UNITED NATIONS SECRETARIAT
ESA/STAT/AC.163

Department of Economic and Social Affairs **June 2008**
Statistics Division **English only**

Report of the UNSD Regional Workshop on Census Data Processing for the English Speaking African countries: Contemporary Technologies for Data Capture, Methodology and Practice of Data Editing

Dar es Salaam, Tanzania, 9-13 June 2008

Table of Contents

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: (1) Presented revised international standards for conducting population and housing censuses, focusing on recommended core topics as identified in the United Nations Principles and Recommendations Revision 2; (2) Discussed ways of improving the management and planning of the census, including outsourcing issues; (3) Presented and discussed contemporary technologies in census data capture, including the use of Optical Mark Recognition (OMR), Optical Character Recognition/Intelligent Character Recognition (OCR/ICR), (4) Discussed the process stages for data capture, (5) Presented an overview of major commercial suppliers for data capture; and (6) Presented 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 53 participants from 20 countries, by three international /regional organizations (UNFPA, UNECA and UNSD), and by two commercial providers (DRS, Betasystems). For complete list see of participants see Annex.

Opening

3. Mr. Kwesigabo, Director of Population and Demographic Statistics of the National Bureau of Statistics of Tanzania welcomed the participants to the workshop. He reminded participants that a population and housing census was a major source of demographic and social information. He, however, lamented past experience in which many countries in the region produced and announced their census results late. Mr. Kwesigabo, therefore, underscored the importance of producing accurate and timely census results. He further reminded the participants to take note that many censuses have failed due to problems in data processing, therefore the need to adequately prepare for data capture and processing.

4. Mr. Christopher Mwaijonga, on behalf of UNFPA, expressed gratitude to the Director General of the National Bureau of Statistics of Tanzania and Director of UNSD Director for inviting UNFPA to participate in the workshop. He stated that UNFPA supported the move by some countries to use population data in assessing poverty and in measuring fertility and maternal mortality. Mr. Mwaijonga also indicated that UNFPA was supportive of South to South technical cooperation and assistance. While encouraging the adoption of new technologies, he cautioned that care must be taken to avoid problems and failure in their proper use. He also alluded to the high cost of scanners and requisite software.

5. He conveyed to all participants greetings from the UNFPA Executive Director, Ms. Thoraya Obaid. Participants were informed that the UNFPA Executive Director attaches high importance to census work and that UNFPA intends to continue its support within the context of its mission which is "to support countries to use population data for policies and programmes to reduce poverty, and to ensure that every pregnancy is wanted, every birth is safe, every young person is free of HIV/AIDS, and

every girls and woman is treated with dignity and respect". He reminded participants about the fact that more than eleven African countries will be involved in the 2010 round of censuses and the need to prioritise the issue of census planning, and to do so early enough to avoid panicking at the end. He emphasised the need to address major constraints in conducting censuses in Africa including lack of comprehensive census plans; limited human and financial resources, as well as lack of sharing census costs; inadequate cost effectiveness of the plans (in terms of collecting data that are relevant, required by the users and can generate results at lower geographical units);inadequate or lack of gender mainstreaming; inadequate storage and maintenance of census materials and equipment; extended data processing period, and lack of statistics broken down by sex, age, geographic location, education, economic status and other variables.

6. Mr. Kidus Mengistu from UNECA underscored the importance of a census as a major source of demographic, social and economic data especially for small administrative domains and their use for objectively developing informed policies. He also pointed out the critical role a census plays in establishing sampling frames. While Africa had a long history of conducting censuses, he noted that the 2000 round witnessed diminished participation. Some of the weaknesses of census programmes in Africa included irregularity in census taking, incompleteness of coverage; limited dissemination of census results; poor archiving and none reporting of metadata. In order to mitigate some of the short comings the UNCEA African programme is geared to reverse this trend. For example, UNECA contributed to the revision of the United Nations *Principles and Recommendations for Population and Housing Censuses* (P&R) supplemented by an addendum specific for use by African countries. In addition, UN ECA will organize workshops in 2008 and 2009 on data processing, planning and other pertinent areas.

7. On behalf of Dr. Paul Cheung, Director of the United Nations Statistics Division (UNSD), Mr. Jeremiah Banda thanked the Tanzania government and National Bureau of Statistics for hosting the workshop and welcomed the participants to the workshop. He then underscored the importance of censuses in Africa as major sources demographic, social and other related data.

8. He explained that this workshop was part of the 2010 World Programme for Population and Housing Censuses, adopted by the United Nations Statistical Commission in March 2005 for the period 2005 to 2014. The three essential goals of the programme were: (i) adoption of agreed set of acceptable international principles and recommendations governing the conduct of a census; (ii) to facilitate and encourage countries to conduct censuses during the period 2005-2014; and (iii) to assist countries in their efforts to disseminate census results in a timely manner.

9. He informed the participants about the role of UNSD played in: updating the revised Principles and Recommendation with respect to developing methods and standards; technical assistance; and conducting census related training workshops. Mr. Banda reiterated the importance and need to generate quality census data for use in objective decision-making. He, in addition, emphasized that it was imperative for all countries to conduct censuses during the 2010 round unlike during the 2000 round of censuses when about 15 African countries did not carry out a census. While other data sources such as household surveys and administrative records are important sources of demographic and social data, these sources are not well developed in most Africa countries, thus a census remains a major source of data. In addition, it is the source of data for small administrative areas.

10. Finally he informed the participants that the main objective of the workshop was to review and discuss available data capture technologies and the highlight good practices in coding and editing. The workshop also would provide a forum for participants to exchange experiences.

11. He then took advantage of the high presence of the press to underscore the indispensable role of the media to promote census activities and help in the dissemination of the census results.

12. Ms. Albina Chuwa, Director General of the National Bureau of Statistics Tanzania, read the opening speech of the Minister for Finance and Economic Affairs hon. Mustafa, Mustafa Mkulo. In his statement the Minister welcomed participants to the workshop and to Tanzania. He welcomed the opportunity for Tanzania to host an important workshop on data capture, coding, editing and related good practices in operationalising census activities. He stressed the importance of improving the planning and management of censuses, in general and data capture in particular. The Minister reminded the participants that the purpose of carrying out a population census was to ensure the availability of reliable and updated benchmark data useful for formulation, implementation, monitoring and evaluating of demographic, social and to some extent economic programmes. He reminded the participants that the workshop gave them an opportunity to learn from each others knowledge and experiences.

13. In highlighting the important role censuses will play in this round, he alluded to the fact that the 2010 census round will provide the much needed data to assist countries to evaluate progress made in the MDGs. The Minister observed that a census was the most expensive statistical programme, therefore there was need to justify its relevance and improve its efficiency. In this regard mapping, questionnaire design, data capture, processing, analysis and dissemination should be carefully planned and executed. He stressed that the public will accept the census results only if they are reliable and credible.

14. He then went on to share with the participants the experience of Tanzania in the area of data capture. Before the scanning technology was used it took more than three years to release census results. In 2002, however, there was great improvement. By applying scanning technology about nine million forms were scanned in 60 days and the general census report was produced within three months. It is against this background that Tanzania will most likely use the scanning technology during the forthcoming census. Our aim is to enhance speed and accuracy in the production of our census results, he emphasized. He, however, cautioned that new technologies bring new challenges, in this connection he advised country participants to learn from each other. In the case of Tanzania, they learned from the experiences of Ghana, Kenya and Zambia in order to effectively and efficiently use the scanning technology.

15. The Minister briefly touched on the issue of outsourcing data capture activities. He advised that hardware and software should be carefully selected and piloted. The minister indicated that Tanzania was eager to learn from countries that have successfully outsourced census data capture.

PRESENTATIONS AND DISCUSSIONS FROM THE VARIOUS SESSIONS

Session 2: 2010 World Population and Housing Census Programme

16. 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.

