UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

ECONOMIC COMMISSION FOR AFRICA

Third Meeting of the Statistical Commission
  for Africa (StatCom Africa-III)

Cape Town, South Africa
21-23 January 2012

Report of the African Group on Statistical Data Management

Background

1.   There is general consensus among African governments and development partners that the quality of statistical data used in decision-making affects the outcome. The need has therefore been established for continuous provision of better statistics in support of sound policy formulation and informed decision-making, as well as for monitoring the achievement of internationally agreed goals, including the Millennium Development Goals (MDGs). However, National Statistical Offices (NSOs) and other government institutions have not been able to provide the full range of statistical products and services increasingly demanded by their respective governments and the international community for better development statistics. One of the areas of weakness is in the management and dissemination of statistical data to make them readily accessible to users.

2.   The African Group on Statistical Data Management (AGSDaM) is one of the working groups of StatCom-Africa, which focuses on policy, technical and operational issues related to the management of statistical data and information, and their organization for easy access by the wider community of users, to support evidence-based decision-making. This report presents the activities of AGDSDaM since the second session of StatCom in January 2010.

1.   Capacity-building activities

1.1   Report on Train-the-Trainer Workshops on Innovative Tools and Techniques for Statistical Data Management and Dissemination

3.   Even though the immediate objective of the first project "is to improve the capacity of UNECA in the collection, management, archiving, documentation and delivery of regional statistical information," the project document emphasizes the principle of "subsidiarity", whereby activities that are better undertaken at the national level should not be done at the centre. This subsidiarity principle was emphasized during the second session of the Statistical Commission for Africa (StatCom-Africa II) when the project was presented. Participants pointed out that the project should build in some direct benefits for member States and that their capacity should also be improved for greater effectiveness.

4.   In line with this observation, during StatCom-Africa II, the African Centre for Statistics (ACS) had an immediate side meeting with the representative of Google foundation, who is the task manager of the Google project with ECA, and it was agreed that ECA and Google Foundation should organize training of trainers for selected member States on innovative tools and techniques for statistical data management and dissemination, as an extension of the project on enhancing development in Africa by provision and use of statistical information. Thereafter, a side meeting of selected member States was organized and they endorsed the proposal to hold these training-of-trainers sessions. Participants at this side meeting were from Nigeria, Morocco, Senegal, and the Arab Maghreb Union (AMU). The delegate from Uganda was invited but could not attend due to other StatCom-related engagements but expressed support for the proposal.

1.1.1 Objectives of the trainings

5.   The objectives were to train nominated staff of statistical and mapping agencies in selected member States on the innovative techniques and tools. It was expected that these trained trainers would, in the first instance, pass on the skills to colleagues in their offices and with the trained colleagues, implement the new methods. Subsequent training sessions would eventually be organized for representatives from other countries, using previously trained trainers who would incorporate lessons learnt in the subsequent training sessions.

6.   Consequently, two training workshops were held in Abuja, Nigeria and Rabat, Morocco, for selected English- and French-speaking countries respectively. Nine African countries, two Regional Economic Communities (RECs), Africa Union Commission (AUC) and Economic Commission for Africa (ECA) were trained on the recent and innovative techniques and tools for statistical data management. The training workshops covered the following topics: ECA Statistical Data Management and Dissemination Platforms; new ideas in statistical data management such as statistical data discovery, special and temporal time-series, community mapping and mobile platforms; and Google technologies for statistical data management such as Google Map Maker, Google Maps, Google Earth, Fusion Tables, and Public Data Explorer.

1.1.2 Abuja Workshop

7.   A training workshop took place in the training room of Nigeria's National Bureau of Statistics (NBS) in Abuja, Nigeria from 23 to 26 March 2010.

8.   Participants were invited from Egypt, Lesotho, Nigeria, Swaziland and Uganda. The AUC was also represented. In the case of Nigeria, several organizations were represented, namely, the NBS, the Office of the Surveyor General of the Federation, the National Space Research and Development Agency, the Nigerian Statistical Association, the University of Lagos (Geography Department) and the Regional Centre for Training in Aerospace Surveys. The full list of participants is annexed (annex I).

