



Economic and Social Council

Distr.: General
12 December 2014

Original: English

Statistical Commission

Forty-sixth session

3-6 March 2015

Item 3 (a) (iii) of the provisional agenda*

**Items for discussion and decision: data in support of the
post-2015 development agenda: big data**

Report of the Global Working Group on Big Data for Official Statistics

Note by the Secretary-General

In accordance with Economic and Social Council decision 2014/219, the Secretary-General has the honour to transmit the report of the Global Working Group on Big Data for Official Statistics. The report presents the highlights of the International Conference on Big Data for Official Statistics, the outcome of the first meeting of the Global Working Group and the results of a survey on the use of big data for official statistics. The report elaborates on the terms of reference and the proposed programme of work of the Global Working Group, covering training, skills and capacity-building; linking big data and the sustainable development goals; advocacy and communication; access and partnership; and cross-cutting issues, such as classifications and frameworks, as well as the exploration of specific big data sources for official statistics, namely, mobile phone and social media data and satellite imagery. The report concludes with points for discussion.

The Statistical Commission is invited to take note of the report.

* E/CN.3/2015/1.



Report of the Global Working Group on Big Data for Official Statistics

I. Introduction

1. Because of widespread and constant use of telecommunications and other devices driven by innovations in technology, digital information is continuously generated, such as data from: Global Positioning System (GPS) devices, automated teller machines, scanning devices, sensors, mobile phones, satellites and social media. The high volume, high velocity and wide variety of these data, which are commonly referred to as big data, and are potentially useful for official statistics, require new tools and methods for capturing, managing and processing them efficiently. However, many statistical offices will still need to make the business case for this investment in innovation by showing how and why big data are useful and relevant. The International Conference on Big Data for Official Statistics, discussed in greater detail below, was encouraging in this respect, as it presented good examples of many ongoing projects on big data.

2. The potential of big data sources resides in the timely — and sometimes real-time — availability of large amounts of data, which are usually generated at minimal cost. The conduct of activities associated with traditional data sources, such as household and business surveys, takes time and those activities are often costly in respect of data production. Utilization of big data could supplement, reduce or replace the use of such types of data collection. However, the statistical community is conscious of the fact that before introducing big data into official statistics and in order to take advantage of these innovative data sources including their application to the monitoring and reporting on the sustainable development goals, it needs to adequately address issues pertaining to methodology, quality, technology, data access, legislation, privacy, management and finance, and provide adequate cost-benefit analyses.

3. Within this context, the Statistical Commission, at its forty-fifth session, held from 4 to 7 March, 2014 recognized that big data constitute a source of information that cannot be ignored and that needs to be evaluated on its own merits.¹ As regards achieving this, the Commission supported the proposal to create a Global Working Group on Big Data for Official Statistics, which would make an inventory of ongoing activities and examples regarding the use of big data, address concerns related to methodology, human resources, quality and confidentiality, and develop guidelines on classifying various types of big data sources.

4. The Statistical Commission stressed that the terms of reference and the mandate of the Global Working Group should be based on strategic considerations, in particular the links to the post-2015 development agenda, the data revolution initiative and the Fundamental Principles of Official Statistics. The Commission requested that the Group also complement and build upon work carried out by the regional commissions and other international statistical agencies, and that the programme of work of the Group pay special attention to the circumstances of developing countries, such as the state of their legal frameworks and any disadvantages in respect of available information technology infrastructure.

¹ See E/2014/24, chap. I.B. decision 45/110.

5. The Global Working Group on Big Data for Official Statistics was created in May 2014, pursuant to Statistical Commission decision 45/110, and held its first meeting on 31 October 2014 in Beijing, immediately after the International Conference on Big Data for Official Statistics, held in Beijing from 28 to 30 October 2014 and co-organized by the Statistics Division of the Department of Economic and Social Affairs of the United Nations and the National Bureau of Statistics of China. Moreover, the Statistics Division and the Economic Commission for Europe (ECE) conducted a joint survey on big data projects and their organizational setting, as part of the input to and preparation for these meetings.

6. The next section provides an overview of the International Conference on Big Data for Official Statistics, followed by a summary of the first meeting of the Global Working Group (sect. III) and the results of a survey on big data projects and their organizational context (sect. IV). The report concludes with the enumeration of points for discussion.

