Statistical Commission
Forty-eighth session
7 – 10 March 2017
Item 3(d) of the provisional agenda
Items for discussion and decision:
Big Data for official statistics

Background document
Available in English only

Proceedings of the Third International Conference on Big Data for Official Statistics

Prepared by the Global Working Group on Big Data for official statistics

Proceedings of the Third International Conference on Big Data for Official Statistics

30 August – 1 September, 2016, Dublin, Ireland

Introduction

The Third International Conference on Big Data for Official Statistics took place in Dublin, Ireland, from 30 August until 1 September 2016. The event was hosted by CSO Ireland, which had secured the beautiful and prestigious Dublin Castle as the venue for this event. The UN Global Working Group (GWG) on Big Data was responsible for the conference programme. More than 200 participants attended coming not only from the statistical community, but also from private sector, civil society, research and academia.

After the conferences in Beijing in 2014 and Abu Dhabi in 2015, this conference had set as objective to take the next steps in the utilization of Big Data in the production of official statistics. The GWG had selected three major themes namely (1) Access to proprietary data and successful partnerships with data providers, (2) Capacity building strategies related to the use of Big Data in the statistical production process, and (3) using Big Data in the compilation of Sustainable Development Goal (SDG) indicators.

The statistical community needs to modernize and strengthen global, regional and national statistical systems to respond more rapidly, effectively and efficiently to the new requirements and challenges including monitoring of the SDGs. This means among others that statistical offices need innovations to incorporate non-traditional sources of data, such as Big Data, which thus far have been underutilized in producing official statistics. These data sources can be leveraged to support the compilation of SDG indicators, so that timely and disaggregated data can be produced and made available to policy makers.

The following gives an overview of the various sessions during the three-day event. This overview is not intended to be comprehensive. It should provide nevertheless a good insight in not only who presented, but more importantly what was presented and discussed.

Day 1 – Access and Partnerships

Tuesday, 30 August 2016

A. Ceremonial Opening

Moderator – Mr. Padraig Dalton, Director-General, CSO Ireland

- Mr. Eoghan Murphy, Minister of State for Financial Services, eGovernment and Public Procurement, Ireland

- Mr. Christian Friis Bach, United Nations Under-Secretary General and Executive Secretary of the United Nations Economic Commission for Europe
- Mr. Lenni Montiel, United Nations Assistant Secretary General for Economic Development (Video Message)

Mr. Eoghan Murphy, Minister of State, reflected on the new ICT environment that has emerged globally including in Ireland in the availability and use of mobile devices and social media. This new environment provides opportunities to create win-win partnerships between the ICT sector and the national statistics system. In particular, he stressed the critical role for the NSIs in collaborating with the ICT partners to harness Big Data in the statistical operations of collection, processing and dissemination of official.

Mr. Christian Friis Bach, USG and Executive Secretary of the UNECE, highlighted that the new technological revolution in the ICT sector has brought us the digitized economy and has changed the way we live, work and make decisions. It has opened up the opportunity to take better and data driven decisions real time allowing us to take early policy interventions in societal, economic and environmental developments. Apart from harnessing the benefits of the data tsunami, he pointed out that we should also manage the downside risk in terms of personal profiling and stigmatization. In order to contain and address the downside risk, we should review and where necessary strengthen our institutional legal framework with respect to confidentiality, privacy and access to data. It would also be necessary to be transparent to the general public and businesses about our scientific methods and operations in the use of the Big Data and create on ongoing debate with the representatives of public and private sector in the use of Big Data. He also indicated that the new data is more granular and detailed in terms of geography, age, sex, minorities, etc., which will lend itself to inform the sustainable development trajectory promoted by the United Nations in the 2030 Agenda for Sustainable Development. The monitoring of the 2030 agenda through the SDG indicators should use these new data sources to inform the progress in in the 17 Goal areas covering poverty eradication, good health and well-being, quality education, climate action, life on earth, or gender equality, while respecting the data as a human rights issue: right to be counted, the ownership over personal data and its use, and the access to quality data for daily decision making.

