3min.) on its way to handle data processing throughout the whole process. It complemented the first presentation provided by Morocco on data capture process stages in session 9.

Session 17: Country Presentations on data Processing (continued)

50. UAE presented their experience on the use of PDAs for their 2005 Census data collection and the technical solution used. They stressed the fact that with this data collection method, they improved the timeliness of census releases (three months instead of a year for previous censuses). The data collection operations were carried out by the use of HP iPAQ devices, running Microsoft technology, and integrated with GPS. Data collected was transmitted to 27 data collection centers throughout the emirate states. The presentation stressed that the validation rules applied on the PDAs made it easier for enumerators to enter clean data. A critical success factor was the technical support team which was highly trained. The presentation also outlined some challenges and the problems they encountered and how these were overcome.

51. Qatar presented its experience on the use of GIS in census operations, including census data collection, analysis and dissemination. The presentation stressed the fact that Qatar has a nationwide GIS policy that helped its use at all the stages of census operations. The presentation provided details on the extensive use of GIS at (i) pre-enumeration (census framework planning, designing forms and coding, preparation of base maps and enumeration areas maps), (ii) enumeration (fieldwork organizing and monitoring, data gathering) and (iii) post-enumeration

(data capturing, processing, tabulation and database; data interpolation and analysis; data dissemination at different administrative hierarchical levels). GIS is particularly used for spatial analysis of census data and for census data dissemination through thematic maps, electronic media, and online (Internet/Intranet). Qatar illustrated the use of GIS and digital mapping by showing their comprehensive paper-based Atlas and the future development of web-based GIS applications and the electronic Atlas.

52. After the presentations made by Egypt, Morocco, UAE and Qatar, a round-table discussion followed, allowing a representative from each of the other participating countries (Iraq, Syria, Saudi Arabia, Bahrain, Jordan, Kuwait, Lebanon and Palestine) to make a brief overview on their census data processing experiences.

53. Iraq explained that in their 1997 Census, they used the keyboard data entry method with FoxPro. They are currently using CPro and Oracle and are planning to use the OMR/OCR/ICR methods for their 2009 Census data capture, in partnership with Egypt.

54. Syria stated that they are using a paper-based approach with key data entry and CPro. They are working on the boundaries of their EAs, and they are exploring for a suitable data processing approach for their 2014 Census.

55. Saudi Arabia highlighted their OMR experience for their 1990 Census and stressed the fact that it was not conclusive due to coding, high-cost and timeliness issues. They used keyboard data entry for their 2005 Census with GIS-based maps.

56. Bahrain stated that they used key data entry for their 2001 Census data capture and are planning to use PDAs for their 2010 Census data collection.

57. Jordan used key data entry for their 2004 Census data capture and Oracle for their census database. They are still exploring the appropriate data processing methods for their 2014 Census, but they are planning to use GIS.

58. Kuwait used OMR for their 1995 Census data capture, but experienced problems due to delivery delays in paper forms and machines. They tried the use of PDAs in their 2005 Census, by acquiring 30 devices for test, but the operation was cancelled because the PDAs were not also delivered on time. They opted for keyboard data entry for 2005 Census, but they are planning to use PDAs in their 2010 Census.

59. Lebanon has been conducting surveys and censuses only for housing and establishments. They used Oracle, SPSS and GIS in their 2003 Census for Housing and Establishments.

60. Palestine used key data entry and CPro for their 2007 Census data processing; they also used GIS for their mapping operations. All these operations were carried out in-house. They are considering the use of PDAs for their next census.

RECOMMENDATIONS & CONCLUSIONS

61. The method of data capture adopted by countries in the ESCWA should be made in relation to the particular circumstances for the countries in question. There is no one mode of collection

that is best for every situation. It may be appropriate to use less technically advanced capture methods such as manual entry. Countries where resources are available the use of more advanced technologies, such as PDAs is encouraged.