II. International Conference on Big Data for Official Statistics

7. The International Conference on Big Data for Official Statistics was attended by approximately 120 participants representing more than 40 countries, various regional and international organizations, and the research, academic and private sector communities. The Conference focused on three clusters of big data sources, namely (a) mobile phones, and GPS and other tracking devices, (b) satellite imagery and other geo-spatial information and (c) Twitter and other social media. The Conference examined the typology of and the specific challenges posed by these sources and the partnerships necessary to exploit them. The final sessions of the Conference were devoted to discussing common benefits and challenges of big data sources, the means of introducing innovation and how to make the business case for big data.

8. The considerable number of projects presented at the Conference reflected the spirit of innovation driving the statistical community. Applications of various big data sources to a wide range of statistical domains demonstrate that big data truly has the potential to improve official statistics. These innovative tools can be added to the portfolio of a statistical system so as to complement existing applications or provide more flexible, short-term solutions to highly relevant policy questions. The different projects also demonstrated how some of the challenges related to methodology, access, privacy and skills can be addressed. Selected highlights of the Conference are set out directly below.

9. The technical details associated with mobile phone records were discussed, including the differences between active and passive mobile phone location data. Several interesting statistical applications and national experiences were presented, encompassing, inter alia, tourism and daytime mobility statistics and estimation of population census data, poverty mapping and tracking of mobility patterns in cases of disease outbreaks. Given the widespread use of mobile devices, including in the developing world, mobile phone data have great potential to provide real-time, low-cost information on pertinent development issues. The key challenges remain the protection of confidentiality, access to mobile phone data and public trust in their use.

10. Satellite imagery has a great potential to provide more frequent and timelier data, especially agriculture statistics, at a highly disaggregated level; however, the

methods of estimation, for instance, as regards crop yields, are still being tested. The ongoing work of the Australian Bureau of Statistics on satellite imagery data, could result in the complementation, or even the partial replacement, of the existing surveys for measuring agricultural crop production. As satellite images are available at a frequency of once every two weeks, their use could lead to a reduction in the frequency (and associated costs) of surveys. Similar applications of satellite imagery to official statistics are being explored in China, Colombia and Mexico, as well as other countries, including in testing and experimentation within the domain of ecosystem accounting.

11. As Twitter, Facebook and other social media sources are in possession of what are arguably the largest data sets on human behaviour, the statistical community is already exploring these sources with a view to devising data-driven social science applications, for instance, on health-related issues. The Netherlands offered a highly promising example of how to derive consumer sentiment estimates from Facebook and Twitter data. Potentially, this could provide early estimates and reduce the need for surveys, since social media estimates can be produced with higher frequency and reduced cost. In China and Italy, web-scraping tools are used to estimate job vacancy rates, which can support current labour statistics through the provision of improved monthly predictions and more finely detailed territorial estimates. The statistical community is making strong progress in this area, even if many methodological challenges remain.

12. The commonalities across data sources were also discussed, in particular in terms not only of methodology and quality concerns, but also of privacy, partnership and information technology. It was recognized that each of the different big data sources encounters existing challenges in one way or another, and that there is an immediate need for a detailed understanding of those challenges. Once the challenges of the big data sources are better understood, the statistical community can leverage their commonalities. For instance, templates of umbrella agreements on access to data with companies operating globally could be established irrespective of the particular data source. The statistical community should collaborate rather than compete with the private sector in order to advance the potential of big data for official statistics. At the same time, it should remain impartial and independent, and invest in communicating the advantages of exploiting the wealth of available digital data in terms of the benefit of the society. Building public trust will be the key to success.

13. Finally, the current and future roles of the community of official statisticians were discussed. 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 topics relevant to the society at large and to policymakers in particular. While innovation is crucial and should therefore be embraced by the statistical community, the traditional core business, which is still needed, will remain. For instance, results derived from big data applications will have to be benchmarked, and it is traditional surveys that will be needed to provide those benchmarks.

14. The statistical community should identify and build business-related cases for big data. Efforts should be made to link the use of big data to the post-2015 development agenda and its sustainable development goals, as these will impose

huge demands on the statistical systems for years to come. Of course, the many proposed targets of the post-2015 development agenda will not all be measurable by indicators based on big data. However, big data are expected nonetheless to play an important role in this regard, given their timeliness and their level of geo-spatial detail. One way forward is to explore proxy indicators based on big data which could provide information more frequently compared with surveys. In this way, big data would serve as a complement to and not a replacement for traditional systems, along the lines of the “flash” estimates of gross domestic product (GDP), for example.

