49th Session of the Commission on Population and Development

11-15 April 2016 United Nations, New York

Strengthening the demographic evidence base for the post-2015 development agenda

www.unpopulation.org #CPD49 #UNPopulation

The Data Revolution in Action: National and International Experiences with Microdata Dissemination and Public Use

Monday, 11 April 2016, at 1.15-2.30 p.m.
Conference Room 4 - Conference Building
United Nations Headquarters, New York

The Population Division of the Department of Economic and Social Affairs (DESA) would like to invite you to a side-event on "**The Data Revolution in Action: National and International Experiences with Microdata Dissemination and Public Use**" on the occasion of the 49th Session of the Commission on Population and Development (CPD).

The objective of this side event is to provide several examples of leading national and international efforts with the dissemination and public use of census, survey and administrative microdata sets for demographic analysis, comparative research, development planning and decision making. In the context of increasing demands for the demographic evidence base for the 2030 Agenda for Sustainable Development, the panelists will share their experience with data archiving, documentation, harmonization, dissemination and use by various types of users to improve evidence-based policy and programme formulation, national or international analytical studies and new innovative ways to use existing data.

Panelists:

- **Pali Lehohla** (Statistician-General of Statistics South Africa) about the experience of Statistics corrections
- **Matthew Sobek** (Research Scientist, Director of Data Integration, Minnesota Population Center, University of Minnesota) about census data in Integrated Public Use Microdata Series (IPUMS)-International, survey data in Integrated Demographic and Health Surveys (iDHS) and geospatial information integrated with census data in TerraPopulus
- **Turgay Unalan** (Household Survey Specialist, MICS team, UNICEF) about the UNICEF Multiple Indicator Survey (MICS) program and experience with dissemination and use of MICS data for secondary data analysis
- **Olivier Dupriez** (Lead Statistician, IHSN coordinator, Development Data Group, World Bank) about the Integrated Household Survey Network (IHSN) coordination experience, microdata dissemination toolkits and National Data Archives (NADA), and IHSN online data navigator

Chair/Moderator: Mr. Patrick Gerland, Population Division, UNDESA

About the panelists

Pali Jobo Lehohla is the Statistician-General of South Africa, a position he held since 2000. He has served as co-chair of PARIS21 and the Chair of the United Nations Statistics Commission. He was the founding chair of the Statistics Commission of Africa (StatCom Africa) and chairs the African Symposium for Statistical Development (ASSD). He was the Vice President of the International Statistics Institute (ISI), and sponsors the Young African Statistician (YAS) movement. He served as one of on the twenty five member panel on Data Revolution appointed by the UN Secretary General, and has recently been appointed to the Independent Accountability Panel for the health of women, children and adolescents. Mr. Lehohla has been a forceful advocate for improving the Civil Registration and Vital Statistics systems in Africa. He was recognized by his alma mater, the University of Ghana for his contribution to the development of statistics in 2015 and was also awarded an Honorary Doctorate by the University of Stellenbosch in the same year.