17. UNECA presented the addendum to the Principles and Recommendations for Population and Housing Censuses. The two census workshops organized by UNSD in 2006 in Mali and Mozambique, recommended that there was a need to prepare an addendum to the P&R. The addendum should complement the global Principles and Recommendations, to give guidelines on issues specific to Africa. The addendum was presented to the 3rd Africa Symposium on Statistical Development (ASSD) in Ghana and the Symposium recommended that African countries follow the addendum for the next census. The addendum complements the list of global core topics, such as specific demographic information particularly relevant for African countries including age of the mother at date of 1st child born alive and orphanhood. The addendum recommends to use for the definition of the place of usual residence the 6 months threshold, and emphasizes some critical issues as census planning and management, and analysis and dissemination. It is also recommended to consider linking the population and housing census with the agricultural census as both are large scale exercises, and may benefit one from the other.

18. On census planning and management the UNECA representative urged countries to prepare a well defined project document for use in soliciting for census funding. He expressed the view that vital statistics could not replace census taking in Africa, partly because they seek limited information and for most countries they are not comprehensive in coverage. He stressed the importance of conducting post enumeration surveys (PESs) to evaluate census coverage and sometimes also content error.

19. With regard to the analysis and dissemination of census data:

    a. He advised countries to develop elaborate data analysis plans including policy oriented analyses, and to cater for other stakeholders, including researchers.

    b. Dissemination should be done according to user's needs, while developing an appropriate mix of dissemination.

Session 3: Preparation of the 2010 round of censuses in the region

20. A round table meeting, which was moderated by UNSD, was organized to allow participating country representatives to shed light on their preparatory activities for the 2010 round of censuses.

21. **Botswana**: It is envisaged that the next census, in Botswana, will be carried out in 2011. The 2001 Census data was captured and edited using IMPS. But the tabulation and analysis of results were done using SPSS. The 2001 Census Data Processing was outsourced. With respect to data processing for the 2011 Population and Housing Census, the Statistics department has not yet reached a decision on whether to outsource the project or develop the system in-house. The decision will mainly depend on the availability of resources, namely, skilled personnel and funds.

22. **Gambia**: The Gambia is planning to carry out a census in 2013. They are currently at the stage of budget preparation. As a background, the following were given as problems and challenges experienced in the previous census conducted in 2003: missing records

of some EAs; missing villages; mixed information from different villages; data from one EA keyed in a wrong EA; compounds skipped; person records skipped; incomplete entry of household records and modules, etc. It is hoped such problems will not recur during the 2013 census.

23. **Ghana**: The next census is scheduled for March 2010. Work on census mapping is currently in progress, however, there are 117 districts to be mapped compared to 100 during the 2000 census. There are plans on the drawing board to conduct a trial census in March 2009. The following are some of preparatory activities in place: A Census Coordinator has been appointed to take full responsibility for the census operations; the Ghana Statistical Service has constituted a nucleus of the National Census Secretariat with the responsibility for the day-to-day census activities and a National Census Technical Advisory Committee has also been formed. National Census Planning Committee and a National Census Publicity and Education Committee were to be formed soon. All the committees would have institutional and sectoral representation and are mandated to provide direction for the effective conduct of the census. About US$ 40 million is earmarked for the 2010 Ghana census.

24. **Kenya**: The last census was carried out in 1999 and the next one is scheduled for 2009. As part of the preparatory activities about 60 per cent of the country has been mapped, a questionnaire has been designed and a pilot test is planned for this year. To avoid the problems encountered (see para. 49) in data capture during the 1999 census, in planning for the 2009 the Statistics Bureau is adequately preparing to maximize the use of data capture technology. Steps being taken include: studying available scanning technologies; gathering information on the best practices by those who had successfully used the technology during the 2000 census round and from experts; training personnel on the use of the scanning technology; planning to pilot the use of the scanning technology and acquiring fully licensed scanning software. Notwithstanding the above, some of the preparatory activities were still lagging behind by the time of workshop.

25. **Mauritius**: The next Housing and Population Census (HPC) is planned for 2010. Censuses in the Republic of Mauritius are based on UN *Principles and Recommendations for Population and Housing Censuses*. Given that the planning and implementation of 2000 Population and Housing Census in Mauritius went out smoothly; it will be used as guide for the 2010 Census. Preparatory tasks such as legal instruments, budgeting, mapping and questionnaire development and preparation has started. The Census unit, whose main task is the planning and execution of Census 2010, has been constituted. The Cartographic work has already started with the updating and digitizing of Enumeration Areas using GIS. The estimated total cost of the 2010 Housing and Population Census is about US $5.5 million. A decision has not yet been made whether to use the scanning technology or not. Most likely, Optical Mark Recognition (OMR) or Optical Character Recognition (OCR) or the Intelligent Character Reader (ICR) will be used. Total funding of the census is expected to come from the national budget.

26. Like the previous censuses, the 2010 Census will be conducted in two phases ; the first phase will be the Housing Census to be taken from February to April and second phase will be the Population Census, from the end June to the first week of July 2010. The main reason for adopting this strategy is the possibility of obtaining during the housing census a list of names and addresses of heads of households which would serve as a frame for the population census.

27. **Namibia**: The next census will take place in 2011. Planning is in the process. In addition, mapping is underway. This involves re-demarcating some EAs and updating the master frame. In the 2000, round scanning technology was used for data capture. SPSS was used for tabulation and DevInfo part of the dissemination. Scanning technology will again be used to capture census data in 2011.

28. **Rwanda**: Census mapping started in 2006. The next census is planned for 2012. Scanning technology will be used for data capturing.

29. **Seychelles**: The country plans to conduct a census in April 2010. Seychelles conducts censuses after every five years. The National Statistical Bureau, meanwhile, has drawn a comprehensive plan. With respect to census mapping the Statistics Bureau acquired orthophoto base maps from the Ministry of Land Use and Habitat. The aerial photo maps were taken in 1999. The base maps are used to update the physical structures (mainly housing) which are drawn manually during field exercises. These are then systematically coded and digitized using Arc-View software. The mapping update has now become a routine exercise.

30. **South Africa**: Statistics South Africa (Stats SA) carried out its first Census in 1996 after the democratic elections of 1994. The second census was conducted after five years, in 2001. The next census will be carried out in 2011 and after which there will be a 10 year interval between censuses. The pilot census is scheduled for October 2009, two years before the main census. Meanwhile Stats SA has established a census permanent structure. In addition, a strategic plan is in place to guide in deciding on the type of data capture technology to be used. A budget of 3 billion Rands has been approved by the government. The demarcation of the country into 100,000 enumeration areas is also underway.

31. **Tanzania**: The next Tanzania Population and Housing Census will be held in August 2012. The reference night for the 2012 census will be the last Sunday of August 2012. As it was in the last censuses a de facto approach with some characteristics of a de jure will be adopted as a method of enumeration. Planning for the next census is complete and preparation is going on although at a slow pace. Preparations started in 2004 with the revision and delineation of enumeration areas (EA) in the Districts. Besides delineation of the country into EAs, revision of the 2002 census geographic frame also involves taking Geographical Positioning System (GPS) coordinates for social and economic features such as schools, hospitals, water points, etc. and digitizing all readings using ArcGIS and ArcInfo software. Based on population projections and other factors, a total of 120,000 EAs are estimated to be eventually delineated throughout the country.

32. As part of preparations, a comprehensive 2012 Census Strategy and Implementation Strategy Handbook has been produced. The document provides a comprehensive strategy for planning, organizing and mobilizing adequate human, material and financial resources required for the census. It was estimated that a total of T.Shs. 87.24 billion (equivalent to US$ 69.3 million) will be spent from financial year 2007/2008 to 2013/14. This amount is equivalent to per capita expenditure of 1.5 US dollars (the Population of Tanzania is projected at 45.6 million by the year 2012). The major constraint the country Statistical Bureau faces is the lack of sufficient funds to support the preparatory activities.

