



UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION
**United Nations Statistical Commission
Global Working Group on Big Data for Official Statistics
First Meeting, 31 October 2014, Beijing, China**

Report

I. Background

Following the International Conference on Big Data for Official Statistics held 28-30 October 2014, the Global Working Group on Big Data for Official Statistics held its first meeting. The meeting was chaired by the Australian Bureau of Statistics, and attended by 34 participants from National Statistical Offices, regional and international organizations.

The International Conference showed that the statistical community is open to innovation, and that there are large opportunities for the use of Big Data for Official Statistics. However, there are a number of challenges relating to methodology, privacy and access to data, partnerships and skills. These challenges are faced in both developed and developing countries.

II. Summary of the 3-day Conference

A. Mobile Phone, GPS and other tracking devices

Mobile Phone positioning data, GPS and other tracking devices were discussed in detail this session. The invited expert, Mr. Margus Tiru, Chief Executive Officer of Positium presented the difference between active and passive mobile phone location data, and showed a number of different statistical applications, such as tourism, day time mobility and estimation of population census data. Following the presentation by the expert, ISTAT (Italy), International Telecommunication Union (ITU), Statistics Netherlands and NBS China presented various applications of mobile phone data. The key challenges remain public trust in the use of and access to mobile phone data, and the protection of privacy and confidentiality.

B. Satellite Imagery and other geo-spatial information

The invited expert Mr. Siu Ming-Tam (Australia Bureau of Statistics) gave a presentation on Satellite Imagery. After presenting the basics of satellite imagery, he focused in particular on the methodology for estimating crop yields. He explained the work of ABS on satellite imagery data in order to replace or complement surveys for measuring agricultural crop production, and

discussed the business case for doing so. Satellite imagery has the potential to provide more frequent and timely data, at a very disaggregated level; however, the methodology remains to be tested. One of the arguments for the use of satellite imagery is the potential to reduce the frequency of surveys, and hence reduce costs. Following his presentation, representatives from DANE (Colombia), NBS China and INEGI (Mexico), presented three different applications of satellite imagery in their countries.

C. Twitter and other social media

This session discussed the use of twitter and other social media for official statistics. The invited expert, Prof. Andrew Schwartz (University of Pennsylvania) discussed the process of acquiring and analysing social media sources, which arguably is the largest dataset on human behaviour, and demonstrated various projects using social media for data-driven social science. He also raised a number of challenges with using social media such as participant consent, data storage, and sample and self-presentation bias. Statistics Netherlands, INEGI (Mexico), ISTAT (Italy), NBS China and the Chinese company Baidu all presented projects that explore the use of social media or the use of data scraping from the Internet. The example of the Netherlands showed a very promising application of obtaining consumer sentiment estimates from Facebook and Twitter data, which may potentially replace the existing surveys, since the social media estimates can be produced with a higher frequency and with reduced costs. The session showed that the statistical community is making strong progress in this area, even if there are numerous methodological challenges. The Statistical Community should be open to work in innovative ways, adding to its portfolio more flexible, short-term solutions to highly relevant policy questions.

D. Commonalities on benefits and challenges of Big Data sources

Having discussed the particularities of three different data sources, this session discussed the commonalities across data sources, in particular in terms of methodology and quality concerns, and issues of privacy, partnership and IT. It was recognized that each of the different Big Data sources encounters all challenges in one way or another, and that there is an immediate need to get to know those in all detail. Once we understand the challenges of the Big Data sources better, we can start leveraging their commonalities.

For instance, templates of umbrella agreements for access to data with globally operating companies could be established irrespective of the particular data source. The national statistical systems (NSS) should collaborate rather than compete with the private sector, in order to advance the potential of official statistics. At the same time, the NSS should remain impartial and independent, and invest in communicating the advantages of exploiting the wealth of available digital data to the benefit of the people. Building public trust will be the key to success.

E. Introducing innovation, especially in developing country circumstances

In this session, a number of new working methods were introduced, that all aim to establish cost-efficient ways of working together. UNECE presented their experience with Sprints, both physical and virtual Sprints, which has proven to be efficient ways of producing concrete

outputs. UNECE also presented the sandbox, which contains both datasets and tools for experimentation by different organizations. Similarly, the World Bank presented their experience with Data Dives and innovation challenges. ITU presented their Big Data strategy, while ESCAP discussed the role of Big Data in less developed national statistical systems.