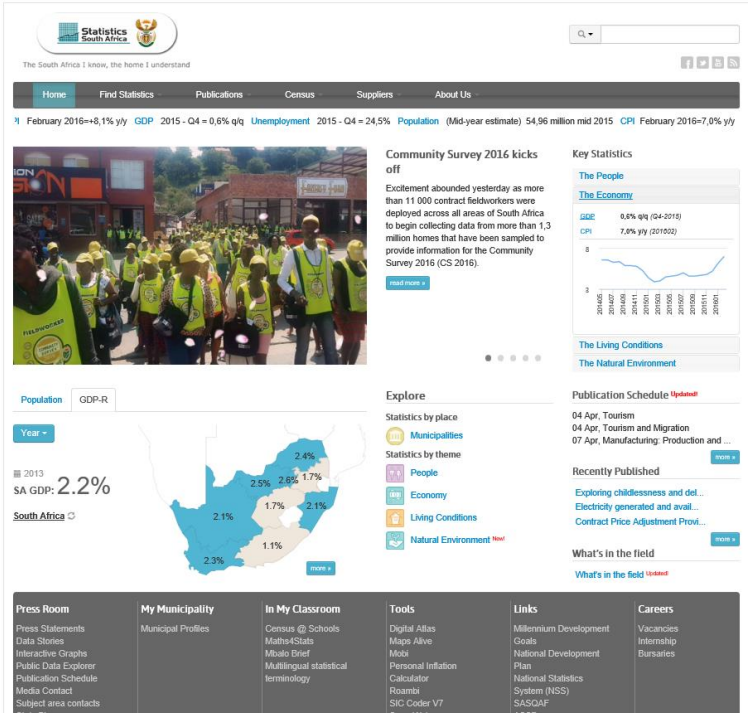
Matt Sobek is the Director of Data Integration at the Minnesota Population Center at the University of Minnesota. For twenty years he has led the development of data infrastructure projects that deliver harmonized microdata to users worldwide free of cost. In the 1990s he co-authored the original IPUMS project with Steven Ruggles. The project combined all existing U.S. public-use census microdata into one consistently coded series and provided a novel web dissemination system to browse and extract data. Joining with Bob McCaa in the 2000s, Mr. Sobek directed the IPUMS-International project, which over the next decade became the world's largest population database. More recently, Mr. Sobek has been the technical lead on the Integrated Demographic and Health Surveys project, which harmonizes that invaluable but sprawling survey data collection to facilitate comparative analysis for sub-Saharan Africa. In addition to varying levels of involvement on a number of other data infrastructure projects, Mr. Sobek was the managing editor and major contributor to the five-volume reference work, the Millennial Edition of the Historical Statistics of the United States.

Turgay Unalan as Household Survey Expert in UNICEF's Data Collection Unit (also known as the MICS Team) provides technical support to MICS implementation globally as well as supporting the development and improvement of survey tools since 2008. Before joining UNICEF, Mr. Unalan worked at Hacettepe University Institute of Population Studies (Ankara, Turkey) as a lecturer and researcher, and was involved in numerous national level household surveys including Demographic and Health Surveys during 1983-2008, and provided support to the 3rd round of MICS as consultant to UNICEF regional offices and Headquarters to support the regional workshops, report writing activities and review of completed summary and final reports.

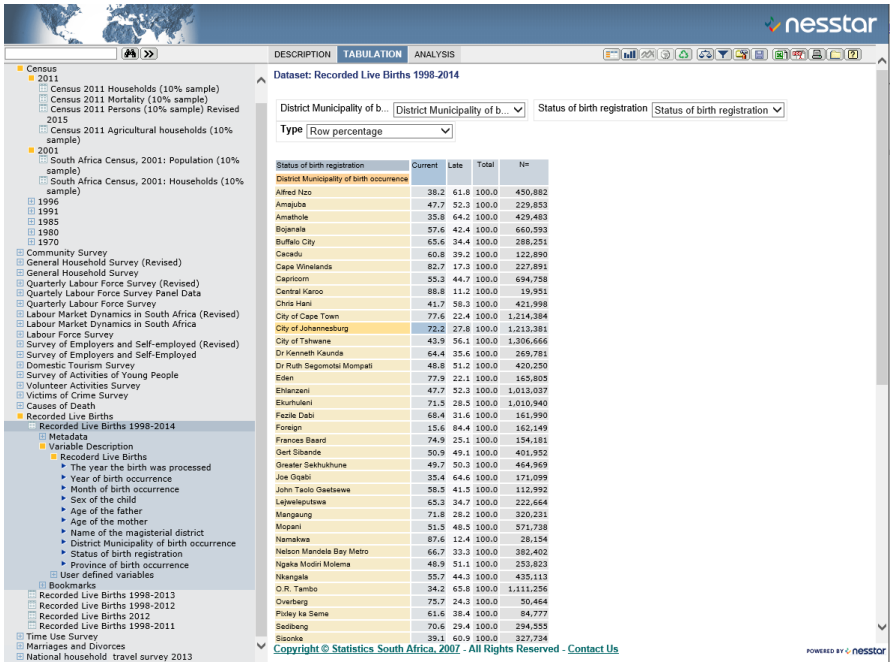
Olivier Dupriez as lead statistician at the World Bank is a co-manager of the World Bank's Survey Unit, and the coordinator of the innovation and research program of the Bank's Development Data Group. He also coordinates the activities of the International Household Survey Network (IHSN). Mr.Dupriez is a demographer, whose work focuses on issues related to harmonization of data collection practices, and on microdata curation, dissemination and use. Before joining the World Bank in 2004, Mr. Dupriez worked with the United Nations as a population census expert in Cape Verde and Chad, and with the Asian Development Bank as a poverty statistician.