33. **Uganda**: The next census will be conducted in 2012 in Uganda. The census mapping exercise will be undertaken to demarcate the country into enumeration areas (EAs). The maps used in the 2002 census will be updated with respect to administrative

boundaries, villages, roads, tracks etc. They are also planning to geo-locate some socio-economic infrastructures like schools, hospitals churches, mosques, and industries using the global positioning systems. The exercise would also create a sampling frame for future surveys. There are plans to use scanning technology with respect to data capture. The budget for the 2012 census is estimated at about US $33 million.

34. **Zambia**: Zambia is planning to conduct its next census in August 2010. Mapping is currently underway although going at a slow pace because of sporadic funding.

35. **Zimbabwe**: Zimbabwe has so far been able to conduct a census decennially since August 1982. The last census was carried out in August 2002. It is hope that the next one will be conducted in 2012. Census mapping is a continuous undertaking, although it is going on a slow pace owing to resource constraints.

36. Highlights of the discussions and issues raised as a result of the country statements were summarized in the conclusions and recommendations.

Sessions 4, 11, 17 and 18: Country presentations and experiences on data processing

37. UNSD presented a summary of the results of the Questionnaire on census data processing, which was sent to the participating countries prior to the Workshop. Nineteen countries replied. With respect to manual data entry, the data capture process is often not outsourced. Several countries develop their own programs written specifically for their capture process (e.g. Seychelles). With respect to OMR & OCR/ICR the data capture process is often done in-house with some countries outsourcing. Regarding the next census, many countries are undecided as of which method to choose but are looking seriously into scanning methods such as OCR/ICR. Manual data entry is proposed for use in countries with previous experience (e.g. Liberia, Seychelles). Many countries store forms both electronically and in hardcopy format. While some countries use electronic means for the storage of forms, many countries have no policy for the storage of forms. Coding is usually done manually with a few countries using little automation methods. When automated, the software is developed in house (e.g. Gambia). Countries have expressed interest in automated methods for speeding up data capture and processing. Almost all countries have an editing system as a part of the census/survey processing. Most countries conduct structural checking for completeness; Range checking and validation; Consistency checking.

38. The Major concerns statistical/census offices have with data capture whether manual and optical capture are: efficiency and consistency of capture speeds, quality control and validation, hardware/equipment delays and maintenance, poor handling of questionnaires and ineligible writing on questionnaires.

39. The discussion raised numerous questions about the choice of having one or two contractors, the verification of the geography, questionnaire preparation before scanning, potential errors introduced by scanning, ways to avoid duplicates and measurement of accuracy.

40. **Ethiopia**: Ethiopia conducted its latest census in 2007. During the previous censuses in 1984 and 1994 they used the keyboard manual data entry method, this was not the case for the 2007. They opted to scan the questionnaires because manual data entry was: time consuming; subject to high non-sampling errors; involved a great deal of human resource management. The need for the scanning technology was also motivated by the quest for timely census results.

41. Before adopting the technology, they learnt from the experience of two African countries namely, Tanzania and Ghana. The data capture in Tanzania was completed in record time and the general tables were produced within three months from the start of the scanning process. In the case of Ghana, they learnt from the difficulties experienced during the 2000 census where: the data capture took more than six months; and there was regular power failure necessitating the installation of a large generator to minimize the power cuts.

42. Ethiopia saw the benefits of the scanning technology from the following perspectives: significant decrease in the time required in capturing data. This helps to get timely data, to policy makers, planners, researchers and other stakeholders. In addition, there was no need to store millions of forms for a long time.

43. However, for scanning to be successful, some prerequisites were necessary to be in place, namely, the need for proper training; a reliable network system; adequate space for forms and data flow; proper file management. Further, pre-census activities were put in place to ensure successful application of the scanning technology. For example, data from the pilot census was successfully scanned (OMR), exported to text format , tabulated and tested; Experts from DRS company assisted in the capturing, validating and exporting the pilot census data; training in scanning technology was done; hardware and software training was conducted; equipment was purchased and installed; census data processing work plan was prepared; the census data processing teams were established; the scanning room was organized with air conditioning in place; a high capacity automatic generator was installed and the batch header database was organized.

44. Despite the above preparatory activities, some problems still crept in. With respect to scanning, some batches slipped through unscanned during data capture; some batches were scanned in parts only; some scanned forms were misplaced in wrong boxes; experienced some limited storage space on scanning machines; the location of scanned images on the storage server at time could not be found; and problems in retrieving scanned images for key correction were encountered; in this case keying took longer time as it was done manually. Notwithstanding some of the above problems, preliminary census results were produced in good time.

45. **Ghana**: There were data capture constraints during the 2000 census which hopefully will not be the case next time around. For example, census forms were designed to cater for a maximum of ten household members, therefore continuation forms were used for large households and enumerators were tasked to mark the form as a continuation and insert a 15-digit reference number which identifies a household. In most cases, Enumerators did not mark the form as a continuation. In addition, sometimes the scanners could not pick the reference numbers correctly. Such problems prolonged the validation process.

46. In addition, power fluctuations, power cuts and low voltage disturbed the flow of work to the extent that it sometimes became impossible to scan during the day. It also led to the destruction of two motherboards of the scanners and caused damage to a couple of computers and printers. This problem was however resolved when a 100kVA generator and a stabilizer were installed.

47. **Lesotho**: Lesotho conducted Population and Housing Census in April 2006. Priority was given to the extraction of population aged 17 and above by districts, constituencies, local community councils and villages. Such information was requested by the Independent Electoral Commission to facilitate the planning and preparations for the

2007 General elections. Coding, therefore, commenced late in September 2006 and was completed in December 2007.

48. Scanning commenced in June 2007 using two Fujitsu scanners and was planned to be completed by June 2008. Data cleaning commenced in July 2008 this will be followed by the tabulation and analysis of census results.

49. **Liberia**: Liberia conducted a population and housing census from 21-30 March 2008. This was after several years of civil war. The last census was carried out in 1994. Data entry had not started by the time of the workshop; however, the editing program is ready. Preliminary results were released on 6 June 2008. The World Bank, through Paris21, has provided some equipment to facilitate the processing of census results.

50. **Kenya**: During the 1999 census data capture presented some considerable challenges. Although the scanning of questionnaires was initially planned to be completed in two months, the process was however, considerably slowed down owing to the following problems: (i) there was poor preparation, there was a mix up of questionnaires in different EAs before the documents were passed on to scanning personnel; (ii) the poor reinforcement on the pencil marks resulted in rescans; (iii) many of the physical batches returned for rescan did not match the ASCII batches of the same system; (iv) poor maintenance of printers introduced errors when they failed to print labels. Some of the substitute handwritten labels were incorrect; (v) Central Bureau of Statistics Staff had limited knowledge of the scanning software (AFPSPRO) and maintenance of the equipment; (vi) the failure to scan household information in the order of serialization introduced errors during the cleaning of data. Many of the above listed problems were observed during and after the scanning process. The problems connived to significantly contribute to the slowing down of the scanning process.

51. Here are some of the lessons learnt from the Kenyan experience: (i) the need for all requisite equipment to be put in place before the actual scanning is started; (ii) questionnaires should be distinct, legible and unbleachable; (iii) paper for printing questionnaires should be of acceptable quality and standard; (iv) proper record of distribution of material to the field with serial numbers should be maintained for future reference; (v) soon after receiving questionnaires from the field microfilming of the same should be done to retain the original data for any other reference; (vi) proper training of supervisors on scanning and network software is imperative;(vii) fast away spares should be brought in together (if coming from outside) with the rest of the equipment; handwritten labeling of batches should avoided; and (viii) work should be scanned sequentially to eliminate errors through guesswork during data cleaning.

52. **Mauritius**: During the 2000 census, activities related to editing, coding and data capture did not proceed smoothly. There was a shortage of staff at various stages of the census operation, causing delays on the time schedule. Because data capture was outsourced to another government department, in most cases, census work was not given priority over other work. It is, therefore, most likely that if the Statistical office will resort to data capture through scanning devices, they will be contingencies to minimize the above mentioned problems.