15. In summary, the International Conference on Big Data for Official Statistics provided an overview of promising applications of big data for official statistics and contributed to a better understanding of the challenges related to methodology, privacy and access. The Conference fostered the sharing of experiences, among national statistical offices and other important stakeholders and partners, on issues of building partnerships, access to data, communication and advocacy, and on the need for training, updating of skills and capacity-building.

III. Meeting of the Global Working Group on Big Data for Official Statistics

16. At its first meeting, held on 31 October 2014, the Global Working Group on Big Data for Official Statistics reviewed its terms of reference (see annex I) and discussed the work programme for 2015. The Global Working Group consists currently of 28 members, representing both developed and developing countries and various international and regional organizations (see annex II).

17. On the margins of the conference, the Statistics Division, ECE, the Economic and Social Commission for Asia and the Pacific (ESCAP), the Statistical Institute for Asia and the Pacific, the International Telecommunication Union (ITU) and Eurostat met to ensure coordination of the work on big data among international organizations. The result was a commitment by all of the organizations to share their information and outputs in this area. In this regard, ECE will create and maintain a calendar of events. It is understood that one means of stimulating cooperation will be to establish close links among the various organizational websites on big data.

A. Terms of reference and mandate of the Global Working Group on Big Data for Official Statistics

18. In general, the statistical community bears the obligation to explore the use of new data sources to meet the expectations of society with regard to enhanced products and improved and more efficient ways of working. However, the terms of reference and mandate of the Global Working Group on Big Data for Official Statistics should of course also be understood within the context of fulfilment of the new data demands posed by the monitoring and reporting requirements of the post-2015 development agenda.

19. The use of big data for official statistics is fully supported by the Fundamental Principles of Official Statistics. By way of example, principles 1, 5 and 6 state, respectively, (a) that official statistics that meet the test of practical utility are to be

compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information; (b) that data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records; and (c) that statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden to respondents and that individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes. Based on these principles, it is not only a possibility but almost an obligation for statistical organizations to investigate and pursue the use of big data sources for statistical purposes, as long as the rights of privacy and confidentiality are strictly observed. Big data have the potential to be relevant and timely and the means of collection has the potential to be more cost-effective than traditional data-collection methods.

20. The terms of reference of the Global Working Group reflect the conclusions of the report of the Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG),² entitled: "A world that counts: mobilising the data revolution for sustainable development" (November 2014). The report emphasizes that statistical offices will need to change, and more quickly than in the past, and continue to adapt, abandoning expensive and cumbersome production processes, incorporating new data sources, including administrative data from other government departments, and focusing on providing data that are human- and machine-readable, compatible with geo-spatial information systems and available quickly enough to ensure that the data cycle matches the decision cycle.³ The report of the Independent Expert Advisory Group also explicitly refers to big data and determines in that regard that national capacity for data science must be developed to leverage opportunities residing in big data and complement official statistics. An increase in domestic resources and international support for developing countries are needed to enable the data revolution to contribute to sustainable development. Through bilateral and international coordination and collaboration, common and standardized applications of big data for official statistics must be developed and scaled up transparently, while demonstrating full compliance with applicable laws.⁴

21. Among the many proposals in the report of the Advisory Group is that regarding the creation of a global "network of data innovation networks" to enable the bringing together of the organizations and experts in the field. This would contribute to the adoption of best practices for improving the monitoring of the new sustainable development goals, enable identification of areas where common data-related infrastructures could address capacity problems and improve efficiency, encourage collaborations, enable identification of critical research gaps and create incentives to innovate.

22. In summary, statistical agencies should choose data sources with regard to their quality, timeliness, costs and response burden. Big data sources fall within the scope of what is desirable in this regard for(as noted above) big data have the potential to be relevant and timely and collection of big data has the potential to be

² The Secretary-General appointed this Independent Expert Advisory Group (IEAG) on 29 August 2014 to draft a report on the data revolution within the context of the post-2015 development agenda. The Advisory Group released the report on 6 November 2014.

³ "A world that counts", p. 9.