62. It is suggested that countries carefully assess their capacity and cost factors before opting for any particular technology for their next census. The use of new technology should not be influenced solely by current trends but rather by national needs.

63. NSOs should focus on their core competencies and may benefit from outsourcing some activities requiring expertise in non-core skills. The feasibility of outsourcing should be objectively assessed. The decision to outsource should be based on identification of technical needs, outputs, timelines, quality levels, and confidentiality protection, and should be precisely specified. The capability of the contractor to delivery should be verified. NSOs should actively monitor the progress and quality of the activity according to agreed upon specifications.

64. NSOs are encouraged to share experiences, approaches and best practices through technical cooperation and study tours for all aspects of census processes and operations including data capture and editing.

65. ESCWA member countries may consider a regional approach to share acquired equipment such as scanners, PDAs, satellite imagery, etc. In doing so, economies of scale may be achieved.

66. The implementation of quality assurance systems is essential to the delivery of census outputs for evidence-based policy making at national levels. Countries are encouraged to adopt appropriate measures to ensure that all processes of census operations, including data capture and editing, meet acceptable quality levels.

67. In order to improve accuracy and timeliness, and lower cost, NSOs are encouraged to consider the use of contemporary technologies for data capture methods such as OMR, OCR, ICR, PDAs and Internet.

68. Analysis based on unedited data can generate biased results. Therefore, NSOs should adopt statistically sound editing and imputation strategies as part of their data processing operations. Caution should be exercised to avoid over-editing as well as manual editing.

69. Both edited and un-edited data sets should be stored to allow for evaluation of the degree and effects of editing and imputation. As part of a sound data quality assurance system, NSOs should consider generating audit trails that document all changes and corrections.

EVALUATION OF THE WORKSHOP

70. Overall, the participants appreciated the workshop's main focus on new technology for data capture and editing. They found the most useful elements to be the presentations on country experiences, since these gave rise to much discussion and afforded the possibility for exchanges of ideas. The participants were not as appreciative of the session given by commercial providers

and felt that the hands-on laptop exercise was too complicated. To improve the workshop overall, it was suggested that more focus be given to country experiences and that these should be spread throughout the workshop to vary the format of the sessions and to encourage more discussion throughout the week.

ANNEXES

Annex 1. Agenda of the Workshop

Annex 2. List of participants

UNSD-ESCWA Regional Workshop on Census Data Processing in the ESCWA region: Contemporary technologies for data capture, methodology and practice of data editing

Doha, State of Qatar, 18-22 May 2008

Provisional Agenda

<i>Time</i>	<i>Topic</i>	<i>Responsibility</i>	<i>Document</i>
<u>Sunday May 18, 2008</u>			
	Opening		
8.30 – 9.15	Session 1 – Opening remarks -welcoming remarks by Host country, UNSD, UNESCWA, administrative matters	UNSD	
	Review of the 2010 World Programme on Population and Housing Censuses and a Discussion on Census Management & Planning Objective: To present a review of the 2010 World Programme on Population and Housing Censuses followed by a presentation on international recommendations on census management and planning.		
9:15 – 10:00	Session 2 – The 2010 World Programme on Population and Housing Censuses Overview on the 2010 World Programme on Population and Housing censuses	UNSD	Pres. 1 (UNSD)
15min.Break	– Presentation by UNSD – General Discussion		
10:15 – 11:15	Session 3 – Census planning and management International recommendations on census management and planning	UNSD	Pres. 2 (UNSD)
15min.Break	– Presentation by UNSD		