53. **Nigeria**: The country conducted its population and housing census in 2006. Scanning technology was used for data capture. Twenty one scanners were used. They were distributed to seven data processing centres located strategically across the country. Immediately after the census, completed OMR/ICR forms (questionnaires) started arriving at the data processing centres. About 35 million forms were scanned and

edited. The scanning and first level editing was completed within nine months of the enumeration period.

54. The following were some of challenges faced during the data capture phases: ensuring that documents for particular geographic locations were archived in sections of the archive and shelves designated for them; making sure that all forms were separated before taking them for scanning; breakdown of loader; rate of getting documents ready for scanning did not usually keep pace with scanning, ensuring that correct batch headers were properly placed on EA batches; and that after scanning making sure that EAs were correctly returned to their marked envelopes; poor field work which resulted in 'missing values' of 'mandatory fields', or outright wrong values for fields; difficulty in linking forms for households with more than 8 persons; integration of the two solution providers: form design and equipment and software solutions were provided by two different companies; dealing with sensitivity of Nigerians to census figure; lack of reliable and uninterrupted power supply.

55. **Sudan**: A census was carried out between 22 April and 6 May 2008. It was part of the peace agreement, between the North and South. The census was needed to provide basic information for the anticipated election and as a basis for distribution of resources. The Optical Mark Reader (OMR) technology was used for data capture. There are working and planning to release the results as quickly as possible, by the end of September 2008. The occupation and industry codes were keyed. Data processing was scheduled to be completed by the end of June 2008. The census was supported by donors including United Nations agencies.

56. **Zambia**: An agreement was reached between Central Statistical Office (CSO) and Examination Council of Zambia that the forms would be scanned using their scanners and personnel. Because of their experience in scanning examination papers for the country it was easy for them to scan the census forms. Thus capacity was already built at ECZ, so instead of CSO inventing the wheel it was prudent for the office to utilize the available capacity from another government agency, this decision saved CSO's time and resources.

57. In this regard the training for Scanner Operators was conducted at ECZ premises from 11th to 12th January 2001. Categories of staff included OMR Operators and Manual Editors. A total of 54 OMR Operators and 10 Manual Editors were trained altogether. The scanning of questionnaires started on 15th January, 2001 and was expected to be completed in April, 2001.

58. Although the 2000 Census was successfully conducted, there were a number of problems/constraints encountered during the field operations and data processing thus:

    i. The postponement of the Census resulted in work continuing into the rain season. This created a great challenge during scanning exercise because of some wet questionnaires;

    ii. Miscoding of the barcode led to a good number of orphan questionnaires;

    iii. The storage facilities were not that conducive for keeping questionnaires, the place was humid

    iv. Dysfunctional rate of Scanning Machines at ECZ was high, it slowed down the work progress;

v. The machines used were not capturing the image of the forms, posing difficulties during editing.

59. In view of the problems/constraints experienced in the field and during data processing the following recommendations are therefore made:

    i. Census exercise should be conducted during dry season to avoid questionnaires getting socked

    ii. That CSO should begin discussions with ECZ on the scanning of 2010 census questionnaires.

60. The scanning of 2000 Census of Population and Housing was conducted successfully despite some of the problems listed above. This success was attributed to t teamwork exhibited by EZC and CSO staff and management.

Session 5: Outsourcing versus in-house Processing

61. 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. A decision on whether or not to outsource should be based on:

    a. Defining the technical needs of the NSO in terms of expected output

    b. Specifying the requirements for the delivery of the output in terms of timeliness, quality assurance, accuracy, confidentiality, etc

    c. An assessment of the market *vis-à-vis* the NSO needs to determine if it would be feasible to undertake the outsourcing

62. Contractor and NSO should have a shared understanding of the requirements of the contract, including objectives, expected outcomes and priorities. Clear Specifications, including standards to be met, are key to ensuring to get what is wanted and that everyone understands what is expected. Specifications should describe in detail the tasks that are the responsibilities of the NSO and for the contractor. Specifications should include detailed milestones with deliverables against which performance should be evaluated. Specifications for the output should also address requirements for timeliness, data confidentiality and security, quality assurance.

63. The discussion focused on the following points:

    a. Transfer of knowledge. It was recognized that outsourcing offers an opportunity for the NSO to acquire knowledge in new technologies, but nevertheless, the contract should explicitly include the corresponding deliverables (e.g. training).

    b. Management. Participants stressed that the management should explain to the regular staff of the NSO the rationale behind the decision of outsourcing to avoid negative reactions.

    c. Outsourcing to another government department, for example the examination council, may be an efficient option as reported by the experience of countries like Zambia (see paragraphs 55-59).

Session 6: Introduction to Data Capture

64. 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 an 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.

65. Choice of method should be part of the overall strategic objective of the census in terms of timeliness, accuracy and cost. The technology used should be decided early in census cycle in order to allow enough time to test and implement the system. When imaging technology is used for data capture, extensive testing is required well in advance of the census.

Session 7: Data Capture: Optical Mark Recognition (OMR)

66. This session consisted of two presentations; one by the UNSD and the other by the representative of the DRS, from UK. The presentation of UNSD mainly dealt with definitions and concepts of the method. OMR is a technology that allows an input device (e.g. imaging scanner) to read hand-drawn marks such as small circles or rectangles on specially designed paper. An OMR works with a specialized document and contains timing tracks along one edge of the form to indicate scanner where to read for marks which look like black boxes on the top or bottom of a form. The advantages of OMR are that this data capture technology does not require a recognition engine. Therefore: it is fast, using minimum processing power to process forms, and costs are predictable and defined. Conversely, OMR is unable to recognize hand-printed or machine-printed characters. Tick boxes may not be suitable for all types of questions. In that regard, questionnaire design and preparation is Critical. Field Operators must take particular care in filling out questionnaires, and training is essential.

67. DRS provided the technical details of the two available OMR technologies, that is, OMR from image and OMR dedicated scanners.

68. During the discussion, a number of issues were raised by the participants. It was specified that the average speed of OMR is around 4000 pages/hour. Participants asked whether all data can be captured through OMR, and it was concluded that OMR is suitable for questions that have limited number of responses. The process for rejected forms was also discussed, as well as the printing requirements and the conditions of storage of the paper.

Session 8: Data Capture: OCR / ICR / IR

69. This session consisted of a presentation by UNSD and by DRS and BetaSystem. UNSD presented the definitions and the main difference between the OCR, ICR and IR technologies. OCR gives scanning and imaging systems the ability to turn images of machine printed characters into machine readable characters, while ICR is able to process hand written characters. OCR/ICR has less strict form design compared to OMR. Forms can be scanned through a scanner and then the recognition engine of the OCR/ICR system interprets the images and turn images of handwritten or printed characters into ASCII data (machine-readable characters). Images are scanned and stored and maintained electronically, therefore, there is no need to store the paper forms as long as you safeguard the electronic files. The relative advantages and disadvantages

were also highlighted in this presentation. Scanning and recognition facilitate efficient management and planning for the rest of the processing workload, but this technology is costly and may require significant manual intervention.

70. The presentations by DRS and BetaSystem were focused mainly on form design, hardware/ software requirements, workflow, accuracy, and relative advantages and disadvantages of the three methods.

71. During the discussion that followed the presentations, participants asked precisions about the "voting machine" system. The principle is to use 2 or 3 recognition engines in order to determine the final result with higher level of accuracy. The ability of OCR/ICR to recognize non roman character, as Arabic or Ethiopian character was also questioned. Images can be stored in raw format or compressed, but the last one entails certain loss of information. On the other hand, big storage capacities are now available.

Session 9: Manual Data Entry

72. This session consisted of three presentations – one by UNSD followed by presentations by Sierra Leone and Uganda. UNSD presented the basics of manual data entry. Computer-assisted keyboard data entry usually uses personal computer data entry programs with built-in logical controls. Some of the tasks accomplished by the programs are: (a) verifying that EA codes are valid and copying them automatically from one record to the next; (b) assigning a number to each person in a household automatically and to each household within an EA; (c) switching record types automatically if required; (d) checking that variable values are always within pre-determined ranges; (e) skipping fields if the logic indicates doing so; (f) supporting keyboard verification of the information entered earlier; and (g) generating summary statistics for the operator and the batch. Decision to use manual entry versus automated entry may take into account: timetable requirements, relativities between staff and hardware costs and possibility to implement more sophisticated technology. Where staff costs are low and computing infrastructure is moderate, keyed entry may be the optimal method.