Mr. Lenni Montiel in his video message also focused his address on the use of Big Data for the monitoring and reporting of the SGD indicators and encouraged the statistical community through the GWG on Big Data for official statistics to work constructively in mutually beneficial public and private partnerships in developing the innovative techniques in multi-source statistical production processes.

B. Introductory remarks by the organizers

- Mr. Padraig Dalton, Director General, CSO Ireland
- Ms. Lisa Bersales, National Statistician, Philippines Statistical Authority
- Mr. Ronald Jansen, Assistant Director, United Nations Statistics Division

The organizers in their introductory remarks stressed the role of the NSIs in harnessing the data deluge for their daily production and dissemination of timely and high quality statistics to allow

for data drive decision making. In the application of the new data, they pointed to the need to maintain trust of and be accountable to the general public. This trust and accountability should be based on full transparency of our methods and standards as reflected in our statistical laws and statistical quality framework. The GWG on Big data should play a critical role in addressing the issues of a) access to Big Data held by the private sector, b) partnership in creating the technological infrastructure, data services and applications, c) in providing capacity building and innovation to the production of official statistics and d) the use of Big Data for monitoring and reporting of the SDG indicators for the 2030 Agenda for Sustainable Development.

C. Keynote address by Mr. Ronan Harris, CEO, Google, Ireland

Mr. Ronan Harris called upon the statistical community to make the cultural change and enter in partnership with the private sector and academia to apply the Big Data in their statistical production processes. He also challenged the statistical community to adopt modern dissemination and visualization practices bringing the knowledge and information for data driven decision making. He further elaborated that the statistical community should embrace the analysis and interpretation of statistics in addition to the production of statistics as part of their products and services delivery to stay relevant. In his closing remarks, he challenged the statistical community whether it has the capability to think BIG and think out of the BOX to stay relevant in terms of the use of new technological infrastructure, new data, and data analytics through innovative public- private partnership relationships.

D. High-Level Panel on Access and Partnership

Moderator – Mr. Padraig Dalton, Director-General, CSO Ireland

Panel

- Mr. Cathal Ryan, Deputy Data Protection Commissioner, Ireland
- Mr. Jan Sonk, Proximus, Belgium
- Mr. Niels Ploug, Statistics Denmark
- Mr. Cosmas Zavazava, ITU
- Mr. Christian Zahorski-Philippe, Amazon Web Services

The panelists all stressed that access to Big Data by the National Statistics Offices, be it mobile phone data, social media, remote sensing information, etc. and working in public-private partnership will bring enormous value to both the private partners and the NSOs in their service delivery. These benefits should build on the technological cloud computing infrastructure and data analytics expertise of the private sector and the trust of the public in the entrusted independent and professional role of NSIs to provide impartial high quality information for daily decision making. Critical for the partnership is that we can formulate the specific statistical outcomes/outputs that the NSIs need to integrate Big Data in the production of official statistics. It was stressed again that the public-private partnership in harnessing Big Data for official statistics should be based on mutual trust.

With trust of the general public, observing full transparency in the collection, storage and use of basic data in accordance with the regulatory and legal frameworks for official statistics and data protection, the enormous value of the use of Big Data for data driven decision making should be unleashed. Successful use of Big Data should be built on transparency, empowerment, privacy by design, and accountability.

E. Best Practices and Partnership

Moderator – Mr. Peter Struijs, Statistics Netherlands

Panel

- Ms. Frances Krsinich, Statistics New Zealand
- Mr. Daniel Quintart, EU GROW
- Mr. Ahmed Almufarji, National Center of Statistics and Informatics, Oman
- Mr. Marc Debusschere, Statistics Belgium.

