UNITED NATIONS

DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS

STATISTICS DIVISION

United Nations Statistical Commission
Global Working Group on Big Data for Official Statistics
Third Meeting, 29 August 2016, Dublin, Ireland

Report

Points for Action

⇒   Review the terms of reference of the management team including description of the relation between the management team and the Task Teams in line with traditional portfolio management

⇒   Create a work plan with milestones for the management team and continue its regular WebEx meetings

⇒   Review the memberships of the Task Teams, approach not so active members and consider additional members. Review the communication tools.

⇒   Formulate (together with the management team) Key Performance Indicators for each Task Team.

⇒   Find funding for a program manager and hire him/her for a longer period of time.

⇒   Draft a proposal and report regarding the recommendations for access to non-traditional data sources. Address the issue for additional consultation.

⇒   Establish a small committee which works on the proposal for a Global Platform – with a strict deadline that will enable us to put a proposal from the UNSC

⇒   Establish a Committee which prepares the 2017 Big Data Conference

Agenda

Opening by Padraig Dalton

Item 1. **Overview report on the work of the GWG and its management team**

Item 2. **Progress report and evaluation of the work of the Task Teams**

      (i)   Access and Partnerships

            a. *Recommendations for Access to Data from Private Organizations for Official Statistics*
            b. Note for Discussion – Peter Struijs, Statistics Netherlands
            c. Note for Discussion – UNSD

      (ii)   Training, Skills, Capacity-building

      (iii)   Satellite imagery, remote sensing and geo-spatial data

      (iv)   Mobile Phone data

      (v)   Social Media data

      (vi)   Big Data and SDGs

Item 3. **The future of the GWG –** Presentation by the ABS and UNSD

Item 4. **The UN World Data Forum** – What is the role of the GWG on Big Data?

Item 5. Other business – New Chair and venue for next Conference

Participants

<u>Countries:</u>   Mr. Padraig Dalton (Ireland - Chair), Ms. Sybille McKeown (Australia), Ms. Sylvie Michaud (Canada), Ms. Hélène Bérard (Canada), Ms. Yuhuan An (China), Ms. Yiajing Cong (China), Ms. Paula Garcia (Colombia), Mr. Neils Ploug (Denmark), Mr. Richard Murphy (Ireland), Mr. Toni Virgillito (Italy), Ms. Hae Ryun Kim (Korea Republic), Ms. Min Jeong Park (Korea Republic), Mr. Juan Muñoz (Mexico), Mr. Mustapha Ait Taleb Ali (Morocco), Mr. Peter Struijs (Netherlands), Mr. Ahmed Al-Mufarji (Oman), Ms. Lisa Grace S. Bersales (Philippines), Mr. Sufyan Daghra (UAE), Ms. Heather Savory (United Kingdom), Mr. Robert Sivinski (United States)

<u>Organizations:</u>   Mr. Albrecht Wirthmann (Eurostat), Ms. Elham Saleh (GCC-Stat), Ms. Esperanza Magpantay (ITU), Mr. Trevor Fletcher (OECD), Mr. Steven Vale (UNECE), Mr. Jong Gun Lee (UN Global Pulse), Mr. Miguel Luengo-Oroz (UN Global Pulse), Mr. Ivo Havinga (UNSD), Mr. Ronald Jansen (UNSD), Ms. Amparo Ballivian (World Bank), Mr. Michael Lokshin (World Bank)

Minutes

0.   Word of Welcome – Padraig Dalton, Director-General of CSO Ireland

Mr. Padraig Dalton, Director-General of CSO Ireland, welcomed the participants to Dublin. He explained that UNSD had asked him to chair the meeting in the absence of Mr. Trevor Sutton, who could not be in Dublin due to urgent matters at the Australian Bureau of Statistics. In his opening remarks, Mr. Dalton stressed the importance of the GWG meeting and the Conference with respect to the direction, which the Global Working Group should be taking going forward. After three years, it is time to reflect on how this group can concretely contribute to the use of Big Data in the daily production of official statistics. The aim for enhanced production of statistics seems clear, namely to use multiple data sources for multiple outputs. A lot of work has been done in various task teams and also in projects at national level, but it may be time to bring it together and somehow make it into shareable services and applications for the benefit of the global statistical community.