73. Sierra Leone described the process used for the 2004 census. The Integrated Microcomputer System (IMPS) developed by US Bureau of the census – which is a system for manual data entry, editing, tabulation and management of census and survey data, was used to capture the data. 120 Data Entry Operators, supervised by 12 data entry supervisors ensured the complete capture of the census forms. The process was organized in order to capture a representative sample of 10% of the population in the first 2 months. The total process lasted 8 months, for a population of less than 5 million people. The main challenges were the duplication of EA codes and household numbers, the possibility for the operators to force values that were out of range, and the data verification was not effective. On the other hand, the benefits from the manual data entry were to bbuild institutional capacity by the purchase of more microcomputers which could be used for other activities after the census, that it created employment facility and training for more people, data can be easily identified and keyed, and inconsistent data can be identify and resolved before keying.

74. Uganda conducted its last census in 2002. The results showed that Uganda population size then was 24.2 millions. The software used was CSPro for capture, CONCOR for editing and CSPro for tabulation. After a study tour to South Africa, Kenya, and Zambia, the Uganda Bureau of Statistics (UBOS) opted for the traditional mode of data capture using keyboards. The UBOS took into consideration several points, as the little time remaining before the census (only 12 months), the budget available for the

equipment and technical expertise, the fact that the pilot census had used data entry. Local expertise was available to manage traditional data entry method, and there was no local technology available to print papers for scanning technology, international procurement for this would have been very slow. About 527 persons working in three shifts for 6 days a week were in charge of the data capture. A team had 10 people and one was a team leader. The process started with the capture of a summary sheet in October 2002, during one month. Then the agricultural module was capture, with 100% verification. Then the Household Module (100% verification and later 20% verification) started in July 2003 and ended in October 2004. All new data entry clerks were subjected to a one month training to perfect their keyboard skills. The focus of the training was speed and accuracy.

75. The discussion raised question about the organization in shifts, and the best way of measuring the performance of the staff. It was concluded that it is more efficient to associate the staff to the evaluation of accuracy and timeliness, and to constantly update training according to the results of the evaluation. The participants also discussed the best way to ensure capture of every EA, without omission or duplicates. It was recommended to constitute as soon as possible a master file with the code of all the EAs.

Sessions 10: Data Capture: Process Stages

76. UNSD made a presentation to highlight the major issues regarding stages of census process, from the field operations to the production of an edited microdata file ready for tabulation. The importance of a complete and updated EAs list was recalled. It was also noticed that reporting population aggregates by EA and capture it quickly gives the possibility to produce first results and to control the data capture process.

77. 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.

78. 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.

Session 12: Data Capture: Overview of Major Distributors/ Commercial Suppliers

79. During this session, three presentations were made by the data processing providers DRS and Beta Systems. Each provider gave an overview on its specific solution to census data processing and some concrete examples illustrating its implementation

Session 13: Data Coding

80. UNSD gave a presentation on coding of data for censuses, which covered the basic concepts and definitions. Coding is the process in which census questionnaire entries are assigned numerical and/ or alphanumeric values. The objective is to prepare data in a form suitable for entry into computer and for further analysis by users. Two methods

were presented: simple coding, limited to reference to one question on the census form, e.g., birthplace; and structured coding, used for complex topics (e.g. occupation, industry, education, etc.), in which reference may be made to more than one question. Regardless of system used, they all rely on coding indexes. The indexes are lists of typical responses likely to be given on a census form that has associated classification code assigned to them. Coding operations may involve one of the three options: (a) assigning numerical codes to responses recorded in words or in a form requiring modification before data entry; (b) rewriting numeric codes recorded say on a questionnaire to a separate coding sheet to facilitate data entry, and (c) use pre-coded entries on questionnaires which may be used directly for data entry. 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 and ISCO) were discussed.

81. Participants discussed the level of detail to be coded regarding the main classification of occupation and industry. It was noted that coding in the most detailed classification entails complexity and requires more skills and time to process. The participants concluded that there is a trade-off between the cost of coding and the value of the information produced, to be discussed with the stakeholders. The possibility of coding in greater detail a sample of the population was also mentioned as a good compromise.

Session 14: Introduction to Data Editing

82. UNSD gave a presentation on the introduction to data editing, commencing with a definition of the terms used. Editing is the procedure for detecting and correcting errors from data, and Imputation is the procedure of assigning values to missing or inconsistent data. The presentation described 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.

83. The discussion highlighted that editing rules must be elaborated by a team composed of demographers, statisticians and IT specialists. Demographers must give and endorse the editing, whereas IT specialists implement these rules into a computer program, and give also rules related to the database structure.

Session 15: Concepts and Methods in Data Editing

84. UNSD gave a second presentation on data editing, going into more detail into the notions of within record editing and across record editing. The presentation gave the definition of the different concepts. “Structure edits” check coverage and relationships between different units: persons, households, housing units, enumeration areas. “Validity checks” are performed to see if the values of individual variables are plausible or lie within a reasonable range, whereas “Consistency checks” are performed to ensure that there is coherence between two or more variables. Two different ways of running the edits were presented: the Top-down Editing approach starts by editing top priority variable (not necessarily first variable on questionnaire) and moves sequentially through all items in decreasing priority; the Multiple-Editing approach uses a set of rules that state the relationship between variables. The process of imputation changes one or more responses or missing values in a record or several records to ensure

internally coherent records result. Two methods of imputation were presented: Cold Deck and Hot Deck. In the Cold-Deck approach, values are imputed on a proportional basis from a distribution of valid responses (e.g., from previous census). In the Hot-Deck approach one or more variables are used to estimate the likely response based on data about individuals with similar characteristics.

85. The participants discussed the number of variables to take into account in the imputation matrix, and the trade between efficiency and simplicity. For example, the neighbouring households are usually processed together, and are usually close in terms of housing characteristics. This should be kept in mind when designing the edits as the addition of too many variables in the Hot-Deck would entail to take a donor far from the household. Participants also stressed the importance of taking into consideration the national social context in the design of the edits.

Session 16: Data Editing (Practical Exercises)

86. UNSD gave an introductory presentation on the basics of CSPRO, a software package freely available on-line, developed by US Census Bureau, and having three main functionalities: data entry, data editing and data tabulation. Then the presentation showed several examples of implementation of edits using CSPro, and detailed the writing of the codes and running on the sample database. Examples covered basic imputation, control between two variables, control among the members of a household, and example of Hot-Deck imputation.

RECOMMENDATIONS & CONCLUSIONS

1. The workshop reiterated the need for African countries to collect data on the core topics presented in the Principles and Recommendations for Population and Housing Censuses Rev. 2 and those listed in the African addendum to the Principles and Recommendations.

2. The workshop took note that disability was a core topic in the revised Principles and Recommendations. In order for African countries to successfully administer the Washington Group proposed census questions on disability; the workshop called upon UNECA and UNSD to provide support to African countries on the implementation of the disability module in their censuses.

3. The participants underscored the importance of the complete coverage of a country with respect to delineation of enumeration areas (EAs), thus, without omission or duplication. The development of a master file, after the completion of a census mapping exercise, comprising a complete list of EAs was recognized as an indispensable step in controlling the process of data capture.

4. Timeliness and accuracy in census data capture were emphasized by participants. Different methods of data capture were accordingly discussed, including OMR, OCR/ICR and manual data entry. In this regard countries were urged to take advantage of the available data capture technology, however, taking cognizance of the countries' circumstances with respect to financial resources, expertise, and outsourcing modalities. On the other hand scanning systems contractors were urged to include, in their solutions, the sustainability of their products.