F. The way forward- making the business case for Big Data

In this final session of the Conference, the current and future roles of the community of official statisticians were discussed. The role of the National Statistical System is changing, and business as usual is no longer possible. For example, official statisticians may be asked more frequently to validate information provided by the market. It was emphasized that the Statistical Community should remain acutely aware of its mission, namely to provide high quality and impartial information on relevant topics to the society at large and to policy makers in particular. Everything that is done should be done in a manner consistent with the role of a national statistical agency.

We should cautiously move forward, while embracing the potential of Big Data at the same time. The Statistical Community must adapt; if not, then users will find other sources. At the same time, it is important to manage the messages we are sending out: Innovation is crucial and the Statistical Community are a part of it, but the traditional core business will remain and is still needed. While innovating, it is important to maintain and improve the efficiency of ongoing operations

All NSOs should identify and build business cases for Big Data. Efforts should be made to link the use of Big Data to the Post-2015 development agenda and the new Sustainable Development Goals, as this will pose huge demands on the statistical systems. The 169 currently proposed targets cannot all be measured by indicators based on Big Data. However, Big Data could play an important role given the high frequency of Big Data and their geo-spatial detail. One way forward is to explore proxy indicators that could provide more frequent information than surveys. In this manner, Big Data is complementary and not replacing traditional systems. This is similar to the “flash” estimates of GDP, for example.

III. Coordination among international organisations

In order to ensure coordination of the work on Big Data among international organisations, a coordination meeting was organised in the margins of the conference, which was attended by UNSD, UNECE, UNESCAP, SIAP, ITU and Eurostat. A summary of that meeting was presented at the Global Working Group meeting. UNECE will create a calendar of events and all agencies will share their outputs. One way to establish that objective is to closely link the websites. It should be noted that the task teams of UNECE will soon complete their work, while the task teams under the global working group are only starting up. This provides the possibility of continuing some of the work already done.

IV. Terms of Reference of Global Working Group

The draft Terms of Reference (TOR) of the Global Working Group, which had been shared in advance with the members, was presented by the United Nations Statistics Division. The TOR needs to be based on some strategic considerations. Firstly, the work on Big Data should be based on the Fundamental Principles of Official Statistics. Secondly, the work of the Global Working Group should relate to the Post-2015 Development Agenda and related initiatives, such as the report of the Independent Expert Advisory Group on the Data Revolution for Sustainable Development. Further, it was stressed that the work of the Global Working Group would complement the work carried out by other regional commissions and organizations. The main points to be considered in the terms of reference are (1) to provide strategic vision, direction and coordination of a global programme on Big Data for official statistics; (2) to promote practical use of sources of Big Data for official statistics, including cross-border data, while building on the existing precedents in Big Data and finding solutions for methodological issues, covering quality concerns and data analytics; legal and other issues of access to data sources; privacy issues, in particular those relevant to the use and reuse of data, data linking and re-identification; and security and management of data, including assessment of cloud computing and storage; (3) to promote capacity building, training and sharing of experience; (4) to foster communication and advocacy of use of Big Data based statistics for policy applications; and (5) to build public trust in the use of private sector Big Data for official statistics including issues of confidentiality and privacy.

The importance of developing a Big Data Strategy was discussed. At the moment only three countries in the GWG have it. These three countries were in principle ready to share the strategy with the other members in order to assist the development of Big Data strategies. It was suggested that the Statistical Community should be seen as an “engine of innovation”, but at the same time remain realistic. We should start small, identify champions, use the right source and be realistic.