⁴ "A world that counts", p. 23.

more cost-effective than traditional data collection methods, and could enable the data cycle to match the decision cycle. The work on big data should contribute to the adoption of best practices for improving the monitoring of the new sustainable development goals under the post-2015 development agenda. Some of the new indicators or proxies of those indicators could be based on big data sources, with improved timeliness and granular social and geo-spatial breakdown.

23. Against this background, the Global Working Group formulated its terms of reference and mandate which would be (a) to provide strategic vision, direction and coordination of a global programme on big data for official statistics, including for indicators of the post-2015 development agenda; (b) to promote practical use of big data sources, including cross-border data, while building on the existing precedents and finding solutions for the many challenges; (c) to promote capacity-building, training and sharing of experience; (d) to foster communication and advocacy of the use of big data for policy applications, especially for monitoring the post-2015 development agenda; and (e) to build public trust in the use of private sector big data for official statistics.

24. The full terms of reference including the mandate are given (as noted above) in annex I.

B. Programme of work and deliverables

25. Pursuant to the discussions held during the International Conference and the Global Working Group meeting itself, the Global Working Group agreed that its programme of work would consist of a number of work streams to be overseen by specific task teams, with one additional team responsible for overall coordination. First, it was agreed that work would continue on the three classes of big data sources focused on by the International Conference, namely, mobile phone data, satellite imagery and social media data. It was understood that advancing the work in each of the three task teams would include engagement with at least one pilot project.

26. Second, as one of the recurring themes of the International Conference had been access to data and building partnerships with the private sector and other communities, a task team was created to explore these linked topics in greater depth. This team will also investigate the possibility of establishing umbrella agreements on access to data with big data providers operating globally. Access to and use of big data are also linked to issues of public trust and communication of the benefits and challenges of big data in general. For the purpose of dealing with these issues, a task team on advocacy and communication was created, which will also investigate fundraising strategies to enable the active participation of developing countries in pilot projects.

27. Third, given the context of the post-2015 development agenda, it was further agreed to allocate to one team the especial task of keeping track of the links between the indicators needed for monitoring the sustainable development goals and the big data applications. Finally, the establishment of two more teams was proposed and agreed: one on training, skills and capacity-building (given that the skills necessary for processing big data are different from those currently available in statistical offices); and the other on cross-cutting issues such as classification and frameworks. This latter team will work, inter alia, on the classification of big data

sources, building on that proposed by the ECE big data group. The overall coordination of the work of these eight task teams will be conducted by a coordination team consisting of the team leaders.

28. It was pointed out that the Global Working Group task teams should build on the experience of the ECE big data project and on the material that is being produced by the ECE big data task teams, inasmuch as the ECE task teams will soon complete their work, while the Global Working Group task teams are only just beginning theirs. This would offer the possibility of a continuation of some of the work flows that have already been initiated. Each Global Working Group task team will need to develop its own work programme, time schedule and outputs, in close consultation with the coordination team.

29. As the deliverables of the Global Working Group are directly related to the work of the eight task teams, it is expected that the Global Working Group will deliver a number of proposals and reports, as detailed below, and will participate in a number of pilot projects as well. The Global Working Group will also continue working on an inventory of big data projects, probably through its coordination team. The results of the first survey are presented in the next section. The concrete deliverables for the year 2015 comprise:

- Proposals and activities for training, skills and capacity-building required to exploit big data sources for official statistics
- A report on links between big data and the sustainable development goals
- A strategy on advocacy and communication in respect of the use of big data for official statistics
- A report on the use of mobile phone data for official statistics, which clarifies and provides solutions to methodological, information technology and privacy challenges, and includes possible uses of those data for monitoring the post-2015 development agenda
- A report on the use of social media data for official statistics, which clarifies and provides solutions to methodological, information technology and privacy challenges, and includes possible uses of these data for monitoring the post-2015 development agenda
- A report on the use of satellite imagery and remote sensing data for official statistics, which clarifies and provides solutions to methodological, information technology and privacy challenges, and includes possible uses of these data for monitoring the post-2015 development agenda
- Proposals for enhancing access to big data sources and partnerships with the private sector
- A report on cross-cutting issues, classification, frameworks and taxonomy

30. In addition to providing the concrete deliverables of the different task teams, the Global Working Group will participate in a number of pilot projects on the implementation of the use of big data for official statistics and will also continue working on an inventory of big data projects, building on the ECE big data project inventory and on the results of the initial survey conducted by the Statistics Division and ECE, as discussed in the next section.