5. Statistical/census offices should encourage its staff to be proactive in acquiring the relevant knowledge from contracted consultants or solution providers in order to build capacity.

6. The meeting discussed, at length, of the possibilities of outsourcing as part of the census process in the area of data capture. In this regard participants recommended that decisions pertaining to outsourcing should be taken early enough during the preparatory stages of a census in order to allow time for the biding process, for testing and implementation of technical specifications. The following aspects should be considered before embarking on out sourcing:

    (a) An assessment of skills available at the national statistical/census office to prepare the tender documents and implement the technology;

    (b) The contract should describe precisely the deliverables expected, the timeframe, and should include strict confidentiality and security requirements, as well as a quality assurance plan;

    (c) National Statistical/census offices should be actively involved in the drawing of contracts and cooperating partners should be responsive in involving NSOs in substantive contract discussions.

7. The workshop recommended that the census data capture process should have a complete quality assurance plan, regardless of the technique used. The different parts of the process should be monitored with few but reliable indicators. If the process is outsourced, the national statistical/census office should control the quality assurance process including security requirements and conformity to the deliverables.

8. The participants discussed different stages pertaining to the processing of census data. They emphasized the importance of early preparation for data capture activities including logistics issues. In addition, it was recommended that the pilot census should include a test of data capture and data editing.

9. Regular backups of material and captured data should be planned for at the earliest possible stage of the census programme. Archiving of the census forms should be undertaken according to the national laws on archiving, and the database should be wholly archived with the relevant metadata. In this regard, the workshop urged UNECA and UNSD to organize training workshops on archiving and to support the development of projects and tools to help countries to safely archive their upcoming census results.

10. In developing the whole process of data capture, national statistical/census offices are to ensure that edit specifications and codes for open-ended questions are developed in collaboration with subject matter specialists.

11. Although editing and imputations may improve the reliability of census results, the workshop recommended adoption of minimalist approaches and the use of sound statistical techniques when imputing. Over-editing and over-imputation should be avoided.

12. The participants expressed, with great appreciation, the benefits they drew from the exchange of relevant experiences during the workshop. In addition, they recommended for enhancement of the knowledge pertaining to census data processing, and urged UNSD to collect and disseminate information through its 2010 World programme website and knowledge base.

13. African countries should mobilize technical expertise from different countries and share knowledge, including providing assistance to each other in census and survey processing.

14. The workshop recommended that national statistical/ census offices contribute to the 2010 World programme on Population and Housing Censuses website in terms of programs used in data processing; reports; questionnaires; experiences and any other material that might be useful to other countries.

EVALUATION OF THE WORKSHOP

15. Overall, the participants appreciated the workshop's main focus on new technology for data capture and editing. The participants emphasized that the most useful element of the workshop is the sharing of experiences, and discussions between the countries. Also, they generally appreciated the Data capture sessions, introduction of new technologies in Data processes and scanning, and presentations of both commercial suppliers and participants. General opinion of the participants is that the workshop was useful and well organized.

ANNEXES

Annex I. Agenda of the Workshop

Annex II. List of participants

Annex I. Agenda of the Workshop

UNSD Regional Workshop on Census Data Processing for the English Speaking African countries: Contemporary technologies for data capture, methodology and practice of data editing

Dar es Salaam, Tanzania, 9-13 June 2008

Agenda

Time	*Topic*	*Responsibility*	*Document*
	Monday June 9, 2008		
	Opening		
9:00-9:30	***Registration of participants***		
9:30 – 10:00	**Session 1 – Opening remarks -** welcoming remarks by Host country, UNSD, UNECA, administrative matters	UNSD	
	Review of United Nations Principles and Recommendations for Population and Housing Censuses and the preparation of the 2010 round of censuses in the region Objective: To present revisions of the United Nations Principles and Recommendations for Population and Housing Censuses followed by a round table on the preparation of 2010 round of censuses in the region.		
10:00 – 11:00	**Session 2 – The 2010 World Programme on Population and Housing Censuses /** UN Principles and Recommendations for Population and Housing Censuses, Rev.2- major issues in the revision, UNECA addendum for the region – Presentation by UNSD – Presentation by UNECA – General Discussion	UNSD-UNECA	Pres. 1 (UNSD)
11:00 – 11:15	***Coffee break***		
11:15 – 12:30	**Session 3 – Preparation of the 2010 round of censuses in the region** – Presentation by UNSD – Presentation by each participant of the situation in his/her country – General Discussion	UNSD - ECA	Pres. 2 (UNSD)
12:30 – 14:00	***Lunch***		
	Introduction to Data Capture Methods and Outsourcing versus in-house processing 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, Manual Data Entry and provide an overview of different process stages		
14:00 – 15:30	**Session 4 – 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 – Presentation of questionnaire results – Presentation by Nigeria – General Discussion	UNSD, countries	Questionnaire

Time	*Topic*	*Responsibility*	*Document*
15:30 – 16:00	***Coffee break***		
16:00 – 17:30	**Session 5 – Outsourcing versus in-house processing.** Is outsourcing required? How to manage outsourcing. Country examples on outsourcing of data capture – Presentation by UNSD – General Discussion	UNSD	Pres. 4 (UNSD)
	Tuesday June 10, 2008		
9:00 – 10:30	**Session 6 – Introduction to Data Capture** Methods of data capturing, advantages and disadvantages of each method, issues for consideration when choosing the method. – Presentation by UNSD – General Discussion	UNSD	Pres. 3 (UNSD)
10:30 – 11:00	***Coffee break***		
	Data Capture: Optical Mark Recognition Technology		
11:00 – 12:30	**Session 7 - Data Capture: Optical Mark Recognition** Construction/Design Characteristics, Hardware and Software Requirements and Scanning/Storage, Advantages and Disadvantages; overview of the major commercial suppliers. – Presentation by UNSD – Presentation by Expert (DRS) – General Discussion	UNSD, Presentation by Experts	Pres. 5 (UNSD) Pres. B (Expert)
12:30 – 14:00	***Lunch***		
14:00 – 15:30	**Session 8 - 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. – Presentation by UNSD – Presentation by Experts (DRS – BetaSystem) – General Discussion	UNSD, Presentation by Experts, Country Presentations	Pres. 6 (UNSD) Pres. C (Expert)
15:30 – 16:00	***Coffee break***		
16:00 – 17:30	**Session 9– Data Capture: Manual Data Entry** – Presentation by UNSD – Presentation by Sierra Leone – Presentation by Uganda – General Discussion	UNSD, Country Presentation	Pres. 7 (UNSD) Pres. I (Country)
	Wednesday June 11, 2008		
9:00 – 10:30	**Session 10 – 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 – Presentation by UNSD – Presentation by Expert (DRS – BetaSystem) – General Discussion	UNSD, Presentation by Experts	Pres. 8 (UNSD) Pres. D (Expert)
10:30 – 11:00	***Coffee break***		
11:00 – 12:30	**Session 11- Country Presentations on Data Processing (II)** – Presentation by Tanzania – Presentation by South Africa – Presentation by Ghana	Presentations by Countries	
12:30 – 14:00	***Lunch***		