9.   The workshop was divided into three components, dealing with the ECA Statistical Database Platform, Brainstorming and General Discussion on Key Concepts, and Google Tools and Applications for Data Management, Dissemination and Mapping. Also, NBS made a presentation of their activities relevant to the theme of the training workshop.

10.   At the end of the workshop, participants indicated that they had learnt the new concepts introduced and acquired skills to apply in their work. Participants indicated interest in installing and using the ECA database platform. This was supported by the workshop evaluation.

1.1.3 Rabat Workshop

11.   The training for selected francophone African countries took place in the training room of Morocco's High Commission for Planning (HPC) in Rabat, Morocco from 1 to 4 June 2010. Participants were invited from Burundi, Cameroon, Democratic Republic of Congo, Senegal and

Morocco. AMU was also represented. In the case of Morocco, several organizations were represented.

12.   This workshop was also divided into the three components, namely, the ECA Statistical Database Platform, Brainstorming and General Discussion on Key Concepts, and Google Tools and Applications for Data Management, Dissemination and Mapping. In addition, HPC made a presentation on Population and Housing Censuses using new technologies.

2.   African Group on Statistical Data Management Meeting

13.   A meeting of the African Group on Statistical Data Management (AGSDaM) was held conjointly with the Ad Hoc Expert Group on Statistical Data Management Platforms. The objectives of the meeting were to agree on and formally accept the terms of reference for the operations of the AGSDaM and to develop a work programme for the working group over the next two years. The meeting also discussed and endorsed the recommendations from the Expert Group for preparation of a handbook. It developed the terms of reference for the working group and agreed to meet regularly via online sessions.

3.   Handbook on Major Statistical Data Management and Dissemination Platforms

14.   One of the recommendations put forward by the Data Management Working Group during the first and second meetings of StatCom-Africa was to set up a group of experts made up of statisticians, data management and geo-information experts to evaluate the major existing statistical data management platforms and compare their features so that member States and their partners could make informed decisions on the selection of platforms for statistical data collection, production and dissemination.

15.   The recommendation was prompted by the plethora of offers for data management platforms that member States receive. Some of these offers are at no or reduced cost as part of assistance projects, while others are at commercial values. Even when there are no financial costs, accepting all offers would result in duplication of efforts with associated wastage of scarce human capacity and with the possibility of data inconsistencies. Such feature comparison will therefore enhance sustainability of the information infrastructure and associated tools for effective management and dissemination of statistical data, applications and services.

16.   In this regard, a meeting of data management experts was held at the ECA headquarters in Addis Ababa, Ethiopia to produce a handbook on data management platforms in Africa to assist countries to select the best platform for managing and disseminating statistical data to users.

17.   The meeting was held jointly with AGSDaM, which is the StatCom-Africa Working Group on Data Management. The production of the handbook was part of the work programme of this group. The meeting was attended by 15 experts. The following countries were represented: Cape Verde, DRC, Ghana, Mauritius, Morocco, Nigeria, Senegal, South Africa, Tunisia, Uganda and USA. Representatives from DevInfo, PROGNOZ, AUC, AfDB and ECOWAS were also in attendance.

18.   The Data Management Working Group was one of those groups formed by StatCom-Africa which had not been active for a while. Hence, the major objective of this meeting was to develop a two-year work programme and reactivate the Group. In addition, as a major activity of the Working Group, the meeting discussed preparation of the handbook that was intended to serve the NSOs of member States as an information source on different statistical data management solutions.

19.   Mr. Chukwudozie Ezigbalike, Data Management Coordinator at ACS, gave opening remarks on the importance and objective of the expert group meeting. In his opening statement, he welcomed all participants to Addis Ababa and the workshop. He noted that the data management working group was one of the StatCom-Africa working groups that had not been active for a while and stated that one of the objectives of the meeting was to develop its work programme. He also underscored the need to adopt terms of reference for AGSDaM during the meeting and the importance of the Statistical Data Management (SDM) handbook to NSOs in Africa. It would serve as a guideline for selecting the right SDM solution, by documenting the features of existing platform solutions.