This session provided excellent insight in successful and potential applications of the use of Big Data for producing timelier and better quality official statistics. Statistics New Zealand uses consumer electronics scanner data, scanner data from supermarkets and web-scraped online price data to estimate the Consumer Price Index using new methodology in which implicit quality adjustments can be made because of the huge volume of transactions. This Big Data project has come to full fruition through collaborative arrangements with private sector and other NSIs. In the partnership of Statistics Belgium with the mobile phone operator Proximus it was seen as very important that the needs for a partnership are mutual, complementary and non-competing. Proximus brought the mobile data, metadata, IT infrastructure and technical expertise into the partnership, whereas Statistics Belgium contributed statistical, methodological and domain expertise, as well as the quality stamp of official data.

Overall, in this session a clear message was provided that using Big Data, being it mobile phone data and geospatial data, will bring value and benefits by a) reducing response burden, b) reducing costs, c) improving the timeliness of statistics and d) improving the quality of statistics. At the same time, the emerging partnerships with the private data holders require further testing of definitions and concepts of official statistics and their transformation in official statistics, the application of new data analytics, the need for new IT infrastructure to store and processes the large data sets and the review of the legal requirements for their access. Moreover, the official statistical community has to recognize that the private sector data holders are profit oriented with a limited social responsibility obligation. Building on these best practices, the statistical community should now draft a new business model for long term public-private partnerships with joint exploration and exploitation and resulting in distinct non-competing outcomes.

F. Best Practices on Partnerships and Data Management Solutions

Moderator – Mr. Albrecht Wirthmann, Eurostat

Panel

- Ms. Hélène Bérard, Statistics Canada,
- Mr. Joe Drumgoole, MongoDB
- Mr. Mark Rowlands, Amazon Web Services

Ms. Bérard presented the latest organizational development at Statistics Canada in adopting a corporate approach to the use of administrative data. A specialized internal service department/division has been created that has centralized a) the acquisition of administrative data from other government departments, b) undertakes on behalf of the other departments of statistics Canada the analysis of new data and c) explores new partnerships with data holders in creating win-win situations. Mr. Drumgoole made the case that the national statistical community will move into a new era of a) open source data away from propriety data, b) of cloud computing away from data center computing and c) of technological infrastructure that can take on processing peaks of petabytes as compared to terabytes. This new era will witness new ways of a) data capture and storage, b) data analytics and c) integration and the creation of new uses and new services. Mr. Rowlands confirmed these developments in the extension of products and services of Amazon Web Services in a secure data environment of cloud computing.

Day 2 – Capacity Building for Innovation in Official Statistics

Wednesday, 31 August 2016

G. Keynote address by Ms. Heather Savory, Deputy National Statistician, Office of National Statistics, United Kingdom

Ms. Savory in her key note addressed the need for the adoption of a multi-facetted corporate approach to the transformation of the National Statistical Office of the United Kingdom to stay relevant in our fast changing data landscape. Business as usual in delivering statistical service to the national and international clients is no longer tenable given the need to manage the downside risk of the external initiatives, competition of other providers of statistics and budget cuts, but also reap the benefits of the changing ICT and data landscape through rapid technological development and data revolution. Transformation being multifaceted covers a) transforming the technology architecture, b) transforming the workforce, c) exploiting the data sources and d) redirecting our products and services. In the new technology infrastructure a unified approach has to be taken with the introduction of the cloud environment and moving away from physical IT equipment on site, the adoption of a common software architecture with common applications for processing, storage, security, etc., the creation of centralized data center, and the creation of a cloud based foundry with distinct platforms for data collection, registers and metadata, data management by economic, social demographic and environmental domains, analysis and research and dissemination. The transformation of the workforce is about creating

