United Nations Regional Seminar on Census Data Archiving for Africa, Addis Ababa, Ethiopia (20-23 September 2011)

Report

Table of contents

Background and justification 3

Organization and participation in the regional seminar 4

Objective and expected outcome of the regional seminar 4

Session 1- Opening Session 4

Session 2 – Results of the pre- seminar questionnaire 5

Session 3 – Introduction to Census archiving 6

Session 4 - Overview of archiving of Micro-data 7

Session 5: Confidentiality and anonymization of micro data 9

Session 6 –Documentation and Cataloguing in data archiving 10

Session 7 -Data storage maintenance and Security 12

Session-8 Access policy and dissemination of Micro data 15

Session-9 Countries Case Studies on Archiving Census Data 16

Session-10 Examples tools for archiving of census data-Use of IMIS 21

Session 11 – Examples of tools for archiving of census data – IHSN toolkit 22

Session 12 – Introduction to Census Info 23

Session 13 – Seminar Conclusions and Recommendations 23

Annex 1: List of Participants 26

Annex 2: Work programme 29

Background and justification

The United Nations Statistics Division (UNSD) as well as the African Centre for Statistics (ACS), through the 2010 World Population and Housing Census Programme, is engaged in improving the statistical capacity of member states to effectively conduct population and housing censuses. To attain this objective, UNSD and ACS are promoting the implementation of international methodological guidelines and standards. They also facilitate the exchange of national experiences and know-how so as to contribute to the efficiency and effectiveness of census operations. UNSD as well as ACS has conducted workshops on various census related topics, including cartography, GIS, census operations management, data processing, evaluation and post enumeration surveys, analysis and dissemination of census data. The next area of focus under the 2010 World Programme has been strengthening the capacity of member states to undertake effective archiving of census data.

National Statistical Offices (NSOs) generate vast amounts of data and information in the course of census operations, including micro- and macro-data, indicators, sample frames, administrative documents, methodological and analytical reports and maps. These assets represent significant investment and constitute valuable and irreplaceable resources of countries which should be managed in a way that sustains their utility for present and future use. Documenting and archiving census data and information are in particular important for preserving institutional memory, planning and implementing future censuses, and conducting comparative temporal analyses. NSOs tasked with producing official statistics are expected to have an effective archiving plan for long-term management of census data and information. However, the reality is such that many developing countries not only lack the institutional capacity for formulating and implementing such a plan, but also do not consider this as a key matter in the census operation.

Archiving encompasses all the activities undertaken by an NSO to ensure that the digital and other contents from data collection operations are maintained in formats that are usable and accessible for current and future uses. The impact of rapid and continued technological changes requires that NSOs take measures to ensure the longevity of data. Information in digital form is at risk of loss from obsolescence of technologies. Digital information is also vulnerable to physical degradations and damages and from improper handling. Computer components and media can physically fail due to human error, natural events, infrastructure failure, malicious destruction and even just the passing of time. As technology upon which digital content relies is replaced by a new one, effective archiving requires managing the impacts due to obsolescence, including through migration of data. To address the challenges of digital archiving, NSOs need to monitor technological developments and systematically consider potential preservation implications.

Other considerations for minimizing loss of and maximizing longevity of data include issues related to data handling protocols, data storage, data security, disaster recovery and emergency rescue of digital content. In addition to technological concerns, NSOs need to address relevant organizational issues to accomplish data archiving objectives. Among these are policies and procedures for data confidentiality and access to and dissemination of archived data. Good documentation is also an essential component of archiving that is critical for ensuring that census data and information are preserved and remain accessible in the long-term. Metadata explain the content and structure of data, provide contextual and explanatory information necessary for properly understanding and using data, and facilitate the discovery, exchange, and archiving and preservation of data.

Organization and participation in the regional seminar

The Regional Seminar was jointly organized by UNSD and ACS. The seminar with 38 participants convened in Addis Ababa, Ethiopia from 20 to 23 September 2011. Representatives from 18 NSOs in Africa—representing both English- and French-speaking countries—that have conducted a census in the 2010 round participated as well as UNSD, ACS, the IHSN and UNFPA (see Annex 1 for the list of participants).

Objective and expected outcome of the regional seminar

The regional seminar provided a forum for identifying and discussing the major issues in archiving census data as well as for exchanging information on national practices and experiences in census data archiving. The seminar aided in identifying the core challenges in census data archiving. It also assisted in identifying the steps to be taken and the considerations that have to be made for formulating and implementing an effective archiving plan suited to the needs and requirements of individual NSOs. The seminar helped in identifying principles and good practices for sustainable, long-term preservation and utilization of census data. The ideas generated by the discussion and the recommendations made by participants contributed towards a potential technical report on archiving census data.

The agenda of activities included items such as: (i) purpose, objectives and scope of archiving, (ii) requirements for archiving census data:- organizational considerations, technological infrastructure, technical capabilities and resources, (iii) data storage, maintenance and security, (iv) documentation, (iv) data confidentiality,(v) data access policy and dissemination of archived data, (vi) country case studies on archiving census data and (vii) recommendations for effective archiving of census data.

Session 1- Opening Session

Opening remarks by UNECA

Opening remarks were made by the Director of the African Centre of Statistics (ACS), Mr. Dmitri Sanga, representing the UNECA in which recognized the teamwork between UNSD and the ACS in organizing the seminar, as co-operation in census taking activities is essential among these institutions. The Director of ACS emphasized the fact that, stakeholders need to ensure that the collected information during the census is processed, analyzed, disseminated and archived. He further indicated that a census is not complete without these processes which are aimed at satisfying the ever growing needs of data users. Mr. Sanga reiterated the purpose of the seminar, i.e., being aimed mainly at discussing critical issues pertaining to census data archiving, and pointed out that the presence of countries from different geographically areas and languages would widen the variety of ideas during the discussions in the seminar. Furthermore, he acknowledged that all the discussions and recommendations during the seminar would go a long way in contributing to the drafting of a report on census data archiving. ACS mentioned the work currently being done or has been completed in support of the 2010 Round of censuses, with undertaken activities including: advocacy campaigns to make countries participate in the current round of the census, workshops, advisory services, and production of technical manuals. The Director of ACS concluded by thanking the colleagues from UNSD and other international organizations present for working together to make the seminar on census data archiving a success.

Opening remarks by UNSD

The opening remarks were delivered on behalf of the Director of UNSD, by Ms. Keiko Osaki-Tomita the Chief of Demographic and Social Statistics Branch at UNSD. Ms. Osaki-Tomita acknowledged the contribution made by UNECA in facilitating the organization of the seminar. She stated that, in the context of the 2010 World Population and Housing Census Programme, UNSD urges countries to conduct a population and housing census at least once in the period 2005 to 2014. She further mentioned that UNSD has undertaken a lot of activities in the current round of censuses with the aim of assisting countries to conduct census activities effectively and efficiently. These support activities included:; seminars/workshops on cartography, management of population census operations, data processing, evaluation, analysis, and dissemination of census data. Ms. Osaki-Tomita articulated that the time had come to shift the focus of the programme of census activities to strengthening the capacity of countries to effectively archive census data. She pointed out that a population and housing census is a massive undertaking which generates vast amounts of data and information. Ms. Osaki-Tomita further stressed that census data represent significant national investment and constitute valuable and irreplaceable resources of a country whose value can further increase through wide-spread and long-term use. She also emphasized the fact that the seminar would provide a forum for sharing of national experiences as well as for identifying and discussing critical issues and challenges associated with effective strategies for efficient archiving census data. Ms. Osaki-Tomita reminded participants that National Statistical Offices (NSOs) are expected to respond to the growing demand of disseminating and sharing census data in variety of forms depending on ever changing technological environment. In conclusion, she noted that ideas generated through discussions at the seminar would be compiled as recommendations for the development of a technical report on census data archiving.

Session 2 – Results of the pre- seminar questionnaire

The presentation by UNSD was focused on country practices on census data archiving and use of census micro-data based on the results of pre-seminar questionnaire. The questionnaire was sent to and answered by countries with the goal of better understanding national and regional practices on census data archiving as well as identifying associated challenges. The questionnaire contained 17 questions and was responded to by 21 countries. Some of the findings of the survey include the fact that 14 countries have a law/regulation for storing paper census questionnaires. Also, seven countries have a law/regulation for archiving of census micro-data and 16 countries have already developed a system for archiving. Meanwhile three countries have not yet developed a system but have a plan for archiving, and 16 countries have a unit responsible for archiving in the NSO. The survey indicates that many countries have developed the system for archiving fairly recently and the tools used or being planned to be used for archiving were: International Household Survey Network Toolkit (IHSN) (17 countries), Redatam-based IMIS (10 countries), CSPro (9 countries) and an in house developed system is being used by 3 countries. Challenges that countries indicated that they face in census data archiving include: lack of specific laws/regulations/policies for archiving, determining anonymization and dissemination procedures, absence of a disaster recovery plan, sustainability of the archiving process, lack of general guidelines for choosing a suitable platform to archive census data, lack of adequate and skilled manpower, and limited funds.