To know more about **Statistics South Africa and its experience with microdata dissemination and public use**

http://www.statssa.gov.za



Experience with Nesstar (http://interactive.statssa.gov.za:8282/webview/)



Additional resources:

- South African Data Archive: http://sada.nrf.ac.za/
- DataFirst: Research Unit and Data Service based at the University of Cape Town, South Africa. https://www.datafirst.uct.ac.za/

To know more about **IPUMS International, iDHS and TerraPopulus**

The Minnesota Population Center (MPC) specializes in the development of data infrastructure for research. The Center's signature activity is the integration of population microdata: converting datasets collected at different times and places into a consistent series with common codes across variables and unified documentation. A web dissemination system enables browsing of the data and creation of customized microdata extracts for downloading and analysis. The **IPUMS-International project** applies this model to international census data, in collaboration with roughly 100 of the world's national statistical offices. With 600 million person records from 277 censuses over the last half century, IPUMS has enabled previously impossible large-scale comparative research on topics such as marriage, fertility, migration and household structure. Using the web dissemination system, researchers can construct a single dataset containing hundreds of millions of person records spanning countries and time periods. IPUMS aims to democratize access for all researchers, and several thousand analysts from more than 100 countries have used the data so far. But IPUMS has an even more fundamental interest in seeing that these irreplaceable data are preserved for the future. The project seeks out data at risk of destruction, assists in recovering them from old media, and serves as a permanent archive for them. For further info: http://international.ipums.org

Over the past five years MPC has been engaged in a project to create an IPUMS-style dissemination system for the Demographic and Health Surveys, one of the most important data sources for the developing world. The first phase of the **Integrated DHS project** focuses on sub-Saharan Africa, and we hope to expand the coverage to North Africa and Southern Asia soon. While access and preservation were the key challenges for census data, the main challenges of the DHS are data discovery and logistics. Prior to IDHS, conducting comparative DHS research could require managing hundreds of files and many thousands of variables, with no central mechanism for exploring the detailed contents of the samples. It was difficult to realize the full potential of these invaluable data without costly and error-prone labor. As with IPUMS-International, the Integrated DHS harmonize all the variables, documents comparability issues, and provides a web system for data browsing and creation of cross-national extracts for downloading. Because the original data agreements with the countries limit dissemination rights to The DHS Program, which manages the survey for USAID, we installed our web system on the DHS servers. The DHS Program manages all user registration and has been extremely cooperative at all stages. For further information: https://www.idhsdata.org/idhs/

A third MPC international project, **Terra Populus**, is in technical terms the most ambitious. It aims to combine population and environmental data in one dissemination system. The system will transform population microdata, summary area data, and raster data (grid cells) into whichever data format is desired by the user. For example, a microdata researcher can summarize rainfall data for the administrative units for a country and attach the results as newly constructed variables on the person records in their data extract. Researchers therefore need not be skilled in all the different data formats to utilize data from domains outside their expertise. The key for combining the data is harmonized geography, and TerraPop has engaged in large scale map acquisition and digitization, particularly of historical geographies. Some key features of the project are still under development, but the full system should be operative in summer 2016. For further information: http://www.terrapop.org/

To know more about the **UNICEF MICS program and activities**

Since its inception in 1995, the Multiple Indicator Cluster Surveys, known as MICS, has become the largest source of statistically sound and internationally comparable data on women and children worldwide. MICS was a major source of data on the Millennium Development Goals (MDG) indicators and will continue to be a major data source during the 2030 Sustainable Development Agenda to measure Sustainable Development Goals (SDGs) indicators.

The information obtained through MICS surveys – on topics ranging from maternal and child health, education and child mortality to child protection, HIV/AIDS and water and sanitation – is fundamental to sound decision-making and advocacy.