multi-disciplinary product teams that are pooled from the different specialized department and include methodologist, IT and software experts, and subject matter statisticians and develop the products in an agile project environment. Exploiting data sources means linking to government data programmes with the rational to modernize the use of administrative, big data and private sector data, fully adopt electronic collection and integrate with the government digitized programme (e.g. open data) and government programme on national data infrastructure and registers. An essential element in redirecting our products and services is to better understand the needs of our clients and working in closer partnership with the private commercial sector, academia, international organisations and privacy groups. An essential element in the transformation is taking the next step in launching a data campus to strengthen the data capability and the research and analytical capabilities. To be launched in September 2017, the campus will develop the next generation of data analysts equipped with the skills to meet the future needs of ONS and UK society as a whole. The campus will also be a world-class center for data science and data engineering, bringing together analysts, data scientists and technologist that will address specifically five research areas: modern economy, sustainability, urban future, UK in the global context and society.

H. High-Level Panel – Innovation and Statistics

Moderator – Mr. Bert Kroese, Deputy Director-General and CIO, Statistics Netherlands

Panel

- Ms. Heather Savory, Deputy National Statistician, ONS, UK
- Ms. Sylvie Michaud, Deputy Chief Statistician, Statistics Canada
- Mr. Padraig Dalton, Director-General, CSO Ireland
- Ms. Elham Saleh, Statistical Centre of the Gulf Cooperation Council
- Mr. Aurel Schubert, Director-General, Statistics Department, European Central Bank
- Mr. Georgy Oksenoyt, Deputy Director-General, Federal State Statistics Service, Russian Federation

The panel addressed the following questions: Why should we use Big Data? What capabilities do we need? How can we build those capabilities? What is the role of the international organisations in creating these capabilities?

The panel agreed that the use of Big Data provides us better quality statistics with faster access to more data. The needed capabilities are multifaceted ranging from IT skills in data science and data engineering, strengthening our legal statistical environment, adopting new statistical methods and adopting a leadership and corporate culture. The capabilities should allow us to adopt a standardized corporate business architecture that is flexible and adaptable to emerging demands and process-based rather than product-based with an increasing use of multiple secondary data sources (away from primary data sources) for multiple statistical outputs. Also our dissemination and communication strategy should target specific users and increasing use of visualizations in dissemination. The capabilities are created through experimentation from data-based methods to model-based methods, introduction of a culture of innovation through performance management (top down) and internal innovation competitions (bottom up).

Moreover the capabilities are created through by operating in a shared national and international IT platforms through collaborative projects with agile multidisciplinary teams. These teams should be based on strategic partnership network between national and international statistical agencies, academia and private commercial sector to reduce the development time from design to final product. The international statistical agencies should play an increasing role in shaping the strategic partnership network and creating effective and efficient platform for collaboration in common projects with the objective of developing common services and application based on standardized business architecture.

I. The role of statistical training institutes

Moderator – Mr. Sufyan Al Barghouti, Federal Competiveness and Statistics Agency, UAE

Panel

- Mr. Oliver Chinganya, Director, African Centre for Statistics, UNECA
- Prof. Innocent Ngalinda, East African Statistical Training Center
- Mr. Albrecht Wirthmann, European Statistical Training Programme, Eurostat
- Mr. Ivo Havinga, Assistant Director, United Nations Statistics Division

The panel indicated that regional statistical training institutes have to address the use of Big Data in the statistical production process through multifaceted training programs. These training programs should address the senior management of the NSIs (in areas of organization and management of the transformation of the national statistical systems to stay relevant), and the technical and IT staff (to create the data engineering, data analytics and data science capabilities). The training institutes should make use of common IT platforms with Big Data processing capacity and include e-learning methods for capacity building. This session led to the request to create an UN-led international network of regional training institutes that would provide a continuous learning environment in new methods, new management and IT techniques and best practices that complement the existing national training institutes, and national data and research centers. Such a collaborative partnership could offer a cost effective solution for continuous training of the next generation of national and international statisticians in common standards and methods in an agreed statistical corporate architecture.