During the general discussion that followed the presentation, some participants wished to know more about the experience of the three countries that indicated having developed their own in house data archiving systems as this experience was seen as vital to be shared with

the rest of the participants. This is because if an in-house system has been developed with more or less the same functionalities as the commercial softwares, then it would be preferable, since the local systems would be more flexible to accomodate emerging changes in data archiving. In response two of the countries indicated they have an established procedure to disseminate their census data micro data, while the third country mentioned that they are presently using two systems namely: the IHSN tool kit and an in-house developed system that can better be called a repository rather than software, that is used for data storage and backup for the ISHN and other data sets been generated.

Participants pointed out that in many countries, census laws are outdated and do not take into account recent developments in archiving particularly in the area of electronic archiving of data instead of hard copies which were accepted as the state of art in the passed. It was concluded that there is a need for the laws to be revised to allow archiving of census data and documents. Countries have a history in census taking, and as such, census instruments, tools, and other documents being developed and used should be archived, with copies of these historical documents in their original paper forms or other formats sent to the national archive for storage and preservation. Some participants mentioned that to take advantage of technological developments, their countries are considering amending their statistical laws concerning archiving of data and documents, including census questionnaires. It was also mentioned, however, that in some countries, although micro data is in the public domain, is it necessary for the public to ask permission through the National Office's to use the data. Safeguarding data confidentiality is the reason most cited for this practice.

It was also mentioned during the general discussion that in attempting to archive past censuses, some countries faced challenges in the form of lack of associated metadata and other documentation that was used during those censuses. This information includes census questionnaires, field staff manuals, sampling procedures, data coding and editing schemes, etc. On the other hand, some of the participants whose countries are using the IHSN toolkit, for instance, have indicated how use of these technologies has facilitated documentation of all the census processes: planning, field operations, data processing, analysis and dissemination, and archiving of data.

Session 3 – Introduction to Census archiving

The session included a presentation by UNSD and also two country presentations by Togo and Seychelles. The presentation by UNSD provided an overview of census data archiving, including the definition, the responsibility of the data producers, need to preserve the data for long-term usage, necessary steps in archiving data, core attributes for a data archiving programme, organizational and technological infrastructure, resource framework and major activities in archiving of micro-data (documentation, anonymization and dissemination). The presentation also highlighted some challenges in the implementation of an archiving programme, such as: resource constraints, lack of specific laws/regulations and policies for archiving, lack of institutional commitment, inadequate IT infrastructure, lack of appropriate skilled personnel, ability to implementing data anonymization procedures, lost data in original census databases, lack of general guidelines for choosing a suitable platform to archive census data, lack of mechanism to cross check the quality of archived micro-data and ability to integrate and harmonize data from successive censuses.

The representative from Togo started the presentation by defining census archiving, purpose of archiving and its benefits, developing an institutional strategy for archiving and

organizational considerations, requirements for archiving in terms of resources, legal framework, technological infrastructure and technical capabilities in archiving census data. The representative indicated that as part of the 4th population and housing census of Togo which was conducted in 2010, the central bureau of the census created an archiving section as part of the division on data processing and archiving. The newly created section is in charge of implementing a plan for archiving for the census and has four archivists and a documentalist. The presentation further highlights expected organizational aspects and archiving plans that the new section has to take into account in order to accomplish its task.

In the presentation on the Seychelles country experience in census data archiving, the presenter looked at aspects such as the procedures to access census micro-data, challenges and benefits of archiving data. The National Population Data base (NPD) of Seychelles, helps partially in the archiving of census data. There are provisions made in Electoral Commission ACT for users to obtain data from census records. This includes micro data from selected variables that are released without identifiers such as: names, address, age, and nationality which are collapsed for security reasons. A formal letter requesting census data is addressed to the commissioner heading the National Bureau of Statistics before data can be obtained. The country faces several challenges in archiving data, including: lack of training in the use of relevant technological tools for archiving census data, inadequate storage arrangements for safe keeping of data, lack of specific laws/regulations, poor backup and restore facilities, high database administration cost, weak system for data upgrading, and reusability of data. The National Bureau of Statistics does not have a system for archiving census data but is planning to establish one in 2011 or 2012.The presenter concluded by stating the following: there is a need to have a law/regulation data archiving; instead of using boxes to store loads of paper form, digital archiving of census forms will save a lot of space and cost less; it is important to maintain data as inexpensively and effectively as possible, as well as ensuring the viability of micro-data in the present as well as in the future; the promotion, acquisition, documentation, dissemination and preservation of micro-data is essential for the production of national statistics, for research and for other use.

During the general discussion, participants pointed out that as part of the archiving strategy for long-term use of census data, countries should take into account not only needs for micro-data, but for aggregated tabulations as well. Participants highlighted the challenges of archiving old datasets and documents using current archiving technology. As a way to avoid this in the future, participants stressed the need to convert archived datasets and documents to the latest storage media for security and compatibility purposes. Participants also pointed out the need to archive the original edited census data prior to anonymization or before any modifications are made.

Session 4 - Overview of archiving of Micro-data

The session included a presentation by UNSD and presentations on country experience by Kenya and Chad. The UNSD presentation offered a synopsis of archiving of micro-data, including the definition of micro-data, the benefits of disseminating micro-data, data files for archiving, preparing data sets, data security, tools for archiving of micro data and the risks of disseminating micro data. It was mentioned in the presentation that the main reason for disseminating micro-data is in support of research by offering flexibility to define variables and modified categories in a way that meets the needs of researchers. It was mentioned in the presentation that micro-data sets can be generated from various data sources such as: censuses, surveys and administrative registers and that a clear acquisition policy that describes scope,

source and mandate for the acquisition of micro data sets is necessary. In this regard, the NSO can play an important role by expanding the scope of the data archive to official sources such as line ministries.

In order to ensure data confidentiality, census data should not contain variables that are direct identifiers. The physical security of the data should also be ensured by controlling access to rooms where data are held. Network security and confidential data on servers or computers connected to an external network with firewall protection and security-related upgrades should be in place to avoid viruses and malicious code.

The presentation by Kenya focused on their specific national experience. Regarding tools used for archiving and dissemination micro data, the Kenya National Bureau of Statistics (KNBS), has used standard and non-standard tools for data archiving and dissemination. Recently KBS has started using Redatam based on the IMIS platforms for disseminating data from the 1989 and 1999 census data and is building capacity for using NADA. Furthermore, Kenya has used the IHSN tool kit to document and archive seven survey datasets. KNBS has not yet undertaken archiving of data from the 2009 census of Kenya. The presenter added that due to advancements in ICT and increased demand for data, KNBS is changing the way of accessing and disseminating information to the users. Traditionally KNBS has been using publications, seminars and workshops to release and disseminate survey and census data. Now, census micro-data can be accessed and queried using IMIS.

KNBS also uses offline dissemination tools such as CDs and DVDs for dissemination of census results, and upon request provides a 5% sample of census data to users. It was also mentioned that Kenya has enable data users to access data from the 2009 census through mobile telephones.

The representative of Chad indicated that the national statistical system (NSS) of Chad of which the INSEED is part, lacks experience in archiving of documents and digital data sets on computer media. The absence of this basic function has failed to capitalize on the experiences within the NSS and has not promoted the development of an analytical system. Thus, the National Strategy for Statistical Development (NSSD) has identified facilitation of use of statistical products as a strategic axis which relates to dissemination, access, data use and promotion of statistical literacy has two strategic objectives: facilitating access to data products and strengthening the role of statistics in support of decision making. According to the NSSD also, the function of archiving in the NSS should be strengthened so as to achieve the following two results: (i) statistical results and metadata are published in a format suitable for users and distributed on time and periods consistent with the GDDS; (ii) a secure national data warehouse is posted online and promotes data analysis by users. The metadata will be developed and updated by a working group to be appointed by the Statistical Programme Committee (SPC) under the coordination of INSEED. Information and communication technology (ICT) will be used to facilitate the archiving and dissemination of the data, and consultants will be recruited to provide guidance for the architecture of the database.

