



UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS

STATISTICS DIVISION
BIG DATA INITIATIVE

UN GLOBAL WORKING GROUP ON BIG DATA FOR OFFICIAL STATISTICS
Abu Dhabi, 19 October 2015

Item 3 of the agenda

“How can we produce Official Statistics with Big Data and what is the role of the GWG to make this happen?”

Discussion note by UNSD and Statistics Netherlands

From the Concept Note of the Conference

1. The global statistical community officially recognized the potential of Big Data, when, in March 2014, the UN Statistical Commission established a global working group (GWG) mandated to provide strategic vision, direction and coordination of a global programme on Big Data for official statistics. The GWG promotes its practical use of Big Data, capacity building and sharing experiences, while finding solutions for the associated challenges.

2. The first global conference on Big Data¹ was organized in Beijing in October 2014 and showcased a variety of examples of Big Data projects for official statistics, such as Mobile phone data for daytime population statistics, Satellite imagery data for agriculture statistics and Social media data for consumer confidence indicators. The second global conference will take place in Abu Dhabi in October 2015 and can take the development of Big Data one step further by moving from examples to guidance. On the basis of detailed descriptions of a significant number of Big Data case studies, some generalized lessons learned can now be brought together as a first step towards guidelines for Big Data. Those guidelines can stimulate training, pilot projects and bringing pilot projects into the production environment. Therefore, the theme of the second global conference is ‘moving from examples to guidance’.

Objective of the 2nd Global Conference on Big Data for Official Statistics

3. Given the need to modernize the statistical production and services, given the need to meet the increasing demands for timelier and more detailed data for monitoring the 2030

¹ See <http://unstats.un.org/unsd/trade/events/2014/Beijing/default.asp>

agenda for sustainable development and given the progress made thus far on a considerable number of Big Data projects, this second Global Conference on Big Data for Official Statistics wants to move beyond the examples of Big Data projects and make the first steps towards developing guidance, which will support training on Big Data issues, support initiatives for Big Data projects and support moving Big Data from pilots to production.

Status of Big Data projects

4. The results of the Big Data survey (Q6) showed that the most used sources are scanner data (26% of the responses), satellite imagery data (21%) and web-scraping data (19%), and that health records (13%), mobile phone data (10%) and credit card data (9%) are less used. Social media data comes twelfth with 4%. As far as partnerships are concerned (Q5), the most common partnership – apart from other statistical agencies – is with other government institutes (25%), followed by satellite imagery providers (21%). Big Data has been mostly applied (Q7) to price statistics (34%), followed by population/migration (17%) and labour statistics (16%), and Big Data is not commonly used for the indicators related to the SDGs.

5. The main reasons to use Big Data for official statistics (Q3) are to provide faster and timelier statistics (86%) and to reduce the response burden (82%). But modernization of the statistical production process (79%), new products and services (78%) and cost reduction (74%) are also considered important or very important. Meeting new demands related to the SDGs is deemed relatively less important (56%). The statistical community expressed a strong need for guidance (Q18) in the areas of skills and training (73%), quality frameworks for Big Data (64%), access to Big Data (64%), and estimation methodologies (58%). There is also demand for guidance on the usage of specific data sources such as web-scraping (48%) and mobile phone data (45%).

Status of the work of the Global Working Group

6. The Global Working Group had its work divided over 8 task teams. Progress has been made, but a general and fair conclusion would be that we are still very much in an initial phase. Guidelines for the three specific categories of Satellite imagery, Mobile Phone data and Social Media data are at a beginning stage, initial inventories of projects have been made by the Big Data and SDGs team and the Cross-cutting team, some documents have been prepared on advocacy, principles for access, good practices for data access and partnerships, MOU templates, and classifications.

7. So work is ongoing, but we may want to evaluate how much more needs to be done to move from pilot to production, and from examples to guidance. What are the factors determining our speed of progress, and what time schedule can we realistically follow? This discussion note is giving some of those factors as input to the discussion.

Factors determining our speed of progress

- Support from top level management
- Having organisational arrangements for Big Data and new techniques
- Enough budget
- Support from external stakeholders
- IT environment
- Engagement in strategic partnerships
- Training, workshops and skills
- Organisational culture
- International cooperation
- Possible need for more pilot projects
- Dissemination of Big Data-based statistics
- Capability to move from research to production
- Data access issues
- Privacy, confidentiality and public image issues
- Availability of appropriate methods

GWG product areas for which a time schedule may be needed

- Guidelines for Satellite imagery
- Guidelines for Mobile Phone data
- Guidelines for Social Media data
- Principles for Big Data access
- Classification of Big Data sources
- Use of Big Data for SDG indicators
- Big Data training modules
- Sandbox:
 - production component?
 - more regional sandboxes?
- Workshops and other capacity building activities
- E-learning modules?
- Big Data projects repository?
- Quality framework
- Big Data methodology
- Guidelines / good practices for Big Data issues (to be specified)
- Other products possibly identified by the GWG in Abu Dhabi

What is the role of GWG to use Big Data for Official Statistics?

8. The current approach of the GWG has been very much oriented on the input side, on Big Data sources and access to Big Data sources. Do we need to balance this approach by also looking at the output side? Another approach would be to look more closely on the needs and demands. If we start from looking at the request for timelier, more frequent

and more disaggregated data to feed the indicators of the 2030 agenda, what would this mean? Would we maybe look at combining satellite imagery, social media and mobile phone data for indicator breakdowns? Should we also approach Big Data more from thematic areas, such as health, education, economic growth and inequalities, employment and environmental issues?

9. Where should we put our priorities and how much work should we and can we realistically take on? What outputs should be produced and when? And how should we organize ourselves? At the end of the GWG meeting we should have tentative answers to these questions, based on how we see the role of the GWG. Since we now have the results of the Big Data survey, we know the actual use of Big Data sources in terms of inputs as well as outputs, and the need of the statistical community for guidance. This should also inform the answers to the questions.