Time	Topic	Responsibility	Document
	– General Discussion		
	<p>Introduction to Data Capture Methods and Outsourcing versus in-house processing</p> <p>Objective: To present an overview of Data Capture management considerations and present and discuss the applications and issues of data capture using Optical Mark Recognition Technology; Optical Character Recognition/Intelligent Character Recognition; Internet/PDA, Manual Data Entry and provide an overview of different process stages</p>		
11:30 –12.45	<p>Session 4 – Introduction to Data Capture</p> <p>Methods of data capturing, advantages and disadvantages of each method, issues for consideration when choosing the method.</p> <ul style="list-style-type: none"> – Presentation by UNSD – Presentation by Expert (Introduction to Forms Processing, by Lockheed Martin) – General Discussion 	UNSD, Presentation by Experts	Pres. 3 (UNSD) Pres. A (Expert)
15min.Break			
13:00 –14.00	<p>Session 5 – Outsourcing versus in-house processing.</p> <p>Is outsourcing required? How to manage outsourcing. Country examples on outsourcing of data capture</p> <ul style="list-style-type: none"> – Presentation by UNSD – General Discussion 	UNSD	Pres. 4 (UNSD)

<i>Time</i>	<i>Topic</i>	<i>Responsibility</i>	<i>Document</i>
<u>Monday May 19, 2008</u>			
Data Capture: Optical Mark Recognition Technology			
8:00 – 9:15	Session 6 - Data Capture: Optical Mark Recognition Construction/Design Characteristics, Hardware and Software Requirements and Scanning/Storage, Advantages and Disadvantages; overview of the major commercial suppliers.	UNSD, Presentation by Experts	Pres. 5 (UNSD) Pres. B (Expert)
15min.Break	– Presentation by UNSD – Presentation by Expert (Overview on OMR, by DRS) – General Discussion		
9:30 – 11:00	Session 7 - Data Capture: Optical Character Recognition/Intelligent Character Recognition/ Intelligent Recognition Construction/Design Characteristics, Hardware and Software Requirements and Scanning/Storage, Advantages and Disadvantages; overview of the major commercial suppliers.	UNSD, Presentation by Experts, Country Presentations	Pres. 6 (UNSD) Pres. C (Expert)
15min Break	– Presentation by UNSD – Presentation by Expert (Overview on OCR, by DRS) – General Discussion		
11:15 – 12:30	Session 8 – Data Collection: PDA-Handheld-computers/Internet	UNSD, Country	Pres. 7 (UNSD)

Time	Topic	Responsibility	Document
15min Break	Different technologies/processes in data collection using handheld devices (e.g. PDAs) and Internet. <ul style="list-style-type: none"> - Presentation by UNSD - Presentation by Country (UAE) - Presentation by Expert (Multi-Channel Data Capture, by Lockheed Martin) - General Discussion 	Presentation	Pres. I (Country) Pres. D (Expert)
12:45 –14:00	Session 9 – Data Capture: Process Stages Scanning, Recognizing, and Verifying Processes associated with data capture and Quality assurance/ management system for data capture and logistic issues as well as how to balance timeliness versus quality <ul style="list-style-type: none"> - Presentation by UNSD (15 min.) - Presentation by Morocco - Presentation by Expert (The Whole Process of Census/Surveys Data Processing, by Intergraph/Beta Systems) 	UNSD, Presentation by Experts	Pres. 8 (UNSD) Pres. II (Country) Pres. E (Expert)
<u>Tuesday May 20, 2008</u>			
8:00 – 9:30	Session 10 Data Capture: Process Stages (cont.)	UNSD	Pres. F (Expert) Pres. G (Expert)
15min Break	<ul style="list-style-type: none"> - Presentation by Expert (Examples of Real Census Project Workflow, by DRS) - Presentation by Expert (Data Quality, by Lockheed Martin) - General Discussion 		