J. Smart investment in innovation in official statistics

Moderator - Mr. Ronald Jansen, Assistant Director, United Nations Statistics Division

Panel

- Mr. Grant Cameroon, World Bank
- Mr. Fessou Lawson, African Development Bank
- Mr. Seb Mhatre, DFID, UK
- Mr. Johannes Jutting, PARIS21

The panel indicated that the rapid changing technological and data landscape require a rethink of the support of the donor community to the national, regional and international statistical systems

with a focus on the strengthening of the capabilities of the NSOs in developing countries. Initial considerations in this fast changing environment is to focus the smart investments in innovation of official statistics on small projects that pilot, test and possibly scale new innovative methods and technologies. These emerging projects and programmes require new collaborative and strategic partnership with academia and private sector at regional and international level that have the capability to scale and take advantage of digital technologies. There seems to be also a shift necessary to funding of recurrent cost of having continuous access to private big data sources and modern cloud based IT infrastructure. Moreover, it seems to require a change in donor perceptions from upfront funding of investments in national statistical systems to investments in common regional and international platforms for access to data, services and applications through strategic partnerships. In developing the funding strategies in smart innovative investments, a broad range of challenges have to considered, including statistical, technical, ethical, commercial, and skill requirements.

K. Innovative approaches to capacity building in the use of Big Data

Moderator – Mr. Steven Vale, UNECE

Panel

- Mr. Niall Wilson, Irish Centre for High-End Computing (ICHEC)
- Mr. Antonio Virgillito, Statistical Office of Italy (ISTAT)
- Mr. Amjad Zaim, Cognitro Analytics

The Irish Centre for High-End Computing (ICHEC) maintains the Sandbox, which is being used by the national statistical offices, which are part of the UNECE-led Big Data team under the high-level group on modernisation of official statistics. The Sandbox provides large data storage capability, high speed connectivity and state-of-the-art Big Data software tools. Collaboration on Big Data, services and applications is possible in this environment. ISTAT showed that the Sandbox can be used as an effective platform for supporting training courses. It runs special software for high performance computing which cannot be installed or run on standard computers. Non-confidential demonstration datasets can be uploaded and shared, and software and the datasets are available from everywhere only with a browser. Cognitro Analytics offers services to evaluate an organization regarding its readiness and gaps to run a Big Data project or bring Big Data into the production process. On basis of such evaluation, necessary training programs are then formulated to build the missing capacity.

L. Sandbox Project on Social Media Data for Sentiment Analysis and Mobility

Moderator – Mr. Steven Vale, UNECE

Panel

- Mr. Juan Munoz, Statistical Office of Mexico (INEGI)

INEGI shared its experience on the use of Twitter as a Big Data source and regarding its collaboration in the Sandbox project. The objective of INEGI's Big Data Project was to

generate experimental indicators for subjective wellbeing from social media data sources. INEGI was able to make mobility analysis, based on the location and time of the tweets. The project used supervised training with support from 5000 students, who each tagged 100 tweets, where each tweet was tagged about nine times. The applications are quite varied from tourism and migration statistics, to mobility and use of roads statistics, to opinion polling and subjective well-being sentiments.

M. Crop Statistics using Satellite Data

Moderator – Mr. Sybille McKeown, Australian Bureau of Statistics

Panel

- Mr. James McBroom, Queensland University, Australia

Mr. McBroom presented some analytical techniques for crop identification using Satellite Data. Different materials reflect and absorb solar radiation differently at different wavelengths. These spectral "signatures" can be used to identify objects such as bare soil, leafy vegetation etc. It has been found that certain combinations of bands (Indices) can be useful in identifying and differentiating types of vegetation. The addition of extra information can aid the differentiation and identification of vegetation types, such as spatial information (neighboring pixels more likely to be similar) or temporal information (growth patterns through time). The approach is to train and verify predictive models based on satellite information against "ground truth" field observations. The predictive model will only be as good as the data they are trained on. A number of mathematical models and machine learning models have been studied with variable rates of success.