V. Work programme 2015-2016

After some discussion on the various activities which could or should be done by the Global Working Group, it was agreed to establish seven task teams plus one coordinating team consisting of the leaders of the seven teams. I am proposing to add one more task team, namely on Social Media data, as we did discuss the importance of this Big Data source at the Conference. UNSD can take the lead on this additional task team. These task teams are then:

1. Task team on Training, skills and capacity building
2. Task team on Linking Big Data and SDGs
3. Task team on Advocacy and communication
4. Task team on Mobile Phone Data
5. Task team on Satellite Imagery
6. Task team on Social Media data
7. Task team on Access and Partnership
8. Task team on Cross-cutting issues, classification, frameworks and taxonomy

During the meeting, a number of countries/organizations expressed their interest in leading or participating in the different teams. The table below summarizes the expression of interest during the meeting.

Task team	Leaders	Participants
Coordination team		<ul style="list-style-type: none"> • Consist of leaders of each task team
Training, skills and capacity building	<ul style="list-style-type: none"> • SIAP • United Arab Emirates 	<ul style="list-style-type: none"> • UNECE • ITU • World Bank • UNSD • Colombia • Philippines
Linking Big Data and SDGs	<ul style="list-style-type: none"> • World Bank • Mexico 	<ul style="list-style-type: none"> • UNECE • ITU • UNSD • Colombia
Advocacy and communication	<ul style="list-style-type: none"> • ESCAP • Bangladesh 	<ul style="list-style-type: none"> • Denmark • UNECE • ITU • World Bank • UNSD • Eurostat • Colombia
Mobile phone data	<ul style="list-style-type: none"> • UNSD 	<ul style="list-style-type: none"> • Tanzania • Egypt • ITU • World Bank • UN Global Pulse • Eurostat • Colombia • Cameroon • Oman • Mexico • Margus Tiru (Expert)
Social Media data	<ul style="list-style-type: none"> • UNSD 	<ul style="list-style-type: none"> • World Bank • ITU • UN Global Pulse • Eurostat • Colombia • Andy Schwartz (Expert)
Satellite imagery	<ul style="list-style-type: none"> • Australia 	<ul style="list-style-type: none"> • Morocco • China • Oman • Mexico

Task team	Leaders	Participants
		<ul style="list-style-type: none"> • UNSD • ITU • World Bank • UN Global Pulse • Pakistan • Eurostat • Colombia
Access and Partnership	<ul style="list-style-type: none"> • Netherlands • Eurostat 	<ul style="list-style-type: none"> • China • ITU • World Bank • UNSD • UN Global Pulse • Cameroon • Mexico • Morocco
Cross-cutting issues, classification, frameworks and taxonomy	<ul style="list-style-type: none"> • Italy • Eurostat 	<ul style="list-style-type: none"> • Netherlands • UNECE • ITU • World Bank • Colombia • Philippines

UNSD will formally send a letter to all participating countries to seek other expression of interests to participate in the various task teams.

Each task teams will need to develop its own work programme, in close consultation with the coordination task team, but a number of different activities were mentioned during the meeting, such as:

- Expand the initial UNSD/UNECE survey to include more projects
- Web portal for Big Data
- Classification of Big Data V2.0
- Global conference, together with global/regional workshops

It was suggested that the task teams would provide an outline of their work within a short period of time to build on the momentum from the conference. In particular it was stressed that the task teams should build on the experience of the UNECE Big Data project and the material that is being produced by the UNECE Big Data task teams. A meeting of the coordination task team will be organized within a few weeks of the Global Working Group meeting, in order to initiate the work of the task teams.

VI. Report to the Statistical Commission

According to Decision 45/110 of the Statistical Commission, the Global Working Group on the use of Big Data for Official Statistics are required to submit a report to the Statistical

Commission at its forty-sixth session in 2015 on the work done. It was suggested that the report would have the agreed upon Terms of Reference in the annex, and that these should be general. The details regarding the work of the Global Working Group would be laid out in the report itself.

Members of the Global Working Group stressed that it is important to note the time line and deliverable of the task teams. Also it was suggested that the programme of work should be organized under three fields of engagements, namely 1) Project pilots and inventory, 2) Technical work and norm setting, and 3) Advocacy and communications activities.

The report will be drafted by UNSD and the chair of the Global Working Group, Australian Bureau of Statistics, and is expected to be shared with the Global Working Group members around 20 November.

VII. Attachments

- Agenda
- List of Participants
- Presentation on Key considerations (UNSD)
- Presentation on Future Work Ideas (UNECE)