Time	Topic	Responsibility	Document
9:45 – 11:00	<p>Session 11 – Data Capture: Overview of Major Distributors/Commercial Suppliers</p> <ul style="list-style-type: none"> – Presentations by Commercial Providers: <ul style="list-style-type: none"> - DRS Census Solutions - LM Census Solutions, including Advances in Form Processing - Beta Systems/Intergraph Census Solutions 	UNSD	<p>Commercial Providers Presentations</p> <p>Pres. H (Expert)</p> <p>Pres. I (Expert)</p> <p>Pres. J (Expert)</p>
15min.Break			
	<p>Data Coding</p> <p>Coding is art of preparing data in a form suitable for entry into computer to facilitate analysis. In this connection, the objective of the session to present an overview of different methods of coding.</p>		
11:15 –12:30	<p>Session 12 – Data Coding</p> <p>Coding systems: Manual (clerical), Computer-assisted, Automatic coding (automatic coding is usually partial) or combination of more than one system. Coding systems in light of the data collection and capture methods planned for the census. Coding of occupations, industry and educational characteristics: At what level of classification? Adapting the international classifications for national use and importance of maintaining international comparability and nationally over time. Coding indexes</p> <ul style="list-style-type: none"> – Presentation by UNSD – General Discussion 	UNSD	Pres. 9 (UNSD)
15min.Break			
	<p>Data Editing</p> <p>Objective: Editing is the procedure for detecting and eliminate errors from data. The objective of the session is to present an overview of the concepts and methods and discuss the application and issues.</p>		

Time	Topic	Responsibility	Document
12:45 –14:00	<p>Session 13 – Introduction to Data Editing</p> <p>Types of Errors (Coverage + Non-response + Content- questionnaire, enumerator, respondent, coding, data entry, etc.). What is editing (concepts of check, control, correct)? Why Edit (give examples of edited and unedited output tables to illustrate potential biasing)? Pitfalls of over-editing. General description of methods of how to edit and how to impute, concepts of manual and automatic edits</p> <ul style="list-style-type: none"> – Presentation by UNSD – General Discussion 	UNSD	Pres. 10 (UNSD)
<u>Wednesday May 21, 2008</u>			
8:00 – 9:30	<p>Session 14 - Concepts and Methods in Data Editing</p> <p>Within Record Editing – Across Record Editing – Concepts of consistency and structural edits, examples of both population and housing edits, example of how edit specifications done. Editing Packages (CSPRO, CANEIS, country-specific packages with their various pros and cons) Process Flow for Capture/Edit – How the entire process hangs together, as illustrated via a flow chart</p> <ul style="list-style-type: none"> – Presentation by UNSD – General Discussion 	UNSD, Country Presentation	Pres. 11 (UNSD)
15min.Break			
9:45 – 11:00	<p>Session 15– Data Editing (Practical exercises)</p> <p>Participant Exercises on CSPRO on laptops</p>	UNSD	
15min.Break			

<i>Time</i>	<i>Topic</i>	<i>Responsibility</i>	<i>Document</i>
Country Experiences in Data Processing			
11:15 –12:30	Session 16 – Country Presentations on Data Processing Results of UNSD pre-workshop questionnaire on capture/ editing (country/ regional analysis) Country/ Regional Presentations on Experiences with Data Processing	CWA, Round-table by Countries	Questionnaire
15min Break	<ul style="list-style-type: none"> – Presentation by ESCWA (on Questionnaire results) – Presentation by Egypt and Demo. By Morocco – Presentations by Countries (Round-table by countries) – General Discussion 		Pres. III (Country)
12:45 –14:00	Session 17- Country Presentations on Data Processing (cont.) <ul style="list-style-type: none"> – Presentation by Qatar – Presentation by Countries (Round-table by countries) - General Discussion 	Presentations by Countries	Pres. IV (Country)
<u>Thursday May 22, 2008</u>			
Final Report, Recommendations & Conclusions			
8:00 – 9:30	Session 18 - Final Report, Recommendations & Conclusions - Final Report, Recommendations & Conclusions: review and adopt report, conclusions and recommendations (Final report lead by Rapporteur, evaluation of Workshop)	UNSD	Final Report
Break			

<i>Time</i>	<i>Topic</i>	<i>Responsibility</i>	<i>Document</i>
10:00 –14:00 (to be arranged by ESCWA)	Session 19 – Census preparation for the 2010 round of censuses in the ESCWA region	ESCWA Representative	ESCWA Doc.