In response to a concern that was raised during the general discussion regarding verifying the quality of raw data, it was stated that this depends principally on the method of collecting data. It was agreed upon that there are difficulties in accurately verifying the quality of raw data as the census is composed of different processes. Participants were also concerned about the multiplicity of tools used for data archiving that could be confusing to countries regarding which

one to use (IMIS, IHSN tool kit, CensusInfo, etc.). Participants were advised to regard the different tools as complimentary each with some unique functionalities.

Based on the presentation by the representative of Kenya, some participants wanted to learn more about the use of mobile telephones in the dissemination of census data as this could be of interest to their countries.

Session 5: Confidentiality and anonymization of micro data

The session included a presentation by the African Center for Statistics (ACS) and also by Zambia and Senegal on their national experience. The presentation by ACS focused on: micro data, micro data confidentiality, anonymization, anonymization techniques and anonymization tools. To ensure confidentiality legal arrangements must be put in place before any micro data are released. The risk of violating data confidentiality can be reduced by the use of numerous annonymization techniques. The downside is that these techniques may reduce the usefulness of the micro data to the data users. Technical arrangements through remote access facilities (RAFs), allow users to have access to the micro data from a remote location and to obtain output over the Internet. Data laboratories/safe sites can effectively control the identification risk whilst enabling users' access to micro-data. It may also be a challenge for the NSO to manage the various anonymization techniques, consisting of removing or modifying the identifying variables contained in the dataset.

It was indicated in the presentation that direct identifiers are variables such as names, addresses or identity card numbers, while indirect identifiers are characteristics that may be shared by several respondents, whose combination could lead to the re-identification of one of them. Anonymizing data consist of determining which variables are potential identifiers. This could rely on making personal judgments while modifying the level of precision of these variables in order to reduce the risk of re-identification to an acceptable level. Synthetic micro data are an alternative approach to data protection - this is produced by using data simulation algorithms. The rationale for this approach is that synthetic data do not pose problems with regard to statistical disclosure. This is because they do not contain real data but preserve certain statistical properties. Generally, users are not keen to work with synthetic data as they cannot be confident of the results of their statistical analysis. Nevertheless, this approach can also help in producing "test micro data sets." In this case, synthetic data files would be released to allow users to test their statistical procedures before having access to "true" micro data in a safe site.

Several computer software have been developed to facilitate anonymization of micro data. The broadly used software is µ-Argus which was created by Statistics Netherlands. Many of the statistical disclosure methods described above can be applied using µ-Argus. The program is interactive and goes through several steps such as, the identification of the dataset (metadata); the selection and computation of frequency tables on which several statistical disclosure control methods are based; the selection of possible recoding and inspection of the results (global recoding). The selection and application of other protection methods and of the anonymisation of the micro data file including the production of a data process report can be carried out using the software. IHSN has also developed tools and guidelines on micro data anonymization. These tools are developed in such a way that other software programs such as Stata, SPSS and SAS can be integrated into the system. Free software, such as the Data Swapping Toolkit (DSTK), can also be used for performing and analyzing data swapping for categorical data.

The presentation from Zambia mentioned the fact that one of the missions of the Central Statistical Office (CSO) is to ensure adequate dissemination of census results to satisfy the needs of different user needs by creating many products from census data and disseminating these products using different channels and media, such as printed copies, website, electronic summaries etc. The presentation also focused on the importance of data confidentiality and how to ensure it. In Zambia, data confidentiality of respondents is guaranteed under the provisions of the Census and Statistics Act of 1967. According to this Act, the Director of the CSO has to authorize access to information and all users have to formally agree to certain conditions such as making no copies of any files, accepting not to use any techniques that could identify any person or establishment, and accepting to hold in the strictest confidence the identification of any establishment or individual that may be inadvertently revealed in any documents or analysis.The CSO ensures that the data that it provides to users is annonymized to a certain extent. In carrying this out, identifiers such as names of respondents, names of establishments, addresses are removed and some geographic areas are collapsed.

The presentation from Senegal focused on the country's statistical law and its implications on the confidentiality and anonymization of micro data, with particular emphasis on methodological issues to annonymize micro data and policies for dissemination. The statistical law was revised in July 2004, ensuring the confidentiality of individual data as well as permitting access to data by all users in a transparent manner. The anonymization of micro data in Senegal includes aspects such as suppressing variables that identify administrative areas and individuals as well as the renumbering households. In Senegal, the general public is allowed limited access to micro data but decision-makers and organizations can access data under strict control, while complete micro data access is allowed to researchers. The policy of dissemination of data to the general public is guided by the law on surveys and censuses. Annonymized micro data sets are provided to all stakeholders of the statistical system, and a micro data bank has been created which can be accessed uniquely by researchers. Currently, the NSO is testing anonymization methods using the DHS results using software such as u-ARGUS and T-Argus.

During the general discussion, participants wanted to know how techniques such as "adding noise", "resampling" and "data swapping" are carried out. They were provided a list of various reference materials for further reading. It was also explained that the importance of documenting the changes made in the anonymization process is to facilitate the tracking down and testing of the imputations being done and to return to the original data set if needed. The software chosen for archiving micro data should allow documentation of the changes undertaken. Documentation is important and can be used as reference material for future statistical operations. It was further mentioned that whatever changes are made on the data, users must be informed about the alterations made. In terms of tools that could be used to give users access to micro–data, IMIS and IHSN toolkits were mentioned as examples.

Session 6 –Documentation and Cataloguing in data archiving

The session included a presentation by UNSD as well as country presentations by Liberia and Burundi. The presentation by UNSD defined what documentation is. It also addressed the role of documentation in census data archiving as well as when it should be undertaken. For instance, it was mentioned that documentation is an incremental process that should be a shared responsibility among various parts of an institution. The presentation went on to say that different types of documentation can be added by different people at various stages of an information object's life cycle. As a result, a common documentation framework, used by different actors so that the actor who is closest to the information to be used as documentation/metadata adds that information to the framework. Three broad categories of material for documentation were

identified in the presentation: (i) explanatory material, (ii) contextual information, and (iii) cataloguing material.

The presentation also included a summary of common metadata standards and the advantage of using these standards. In this context, the following XML-based standards were discussed: (i) the Data Documentation Initiative which is based around the data lifecycle model and provides specifications for a structured framework for organizing the content, presentation, transfer and preservation of metadata in the social and behavioural sciences, and (ii) the Dublin Core Matadata Standard which is a general purpose metadata standard for describing digital resources related to micro-data, such as questionnaires, reports, manuals, data processing scripts, and programs.

The presentation also provided a definition of cataloguing, its benefits and also examples of the items that are included in catalogues as well as characteristics of a good catalogue from the users' and also the data producers' point

In the presentation on the experience of Burundi, it was stated that ISTEEBU has created a unit for archiving data from censuses and surveys. Since 2000, ISTEEBU has used the International Household Survey Network (IHSN) toolkit for documentation and archiving, and intends to use IMIS for archiving data for the 2008 census of Burundi. The documentation by ISTEEBU using the toolkit covers the following elements: data quality assessment related to the methodology for sampling, data collection, extrapolation of the results and weighting, data processing, geographical coverage and non-response rate. Documentation in the toolkit also includes variables and their definitions in the form of a dictionary that describes each variable and provides a clear definition for it. For the documentation of data sets, the toolkit permits the creation of directories containing archived documents that are classified by subject for easy referencing by researchers in the catalogs. The toolkit also enables archiving documentation on policies on data confidentiality and procedures for data anonymization.

According to the presentation from Liberia, the Liberia Institute for Statistics and Geo-Information Services (LISGIS) was using manual documentation before 2007 and started using the IHSN metadata Toolkit to document the Censuses and surveys in 2008. LISGIS has a Division for Coordination and Dissemination that is charged with the responsibility of documentation, achieving and dissemination of all censuses and surveys conducted by the institute. Data from the 2008 census of Liberia has been archived based on 20% of the total population. As part of the archiving process, the following are identified: methodology in terms of methods for the census, variables, definitions, and data capture information, uploaded census questionnaire and the associated documentation. The data are edited before the documentation is completed and uploaded. LISGIS has put into place procedures for the assessment of data before they are disseminated through website and other dissemination tools. The agency has also instituted anonymization procedures to remove individual and household identifiers.

LISGIS is also using the NADA Data Cataloguing which is being customized and uploaded to the agency's website. Also for preservation and dissemination of data, LISGIS is using IMIS which contains a set of relevant indicators at national and local levels as well as information on methodologies used for data collection, processing and analysis.