Time	*Topic*	*Responsibility*	*Document*
14:00 – 15:30	**Session 12 – Data Capture: Overview of Major Distributors/Commercial Suppliers** – Presentations by Commercial Providers	UNSD	Commercial Providers Presentations
15:30 – 16:00	***Coffee break***		
	Data Coding 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.		
16:00 – 17:30	**Session 13 – Data Coding** 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 – Presentation by UNSD – General Discussion	UNSD	Pres. 9 (UNSD)
Thursday June 12, 2008			
	Data Editing 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.		
9:00 – 10:30	**Session 14 – Introduction to Data Editing** 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 – Presentation by UNSD – General Discussion	UNSD	Pres. 10 (UNSD)
10:30 – 11:00	***Coffee break***		
11:00 – 12:30	**Session 15 – Concepts and Methods in Data Editing** Within Record Editing – Concepts of validity and consistency checks, examples of both population and housing edits, example of how edit specifications done, concept and methods of imputation – Presentation by UNSD – General Discussion	UNSD	Pres. 11 (UNSD)
12:30 – 14:00	***Lunch***		
14:00 – 15:30	**Session 16 - Data Editing (Practical exercises)** Exercises on CSPRO	UNSD	Pres. 12 (UNSD)
15:30 – 16:00	***Coffee break***		
16:00 – 17:30	**Session 17- Country Presentations on Data Processing (III)** – Presentation by Kenya – Presentation by Mauritius – Presentation by Zambia – Presentation by Swaziland	Presentations by Countries	
Friday June 13, 2008			
	Final Report, Recommendations & Conclusions		
9:00 – 10:30	**Session 18- Country Presentations on Data Processing (IV)** – Presentation by Ethiopia – Presentation by Gambia – Presentation by Sudan	Presentations by Countries	

Time	*Topic*	*Responsibility*	*Document*
10:30 – 11:00	***Coffee break***		
11:00 – 12:30	**Session 19 - 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

Annex II: List of participants

UNSD Regional Workshop on Census Data Processing for the English speaking African Countries: *Contemporary technologies for data capture, methodology and practice of data editing*

Dar es Salaam, Tanzania, 9-13 June 2008

No.	Country Name	Contact Person/Address
1.	Botswana	Mr. Maclean Gwafila Principal Statistician Central Statistics Office Private Bag 0024 Gaborone, Botswana Tel. No.: 267 367 1300 Fax No.: E-mail: mgwafila@gov.bw
2.	Botswana	Ms. Kelebetse Mbiganyi Principal Statistician Central Statistics Office Private Bag 0024 Gaborone, Botswana Tel. No.: 027 367 1394 Fax No.: E-mail: kmbiganyi@gov.bw
3.	Botswana	Mr. Moshe Raletsatsi Statistician II Central Statistics Office Private Bag 0024 Gaborone, Botswana Tel. No.: 267 3188500 Fax No.: 267 3188610 E-mail: msraletsatsi@gov.bw
4.	Ethiopia	Mr. Kifle Gebre Wolde Systems and Programming Team Leader Central Statistical Agency P.O. Box 1143 Addis Ababa, Ethiopia Tel. No.: 251 0911409817 (cell phone) Fax No.: 251-011553011 E-mail: kgebre@yahoo.com
5.	Gambia	Mr. Lolley Kah Jallow Principal Information Analyst Central Statistics Department Ministry of Finance Mamadi Manjang Highway Kanifing Municipal Area Gambia Tel. No. :(220) 4377845 / 7827395 Fax No.: E-mail: lolsaka@yahoo.com

6.	Ghana	Mr. Kwadwo Bamfo Danso-Manu Director of ITC Ghana Statistical Service P.O. Box GP1098 Accra, Ghana Tel. No.: 233-277-226- 227/ 233-21-671737 Fax No.: 233-21-664304 E-mail: dansomanu@yahoo.com
7.	Kenya	Mr. Ohmond Omenya Nyahul Statistician I Kenya National Bureau of Statistics P. O. Box 30266 00100 Nairobi GPO, Kenya Tel. No.: 254 720 770983 Tel. No.: 254-20-315977 E-mail: omenya@knbs.go.ke
8.	Kenya	Mr. Christopher Omolo Principal Economist/Demographer Kenya National Bureau of Statistics P. O. Box 30266 00100 Nairobi GPO, Kenya Tel. No.: 254-020-244079; 254-722632692 Fax No.: 254-020-244079 E-mail: cndayara@yahoo.com
9.	Kenya	Mr. Mutua Kakinyi Senior Statistical Officer Kenya National Bureau of Statistics P. O. Box 30266 00100 Nairobi GPO, Kenya Tel. No.: 254 20 317588; 254 20 317651 Fax No.: 254 20 315977 E-mail: kakinyim@kubs.go.ke ; kakinyim@yahoo.com
10.	Lesotho	Mr. Emisang Tsosane Data Processing Manager Bureau of Statistics P. O. Box 455 Maseru 100 Lesotho Tel. No.: 266 22 323852 Fax No.: 266 22 310177 E-mail: ei.tsosane@bos.gov.ls
11.	Liberia	Mr. Joseph Nyan Director/Data Processing Liberia Institute for Statistics and Geo-Information Services (LISGIS) Tubman Boulevard, Sinkor P. O. Box 629 Monrovia, Liberia Tel. No.: 231-6556174 Fax No.: E-mail: jnyan71@yahoo.com
12.	Malawi	Ms. Angela Msosa Chief Statistician National Statistical Office

		P. O. Box 333 Zomba, Malawi Tel. No.: 265-1-524 377 Fax No.: 265-1-525 130 E-mail: amsosa@statistics.gov.mw angela.msosa@yahoo.com
13.	Mauritius	Ms. Narainee Devi Gujadhur Statistician Central Statistics Office Life Insurance Corporation Centre 1 John Kennedy Streets Port Louis, Mauritius Tel. No.: 230 2122316/17 Fax No.: 230 2114150 E-mail: cso@mail.gov.mu ; ngujadhur@mail.gov.mu
14.	Namibia	Ms. Clothilde Narib Senior Statistician Assistant National Planning Commission Secretariat Office of the President Government Office Building Luther Street, Private Bag 13358 Windhoek, Namibia Tel. No.: 264 81 361 1881 Fax No.: 264 61 239376 E-mail: cnarib2004@yahoo.co.uk
15.	Namibia	Mr. Wilhelm Uushona Shipingana Chief systems Administrator National Planning Commission Secretariat Office of the President Government Office Building Luther Street, Private Bag 13358 Windhoek, Namibia Tel. No.: 264 61 283 4111 Fax No.: E-mail: wshipingana@npc.gov.na
16.	Nigeria	Ms. Adesola Taibat Fatilewa Assistant Director National Population Commission Information Technology Department Plot 2031 Olusegun Obasanjo way Wuse Zone 7 P.M.B. 0281. Garki Abuja Tel. No.: 234 08037868869; 234-08055500264 Fax No.: E-mail: adesolaf@yahoo.com
17.	Rwanda	Mr. Augustin Twagirumukiza Data Processing Manager Rwanda National Institute of Statistics Ministere des Finances et de la Planification Economique P.O. Box 6139 Kigali, Rwanda Tel. No.: (250) 0864626

		Fax No.: (250) 0864626 E-mail: twaugustin@yahoo.com
18.	Seychelles	Mr. Michel Jean Mellie Principal Statisticien National Statistiques Bureau Caravelle House P.O. Box 206 Victoria, Mahe Seychelles Tel. No.: (248) 6111668 Fax No.: (248) 225634 E-mail: michel@nsb.gov.sc
19.	Sierra Leone	Mr. Thekeka Moses Conteh Programmer and Officer-in-Charge, Data Processing Division Statistics Sierra Leone A. J. Momoh Street Tower Hill, PMB 595 Freetown, Sierra Leone Tel. No.: 232-33-399-883 Fax No.: 232 22 223 897 E-mail: thekekamoses@yahoo.com
20.	Sierra Leone	Mr. Abdul Sawar Bakarr Dissemination and Archiving Officer Statistics Sierra Leone A. J. Momoh Street Tower Hill, PMB 595 Freetown, Sierra Leone Tel. No.: 232 33 880 960/+232 77 831 997 Fax No.: 232 22 223897 E-mail: al_shamin@yahoo.com
21.	South Africa	Ms. Margaret Africa Statistics South Africa De Bruyn park Building Private Bag X44 Pretoria, 0001 South Africa Tel. No.: 2712 310 4700 Fax No.: 2712 310 8125 E-mail: margareta@statssa.gov.za
22.	South Africa	Mr. Meshack Ramoenyane Data Processing Professional Statistics South Africa De Bruyn park Building Private Bag X44 Pretoria, 0001 South Africa Tel. No.: (012) 310 4830 Fax No.: E-mail: meshackr@statssa.gov.za
23.	South Africa	Mr. Gideon Sello Mosamai Data Processing Operational Manager Statistics South Africa De Bruyn park Building Private Bag X44