Final List of participants

UNSD Regional Workshop on Census Data Processing in the ESCWA Region: *Contemporary technologies data capture, methodology and practice of data editing*

Doha, Qatar, 18-22 May 2008

No.	Country Name	Contact Person/Address
1.	Bahrain	Ms. Duaa Sultan Al-Harban Senior Researcher General Directorate of Statistics & Population Registry Central Informatics Organisation P.O. Box 5835 Bahrain
2.	Egypt	Ms. Eman Mohamed Al-Asmar CAPMAS) P. O. Box 2086 Nasr City, Cairo, Egypt
3.	Egypt	Ms. Nabila Taha El-Mahdy General Manager CAPMAS) P. O. Box 2086 Nasr City, Cairo, Egypt
4.	Iraq	Ms. Nuha Khudor Yousif

		<p>Director of Population and Labour Force Central Organization for Statistics And Information Technology (COSIT) Ministry of Planning P. O. Box 8001 Baghdad, Iraq</p>
5.	Iraq	<p>Ms. Amal Hasoon Zghar Al-Saadi Chief Statistician Assistant Population and Labour Force Central Organization for Statistics And Information Technology (COSIT) Ministry of Planning P. O. Box 8001 Baghdad, Iraq</p>
6.	Jordan	<p>Ms. Fathi M. Nsour DG Assistant for Censuses and Surveys Department of Statistics of Jordan P. O. Box 2015 Amman 11181, Jordan</p>
7.	Jordan	<p>Ms. Ikhlas Salim Aranki Director of Household Survey Department of Statistics of Jordan P. O. Box 2015 Amman 11181, Jordan</p>
8.	Jordan	<p>Ms. Ghaida Khasawneh Statistician Department of Statistics of Jordan</p>

		P. O. Box 2015 Amman 11181, Jordan
9.	Jordan	Ms. Safa'a M. Al-Zuobi Programmer Department of Statistics of Jordan P. O. Box 2015 Amman 11181, Jordan
10.	Kuwait	Mr. Essa M. B. Al-Sheikh Supervisor of Census & Sampling Design Central Statistical Office P.O. Box 26188 Zip Code: 13123 Safat, Kuwait
11.	Kuwait	Mr. Hammad Fnakher Hamed Alenezi Statistical Analyst Central Statistical Office P.O. Box 26188 Zip Code: 13123 Safat, Kuwait
12.	Lebanon	Mr. Vicken Ashkarian Expert in Geography Central Administration of Statistics Kantari, Army Street Trade and Finance Building, 5th Fl. Beirut, Lebanon
13.	Lebanon	Mr. Ziad Abdallah Head of IT Department

		<p>Central Administration of Statistics</p> <p>Kantari, Army Street</p> <p>Trade and Finance Building, 5th Fl.</p> <p>Beirut, Lebanon</p>
14.	Morocco	<p>Mr. Oussama Marseli</p> <p>Chief of Census and Population and Housing Services</p> <p>Direction de la Statistique</p> <p>Haut Commissariat au Plan</p> <p>B.P.178, Rue Mohamed Belhassan</p> <p>Elouazzani-Haut Agdal</p> <p>10001 Rabat, Morocco</p>
15.	Palestine	<p>Mr. Hussein Husni Mugamis</p> <p>Fieldwork Coordinator</p> <p>Palestinian Central Bureau of Statistics</p> <p>P.O. Box 1647</p> <p>Ramallah, West Bank (via Israel)</p>
16.	Palestine	<p>Mr. Jaafer T. Q. Qadous</p> <p>Programmer</p> <p>Palestinian Central Bureau of Statistics</p> <p>P.O. Box 1647</p> <p>Ramallah, West Bank (via Israel)</p>
17.	Qatar	Mr. Nasser Saleh AlMahdi
18.	Qatar	Dr. Ahmed Hussein