IV. Results of the survey on big data strategies and project inventory by the Statistics Division and ECE

31. In September 2014, the Statistics Division and ECE conducted a joint survey on big data projects in official statistics. The goal was to provide an overview of active big data projects in order to facilitate a more informed discussion within the community at large and to further develop the programmes of the Global Working Group. The survey was sent to all offices participating in the ECE big data project and to the members of the Global Working Group on Big Data for Official Statistics.

32. The survey had two objectives: collecting information on completed, ongoing or potential big data projects, including information on partnerships, data sources and tools, and on the organizational setting of those projects. Two distinct questionnaires were distributed. The aim of the first was to collect information on overall big data organization, strategies and governance structures, while the focus of the second was on actual big data projects. For the purposes of this survey, a fairly broad definition of big data was adopted, according to which they were determined to be “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”.

33. In total, the survey was sent to 78 national statistical offices and 28 international organizations.⁵ Thirty-two countries responded to the questionnaire on the organizational context, while 24 countries responded to the individual project questionnaire, submitting information on a total of 54 projects. In addition, international organizations submitted information on three projects. It is fair to assume that these countries and international organizations probably account for a substantial number of completed and ongoing big data projects worldwide. Nonetheless, the Global Working Group considers conducting a global follow-up survey with a view to collecting information on more projects for the big data project inventory. The full analytical report of the findings of this survey is provided as a background document. Below is a summary encompassing some of the main findings.

A. Main findings: organizational context

34. The questionnaire on the organizational context focused on the general experience of and frameworks for handling big data, in terms of big data strategy, governance and management structures, assessment of quality, privacy and confidentiality issues and skill shortage. The responses collected support the conclusions of the International Conference on Big Data for Official Statistics, regarding the major challenges facing national statistical offices.

35. The survey revealed that while only a few countries have developed a long-term vision for the use of big data, a number of countries are currently on the verge of formulating a big data strategy. Furthermore, a number of countries have established internal labs, task teams or working groups to carry out pilot projects designed to determine if and how big data could be used as a source of official

⁵ International organizations were e-mailed the format in large part to inform them that the survey was being conducted.

statistics, but most countries have not yet defined business processes for integrating big data sources and results into their work and do not have a defined structure for managing big data projects.

36. In many countries, steps are being taken to build partnerships for the purpose of exploring big data opportunities as part of the formulation of a big data strategy. In order to minimize the risk of failure in exploring these new technologies, countries have decided to actively participate in international collaborations such as the ECE Sandbox project, the Eurostat Task Force on Big Data and the Global Working Group on Big Data for Official Statistics established pursuant to a decision of the Statistical Commission.

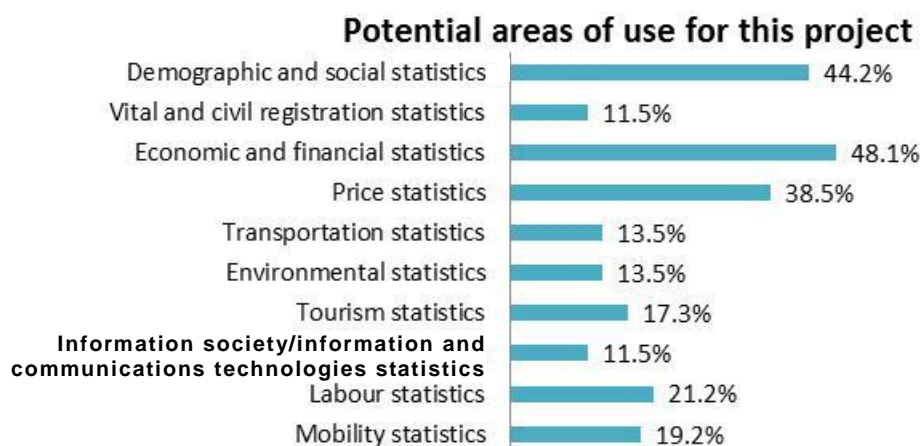
37. Only a few countries have a specific privacy framework for big data, while most survey respondents indicated that the privacy framework for traditional statistics is currently applied to big data as well. Given the importance of their public image, many organizations stress the importance of protecting privacy and confidentiality when dealing with big data, to an extent that exceeds even the bounds of what is strictly required by law. Some organizations avoid the issue of privacy and confidentiality by conducting all the testing and manipulation of the data at the location of the big data provider, and then transmit only the aggregates to the national statistical office. Some organizations argue that while strict data protection regulations are needed, they often erect a barrier to securing access to the data.