The challenges being encountered presently in archiving data sets are: lack of motivation of current staff, insufficient resources, lack of cooperation among ministries and agencies to provide information for documentation.

During the general discussion, participants emphasized that documentation should be undertaken throughout the census operation and not be left until the end.

Session 7 -Data storage maintenance and Security

The session comprised of four presentations, one by the ACS, and three national presentations by Ghana, Ethiopia and Cape Verde. The presentation by the ACS was composed of the following aspects: strategies for data storage, institutional back-up policy, procedures to safeguard and secure data, procedures for data transmission and encryption. In the presentation, data storage was defined as the holding of data in an electromagnetic form for access by a computer processor. Two main kinds of storage were identified: (i) primary storage where data is held in random access memory (RAM) and other memory devices that are built into computers and (ii) secondary storage where data is stored on external storage devices such as hard disks, tapes and CD's. It was pointed out in the presentation that strategies for data storage depend on the information environment and that the success of any census archiving system is highly associated with its ability to store and manage information. Furthermore, storage systems are a critical part of the NSOs network infrastructure that must keep pace with the ever increasing availability of data in the holdings of the NSO.

In the presentation, it was stated that in designing a viable storage strategy for census data archiving, the choice of the right technology for the primary storage as well as a solid backup procedure that ensure system management must be guaranteed. The storage strategy design issues to be considered in constructing a the system are: preventing data loss, offering adequate capacity that can be easily scaled as storage needs grow, providing fast access to data without interruptions, being prepared for equipment failures, and using cost-effective technologies.

In the ACS presentation, it was also pointed out that backup copies are an insurance policy against the possibility of data being lost, damaged or destroyed. In cryptography, encryption is the process of transforming information (referred to as plaintext) using an algorithm (called cipher) to make it unreadable to anyone except those possessing special knowledge of it.

The presentation from Ghana elaborated their strategy regarding data storage, maintenance and security. In terms of archiving strategy and management, the GSS has setup the National Data Archive, which has adopted the Data Documentation Initiative (DDI) and the Dublin Core (DCMI) international metadata standards, since 2008. Census micro data is archived by adopting the International Household Survey Network (IHSN)'s standard procedures and recommendations for data archiving. Data is anonymized by altering or suppressing variables which could potentially identify a respondent or establishment. The presentation also highlighted some of the challenges faced in terms of (i) unavailability of census related documents (questionnaires, manuals, codebook, etc.) at a centralized location, and (ii) lack of consistent or harmonized definitions, categorization, and classifications of variables among different censuses/surveys.

The presentation also covered data storage technology mapping which is a system that is virtualization environment aware, application aware, and can determine how data is being

accessed. In terms of storage data protection, it was mentioned during the presentation that GSS has not yet implemented a de-duplication solution but plans to combine this solution with their enterprise backup solutions. Also covered was the GSS data center solution architecture as well as the server and storage infrastructure.

Regarding data security, it was pointed out that basic security issues faced by the GSS include physical security (e.g. stolen laptops), internal security (e.g. file backups), external security (e.g. Internet security), and integrity (e.g. audit trails). One other security-related issue that could be considered is whether smart cellular telephones should be banned from data centers.

In the presentation, it was mentioned that for backup and disaster recovery, GSS performs daily backup of their virtual machines and keeps an encrypted copy of the process export on an external drive that is kept at the datacenter of the Ministry of Finance. The eGovernment Network Infrastructure will be used as the backbone network to link to remote branches. In addition, GSS has made plans to use the disaster recovery site of the eGovernment Datacenter that is about 400km away.

It was stated that for data their data access policy, the GSS as a public institution has the obligation to promote data dissemination to facilitate governance and national development. GSS has set up three levels of access to archived census or survey micro-data (i) public use files which are free on the internet; (ii) licensed datasets which require a signed agreement; and (iii) datasets only accessible on location. Customized tables are provided at a token fee while 1% of census raw data are provided upon formal request and payment of a processing fee.

A representative of the Central Statistical Agency (CSA) made a presentation on the experience of Ethiopia with regard to census data archiving. The presentation included an overview of census data archiving that is undertaken by the CSA. In addition, the methodology adopted with regards to data storage was also discussed with emphasis made on the procedures/steps taken for safe-guarding the security of the census data. The presentation also covered challenges faced by the CSA in census data archiving.

Cape-Verde presented on their experience regarding the DataCenter project whose objective is to consolidate the INS computer system in response to the obsolescence of old infrastructure that does not guarantee the security of data, in particular in the following areas: (i) migration of existing physical servers to the "cloud" environment; (ii) increasing volume available to store data; and (iii) improving security, availability and breakdown tolerance in terms of backup data to ensure business continuity and also to implement a more robust infrastructure to allow access to information by researchers.

The INS is instituting “virtualization” which is a technique that allows sharing and using resources from a single computer system composed of several other virtual machines. Each virtual machine provides a comprehensive computer system very similar to a physical machine. Thus, each virtual machine can have its own operating system, applications and network services. It was mentioned in the presentation that virtualization allows (i) more centralized management whereby resources are managed, distributed and shared (server) with easy resource allocation of CPU and memory; (ii) migration of existing servers (creating new virtual machines; reduction of hardware costs and maintenance, resulting in a lower power consumption; fast disaster recovery; and simplified and centralized administration).

In terms of data storage and maintenance at the INS, the hard disk management system is based on the RAID6 (standard) system to maintain data integrity in case of failure of at least 2 disks. There is therefore need to set up a SAN (Storage Area Network) with the benefits of (i) sharing the whole storage capacity for all existing and future servers; (ii) implementation of high availability (shared storage and replication of data); and (iii) flexibility to increase capacity.

The INS system also ensures business continuity (fault tolerance and replication) through (i) implementation of the Cluster - system with two servers (called nodes) in which they work together, ensuring the availability of services and resources on an ongoing basis by using the implicit redundancy in the system; (ii) achieving high availability through local redundancy in case of failure of a server, and also remote redundancy in case of loss of the primary location. Assuring business continuity is supported by the central storage of data and its copy on a remote site.

In the presentation, it was mentioned that in order to promote democratization of access to information, the INS created the INE-Labstat, a laboratory for the analysis of statistical data, allowing researchers to manipulate data, but without physical access to micro-data, and making it impossible to copy from the desktop. INE-Labstat has a server hosted in the data center and terminals without local resources such as OS, CPU and memory and peripherals writing (USB, CD / DVD etc.). This set-up minimizes vulnerability to viruses and eliminates the possibility of direct data access.

The presentation also covered aspects of security and access to data (Ine-Labstat). In this regard, it was stated that to ensure the physical Security of the data, the gateway to the DataCenter is prohibited except to authorized persons. It is planned that in the future a biometric test and video surveillance will be installed. Logical security is provided by the system manager whereby everyone has a username and a password and a "homefolder" to store data.

During the general discussion, participants appreciated the use of new technologies like cloud computing. Participants also, however, raised an issue about the format in which the data and metadata to be preserved should be. It was indicated that it is highly recommended to use the ASCII file format or the XML since these are readable in any software. Secondly, it is safe to archive the very raw data with possibly different versions of it for different user groups. Also, it was underlined that backups must be done automatically to avoid tragic mistakes. Some participants wanted to know the criteria used to categorise data files as public and licensed. It was suggested that public data are posted on the website and are tabulated while licensed datasets are for registered users only.

Participants also stressed the importance of establishing a backup system as well as the measures taken in building an offsite backup system. It was pointed out that this measure is extremely important because if there is a disaster and the data is lost, there would be a backup system for data retrieval.

Session-8 Access policy and dissemination of Micro data

The session had four presentations, one by UNSD and national presentations by Botswana, Egypt, and Rwanda. It was emphasized in the presentation by UNSD that "A census is not complete until the information collected is made available to potential users in a form suited to their needs". The presentation defined accessibility as a census strategic objective and also discussed issues about whom micro-data should be made available to. It further talked about why national statistical offices should support the research community including the perspectives of data producers and of the research community regarding access to micro-data. The presentation also discussed how access to micro-data should be managed as well as the importance of supporting legislation and the attendant responsibility of national statistical offices as custodians of individual data confidentiality.

Also covered in the presentation was the issue of how the seemingly opposing views of researchers and the national statistical offices can be managed. The presenter also presented suggestions for how national statistical offices can manage risks of micro-data access and also some management issues associated with the release of micro-data. Also discussed in the presentation are examples of policies for access to archived data and arguments for and against providing micro-data for a fee and copyright issues.