		P.O. Box 7283 Doha, Qatar
19.	Qatar	Mr. Mansour Almalki
20.	Qatar	Mr. Mohammed Said Al-Mohannadi
21.	Qatar	Mr. Mohammed S. AlBoainain
22.	Qatar	Ms. Noora Jamaan Al-Abdullah
23.	Qatar	Mr. Mark Grice
24.	Qatar	Dr. R.C.S. Taragi
25.	Qatar	Ms. Moza Soud Al-Musallam
26.	Qatar	Ms. Marjorie Cordett
27.	Qatar	Ms. Chethana Amai
28.	Qatar	Mr. Khalis Mahmud Khalis Qatar Statistical Authority
29.	Saudi Arabia	Mr. Othaim Mohamed Al-Othaim Director of Data processing for Census Project Operation Department Manager Central Department of Statistics Ministry of Planning P.O. Box 3735 Riyadh 11481, Saudi Arabia
30.	Saudi Arabia	Mr. Fahad A. Al-Fhied Director of Population and Vital Statistics Central Department of Statistics Ministry of Planning P.O. Box 3735 Riyadh 11481, Saudi Arabia

31.	Syria	Mr. Chafik Arbash General Director Central Bureau of Statistics Nizar Kabbani Street Damascus Syrian Arab Republic
32.	Syria	Mr. Ayman Al-Mobayed Assistant to Director of IT and GIS Centre Central Bureau of Statistics Nizar Kabbani Street Damascus, Syrian Arab Republic
33.	United Arab Emirates	Mr. Adeb Hassan Mohamed Al-Hammadi Programmer Central Statistical Department Ministry of Economy and Planning P.O. Box 904 Abu Dhabi, United Arab Emirates
34.	United Arab Emirates	Mr. Ali Abdulla Al-Khuraim Al-Zaabi Director of Population and Social Statistics Section Department of Planning and Economy P.O. Box 12 Abu Dhabi, United Arab Emirates
35.	United Arab Emirates	Mr. Saif Eldin Mohamed Abbas Senior Statistician Department of Planning and Economy P.O. Box 12 Abu Dhabi, United Arab Emirates

Private Companies

36.	Beta Systems Software AG	Mr. Richard Josef Lang Director Consulting International Beta Systems Software AG Huebnerstr.03 D 86150 Augsburg
37.	DRS Data Services Limited	Mr. Andy Tye International Manager DRS Data Services Ltd. 1 Danbury Court, Linford Wood Milton Keynes, Bucks MK14 6LR, England, UK
38.	DRS Data Services Limited	Mr. Patrick Hickmott Sales Manager Middle East DRS Data Services Ltd. 1 Danbury Court, Linford Wood Milton Keynes, Bucks MK14 6LR, England, UK
39.	Intergraph	Dr. Jens Hartmann
40.	Lockheed Martin	Mr. Fred Highland

United Nations Statistics Division (UNSD)

41.	UNSD	<p>Mr. Amor Laaribi Statistician Demographic Statistics Section Statistics Division DC2-1568 United Nations New York, NY 10017</p>
42.	UNSD	<p>Ms. Diane Stukel Inter-Regional Adviser Demographic Statistics Section Statistics Division DC2-1546 United Nations New York, NY 10017</p>

United Nations Economic and Social Commission for Western Asia (UNESCWA)		
43.	ESCWA	Mr. Alope Kar Regional Adviser on National Accounts and Economic Statistics Statistics Division P.O. Box 11-8575 Beirut, Lebanon
44.	ESCWA	Ms. Fathia AbdelFadil Team Leader, Statistics Division P.O. Box 11-8575 Beirut, Lebanon
45.	ESCWA	Ms. Amal Nicola Administrative Assistant Statistics Division P.O. Box 11-8575 Beirut, Lebanon