1.   Report on the work of the GWG and its management team

(i)   Conclusions from GWG meeting in Abu Dhabi

It was concluded at the GWG meeting of 2015 in Abu Dhabi that the GWG needs (i) to be more engaged in pilot projects, of which some with application to the SDG indicators; (ii) to make compelling business cases for Big Data partnerships shown as win-win situations, and (iii) to make the inventory of Big Data projects available. The GWG should have special focus on the use of Big Data for SDG indicators and developing countries.

(ii)   Progress made since Abu Dhabi

Progress made by the Task Teams is reported under agenda item (2). The GWG made progress – among others – through collaboration in pilot projects in the UNECE Sandbox and the upgrade of its website, including the availability of the inventory of Big Data projects. The work of the GWG was supported by a program manager, who was contracted via the Australian Bureau of Statistics. The program manager was able to stimulate the work of the management team in a positive way which resulted in an increased number of teleconferences over the course of the year. By contrast the meetings of the Task Teams were irregular.

(iii)   Management Team

The GWG was restructured by abolishing the Task Teams on advocacy and communication, and on cross-cutting issues and moving their objectives and deliverables to the management team. This means that the terms of references of the management team need to be revised and in doing that come up with a description of the relation between the management team and the TTs that are in line with traditional portfolio management.

Whereas the management team held regular teleconferences, most of the time was taken up by programmatic and organizational matters around the Big Data Conference. The themes of

advocacy, communication or Big Data quality did not get the attention, they would deserve. Going forward we have to separate the Conference program from the work of the management team, by which we can refocus the work of the management team. A work plan with milestones for the rest of 2016 needs to be established and regular WebEx meetings need to continue. This means also that a new program manager needs to be found and funded. The main task of the program manager will be to facilitate and stimulate the work of both the management team and the Task Teams. In this regard, the memberships and how active member are can be reviewed and appropriate actions taken with regard to not so active members. A call for additional members for some TTs could also be considered. The program manager and the TTs together with the management team should formulate Key Performance Indicators for each TT. Moreover, effective communication needs to be ensured, which may mean to replace or enhance the Trello as communication tool.

2.   Progress reports of Task Teams

(i)   Task Team on Access and Partnerships

Under this agenda item three documents were submitted on recommendations for access to Big Data sources, an accompanying note for discussion, and a discussion note by UNSD. The recommendations have the objective of helping statistical organizations in getting access to the new data sources, while striking a balance between the legitimate interests of organizations holding Big Data and the public need for better official statistics. The note argues that the access recommendations have been established to meet a demand by national statistical agencies for a reference in their contacts with legislators, to meet a demand by data providers for a reference point and justification for data sharing, and to meet the demand of international agencies for a reference point. These access recommendations are linked to the Fundamental Principles of Official Statistics and are appealing to the social responsibility of the data providers.

The GWG discussed what the next steps should be in making these recommendations available to and acknowledged by the global statistical community and by other stakeholders. It was agreed that the issue needs to be handled with some urgency, that more consultation with stakeholders would be undertaken through a targeted survey, and that the recommendations would be reported to the Statistical Commission as part of a package. This package would include (1) motivation for recommendations for access to Big Data, (2) outlining the shortcomings – in this regard – in the existing Fundamental Principles of Official Statistics and in the existing Statistics Codes of Practice, and (3) requesting the Commission to assess the need for a revision of the Statistics Code of Practice with respect to obtaining access to new non-traditional data sources (such as Big Data) for statistical purposes.

(ii)   Task Team on Training, Skills and Capacity-building

Sufyan Barghouti (UAE) gave an overview of the progress made by the Task Team. The second day of the Conference shows a rich program around the topic of Capacity Building for innovation in official statistics. This reflects the points of attention of this Task Team. The role of training institutes highlights the formulation of course curricula around Big Data, which has

taken shape in Europe within the European Statistical Training Program (ESTP), and which the TT wants to bring together at the global level. Other topics which receive attention from the TT are the pilot projects in the Sandbox, new data visualization tools, new learning tools, like promoted by FutureLearning.com and agile project management. Finally, the TT likes to stay actively involved in the workshops, which are set up under the umbrella of the GWG, like the workshop on Satellite imagery data organized for a large part by the Australian Bureau of Statistics, and like the training courses organized under ESTP, for which developing countries could be invited.