N. Web scraping for Job and Price Statistics

Moderator – Mr. Niels Ploug, Statistics Denmark

Panel

- Mr. Boro Niki, Statistics Slovenia
- Ms. Paula Garcia, Statistical Office of Colombia, DANE

In the short term Statistics Slovenia is trying to improve the existing surveys on Job Vacancy statistics by adding some web scraping information from job portals, enterprise websites in addition to other administrative data. In the longer term, Statistics Slovenia would derive the job vacancy statistics completely via web scraping, and use the traditional survey only for benchmarking at certain periods in time. In Colombia, web scraping is used to measure variations in the consumer price index. Prices are scraped from supermarket websites in Bogota. DANE is in the process of creating and validating algorithms which will scrape the web according to certain specifications, such as brand, variety, unit, primary and secondary characteristics. Some results obtained so far were (1) the websites of selected supermarkets were viable to apply web scraping techniques; (2) the used method was valid to obtain prices and specifications of several products from the basket of the CPI, including food products, that represent 28,1% of the CPI; (3) the method was appropriate to complement the traditional

collection process of the CPI; and (4) the use of free software as R is a good alternative to explore new methods in a cost-effective way.

O. Data Analytics and Visualization solutions

Moderator – Mr. Ronald Jansen, United Nations Statistics Division

Panel

- Ms. Julie Whipple, Global Director, Corporate Responsibility, Qlik
- Mr. Emmet Dowling, SAS, Ireland
- Mr. Miguel Luengo-Oroz, UN Global Pulse, New York, USA

The presenters showed all three how visualizations of large amounts of data can help the decision making process, especially in emergency situations. Qlik provided examples of the Zika outbreak and how visualization can help not only to show where the incidence of the disease has been shown to be high, but also how it helps to communicate to the public in last what and where the threads are, and how precautions can be taken. SAS called this: data visualization as a gateway to actionable analytics. The example, which SAS showed, was the use data analytics for obtaining rapidly and economically sufficient basic building materials to strengthen the temporary living facilities for those without a home in Nepal due to the recent earthquake. Finally, UN Global Pulse presented a large number of examples of Big Data projects, in which they have been active and in which they use visualizations and analytics to derive meaning and policy implications.

P. Future Learning

Moderator – Mr. Niels Ploug, Statistics Denmark

Panel

- Mr. Nigel Smith, FutureLearn.com

Mr. Nigel Smith presented FutureLearn, which is a pioneer in distant learning and founded by the Open University. Based in the UK, FurtureLearn has 4 million registered users from all over the world. Given the large platform, FutureLearn is interesting as a partner. It offers partners new, innovative digital marketing and brand development services. It has established partnerships with the world's top universities and institutions, sharing knowledge and building a network. Beyond video lecture recordings, this platform is about telling stories while using interactive content, the best of the web, and engaging video. As an example, a demographics course on occupational health in developing countries drew an audience where 41% came from African countries, and 45% were female. The platform can and has been used also for some free online courses, as for the Ebola outbreak, a few years ago, to understand transmission, response and control.

Q. Agile Project Management

Moderator – Mr. Ronald Jansen, United Nations Statistics Division

Panel

- Mr. Simon Reindl, Agile

Mr. Simon Reindl gave an introduction to Agile project management. The Agile framework is based on acting on feedback and a continuous improvement approach. It is about responding to change instead of following a plan. The focus is on delivering business value and on delivering a working product. With Agile you make process and policies explicit, implement Feedback loops, improve collaboratively, and evolve experimentally. This new project management approach is especially useful for innovations, such as Big Data projects, since outcomes may not be known from the start, which should not prevent you from starting on it.