38. While most respondents recognize the challenges associated with information technology, skills, legislation and methodology, the biggest challenge for most big data projects is the limited or restricted access to potential data sets. Since big data are, to a large extent, owned by the private sector (e.g., online companies, mobile phone operators and banks), achieving close collaboration with the private sector becomes highly important. Many of these players are global companies; hence, the global statistical community could use their collective bargaining power to obtain access to these data sets.

B. Main findings: individual projects

39. The questionnaire on individual projects contained more detailed questions on the potential areas of use, status of project, outcomes and lessons learned, as well as detailed questions on the partnership arrangements, data sources and data analysis, tools and skills used. Many of the projects for which information was collected by the survey were ongoing and conducted for purposes of testing and experimenting with the use of big data. Some, on the other hand, had been completed. The other projects were in a more or less advanced exploratory stage.

40. In terms of potential areas of use, most of the projects encompass economic and financial statistics, demographic and social statistics, and price statistics (see the figure below). It is worth noting that many projects are often applicable to several statistical areas. Such projects exploit one type of data sources for use in different statistical areas. For example, the National Institute of Statistics and Geography of Mexico collected a sampling of messages posted on Twitter, and utilized the data to produce different types of statistics, i.e., indicators of subjective well-being, and tourism and border mobility statistics.



41. Just over half of the projects are based on partnerships, the most common partner being a commercial enterprise, followed by other government agencies. The most common type of partnership is one with a data provider for the purpose of securing access to data. This category includes partnerships with mobile phone companies, smart-meter data providers and social media data aggregators. For example, for the purpose of conducting a project on the production of daily mobility patterns, the National Statistical Institute of Italy partnered with a mobile phone data provider. The Federal Statistical Office of Switzerland engaged in a partnership with the four major retail stores in that country to be able to receive scanner data twice a month in order to produce consumer price indices for food and near-food commodities.

42. Besides creating partnerships with commercial data providers, national statistical offices have also established partnerships with analytical partners for many projects. This could be for the purpose of acquiring the specific know-how of the big data source, transforming raw data into the intended data structure or applying analytical techniques to produce certain outputs. However, this kind of partnership could also be set up to secure human resources capable of performing labour-intensive tasks. For example, the National Statistical Institute of Italy partnered with a university in Mexico to obtain assistance from a large number of students in classifying the content of tweets so as to enable the analysis of subjective well-being.

43. Overwhelmingly, survey participants chose internal hosting solutions over the purchasing of external hosting services. Privacy concerns were cited as the primary reason for internal hosting. External hosts were used more often when the data were deemed to be already in the public domain. On the other hand, internal hosting could become a limiting factor in respect of the choice of a big data source, inasmuch as investments in internal hosting would need to cover hardware, software and human skills, which could represent a significant hurdle to the commencement of a big data project, especially given the fact that technology is rapidly changing.

V. Conclusions and points for discussion

44. The statistical community has started to explore in earnest the opportunities offered by big data sources for application in official statistics. The International

Conference on Big Data for Official Statistics held in Beijing showcased applications of several big data sources for a number of different types of official statistics, and fostered experience-sharing between developed and developing countries. Many of the challenges facing national statistical offices are similar — such as achieving access to data, privacy concerns, methodological issues, expanded computing resources requirements, and lack of in-house skills for big data analytics — and underline the need for global collaboration. The Global Working Group has begun work by formulating its operating modalities and focal areas. It has also initiated the process of preparing an inventory of ongoing activities regarding the use of big data for official statistics. Through the survey on big data projects, a number of interesting examples were collected which may offer other organizations important lessons learned. The Global Working Group in collaboration with regional working groups needs to develop further guidance and conduct a number of pilot projects with the objective of consolidating and accelerating progress in the use of big data for official statistics.

