



# UN Global Working Group on Big Data

**UNECE Workshop on Statistical Data Collection**  
*Washington, DC 29 April – 1 May 2015*



**United Nations Statistics Division**

**Nancy Snyder, Statistician, International Merchandise Trade Statistics Section**

## Global Working Group on Big Data: Background

Technological Innovations, broad use of electronic devices and the all-around generation of digital information is fundamentally changing the availability of real-time information, such as data from GPS devices, mobile phones and from social media.



The statistical community needs to adequately address issues pertaining to **methodology, quality, technology, data access, legislation, privacy, management** and **costs**, and provide adequate **cost-benefit analyses**.

# Global Working Group on Big Data



Recognizing the need for further investigating the benefits and challenges of Big Data for official statistics, the Statistical Commission agreed at its 45th meeting in March 2014 to create the Global Working Group (GWG) on Big Data for Official Statistics, which is to deliver a global work programme for Big Data for the coming years.

**Members:** 18 countries, 9 international organizations, academia and the private sector

Australia (Chair), Bangladesh, Cameroon, China, Colombia, Denmark, Egypt, Indonesia, Italy, Mexico, Morocco, Netherlands, Oman, Pakistan, Philippines, Tanzania, UAE, USA, UNSD, UNECE, UNESCAP, UN Global Pulse, International Telecommunication Union (ITU), OECD, World Bank, Eurostat, GCC-Stat

**Partners:** Positium, University of Pennsylvania

# Global Working Group on Big Data

At its first meeting in October 2014 in Beijing, the GWG established 8 task teams:

- Advocacy and Communication
- Big Data and SDGs
- Access and Partnerships
- Training, Skills and Capacity Building
- Cross-cutting Issues
- Mobile Phone Data
- Satellite Imagery
- Social media

The teams have started working and will deliver first results by the end of 2015.



<http://unstats.un.org/unsd/trade/bigdata/>

UN Big Data for Official Statistics

The statistical community has the obligation of exploring the use of new data sources, such as Big Data, to meet the expectation of the society for enhanced products and improved and more efficient ways of working. Big Data could also support the monitoring of the Post-2015 development goals by improving timeliness and relevance of indicators without compromising their impartiality and methodological soundness.

The banner features a central text area with a dark background and white text. The text is flanked by a grid of satellite imagery from various cities, including New York and Tokyo. The overall design is modern and data-oriented.

### **Draft Terms of Reference of UN GWG for Big Data**

- Sharing of best practices of Big Data for official statistics;
- Improve the complementarity and effectiveness of the roles of international and regional organizations, donors and recipient countries
- Working on concrete and practical examples of Big Data sources,
- Addressing the issue of Cost of Big Data
- Developing an adequate communication strategy
- Reaching out to Big Data initiatives, civil society, private sector, research and academic communities

## Global Working Group on Big Data: Deliverables

### Planned deliverables:

- Proposals for training, skills and capacity building to use Big Data for official statistics
- A report on links between Big Data and the sustainable development goals
- A strategy on advocacy and communication of the use of Big Data for official statistics
- Reports on the use of mobile phone data; social media data; and satellite imagery and remote sensing data for official statistics
- Proposals for enhancing access to Big Data sources
- A report on quality assurance & methodological frameworks
- Identify and share good practices from projects using Big Data in official statistics
- The GWG will participate in a number of pilot projects on the implementation of Big Data for official statistics



## Global Working Group on Big Data: Survey

The GWG will build on the results of the initial survey conducted by UNSD and UNECE in September 2014.

Initial UNSD/UNECE Survey on Big Data – Objective:

- To provide an overview of active Big Data projects
- We used the following definition of Big Data:

Big data are data sources with a high volume, velocity and variety of data, which require new tools and methods to capture, curate, manage, and process them in an efficient way

The survey was sent to all participating offices in the UNECE Big Data Project and to the members of the GWG -- 78 national statistical offices and 28 international organizations. Replies to the questionnaire on the organizational context were provided by 32 countries and 3 organizations, while 24 countries and 3 organizations were able to report on a total of **57 Big Data projects**.

# Summary of questionnaire on Big Data Organization and Strategy

- **Big Data strategy** - Only a few countries have developed a longterm vision for the usage of Big Data, some are formulating a Big Data strategy.
- **Quality** - The work of international organization is deemed important, in particular many countries mention the UNECE quality task team that aims to establish a quality framework
- **Privacy** - Crucial to protect privacy, even beyond what is defined by the rule of law, in order to protect public image
- **Skills** - So far NSOs have relied more on providing training to existing staff rather than hiring of a new type of staff (“data scientists”).