Day 3 – Big Data and the SDG indicators

Thursday, 1 September 2016

R. High-Level Panel – Big Data for SDG indicators

Moderator – Mr. Ivo Havinga, United Nations Statistics Division

Panel

- Mr. Niels Ploug - Statistics Denmark and Chair of the GWG on Big Data
- Ms. Lisa Bersales - Philippine Statistics Authority and co-chair of the Inter-Agency Expert Group on SDG indicators
- Mr. Aboubacar Beye - Director General, National Institute of Statistics and Demographics, Senegal
- Ms. Aberash Tariku Abaye - Deputy Director General, Central Statistical Agency, Ethiopia
- Mr. Mohammad Abdul Wazed - Director General, Bangladesh Bureau of Statistics

This high-level panel was about how Big Data could help in the new demands for frequent and granular data to monitor the progress in achieving the sustainable development goals. Whereas these new demands are already challenging in developed economies, they are even more daunting in a developing country context. Progress has been made in assessing the needs in terms of broadening of the collection and sharing source data, the strengthening of statistical registers, the use of innovative technology and the broadening of the profile of the statistical and data analytical skills of the statistician. These assessments will contribute to the development of the strategies and related work programmers for modernizing and transforming the national statistical systems, include the reform of the statistical legislation to provide a broader access to and use of administrative and big proprietary data sources while ensuring data privacy, security

and protection. Concreate steps are taking to introduce new technologies in the collection of household surveys and undertaking population and housing censuses such as the use of computer assisted handheld devices. Moreover, explicit requests are made to the global and regional statistical agencies to provide access to innovative technologies and related capacity building at the country level to advance the use of Big Data in the official statistics production process, including the generation of (disaggregated) SDG indicators.

Bangladesh made an explicit call to make more funds available to help statistical offices in developing countries with the implementation of the new demands. It stated that a robust follow-up and review mechanism for the implementation of the new 2030 Agenda for Sustainable Development will require statistical data, big data, and administrative data to monitor progress, inform policy and ensure accountability of all stakeholders. Good, available, and well-used data can be of a big help for SDGs monitoring and evaluation. For the Bangladesh Bureau of Statistics more budget to handle big data is needed for investment in IT equipment, technologies and skills training. The statement of Bangladesh is available on the website of the Conference, see https://unstats.un.org/unsd/bigdata/conferences/2016/default.asp

S. Session 1 – Big Data for SDG indicators

Moderator – Ms. Amparo Ballivian, World Bank

Panel

- Mr. Arnold Dekker, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
- Mr. Frederic Pivetta, Real Impact Analytics, France
- Mr Sainarayan Ananthanarayan, International Civil Aviation Organization (ICAO), Canada
- Mr. Sumair Sayani, Project 8, Demand Institute, USA / Canada

These panelists represented public and not-for-profit institutes, which are making efforts to support the 2030 agenda for sustainable development by improving the measurement of the SDG indicators. CSIRO tries to do this for SDG 6: Water and Sanitation. Via earth observations (EO), CSIRO wants to advance the global monitoring of water quality and develop a strategic plan for incorporation of current and future EO information into national and international near-coastal and inland water quality monitoring efforts. Real Impact Analytics focuses on the use of Call Data Records (CDRs) and providing algorithms to use these mobile phone data. Application of CDRs has shown most successful in the area of "Good Health", SDG 3. ICAO endeavours to use aviation data for the measurement of indicators in relation to SDG 9, especially targets 9.1 (with respect to transportation) and 9.4 (with respect to CO2 emissions), while Project 8 has started working on bringing together academics and researchers in trying to work together in a better assessment of food needs as part of the global effort to end hunger and achieve food security (SDG 2).