In the presentation on the experience of Botswana, it was indicated that the Central Statistical Office has not started archiving of census data and hence had no explicit experience on access policy for archived data to share. The presenter indicated that data are stored on personal computers and the server and are accessed as and when there is a request for it. Data is provided free of charge. The statistical office has a statistical law that ensures that confidentiality of individual's micro-data is maintained. To put more emphasis on the law, every new officer at the office has to take an oath of secrecy when they join the statistics office.

The presentation on the experience of Egypt covered the following: the International Household Survey Network (IHSN) Micro-data Management Toolkit in CAPMAS, micro-data access, data dissemination, future of micro-data access by the public, future of dissemination, and key challenges.

In Egypt, data are dispersed across agencies sometimes making it hard to locate proper data sets. Also, explanations of methodologies underlying the data are hard to find. CAPMAS has adopted the IHSN micro-data management toolkit for the documentation and archiving of census and survey data. The process is on-going.

In Egypt, methods of data dissemination are evolving and the Internet has revolutionized data dissemination and analysis. In terms of data access and dissemination policy, it was indicated during the presentation that the statistical law in Egypt states that CAPMAS collects information for statistical purposes and it is never used to identify individual. Also, information released to users should not be sold to others, and violating this rule is a crime with severe penalties. When they utilize the data, users should indicate that the source of the information is CAPMAS.

For the future, Egypt is interested in creating a centralize data warehouse that would allow data users to customize the data they need with options for delivery of update notices. The central dissemination portal would include clear, easy to-understand explanations of concepts underlying data as well as up-to-date technical documentation.

According to the presentation by Rwanda, the National Institute of Statistics of Rwanda (NISR) believes that micro-data needs to be disseminated because it plays an important role for further analysis that may be required. The institute has undertaken micro-data arching and uses the following tools: (i) online IMIS based on Redatam R+SP which includes 5 surveys and the 2002 census; (ii) downloadable online datasets that include statistical data, metadata, census document and reports; (iii) NADA is under development and will include all surveys and census conducted at NISR.

In term of the access policy, it was stated that, upon request, NISR provides access to data on the institute's website free of charge to all users.

It was indicated in the presentation that NISR faces the following challenges in its efforts to archive and disseminate micro-data: (i) creating awareness and interest in the data; (ii) having a huge database which requires good bandwidth; (iii) Internet Service Provider not being able to host the database; and (iv) choosing the tool to be used in micro-data dissemination.

One of the issues discussed during the general discussion was whether or not data should be provided at a fee. Some participants supported charging a fee for data access and reasoned that this increases the use of statistics. They also stated that charging fees also covers administrative costs that were not initially included in the census budget, but recognized that in certain situations, the fees are waved, such as in the case of a students. Some participants recognized that with increasing demand for statistical data from the private sector, some companies with enough resources/funds should pay for access to these micro data. Participants also wanted UNSD to inform them if there are international guidelines regarding data for at fee or free of charge. Participants were informed that the decision on whether or not to sell micro data was solely up to the country.

Session-9 Countries Case Studies on Archiving Census Data

Country case study presentations were made the Gambia, Morocco, Malawi and Uganda. A representative of the Gambia Bureau of Statistics (GBOS) made a presentation on their national experience on the archiving of census data. In the presentation, it was mentioned that the Gambia National Data Archive (GNDA) has been established in order to: (i) promote best practice and international standards for the documentation of micro-data amongst data producers in the country; (ii) provide equitable access to micro-data in the interest of all citizens, by protecting confidentiality and following international recommendations and good practice; (iii) promote the effective use of existing census data for statistical and research purposes thereby encouraging a diverse range of analytical work through secondary research; and (iv) ensure the long term preservation of micro-data and the related metadata, and their continued viability and usability in the future.

It was also indicated in the presentation that the Gambia National data Archive pursues the above objectives within the framework of the Statistical Act and the United Nations Fundamental Principles of Statistics. Where micro-data cannot be released due to confidentiality or other reasons, the National Data Archive provides the public with detailed metadata and other publicly available materials. The Gambia National Data Archive has adopted the Data Documentation Initiative (DDI) and the Dublin Core (DCMI) international metadata standards.

Documentation activities at GBOS include:

(i) Acquisition which is primarily established to archive census micro-data produced by GBOS and other official data producers. It also serves as a repository for non-official datasets. Data producers interested in depositing data in the data archive are invited to contact GBOS.

(ii) Data documentation which serves several important functions. It helps data producers build institutional memory, and helps researchers to: (a) find the data they are interested in; (b) locate the datasets and variables that meet their research requirements; (c) understand what the data are measuring and how the data have been created, and assess their quality; and (d) understand the survey design and the methods used when collecting and processing the data, thereby reducing the risk that data will be misunderstood or misused.

(iii) Anonymization - Gambia Bureau of statistics is charged with legal and ethical obligations to protect the confidentiality of census respondent. The Gambia National Data archive aims to protect confidentiality of the data by: (a) restricting access to data that present a potential disclosure risk to scrutinized users only, under formal conditions; and (b) anonymizing data when necessary, by altering or suppressing variables which could potentially identify a physical or legal individual. This may make the data less useful for analysts. The National Data Archive seeks to minimize the information loss while ensuring an acceptable level of disclosure risk. Principles and methods applied for measuring the risk and for anonymizing data are those provided or recommended by the International Household Survey Network.

(iv) Dissemination which increases the quality, use and potential impact of data, by: (a) making it possible for analytical work to be replicated, a critical step to good science; (b) creating the potential to use old data to test new ideas; (c) reducing the costs of data collection and the burden on respondents, by avoiding the need for researchers to undertake their own surveys; (d) demonstrating transparency and credibility in data production, which are at the heart of good governance; and (e) improving the relevance and quality of data by incorporating users feedback in future data collection. The GBOS understands that making micro-data available also has down sides as it exposes data producers to criticism, increases the risk of breach to confidentiality, and can result in conflicting outputs being generated. Having faith in the ethical conduct of data users and in their willingness to contribute to the quality and usefulness of the data, The Gambia National Data Archive considers that the benefits outweigh the disadvantages. The GBOS insists, however, that access to micro-data must not be

seen as a right and will only be permitted to bona fide users, and for statistical and research purposes only.

(v) Preservation - Micro-datasets can be damaged or lost because of human error, because of technical problems, or because of disasters such as fire or flood. New technologies can also render old data unreadable, because of either hardware or software advances. The Gambia National Data Archive is implementing standard procedures for ensuring the physical security and long-term usability of its resources, together with associated backup arrangements for minimizing the impact of adverse events.

In terms of policies and procedures at the GBOS, micro-datasets are categorized into three groups, according to the sensitivity of their content and their inherent disclosure risks: (i) public use files which are made available on-line to all interested users, for research; (ii) licensed files which involve a signed agreement between GNDA and external trusted users, to permit them to access semi-anonymized data files; (iii) files accessible on-site (data enclave), representing sensitive data for which access is only provided on site under strict conditions, and only for research purposes; and (iv) GNDA –the GBOS scrutinizes the generated outputs in a full disclosure review before they are released.

A representative of the Direction de la Statistique (DS) presented on Moroccan experience regarding arching for the census. According to the presentation, Morocco established the Law of November 30, 2007 to regulate the archiving, preservation, confidentiality and dissemination of all documents belonging to the state including questionnaires and documents from the censuses and statistical surveys. Documentation of source documents serves the following purposes: (i) to better exploit and use the micro data; (ii) useful experience for future censuses; (iii) for historical purposes. Archiving of source documents is done through the Toolkit and also through web pages stored on a DVD and viewed from an Intranet. It was stated in the presentation that archiving of the source documents for the last census began only after the data processing and release and publication of results. This was deemed a real constraint to collect all documents. Consequently, for the next census which is planned for 2014, the DS anticipates that the archiving will start at the beginning of preparations of the census and shall appoint persons to be responsible for this task.

The DS has undertaken archiving of data and digital documents of censuses and surveys for a long time. But this archiving is not exhaustive especially in terms of documents, and not based on any system or procedure. The archive, however, has preserved almost all of the data and only a part of the documents. Since 2009, DS is using the IHSN Toolkit but has not migrated all data and documents in this system. The last three censuses (1982, 1984 and 2004) and some surveys are still archived on DVD in a non-structured manner.

All the micro-data are stored in ASCII format. Also, all electronic archives are stored in (i) the data center, (ii) on a set of 2 magnetic tapes one of which is kept on a remote location and (iii) on 2 DVDs one of which is kept at a remote location. Each time the DS acquires new software or a new version of software, all files get converted.