		Pretoria, 0001 South Africa Tel. No.: 27123104853 Fax No.: 27 866 55 4025 E-mail: sellom@statssa.gov.za
24.	South Africa	Mr. Andrian Twala Statistics South Africa De Bruyn park Building Private Bag X44 Pretoria, 0001 South Africa Tel. No.: 27123104839 Fax No.: 2786 600 5334 E-mail: andriant@statssa.gov.za
25.	South Africa	Mr. Antoine Roger Pakutudila Muanza System Developer Statistics South Africa De Bruyn park Building Private Bag X44 Pretoria, 0001 South Africa Tel. No.: 27 123 04709 Fax No.: 27 123 104865 E-mail: AntoineM@statssa.gov.za
26.	Sudan	Mr. Amin Ahmed Daoud Director, Central Bureau of Statistics P. O. Box 700 Khartoum, Sudan Tel. No.: 249 122 316121 Fax No.: 249 183 771860 E-mail: amin_daoud@yahoo.com
27.	Swaziland	Mr. Muziwandile Dube Senior Statistician Central Statistical Office P.O. Box 456 Mbabane H100, Swaziland Tel. No.: 268 628 0173 Fax No.: 268 404 3300 E-mail: muzi_dube@yahoo.com statistics@africaonline.co.za
28.	Uganda	Mr. Anthony Matovu Principal Systems Analyst, Data Processing Uganda Bureau of Statistics Statistics House Plot 9 Colville Street P.O. Box 7186 Kampala, Uganda Tel. No.: 256 041-706000, 256-772-438903 Fax No. 256-41-237553 E-mail: anthony.matovu@ubos.org anthony.matovu@yahoo.com
29.	Zambia	Mr. Frank Kakungu IT Manager Central Statistical Office

		P. O. Box 31908 Lusaka 10101, Zambia Tel. No.: 260 211 253468 Fax No.: 260 211 253468 E-mail: fkakungu@zamstats.gov.zm
30.	Zimbabwe	Mr. Wish Mwanbasheuta Chipiro Chief, Computer Programmer Central Statistical Office P. O. Box CY 342, Causeway Harare, Zimbabwe Tel. No.: 263-04-793967 Fax No.: 263-04-728529 or 263-04-744757 E-mail: chippy_w@yahoo.co.uk or director@csp/zarnet.ac.zw
31.	Tanzania	Mr. Irenius Joseph Ruyobya Principal Statistician National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: iruyobya@hotmail.com
32.	Tanzania	Ms. Aldegunda Komba Manager, Demographic and Social Statistics Depart. National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: aldekomba@nbs.go.tz
33.	Tanzania	Mr. William D. Mabusi Manager, IT Department, National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: wdmabusi@nbs.go.tz
34.	Tanzania	Ms. Beatrice Rwegoshora IT Specialist National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: beaty-gift@hotmail.com
35.	Tanzania	Mr. Anthony Kinyage IT Specialist National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: kinyage@nbs.go.tz

36.	Tanzania	Mr. Vincent Chacha Mugaya Head of Geoinformation Section, National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: mugaya@yahoo.com & vinmugaya@nbs.go.tz
37.	Tanzania	Mr. Muhoji Kitambara IT Specialist National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: kitambara@nbs.go.tz
38.	Tanzania	Mr. Kuchengo Seif Statistician National Bureau of Statistics Kivukoni Front P.O. Box 796 Dar Es Salaam, Tanzania Tel: +255 22 2122722/3/4 Fax: +255 22 2130852 E-mail: skuchengo@yahoo.com
39.	Tanzania	Ms. Mayasa M. Mwinyi Senior Statistician Office of the Government Statistician, Zanzibar P.O. Box 2321, Zanzibar Tel: +255 24 2231869 Fax: +255 24 2231742 E-mail: mmayasa@@hotmail.com
40.	Tanzania	Mr. Amour Bakari Senior Statistician Office of the Government Statistician, Zanzibar P.O. Box 2321, Zanzibar Tel: +255 24 2231869 Fax: +255 24 2231742 E-mail: economicstat@ocgs.go.tz
41.	Tanzania	Mr. Haroub A. Othman Head of Statistics Office, Pemba. Office of the Government Statistician P.O. Box 2321, Zanzibar Tel: +255 24 2231869 Fax: +255 24 2231742 E-mail: aliharoub@yahoo.co.uk
42.	Tanzania	Mr. Abdallah Othman Statistician Office of the Government Statistician, Zanzibar P.O. Box 2321, Zanzibar Tel: +255 24 2231869 Fax: +255 24 2231742 E-mail: abdullaothman@hotmail.com

43.	Tanzania	Girson Ntimba Economist Planning Commission P.O. Box 9242 Dar Es Salaam, Tanzania Tel: +255 22 2112681/2. E-mail: gntimba@yahoo.co.uk
44.	Tanzania	Mr. Robert Masatu Masingiri Economist Planning Commission P.O. Box 9242 Dar Es Salaam, Tanzania Tel: +255 22 2112681/2. E-mail: masingiri2000@yahoo.com
Private Companies		
45.	DRS Data Services Limited	Mr. John Gomersall DRS Regional Manager DRS Data Services Limited Danbury Court, Linford Wood, Milton Keynes, MK14 6LR, England. Registered in England No.5568337 Tel. No: +44(0) 1908 666088 Fax No: +44(0) 1908 607668 E-mail: john.gomersall@drs.co.uk
46.	DRS Data Services Limited	Andy Tye International Manager DRS Data Services Limited Danbury Court, Linford Wood, Milton Keynes, MK14 6LR, England Tel. No: +44(0) 1908 666088 Fax No: +44(0) 1908 607668 E-mail: andy.tye.drs.co.uk
47.	Beta Systems Software AG	Richard J. Lang Director Consulting International Hubnerstrabe 3, D-86150 Augsburg Tel: +49(0) 821 794 183 Fax: +49(0) 171 353 75 79 E-mail: richard.lang@betasystems.com
UN Agencies		
48.	UNFPA	Christopher Mwaijonga UNFPA Assistant Representative 11 Ocean Road, Sea View, P. O. Box 9182, Dar es Salaam Tanzania Tel. Gen.: 255-22-2163500 Ext. 505 Direct: 255-22-2163505 Fax: 255-22-2132001 E-mail: mwaijonga@unfpa.org
49.	UNFPA	Mr. Mwiru Sima UNFPA Monitoring and Evaluation Officer 11 Ocean Road, Sea View, P. O. Box 9182, Dar es Salaam Tanzania Tel. Gen.: 255-22-2163500 Ext. 505 Fax: 255-22-2132001 E-mail: sima@unfpa.org

50.	UNFPA	Mr. Rutasha Dadi UNFPA Assistant Representative. 11 Ocean Road, Sea View, P. O. Box 9182, Dar es Salaam Tanzania Tel. Gen.: 255-22-2163500 Ext. 505 Fax: 255-22-2132001 E-mail: dadi@unfpa.org
51.	ECA	Mr. Kidus Mengistu African Centre for Statistics UN Economic Commission for Africa (UNECA) P. O. Box 3001 Addis Ababa, Ethiopia Tel. No.: 251 11 544 3235 Fax No.: 251 11 551 0389 E-mail: kmengistu@uneca.org
United Nations Statistics Division (UNSD)		
52.	UNSD	Mr. Jeremiah Banda, Chief Demographic and Social Statistics Branch Statistics Division DC2-1554 United Nations New York, NY 10017 Tel. No.: (212) 963 8338 Fax No.: (212) 963 1940 E-mail: banda@un.org
53.	UNSD	Mr. Jean-Michel Durr, Chief Demographic Statistics Section Statistics Division DC2-1556 United Nations New York, NY 10017 Tel. No.: (917) 367 9165 Fax No.: (212) 963 1940