45. The Statistical Commission is invited to express its views on:

(a) The terms of reference and mandate of the Global Working Group on Big Data for Official Statistics;

(b) The proposed task teams of the Global Working Group and corresponding deliverables on training, skills and capacity-building; linking big data and the sustainable development goals; advocacy and communication; mobile phone data; social media data; satellite imagery; access and partnership; and cross-cutting issues, such as classifications and frameworks;

(c) The results of the Statistics Division/Economic Commission for Europe survey and the plans to expand this into a global survey;

(d) The pilot projects, which the Global Working Group intends to conduct in collaboration with the regional working groups, with a view to consolidating and accelerating progress in the use of big data for official statistics.

Annex I

Terms of reference and mandate of the Global Working Group on Big Data for Official Statistics

The statistical community has the obligation of exploring the use of new data sources to meet the expectation of the society for enhanced products and for improved and more efficient ways of working. The terms of reference and mandate of the Global Working Group on Big Data for Official Statistics are based on this principle, but should certainly also be understood within the context of fulfilling the new data demands posed by the monitoring and reporting requirements under the post-2015 development agenda.

The main deficiency of the indicators for monitoring the Millennium Development Goals was lack of timeliness (or availability). To improve on that situation for the goals of the post-2015 development agenda, the High-level Panel of Eminent Persons on the Post-2015 Development Agenda, in its report of 30 May 2013 (“A new global partnership: eradicate poverty and transform economies through sustainable development”), called for a data revolution, which would draw on existing and new sources of data in order to fully integrate statistics into decision-making, promote open access to, and use of, data and ensure increased support for statistical systems.

The present terms of reference also refer to the report of the Independent Expert Advisory Group on a Data Revolution for Sustainable Development, which emphasized that statistical offices will need to change, and continue to adapt, abandoning expensive and cumbersome production processes, incorporating new data sources, including administrative data from other government departments, and focusing on providing data that are human- and machine-readable, compatible with geo-spatial information systems and available quickly enough to ensure that the data cycle matches the decision cycle.^a

Within this context, big data sources are recognized as constituting an important part of the data revolution needed to support the monitoring of the post-2015 development goals. Big data could contribute to improving some aspects of the quality of statistics, such as timeliness and relevance, without compromising their impartiality and methodological soundness.

Also, the Fundamental Principles of Official Statistics^b encourage the use of new data sources such as big data, as they state that:

- Official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens’ entitlement to public information (principle 1)
- Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents. (principle 5)

^a IAEG, “A world that counts: mobilising the data revolution for sustainable development”, p. 9.

^b See General Assembly resolution 68/261.

- Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes. (principle 6)

Guided by these strategic considerations within the context of the Fundamental Principles, the post-2015 development agenda and the data revolution as a basis, and with reference to Statistical Commission decision 45/110, the mandate of the Global Working Group is formulated as follows:

(a) To provide a strategic vision, direction and coordination for a global programme on big data for official statistics, including for indicators of the post-2015 development agenda;

(b) To promote practical use of big data sources, including cross-border data, while building on existing precedents and finding solutions for the many existing challenges, including:

- Methodological issues, covering quality concerns and data analytics
- Legal and other issues in respect of access to data sources
- Privacy issues, in particular those relevant to the use and reuse of data, data linking and re-identification
- Security, information technology issues and management of data, including advanced means of data dissemination, assessment of cloud computing and storage, and cost-benefit analysis

(c) To also promote capacity-building, training and sharing of experience;

(d) To foster communication and advocacy of the use of big data for policy applications, especially for the monitoring of the post-2015 development agenda;

(e) To build public trust in the use of big data for official statistics.

Annex II

Members of the Global Working Group on Big Data for Official Statistics

Countries

Australia
Bangladesh
Cameroon
China
Colombia
Denmark
Egypt
Indonesia
Italy
Mexico
Morocco
Netherlands
Oman
Pakistan
Philippines
United Arab Emirates
United Republic of Tanzania
United States of America

Organizations

Economic and Social Commission for Asia and the Pacific/Statistical Institute for Asia and the Pacific
Economic Commission for Europe
Eurostat
International Telecommunication Union
Organization for Economic Cooperation and Development
Statistical Centre for the Cooperation Council for the Arab Countries of the Gulf
Statistics Division of the Department of Economic and Social Affairs of the United Nations
United Nations Global Pulse
Universal Postal Union
World Bank