Standard MICS questionnaires are customized by implementing agencies, based on an assessment of a country's data gaps and needs. All survey activities, from fieldwork to report writing are carried out by the implementing agencies – with continuous technical support from UNICEF.

MICS surveys are designed to be representative. The average sample size in the 5th round is around 12,000 households, although it varies greatly from one survey to the other. Interviewers administer the household questionnaire as well as individual questionnaires to women and men aged 15 to 49 years and to mothers or caretakers of all children under 5 years of age. The number of topics covered has increased substantially over the years as demand for data has grown. In the 5th round, MICS is providing data on more than 130 internationally agreed-upon indicators.

In addition to collecting information on intervention coverage, MICS also explores knowledge of and attitudes to certain topics, and specific behaviors of women, men and children, enabling analysts to gain insights into behaviours that may affect women's and children's lives. MICS routinely disaggregates data so that disparities associated with age, gender, education, wealth, location of residence, ethnicity and other characteristics are revealed.

UNICEF provides technical support for MICS surveys every step of the way. Central to capacity strengthening efforts are a series of three regional workshops – on Survey Design, Data Processing, and finally, Data Interpretation, Further Analysis and Dissemination - at critical stages of the MICS process. Effective on-site support is provided by country MICS coordinators and UNICEF focal points, backed up by technical assistance at the regional and global levels. High-quality data are obtained thanks to thorough and tested field procedures combined with rigorous data verification.

Over the years, MICS has pioneered the development of new measurement tools in areas including early childhood development, child discipline, hand washing, post-natal health care and low birthweight. Many of these tools have been adopted by other international survey programmes. Development of a number of new modules and questionnaires (social protection, child disability, victimization, post-emergency assessment, rapid water quality testing and learning assessment) are underway, many of which are intended to capture new indicators pertinent to the Sustainable Development Goals (SDGs).

To know more about UNICEF MICS program, data and tools:
http://mics.unicef.org

To know more about the **International Household Survey Network (IHSN)**

The International Household Survey Network (IHSN) is an informal network of international agencies. The mission of the IHSN is to improve the availability, accessibility, and quality of survey data within developing countries, and to encourage the analysis and use of this data by national and international development decision makers, the research community, and other stakeholders.

To support this mission, the key objectives of the IHSN are:

- Improved coordination of internationally sponsored survey programs, with emphasis on timing, sequencing, frequency, and cost-effectiveness
- Availability of coordinated and practical technical and methodological guidelines for all stages of the survey life cycle
- Availability of a central survey data catalog which would inform data users of the availability of survey and census data from multiple sources
- Availability of standards, tools, and guidelines that would allow data producers to document, disseminate, and preserve microdata according to international standards and best practices
- Improved collaboration between data producers and users

For further information about IHSN resources:

- International Household Survey Network: http://www.ihsn.org
- Accelerated Data Program for country support: http://adp.ihsn.org
- IHSN Guidelines and best practices for data archiving and dissemination: http://www.ihsn.org/home/node/115
- IHSN Microdata Cataloging Tool (NADA) for web-based portal for researchers to browse, search, compare, apply for access, and download relevant census or survey information. http://www.ihsn.org/home/software/nada
- IHSN inventory of national and international micro-datasets public archives (NADA): *86 repositories by national statistical offices and research organizations in more than 60 countries or areas* as of March 2016: http://adp.ihsn.org/survey-catalogs

- IHSN Survey catalog: inventory of national household surveys http://catalog.ihsn.org/index.php/catalog
- IHSN Gender Data Navigator: searchable inventory of gender-related questions found in survey and census questionnaires from low- and middle-income countries. Available from http://www.ihsn.org/home/gender-data-navigator
- IHSN DDI Metadata Editor (Nesstar Publisher 4.0.9), a feature-rich editor for the preparation of metadata and data for publishing in an online catalog, such as the IHSN-developed National Data Archive (NADA). The metadata produced by the Editor is compliant with the Data Documentation Initiative (DDI) 2.n and the Dublin Core XML metadata standards. http://www.ihsn.org/home/software/ddi-metadata-editor
- Statistical Disclosure Control (SDCMicro), a free, R-based open-source package for the generation of protected microdata for researchers and public use. Available from http://www.ihsn.org/home/software/disclosure-control-toolbox