3.1   Handbook on SDM platforms

20.   The presentation on the handbook started with definitions covering the rationale, objectives, mode of operation and scope of work for preparing the draft. This was followed by a description of the required features of a typical SDM platform. A review of currently available SDM platforms was also presented. The presenter noted the factors affecting SDM platform selection and outlined major steps to be followed in selecting such a platform.

21.   The scope of the handbook was outlined and it was pointed out that it would only deal with the platforms for handling macro-data and would focus on the data management platforms currently being used in African countries. It was reaffirmed that the main objective of the handbook produced was to document the features of major statistical data management platforms, to guide selection and use of such systems in NSOs and other statistical offices of member States.

22.   Experts thoroughly discussed the draft handbook and made suggestions for its finalization, including:

- The SDM platform should not allow data entry without data sources; data should always be incorporated within the data source. The area, period and unit of measurement should also be included when entering the data source;

- The current SDM platforms systems should be able to work with the common existing statistical data manipulation software;

- Keeping up with the latest technology for dissemination must be a priority requirement for SDM platforms. On the other hand, considering the lack of infrastructure in Africa, the print medium must also be maintained as a required feature for data dissemination;

- SDM platforms should possess multilingual support;

- Mapping and visualization should not be considered as an option but rather as a core requirement of SDM platforms;

- No single solution could fit the requirements of all States and countries should not be forced to take a single solution as a prescription. The meeting thus recognized that Africa was diversified and that solutions should address this fact;

- Multiple specialized systems which could be integrated to offer an end-to-end solution should be considered since statistical production and dissemination has different phases and there are systems specializing in these different components;

- Strong collaboration with schools and universities should be established to build the capacity of students and researchers in using data management systems;

- Existing statistical database should be widely disseminated especially among decision-makers, public institutions, civil society and private sector, financial and technical partners as well as among researchers and students, in order to increase statistical data use;

- SDM platforms should incorporate proper data validation and quality assurance frameworks. Various validation techniques for error prevention, detection and correction should be put in place. SDM platforms should also have a facility to measure the accuracy of data, using various statistical techniques;

- SDM systems should NOT change the hardware requirements frequently, considering the constraints that NSOs in Africa have in acquiring new hardware;

- There should be standardization and harmonization of concepts, definitions, and terminologies within the metadata for SDM platforms; and

- SDM platforms developed based on open source and open software architecture should be encouraged. This insures the availability of software and its sustainability for user needs.

3.2   Current practices on SDM in selected countries and organizations

23.   A representative from DataFirst at the University of Cape Town, South Africa, and representatives from NSOs in Ghana and Nigeria gave presentations on current data management practices in their respective organizations. A representative from the Economic Community of West African States (ECOWAS) discussed the regional data management platforms. In various presentations, experts pointed out that lack of standardization and harmonization of concepts, definition and classification among platforms, and also inconsistent and conflicting data on the

same indicator among various platforms were some of the major shortcomings that were engulfing the current SDM platform.

24.   According to the presentations, platforms had been customized to suit the needs of their organizations and stakeholders. Some of the major strengths of the current SDM platforms were capability to generate various reports on data usage and platform activity level monitoring; availability of data access levels and user roles to prevent unauthorized access or illegal data usage; and availability of in-built feedback mechanisms for interaction with users. It was noted that quite intensive network infrastructures had been designed and implemented to interconnect branches of NSOs as well as other data sourcing and user organizations to facilitate and expedite data capturing, production and dissemination. Major SDM platforms such as StatBase, DevInfo, EuroTrace, Nesstar, and CountryStat were currently in use in these statistical offices.