T. Session 2 – Big Data for SDG indicators

Moderator – Mr. Misha Lokshin, World Bank

Panel

- Mr. Luis Crouch, RTI International,
- Mr. Brant Zwiefel, Microsoft
- Mr. Paul Dommel, IBM
- Mr. Patrick Dunagan, Google / Skybox

In this panel the private and research sector presented how they are involved and would like to be further involved in the measurement of the SDG indicators. RTI International is an independent, nonprofit research institute dedicated to improving the human condition. For example, to improve the health information systems (SDG 3) in countries RTI promotes to combine paper and cell phones, apps and radio, and traditional surveys and SMS surveys. Microsoft presented in detail a trusted collaborative data project it recently executed in California for safe and sustainable water provision (SDG 6). The collaboration is challenging, especially because of the need for compliance with security and privacy policies. In California, a number of agencies collaborated to monitor water conservation. Microsoft works with the concept of data vault, which is a hyper-scalable repository of potentially sensitive data from multiple parties that facilitates secure data sharing, and provides mechanisms for data ingestion, policy tagging, encryption, storage and strong access to data, while maintaining compliance with a diverse set of effective policies. The principles of a trusted data collaborative consist of transparency, accountability and fair value exchange.

IBM focused in its presentation on monitoring youth unemployment (SDG 8). The factors influencing youth unemployment are education, personal stability, community experience, and labor market fluctuations. In the UK, IBM supported local government to collect social media data and to analyze these with respect to the mentioned factors in an attempt to predict the chances of youth on the labor market. Finally, the Skybox affiliate of Google demonstrated its use of satellite data for estimation of economic indicators (SDG 7 and 8). Small, high performance satellites allow Skybox to scale its constellation and enable timely image collections. Relevant information is extracted from the imagery to create a timely and granular datasets. Imagery and data are thereupon delivered in an easy to use web application to explore patterns and support decision-makers.

All 4 presenters showed the willingness to build public-private partnerships in an effort to use new data sources and modern technologies for better decision making.

U. Concluding Panel

Moderator – Mr. Ronald Jansen, United Nations Statistics Division

Panel

- Mr. Padraig Dalton, CSO Ireland
- Mr. Bert Kroese, Statistics Netherlands
- Mr. Ivo Havinga, United Nations Statistics Division
- Mr. Oliver Chinganya, African Centre for Statistics, UNECA

The moderators of the high-level panels gave their views on the important takeaways during the three days of the conference. UNECA was added to the panel to provide the perspective of countries with less developed statistical systems.

While the national statistical offices should explore the access to and use of the new proprietary (big) data sources, the moderators pointed out that data privacy and confidentiality remain paramount for the community of official statistics because a breach of these principles will have immediate negative impact on the trust of the general public in official statistics. This trust and accountability should be based on full transparency of our methods and standards and reflected in our statistical laws and statistical quality framework.

The moderators also stressed that national statistical offices should adopt a multi-facetted corporate approach to its needed transformation and modernization to stay relevant in a fast changing data landscape. Business as usual in delivering statistical service to the national and international clients is no longer tenable given the need to manage the downside risk of the external initiatives, competition of other providers of statistics and budget cuts, but also reap the benefits of the changing ICT and data landscape through rapid technological development and data revolution. National statistical offices should consider the introduction of the cloud environment while moving away from physical IT equipment on site and the adoption of common open source software architecture with common applications for storage, processing and dissemination in a trusted and secure technological environment. The introduction of this new technology infrastructure should also facilitate the sharing of data, technology and software applications among the official statistical community.

The moderators also underlined that emerging projects and programmes on the introduction of the new technological environment for the statistical production process and functions of the national statistical institutes require new collaborative and strategic partnership. These partnerships should extend to the academia and private sector at regional and international level that have the capability to scale and take advantage of digital technologies. There seems to be also a shift necessary to funding of recurrent cost of having continuous access to private big data sources and modern cloud based IT infrastructure. Moreover, it seems to require a change in donor perceptions from upfront funding of investments in national statistical systems to investments in common regional and international platforms for access to data, services and applications through strategic partnerships

The GWG on Big Data for official statistics should play a critical role in addressing the issues of (a) access to Big Data held by the private sector, (b) partnership in creating the technological infrastructure, data services and applications, (c) in providing capacity building and innovation to the production of official statistics and (d) the use of Big Data for monitoring and reporting of the SDG indicators for the 2030 Agenda for Sustainable Development.