(iii) Task Team on Satellite imagery data

Sybille McKeown (Australia) provided an overview on the progress made by the Task Team on Satellite Imagery, Geospatial Data and Remote Sensing. The Task Team will complete a report which includes the business case for using big data for official statistics to monitor the 2030 agenda, information about sources of earth observations data, methodologies for producing statistics, outlines of the Task Team's pilot projects and guidance for practitioners exploring the use of EO data for the first time. These pilot projects are the application of satellite imagery data in the production of agricultural statistics (Australian Bureau of statistics), Skybox (Google) commodity inventory assessment using proprietary satellites, environmental analysis of climate scenarios in various regions in Mexico (INEGI, Mexico), and use of satellite images to calculate statistics on land cover and land use (DANE, Colombia). The Task Team also collaborates in developing a hands-on workshop to teach methods for using earth observations data to produce crop statistics. This workshop will be held in the near future in Asia and Latin America. ITU pointed out that the Task Team could consider also involving the space agency regulators in its work.

(iv) Task Team on Mobile Phone Data

The Task Team has started the preparation of a Handbook on the use of Mobile Phone data for official statistics, which builds on experiences learned from a project in Oman with involvement of Positium, and a few projects of Eurostat. It can further build on the work done at Eurostat mostly through its ESSNet with projects in the Netherlands, France, Belgium, Slovenia, and Spain. Several papers have already been drafted regarding mobile phone data, such as the Big Data Action Plan and Roadmap preparing the European Statistical System for integration of big data sources into the production of official statistics (adopted in 2014) and a feasibility study on the use of mobile phone data for tourism statistics (2012-14).

The Handbook will include partnership business models between national statistical offices and mobile operators and applications for (i) tourism Statistics (international, domestic and sub-national), (ii) day-time population statistics (same day tourism statistics), (iii) event statistics, (iv) transport statistics, and (v) population density statistics. This Handbook will provide practical guidance and good country practices which should prove useful particularly for developing countries.

ITU pointed out the work it was doing with a number of pilot countries, which could be useful for the Task Team. Similarly, UN Global Pulse also had materials available. The discussion on the partnership models should include a financial model as well, either paying for services or paying for data.

(v)   Task Team on Social Media Data

In a very similar manner, the overview of the Task Team on Social Media data focused on the delivery of a Handbook, which explains how to use social media data with Big Data techniques for producing official statistics. Each chapter of the handbook has a leading institution which is responsible of defining the layout of the chapter, creating teams to work on, organizing work sessions and ensuring coherence and consistency among the different chapters. The content of the handbook will include a description of the data coming from some of the more popular Social Media Networks, and of a proposed method to collect and store data of such networks, with step by step examples. It will also include Social media methods and techniques including Natural Language Processing (NLP) and extracting information from non-message part of social media.

UN Global Pulse pointed out that it had just reached an agreement with twitter on the use of its data. It should be further discussed how this can be used by other GWG partners and how this can be reflected in the Handbook. Maybe the Twitter data can be fed more easily into the Sandbox.

The Open Algorithm initiative should also be discussed in the Handbook (and maybe also in the Handbook on Mobile Phone data), which can touch upon the underlying issue of white-box versus black-box approaches.

(vi) Task Team on Big data and SDGs

A brief overview was given of the work of the Task Team on Big Data and SDGs. The TT surveyed a variety of institutions on the use of Big Data for SDGs in addition to the 2015 GWG survey and the obtained information was fed into the inventory of Big Data projects. The World Bank and UNSD also worked out a document for the governance and management of the GWG Big Data Inventory. This inventory is a catalogue of Big Data projects that are relevant for official statistics, SDG indicators and other statistics needed for decision-making on public policies, as well as for management and monitoring of public sector programs/projects. For any new addition of a Big Data project to the inventory, permission to publish needs to be obtained from the custodian of that project, before entering the project information into the existing fields. Any organization or individual will be able to propose new entries to the inventory by filling out the Google form linked to the inventory website.