3.3   Contributed presentations on existing SDM platforms

25.   Experts presented three existing SDM solutions, namely StatBase of UNECA, DevInfo of UNICEF and PROGNOZ of AfDB.   All three presentations focused on the capability of their respective systems in terms of the features that functioned for capture, validation, production and dissemination of statistical data. Metadata management support and standard compliance were also addressed in these presentations. Mapping features were intensively discussed as all three SDM platforms support this functionality.

26.   Participants discussed various issues with regard to these SDM platforms including:

- Level of customization requirement to use the systems;

- Micro-data support;

- Multilingual support;

- Awareness-raising strategy employed;

- Mapping and boundary creation support; and

- Data exchange standard support.

27.   The free and open source platform concept was presented by the ECA. The presenter highlighted the advantages of the architecture and how vital it was for developing new tools, and for upgrading or changing existing tools. A decentralized software evolution was occurring, through the contributions from a strong community network of software developers in Africa. The presenter noted the challenges and opportunities of the free and open source platform for developing SDM platforms and outlined the development community formation models. Participants discussed options in detail and requested further investigation of the platform.

3.4 Breakout sessions

28. Three breakout groups were formed and breakout sessions were conducted in parallel to discuss the following topics:

Group 1: Data capturing and data exchange, metadata, constraints/opportunities in adopting SDMs

29. This group discussed different aspects of data capturing and exchange as well as the importance of metadata management. Accordingly, the group reached consensus on the following items:

- A good data management system should allow validation at the capturing stage and keep a history of data changes;

- Data exchange should not compromise data security. Standardization and harmonization of concepts is crucial;

- Data distributors need to be able to produce hard copies as well as digital products;

- Existing international metadata standards should have the basic information within the metadata including the definition, description and meaning. It was necessary to have the ability to transpose the major existing international standards into statistical data management systems; and

- The constraints with regard to data capturing and exchange should be acknowledged and addressed, including cost of SDM implementation, high turnover of skilled personnel in data- producing organizations, and lack of skilled human resources, proper documentation, inter- linkage of departments, standardization guidelines and procedures, linkages with other data producers and systems and of communication tools.

Group 2: Data production and validation, and open platforms

30. Participants stressed that guidelines and procedures were crucial components that helped to assure quality data production and validation. SDM platforms also needed built-in features to validate data at different stages and clear implementation of standard quality frameworks. Group members deliberated in detail on the free and open source platform, which encompassed a type of software architecture that allows for adding, upgrading and swapping components/modules. SDM systems using the open software architecture could benefit from the decentralized software evolution, with different parties contributing in developing new tools and upgrading or changing existing tools.

31.   The concept of the free and open source platform was well discussed and its advantages highly appreciated by all the participants. Various implementation strategies deliberated included the following points:

- Free user community contribution: individuals or organizations could develop open source software and on a voluntary basis contribute components to be plugged into it;

- Hire software companies: the organization responsible for developing the open software could hire software companies to develop the components based on the specific requirements of end users;

- Development partner support: development of components could be sponsored by partners to meet end-user requirements. Sponsors could hire software development companies for component development; and

- Software companies selling components: they would develop the components and make them available for sale to end users.

Group 3: Data Dissemination, mapping and visualization

32.   Group 3 agreed that dissemination was a key stage in data management and recommended that data dissemination choices be made with the end user in mind. Participants reached consensus that dissemination encompasses the systems and methods we use to make the data available for use. Hard and soft-copy formats and online production were important as data users in African countries did not all have the same level of access to technology.

33.   The following challenges to data dissemination were noted:

- Lack of training in data dissemination tools;

- Lack of distribution channels within organizations and for public dissemination;

- Lack of means to keeping data current;

- Lack of data for some countries (this was seen as an issue for the regional organizations especially). In some areas this was due to lack of coordination within the national statistical systems;

- Validation of data before distribution– data quality assurance;

- Data harmonization constraints;

- Lack of policies on data dissemination – if there were no policies there could be no permissions and without permissions the sharing of country data was not possible;

- Lack of links from one data disseminator to another; and

- The attitude that what is important is maximum features with minimal training requirements and minimal cost, preferably free.