In terms of the dissemination of micro-data, it was indicated in the presentation that all the results of the 2004 census (tables, reports, thematic maps, graphs, etc.) were disseminated through all possible means, including website, paper and CD Rom. With regard to micro-data, a complete copy was made available to the d'Étude et de Recherche Démographique at the Haut Commissariat au Plan. And 16 partial copies, each containing data relating to a region was given to the corresponding regional Directorate. It was mentioned that in general, requests for micro-data made by an administration, an association, a research company, a researcher, or a university are fulfilled. On the other hand, source documents, managed by NADA, are published on the Intranet of the DS.

It was stated in the presentation that, except for direct identifiers (names and addresses), no anonymization has been made to the data. To ensure the security of the data, DS has instituted the following measures: (i) installed a centralized anti-virus system; (ii) installed an intrusion and detection and protection system, and a Fire Wall; and (iii) authentication requirement for entry to the network; and (iv) access to the database enabled through granting of permission.

The representative of Malawi made a presentation on the experience of the Malawi National Statistics Office (NSO) on census data archiving. The presenter stated that accessibility to archived data at the NSO is granted upon official request through the Commissioner and an agreement is signed indicating that data will be used for research purposes only. It was also mentioned that in Malawi, the NSO highly values the confidentiality of the census and survey data and therefore requires users of micro-data to take an oath of secrecy agreeing to use the data for research purposes only.

It was stated in the presentation that for arching purposes, data are converted to IHSN compatible formats. In this regard, micro-data from the 2008 census were converted to SPSS. The presenter also mentioned that currently, NSO has no tool/software for anonymizing the micro-data, but is planning to develop an in-house tool. Also, NSO has 100% of the micro-data on its website although not in the public domain.

In the presentation, it was mentioned that policies on archived data and on data confidentiality have not been finalized. According to planned policies: (i) following the launch of census/survey results, NSO will comprehensively archive micro-data and metadata in a data infrastructure that prevents loss or damage; (ii) encourage data sharing while acknowledging the legal rights and complying with the release/sharing conditions; and (iii) NSO will disseminate 10% of the micro-dataset.

The presentation also highlighted some challenges that NSO faces in the archiving of micro-data, including: (i) lack of anonymization system for crumbling the unique identifiers; (ii) advanced training of staff; (iii) technological tools; (iv) insufficient computers and personnel; (v) computer viruses; (vi) data archiving software or license. For the future, NSO plans to use IHSN tool kit for archiving of micro-data from surveys and censuses.

In the presentation by the representative of Uganda, it was stated that documentation at the Uganda Bureau of Statistics (UBOS) serves to: (i) improve institutional memory; (ii) disseminate standardized documentation to users; (iii) raise the profile of UBOS; (iv) improve data retrieval and analysis; (v) improve time series analysis; and (vi) provide survey information to users globally via the National Data Archive (NADA).

In terms of current dissemination policy, UBOS releases micro-data files for use by researchers for scientific research purposes when: (i) the Executive Director is satisfied that all reasonable steps have been taken to prevent the identification of individual respondents; (ii) the release of the data will substantially enhance the analytic value of the data that have been collected; (iii) for all but purely public files, researchers disclose the nature and objectives of their intended research; (iv) the researchers have signed an appropriate undertaking; and (v) all legal provisions have been adhered to by the researcher. A new dissemination policy which is under development and not yet legalized will give users access to the micro-data once they register on the UBOS website.

The process of making micro-data accessible through the UBOS website involves the following steps: (i) creation of an inventory of existing surveys including collection of documents and datasets; (ii) validation and verification of the datasets; (iii) documentation using the DDI and DCMI templates in the IHSN Metadata Editor; (iv) quality control using checklists; (v) creation of CDs using the CD ROM builder; and (vi) uploading of data and documentation onto the National Data Archive (NADA).

The documentation template is customized by the IHSN and is based on the international DDI and DCMI and contains the following elements: (i) document description; (ii) study description; (iii) datasets; (iv) variable groups; and (v) external resources. Documentation description is used to capture information on the DDI file and helps with versioning control of the DDI file with information on: study title, metadata producer, date of production, DDI document version, and DDI document ID number. In terms of study description, elements that are used to describe the survey or census and includes: identification, version, overview, scope (abstract, kind of data, unit of analysis), coverage, producers and sponsors, sampling, data collection, and data processing. Datasets are also described using the followin elements: (i) file description – desribes who authored the datafile, its contents, missing codes; (ii) key variables and relations - describes the relationship between datasets; and (iii) variables - describe each variable statistics, questions and weighting. The template element on external resources is based on the DCMI template and used to document: (i) survey report and questionnaire; (ii) list of codes and classifications; (iii) statistical tables - report tables and aggregates; (iv) technical documents - methodologies and manuals; (v) administrative documents - budget and accounting documents; (vi) maps and pictures; and (vii) scripts and programs.

UBOS currently has 22 surveys documented and NADA allows online searches of variables and provides detailed information on existing suveys. More information on NADA is available at http://www.ubos.org/nada3/index.php/catalog.

Session-10 Examples tools for archiving of census data-Use of IMIS

The presentation by UNFPA started off by providing a context for the rational for IMIS, such as: the limited interaction between data producers and data users; poor integration of data systems; inadequate knowledge of existing data and information; poor accessibility to data; poor dissemination of information; and inconsistencies of data systems. It was stated that in terms of vision, IMIS is intended to be a reliable one-stop shop national data source with multiple purposes including generation of indicators for monitoring development programmes and the MDGs. The mission of IMIS is: (i) strengthening the partnership between producers and users of data; (ii) consolidation of data files available in an integrated and centralized database; (iii) quick and easy access to data by all potential users; (iv) monitoring development through the generation of indicators (MAP) including the MDGs; (v) assistance to the NSI to fulfill their role as depositary of data. Specific objectives of IMIS are to : (i) create integrated databases complete and coherent; (ii) create a reliable network for sharing data and information; (iii) develop a strong interface (but easy) of indicators calculation; (iv) strengthen national technical capacity in generation and maintenance of databases; and (v) facilitate the development of the culture of using Databases.

The presentation outlined the following as benefits of IMIS: (i) easy and rapid access to voluminous micro and macro data; (ii) compatibility with most usual software; (iii) export to dissemination software such as DevInfo; (iv) security in terms of data stored in compressed encrypted format; (v) original file no longer needed, hence removing the fear to share original data; (vi) selection of universe and a combination of geographical areas for specific analysis; (vii) enhancement of existing data systems; and (viii) assistance to NIS to play its role of repository of statistical data. The following were identified as some of the features of IMIS: (i) integrated statistical database; (ii) web-based database (intranet/internet); (iii) set of relevant indicators at national and decentralized levels; (iv) large number of partners using the database (rapid and easy access). In terms of by-products, IMIS enables harmonization of methodologies of data collection, processing and analysis, and also preservation and dissemination of data.

The presentation also gave a listing of countries in the African region in which activities related to the IMIS project have been initiated. In terms of country experiences with IMIS, the presentation provided some similarities, differences and also challenges. Identified challenges include the fact that for the uploaded IMIS in the countries, the system is developed in some countries but is not yet available online, especially for technical, material and organizational reasons. Also, updating of existing applications is a challenge.

The presentation was conclusion with a summary of some lessons learned which included the fact that in order to implement IMIS in the countries, it is important to have a local team including computer scientists, statisticians and demographers in charge of calculating and interpreting the indicators, the project leaders and all those resources with knowledge of the surveys or censuses to incorporate into the system. Also, it has been found out that the posting of IMIS on the web is a bit difficult in some cases. On the other hand, IMIS is very useful and appropriate for integrity checking and data

The presenter also encouraged south-to-south cooperation in implementing the system and solving problems arisen in due course.

During the general discussion, a questioned as to whether it was possible to compare the advantages and disadvantages of different dissemination tools, including Census Info, DevInfo, IMIS, etc. In response, the fact that these dissemination software are complementary was stressed. In this context, UNSD informed participants that a table has been compiled to show in a comparative way the characteristics and features of the different census data dissemination software packages (CensusInfo, DevInfo, Redatam, CsPro, SPSS and SAS). The table makes it possible to compare the tools with respect to different characteristics in order to assess what they can and cannot do, thereby, assisting countries to make an appropriate choice of what software to us.