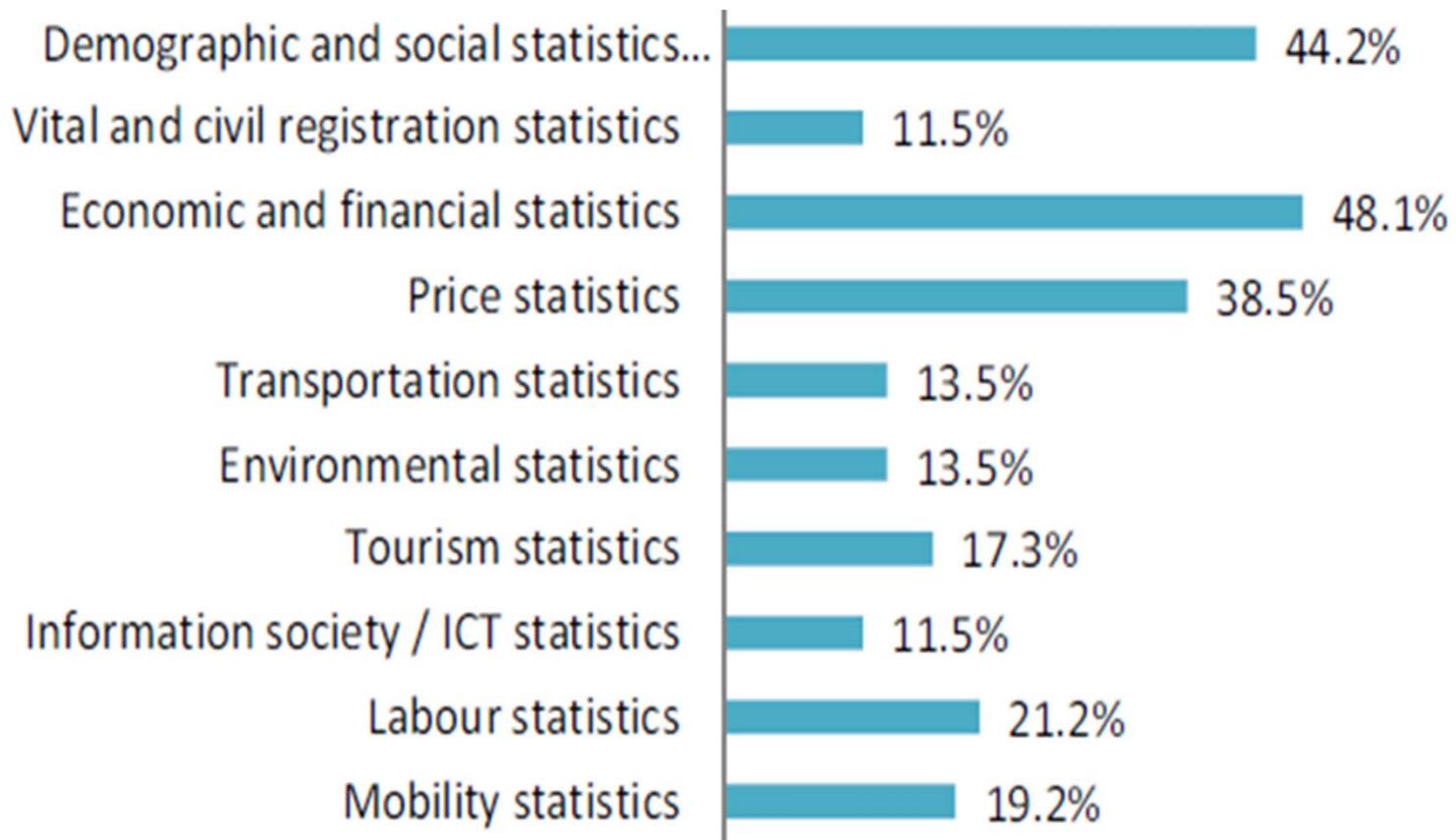
**Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data**

<http://unstats.un.org/unsd/statcom/doc15/BG-BigData.pdf>



## Summary of Results from Initial Big Data Survey

### Potential areas of use for the projects cited:



## Summary of Results from Initial Big Data Survey

### Challenges

- Most respondents agree that the biggest challenge for most Big Data projects are the limited access to potential datasets.
- Big Data are, to a large extent, owned by the private sector and many of these players are global companies;
- Thus, the global statistical community should use collective bargaining power to obtain the access to these data sets.

## Summary of Results from Initial Big Data Survey

### Technical tools

- Overwhelmingly survey participants chose to use internal hosting solutions rather than purchasing external hosting services
- Often due to privacy concerns
- Having to invest significant resources in a cluster before beginning a project is a significant hurdle
- Data analysis mainly conducted using R and SAS
- However, a large variety of tools were used in the projects, making it clear that no significant standard has emerged

### Skills & Training

Overwhelmingly the survey respondents indicate that they so far have relied more on providing training to existing staff rather than hiring of a new type of staff (or “data scientists”).

Some training in Java, SAS, SQL, SPSS, Stata and R

Due to budget constraints in a number of countries, hiring of data scientists is currently not planned and therefore further training of the current staff is a more realistic step for the future.

## Global Working Group: Timeline

### The Global Working Group:

- Will conduct a follow-up to this survey in summer 2015 to all countries
- Will update Terms of Reference for all Task Teams on the website
- Is currently working on the deliverables
- Will report on its work to the Statistical Commission for its upcoming 46th session in March 2015
- Will meet for the 2<sup>nd</sup> time in October 2015



The banner features a grid of satellite and aerial images of various cities, including New York and Tokyo. A central dark grey box contains the title and a paragraph of text. The UN logo and 'BIG DATA' text are in the top left, and a URL is in the top right.

 **BIG DATA**

<http://unstats.un.org/unsd/trade/bigdata/>

### UN Big Data for Official Statistics

The statistical community has the obligation of exploring the use of new data sources, such as Big Data, to meet the expectation of the society for enhanced products and improved and more efficient ways of working. Big Data could also support the monitoring of the Post-2015 development goals by improving timeliness and relevance of indicators without compromising their impartiality and methodological soundness.

*Thank  
You*

Nancy Snyder

[snydern@un.org](mailto:snydern@un.org)

United Nations Statistics Division