34. It was also observed that:

- User management should be an integral part of data dissemination when sharing micro-data but be placed in the public domain when sharing general socio-economic macro-data;

- User subscription services – when systems notify key groups once new data are available - were a useful but not necessary feature;

- Maps and graphs should be used in place of words and tables in data interfaces, particularly where the data are being presented to students or government decision-makers. Mapping of data was important because regional disparities in resources could easily be identified in this manner and their correction fed into government policy and support to regional development; and

- There were specific mapping programmes for improving use. Therefore, a mapping facility or at least data management software compatible with GIS software such as ArcGIS should be a requirement for a data management system. Good mapping of government data would require cooperation between the government departments responsible for data collection and mapping.

3.5   Recommendations

35.   In addition to the recommendations regarding the SDM platforms, the following recommendations were also made by participants:

- As the mission of ECA was to support African countries for their economic and social development and regional integration, availability of up-to-date and accurate statistical data was critical. It was recommended that ACS, which is in charge of statistics at ECA, should elaborate its implementation plan for the SDM platform (StatBase) and present it to African member States;

- NSOs should improve collaboration with institutions involved in producing statistical databases, to maintain complete and quality data and to have objectively verifiable indicator definitions;

- African countries should elaborate statistical data management and dissemination policies; and

- In order to strengthen the Ad Hoc Committee, a representative from Africa Statistical Association (AfSA) should be included.

4. Support to Member States and Regional Economic Communities: Implementation of National Subregional Databases

36.   ECA has facilitated member States and RECs in dissemination of statistics through support to development of national and subregional statistical databases. As a result, ECA staff have undertaken four data management missions to Yaoundé, Rabat, Libreville and Khartoum to conduct training in use of StatBase, a statistical data management platform developed by ECA. The selected staff trained in the use and management of the StatBase statistical data management platform came from NSOs in Cameroon, Gabon, Morocco, the Sudan, ECA Subregional Offices (SROs) for Central Africa, and for North Africa, as well as from AMU and the Economic Community of Central African States (ECAS).

5. Joint African Statistical Yearbooks

37.   Following the recommendation of the African Statistical Coordination Committee, ECA, AfDB, and AUC continued collaboration in production of the joint African Statistical Yearbook (ASYB). The 2010 and 2011 issues have been produced and the 2012 issue is currently under production. The contents of the yearbooks highlighted contemporary issues such as the MDGs, environment statistics and the Information Society. Joint production and publishing by the three institutions has resulted in reduced data inconsistency and has greatly reduced the burden of country responses to data requests.

6. African Integrated Census Microdata

38.   In collaboration with the Minnesota Population Center, ACS has developed a web portal to disseminate anonymized and high-precision African Census Samples representing 10 African countries. The African Integrated Census Microdata (AICMD) portal was launched at the Africa Symposium for Statistical Development in November 2010. The portal can be accessed at Http://ecastats.uneca.org/aicmd. The AICMD samples hold records of 51,524,324 persons (as of August 2011) and are freely available to researchers and policymakers at no cost by simply registering and agreeing to the conditions of use in place to protect statistical confidentiality and ensure sharing of research results from use of the samples. The portal also includes over 500 original source documents—census forms and instructions manuals that are also freely available for viewing and downloading from the AICMD searchable metabase.

The commission is invited to consider the following actions:

- Endorse the handbook on statistical data management platforms produced by experts in Africa

- Request ECA and partners to study data flow within the national statistical system to devise a an optimal mechanism in a country

- Study and recommend an optimal statistical information system at national statistical office to link the information flow in various units of NSO

- Request the working group to advocate and build capacity in the use of geo-spatial tools in census and survey undertaking

- Request the working group to coordinate the data collection attempts by regional, sub-regional organization to reduce response burden of countries.

- Request ECA and partners to develop an open-source platform for data management and dissemination which will be collaboratively developed by experts in Africa and partners in statistics and owned by African NSO