Session 11 – Examples of tools for archiving of census data – IHSN toolkit

The presentation by the World Bank covered the "IHSN Microdata Management Toolkit and related standards and good practices". It was stated that there are two main components that make up the IHSN toolkit :(i) Metadata Editor-a specialized software for documenting any kind of micro data (surveys, censuses, administrative records), and (ii) National Data Archive (NADA)-an open source application for cataloguing and dissemination (CD-Builder for dissemination). The IHSN Toolkit is compliant with the Data Documentation Initiative (DDI) and the Dublin Core Metadata Initiative (DCMI) standards. DDI and DCMI are XML metadata standards with standard checklists of what you need to know about a study and its dataset (DDI), and about the related resources (DCMI). DDI was developed by academic data centers but is now used in most countries in the world, and by various software applications (e.g. DevInfo, CsPro). There are two versions of DDI: Version 2.n consists of the DDI codebook and is used by the Toolkit while Version 3.n consists of the DDI life cycle application. The advantage of XML is that (i) it can be transformed into many kinds of outputs such as HTML, PDF, Databases and Others, and (ii) plain text files are not specific to any operating system or application and are therefore "durable" metadata. In terms of the development for the IHSN Toolkit, the metadata Editor was developed by Nesstar Ltd ("Nesstar Publisher") with IHSN support, but is now a freeware. The development of the software benefited from many users' feedback. The NADA CD-builder was developed by the World Bank/IHSN, is available at www.ihsn.org/nada.

In the presentation, the following were identified as benefits of the IHSN Toolkit to national statistical offices: (i) replicability and transparency, (ii) visibility, (iii) credibility, (iii) institutional memory, (v) knowledge generation (if micro-data is disseminated) leading to an increase and demonstrate the value of data resulting in more funding, (vi) satisfy a legal requirement in some countries, and (vii) participate in Open Data / Data Liberation movement.

It was also mentioned in the presentation that IHSN has also developed guidelines for use by countries. These include guidelines: (i) for documenting a dataset using the IHSN Toolkit, (ii) on formulating an access policy and procedures, and (iii) on long term preservation of data and metadata based on OAIS "standard". IHSN is currently working on two manuals, one on producing public use census sample files, and the other on anonymizing micro-data.

The presentation concluded with some recommendations for countries, namely, to: (i) comply with the DDI standard, (ii) produce sample dataset (n%) for public (free) dissemination of

micro-data, (iii) publish a formal micro-data management and dissemination policy, (iv) assess their preservation policy/procedures, and (v) preserve all versions of their census data.

Session 12 – Introduction to Census Info

The presentation by UNSD covered the following topics: (i) what is CensusInfo, (ii) objectives of CensusInfo, (iii) main features of the CensusInfo software, (iv) UNSD activities related to CensusInfo, and (v) the Censusinfo website.

In terms of what CensusInfo is, it was mentioned that it is a software tool to disseminate census data on the web and on CD-ROM. It was developed by UNSD in partnership with UNICEF and UNFPA to disseminate census results. This is because from past rounds of population and housing censuses, dissemination of the results was the weakest part. The objectives, therefore, was to develop an easy to use and customizable software solution to facilitate the dissemination of census results at any geographical level. Some of the main features of CensusInfo that were mentioned in the presentation include the fact that it (i) disseminates census results at any geographical level and for multiple censuses; (ii) provides a suggested list of census topics, data and indicators with accompanying metadata (based on UN recommendations); (iii) easy to customize to meet country-specific needs; (iv) supports to disseminate census data with multiple languages; (v) generates user-defined tables, maps, graphs and reports; and (vi) allows importing data from other software applications such as CSPro, Redatam, SAS and SPSS. The presentation provided a brief overview of UNSD activities to assist countries to adopt CensusInfo, including conducting of (i) regional workshops at which participants get an introduction and overview of CensusInfo software, and (ii) national training on site for advanced and user application training. There was a demonstration of the CensusInfo website.

During the general discussion, participants expressed a keen interest in the CensusInfo application but raised concern over the multiplicity of available software for census data dissemination and the resulting confusion for the countries regarding which one to choose. In response, the complimentary nature of the available software was discussed.

Session 13 – Seminar Conclusions and Recommendations

The participants of the seminar agreed on the following conclusions and recommendations which would also apply to surveys and administrative records produced through the National Statistical System (NSS), where appropriate:

General

- Census data represent significant investment and constitute valuable and irreplaceable resources of a country. The long-term value of census data is greatly enhanced through proper preservation of data and accompanying documentations.

- Archiving encompasses a broad range of processes thereby requiring a clear and comprehensive strategy, starting from the census planning stage, as well as a sustainable organizational setting with proper supporting technology and adequate resources.

- It was recommended that NSOs include in their national census budgets resources for dissemination and archiving of micro-data. This would provide NSOs with the opportunity to explore the possibility of providing micro-data free of charge

Legal Framework/statistical laws

- Although many countries in the region have statistical laws, they are out-dated in many ways. Out-dated laws need to be revised to better reflect current methods of census taking and use of modern technologies. The revised laws should address:
    - dissemination and archiving of micro-data with proper documentation;
    - retention of completed paper census questionnaires for a limited period of time depending on the situation in the country;
    - the archiving a copy of materials used during census operation
    - data anonymization to ensure data confidentiality

Technological infrastructure

- Participants pointed out the difficulties faced by their countries in keeping up with changes in technologies, given the rapid advancement of information technologies. Limited skilled manpower and resources allocated for data archiving are among the common challenges to countries in Africa.

- Participants recommended that NSOs monitor changes in technology and when necessary update software and hardware for storing census data, as some countries have not been able to access and archive datasets that were stored using old technology. Countries requested that international and regional organizations, such as UNSD, UNECA, UNFPA, and the IHSN look into this issue.

Confidentiality

- The majority of countries have statistical laws which include clauses to ensure confidentiality of census micro-data. However, increasing demand for micro-data and advances in technology, pose challenges for provision of the data while maintaining data confidentiality.

- Participants recognized the mandate of the NSO in maximizing accessibility to and utilization of data, including micro-data, but stressed the need to maintain data confidentiality and not release individual identifiable data as this could erode public trust in the NSO.

Anonymization of microdata and dissemination

- As countries are increasingly providing access to census micro-data, appropriate methods or tools for anonymization of data need to be developed to reduce the risk of disclosure of confidential information. Participants, however, pointed out that due to the diversity of anonymization tools and procedures, adequate guidelines are required in order to evaluate and compare them. Countries requested that international and regional organizations, such as UNSD, UNECA, UNFPA, and the IHSN look into this issue.

- Participant drew attention to the archiving and dissemination of data on sensitive and potentially sensitive topics, such as religion and ethnicity. Depending on the national context, countries should identify such sensitive topics and through appropriate anonymization procedures ensure non-disclosure of persons based on such sensitive variables.

Documentation

- Adequate documentation and metadata are necessary for data interpretation and utilization. Census archiving should include long-term storage of both electronic data and its affiliated documentation.

- Documentation covers the entire lifecycle of the census operation. It is therefore important to include it in the overall plan of the census and to undertake documentation at each stage of the census operation and not to wait until the end.

- It was suggested that countries use international data documentation standards for archiving thereby bringing together statistical data and accompanying documentation for better utilization and interpretation of the data. In this connection, the experience of a number of countries in the region that are using the IHSN toolkit as well as the Redatam-based IMIS for archiving of data is important. Use of these tools helps to standardize the information to be archived both within the country and among countries.

Data Storage and Security

- For long-term archiving of census data and documentation, countries are advised to put in place strategies for adequate storage and also for ensuring the physical and technological security of the data. Data security strategies should also cover plans for back-up, migration of data and meta-data files to new formats and media, use of off-site back-ups and also for disaster recovery.

- In order to ensure long-term accessibility and readability of census micro-data, it was suggested that a copy of the raw and edited versions of the census micro-data be stored in ASCII format, which is easily read by all statistical packages available in the market.

Data Dissemination

- It was recommended that NSOs plan for and secure adequate resources as part of the data dissemination strategy at the beginning of the census operation.

- It was observed that, in the region, there is a trend toward data utilization by the private sector. Data dissemination strategies should take this new trend into account.

- Participants recommended that NSOs develop clear data access policies so as to ensure adequate utilization of census data. These policies should be periodically reviewed and revised as necessary.