Mapping of Big Data projects to the SDG goals and targets is an ongoing exercise. The inventory has provision for this, but is not been fully filled yet. Further work on linking to SDG indicators is needed, including references to the existing methodological work on those

indicators. It is recognized that the Big Data Inventory can be an excellent tool for promoting the work of the GWG, and advocacy should include reference to the inventory.

3.   Future of the GWG

Over the last three years the GWG identified many promising public and private Big Data initiatives at the national and regional level (held by both public and private agencies), which aim to make data, services and applications accessible and aim to accelerate their synergies for research and capacity building. A proposal is now made in which the GWG brings all those initiatives together at the global level. This is a proposal for a Global Platform for Big Data for Official Statistics to provide scale and scalability for data, services, applications and infrastructure for the use of Big Data and its integration with administrative sources, geospatial information and traditional survey and census data. The platform would be developed under auspices of the UN Statistical Commission.

The business model for the Global Platform includes the following four components: (i) products and services, (ii) business institute, (iii) customers and (iv) financial approach. The products and services consist of the suite of assets delivered to customers including data, services, applications and infrastructure, where data include trusted and transparent individual datasets and integrated data that are provided by private and public partner organisations and services to customers will include technical statistical services (e.g. stand alone or as building blocks to develop applications) as well as capacity building.

The business institution, which could be developed in a Global Centre for data, services and innovation, consists of (a) a managing partner, (b) a partner network and (c) a governing organization. The Managing Partner will coordinate the Platform's activities and develop, deploy, operate, and manage the products and services, and will need full-time resources to carry out its functions. Important functions of Managing Partner are establishing and managing relationships with partners; managing the portfolio; customer relationship management; and outreach and marketing efforts. Partners provide content to the Global Platform.

The GWG pointed out that a number of issues need further attention, including the position of the Sandbox with respect to the Global Platform, the handling of data (anonymization, restricted use, macro-data versus micro-data etc.) on the Platform, the organization of data sharing, relation of Platform to research communities, the position and role of the GWG with respect to the Global Platform, what would be the legal basis of the Platform, and what would be the incentives for the various partners to become part of the Platform.

The GWG agreed that this Global Platform is a good idea and agreed also to establish a small committee of members of the GWG to further explore the components of the business model and prepare a request to the Statistical Commission at its upcoming session in March 2017.

4.   UN World Data Forum

The UN World Data Forum will take place on 15-18 January 2017 in Cape Town, South Africa, see http://undataforum.org/ The conference will cover a wide range of themes and topics including innovations in household surveys, censuses, administrative data, and civil registration and vital statistics, unleashing the power of new data sources (e.g., big data), new approaches in data collection (e.g., citizen generated data and perception surveys), integrating geo-spatial and statistical data and information, bridging the scientific community with the data community, and cooperation strategies between official statistical systems and the private sector. It will also include data in post-conflict situations, data and human rights, empowering users through data literacy, data and technology in everyday decision making, innovations in data communication and visualization, data needs at the local level, and data journalism.

The same committee working on the proposal for the Global Platform will look into the question how the GWG can contribute to the UN World Data Forum.

5.   Other business

Two more issues were brought to the attention of the GWG.

First, Trevor Sutton had conveyed the message that he could no longer assume the responsibilities of the function of Chair of the GWG. Padraig Dalton as acting Chair sincerely thanked Trevor on behalf of the whole working group for his efforts and dedication to the work of the GWG over the last three years.

Niels Ploug of Statistics Denmark was proposed as the new Chair. The GWG unanimously agreed with this proposal. The term of the chairmanship will be for two years.

Secondly, a venue for the Big Data Conference of 2017 will need to be decided upon. Four proposals have been received by Japan, Republic of Korea, Mexico and Colombia, respectively. Some internal discussions will be held and a proposal will thereupon be circulate to the GWG. We may want fix the venue for at least the coming two years.

It was finally also agreed that a Conference Committee should be established under the leadership of the GWG Chair for the organization of the next Conference.