Annex 1: List of Participants

No	Country Name/ Organization	Count	Contact Person Information
1.	Botswana	1.	Mr. Kealeboga Rampeethwane Assistant System Analyst General Statistics Office
2.	Burundi	2.	Ms. Spec Nimbona Cadre ONI
3.	Cape Verde	3.	Mr. Joao Baptista Lopes De Pina Coordinator of Div. Informatics National Institute of Statistics
4.	Chad	4.	Mr. Riradjih Madnodji Chief of Department INSEED/HEP
5.	Egypt	5.	Mr. Khaled Eldeeb Director of Decision Support System CAPMAS
6.	Ethiopia	6.	Mr. Abas Furgasa Programmer Central Statistical Agency
		7.	Ms. Eleni Kebede Zewdie Automation and date dissemination expert Central Statistical Agency
		8.	Mr. Gezaheen Taddesse Senior Researcher Central Statistical Agency
		9.	Mr. Girum Wordofa Statistician Central Statistical Agency
7.	Gambia	10.	Ms. Lolley Kah-Jallow Principal Information Analyst Gambia Bureau of Statistics
8.	Ghana	11.	Mr. Kwadwo Bamfo Danso-Manu Data Processing Manager Ghana Statistical Service
9.	Kenya	12.	Mr. Silas Mulwah Senior Statistician Kenya National Bureau of Statistics
10.	Liberia	13.	Mr. Yusuff M. Sarnoh Senior Research Officer LISGIS
11.	Malawi	14.	Mr. Richard Annuel Paul Phiri Statistician National Statistical Office

No	Country Name/ Organization	Count	Contact Person Information
12.	Morocco	15.	Mr. Azzouz Taroua Head of IT Department
13.	Rwanda	16.	Mr. Didier Uyizeye Data Processing Programmer NISR
14.	Senegal	17.	Mr. Kane Mamadou Chief de bureau ANSD
15.	Seychelles	18.	Mr. Bertrand Louis-Marie Senior Statistical Assistant National Bureau of Statistics
16.	Togo	19.	Mr. Komi Kolagbe Database administrator DGSCN
17.	Uganda	20.	Ms. Margaret Hellen Muriel Atiro Information officer Data Processing Uganda Bureau of Statistics
18.	Zambia	21.	Ms. Etambuyu Lukonga Documentalist Central Statistical Office
19.	United Nations Statistics Division	22.	Ms. Keiko Osaki-Tomita Chief Demographic and Social Statistics Branch United Nations Statistics Division
		23.	Ms. Margaret Mbogoni Statistician Demographic Statistics Section United Nations Statistics Division
		24.	Ms. Meryem Demirci Interregional Adviser Demographic Statistics Section Demographic and Social Statistics Branch United Nations Statistics Division
		25.	Mr. Yacob Zewoldi Resource Person
20.	UNFPA	26.	Mr. Boubacar Sow Regional Technical Advisor Population and Development, Census, Data & Research Western and Central Africa SRO Dakar, Senegal
21.	World Bank	27.	Mr. Olivier Dupriez The World Bank, Development Data Group
22.	UNECA	28.	Mr. Dimitri Sanga Director, African Centre for Statistics (ACS) UNECA

No	Country Name/ Organization	Count	Contact Person Information
23.		29.	Mr. Raj Gautam Mitra Chief of Demographic and Social Statistics Section African Centre for Statistics (ACS) UNECA
24.		30.	Mr. Ayenika Godheart Mbiydzenyuy Statistician Demographic and Social Statistics Section African Centre for Statistics (ACS) UNECA
25.		31.	Mr. Molla Hunegnaw Statistician Statistical Development Section African Centre for Statistics (ACS) UNECA

Annex 2: Work programme

Time	Topic	Responsibility	Document
Tuesday 20 September, 2011			
	Opening		
9:00 – 9:30	***Registration of participants***		
9:30 – 10:00	**Session 1 – Opening session** - Opening remarks by UNECA - Opening remarks by UNSD - Administrative matters	UNECA UNSD	
10:00 – 11:00	**Session 2 – Results of pre-seminar questionnaire -** Present results of the pre-seminar questionnaire on strategies and experiences on census archiving for countries in the region - Presentation by UNSD - General discussion	UNSD	Pres. 1
11:00 – 11:30	***Coffee break***		
11:30 – 12:30	**Session 3 – Introduction to census archiving** – Session discusses what census archiving is, purpose of archiving and its benefits, developing an institutional strategy for archiving, and requirements for archiving in terms of resources, legal framework, organizational considerations, technological infrastructure, technical capabilities, etc. - Presentation by UNSD - Presentation by countries o Togo o Seychelles - General discussion	UNSD Countries	Pres. 2
12:30 – 13:30	***Lunch***		
13:30 – 15:00	**Session 4 – Overview of archiving of microdata** – Session will cover definition of microdata, why disseminate microdata, acquisition and preparation of data, tools for archiving of microdata, dissemination of microdata, risks of disseminating microdata - Presentation by UNSD - Presentation by countries o Kenya o Chad - General discussion	UNSD Countries	Pres. 3
15:00 – 15:30	***Coffee break***		
15:30 – 17:00	**Session 5 – Confidentiality and anonymization of microdata –** Session covers what data confidentiality is and why it is important to maintain confidentiality of individual data, how to identify risk to disclosure, what is data anonymization and examples of procedures for anonymization of microdata, challenges in anonymization of microdata - Presentation by UNECA - Presentation by countries o Zambia o Senegal - General discussion	UNECA Countries	Pres. 4

Time	Topic	Responsibility	Document
	Wednesday 21 September, 2011		
9:00 – 10:30	**Session 6 – Documentation and cataloguing in data archiving** – Session covers what data documentation is and its role in context of archiving, role of metadata. Documentation includes metadata and other information on data sources, methods of enumeration, variables and their definitions, coding schemes, editing, assessment of data quality, confidentiality and anonymization procedures, geography, census year, etc. - Presentation by UNSD - Presentation by countries o Burundi o Liberia o Mali - General discussion	UNSD Countries	Pres. 5
10:30 – 11:00	***Coffee break***		
11:00 – 12:30	**Session 6 – Documentation and cataloguing in data archiving (contd.)** - Presentation by countries - General discussion	Countries	
12:30 – 13:30	***Lunch***		
13:30 –15:00	**Session 7 – Data storage, maintenance and security** – Session discusses strategies for data storage, institutional back-up policy, procedures to safeguard the security of the data, procedures for data transmission and encryption of the data - Presentation by UNECA - Presentation by countries o Ghana o Ethiopia o Algeria o Cape Verde - General discussion	UNECA Countries	Pres. 6
15:00 – 15:30	***Coffee break***		
15:30 – 17:00	***Session 8 – Access policy and dissemination of microdata – Session discusses***		Pres. 7
	Thursday 22 September, 2011		
09:00 – 10:30	**Session 8 – Access policy and dissemination of microdata (contd.)** - Presentation by countries - General discussion	Countries	
10:30 – 11:00	***Coffee break***		
11:00 – 12:30	**Session 9 – Country case studies on archiving census data** - Presentation by countries o Gambia o Morocco - General discussion	Countries	
12:30 – 13:30	***Lunch***		
13:30 – 15:00	**Session 9– Country case studies on archiving census data (contd.)** - Presentation by countries o Malawi o Uganda - General discussion	Countries	

Time	Topic	Responsibility	Document
15:00 – 15:30	***Coffee break***		
15:30 – 17:00	**Session 10 – Examples of tools for archiving of census data – Use of IMIS** – For the 2010 round of population and housing censuses, some countries are using Redatam-based IMIS for the dissemination of their census results. The session will highlight how IMIS is being used for archiving of data including its functionalities - Presentation by UNFPA - General discussion	UNFPA	Pres. 8
Friday 23 September, 2011			
9:00 – 10:30	**Session 11 – Examples of tools for archiving of census data - International Household Survey Network (IHSN) tool kit** – IHSN has developed a tool kit for data documentation and dissemination. Presentation will cover overview of the tool kit and its functionalities for data archiving, including the DDI metadata standard. - Presentation by World Bank - General discussion	World Bank	Pres. 9
10:30 – 11:00	***Coffee Break***		
11:00 – 12:30	**Session 12 – Introduction to CensusInfo** – UNSD in collaboration with UNFPA and UNICEF has developed the CensusInfo software for the dissemination of census data. Presentation will cover overview of CensusInfo and how it can be used to present and disseminate census results - Presentation by UNSD - General discussion	UNSD	Pres. 10
12:30 – 13:30	***Lunch***		
13:30 – 14:30	**Session 13 – Seminar conclusions and recommendations** - Presentation by UNSD - General Discussion	UNSD	
14:30 – 15:30	**Session 14 – Closing session**	UNSD UNECA	