Economic and Social Council

Statistical Commission
Forty-eighth session
7-10 March 2017
Item 3 (d) of the provisional agenda*
Items for discussion and decision: big data for
official statistics

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 2016/220 and past practices, the Secretary-General has the honour to transmit the report of the Global Working Group on Big Data for Official Statistics. Over the past three years the Working Group has identified many promising public and private big data initiatives at the national and regional levels, of both public and private agencies, which aim to make data, services and applications accessible and to accelerate their synergies for research and capacity-building. The progress made by the Working Group is presented in the report, together with ideas and proposals for future work. The report presents a proposal for a global platform to bring together all the initiatives mentioned that are of relevance to furthering the use of big data and alternative data sources in a broader context at the global level. The platform could 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 the auspices of the Statistical Commission. The Commission is invited to comment on the proposal for the global platform and on the progress and recommendations made by the Working Group. Points for discussion are contained in the last section of the report.

---

* E/CN.3/2017/1.

Please recycle

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

I. Introduction

1.   The Statistical Commission created the Global Working Group on Big Data for Official Statistics at its forty-fifth session, held in 2014. In accordance with its terms of reference (see E/CN.3/2015/4) and decision 46/101 of the Commission (see E/2015/24-E/CN.3/2015/40), the Working Group provides strategic vision, direction and coordination of a global programme on big data for official statistics, including for indicators in the 2030 Agenda for Sustainable Development. It also promotes capacity-building, training, the sharing of experiences and the use of big data for policy applications.

2.   The Working Group reported to the Statistical Commission at its forty-sixth and forty-seventh sessions, held in 2015 and 2016, respectively. The Working Group presented highlights of its first and second global conferences, held in Beijing and Abu Dhabi, respectively, and of its programme of work, covering training, skills development and capacity-building; the linking of big data to 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, including mobile phone data, social media data and satellite imagery.

3.   In 2016, the Commission requested the Working Group to: complete a package of initiatives to improve access to proprietary data, such as big data; develop training courses for the use and application of big data in cooperation with regional training institutes, with a view to developing countries in particular; and further develop the big data quality framework to ensure trust in the compilation of official statistics;

4.   The Commission also requested that the following be added to the work programme: a review of web scraping for use in the collection of big data; the improvement of the group website and the big data repository; advocacy for big data in developing countries; and the linking of the work of the Working Group to that of the High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development.

5.   The Working Group acted on the requests of the Commission, as shown in the overview of the progress made by the Working Group in 2016 in section II and in the outcome of the third Global Conference on Big Data for Official Statistics, described in section III. Section IV contains a special focus on access to proprietary data. Section V presents a proposal for a global platform for big data, services and applications. The report concludes with several points for discussion.

II. Progress of the Working Group

6.   The Working Group held its third Global Conference on Big Data for Official Statistics in Dublin in 2016. It formulated a proposal for a global platform for collaboration on data, services and applications and reached agreement on bringing

access to proprietary data within the overall review of the quality framework for official statistics. It also achieved progress on other issues, including collaboration on pilot projects in the "sandbox" of the Economic Commission for Europe; the development of a number of handbooks on the use of satellite data, mobile phone data and social media data; and the upgrading of its website, including with respect to the availability of the inventory of big data projects. Details on the progress of the various task teams in their specific areas of work are provided in the current section. The summary of the main outcomes of the third Global Conference, the proposal for a global platform and the recommendations on access to proprietary data are discussed in subsequent sections.

A. Satellite data

7.   The Working Group task team on satellite imagery, geospatial data and remote sensing developed a handbook that contains information on sources of Earth observation data, methodologies for producing crop statistics and other statistics through the use of satellite imagery data, outlines of the task team's pilot projects and guidance for national statistical offices) in exploring the use of Earth observation data for the first time. The pilot projects include the application of satellite imagery data in the production of agricultural statistics by the Australian Bureau of Statistics, the use of proprietary satellites for commodity inventory assessment by Terra Bella (a subsidiary of Google), the environmental analysis of climate scenarios in various regions of Mexico by the National Institute of Statistics and Geography of Mexico INEGI) and the use of satellite images in compiling statistics on land cover and land use by the National Administrative Department of Statistics of Colombia (DANE). The handbook is expected to be finalized in 2017, after it has been reviewed during an expert group meeting with crop production and land cover experts from the Food and Agriculture Organization of the United Nations and the European Commission.

8.   The task team is also developing a hands-on course to teach methods for using Earth observation data in generating agricultural crop statistics and provides case studies of applications of Earth observation data for measuring progress towards the Sustainable Development Goals. The course material provides a background in the use of satellite imagery data for official statistics and advice on selecting data sources and methods for generating statistical outputs. It includes a combination of presentations, discussion sessions, demonstrations of how to use the software package "R", an example of the processing of freely available satellite data for analysis within Google Earth Engine and hands-on assistance in producing statistical outputs from satellite imagery data. The material will be further attuned to use in ecosystem accounting for agricultural areas. The course material will be reviewed at a workshop in January 2017, prior to the holding of other workshops in Asia and Latin America.

B. Mobile phone data

9.   The task team on mobile phone data started its preparation of a handbook, expected to be finalized in 2017, that builds on experiences from several Eurostat

projects and a project in Oman involving the company Positium. It can further build on the work done within the Centres and Networks of Excellence of the European Statistical System (ESSNet) on projects in Belgium, France, the Netherlands, Slovenia and Spain. The handbook will include business models for partnerships between national statistical offices and mobile operators and applications for tourism statistics (at the international, domestic and subnational levels), daytime population statistics (same-day tourism statistics), event statistics, transport statistics and population density statistics. The handbook will provide practical guidance, information on country practices and numerous references, which should prove useful for developing countries in particular.

C. Social media data

10.   Along similar lines, the task team on social media data has started to develop a handbook that is expected to be completed in 2017. It explains how to use social media data with big data techniques for producing statistics. The handbook will include descriptions of various social media data sources (such as Twitter and Facebook), of corresponding methods and techniques, including natural language processing, and of methods in extracting information from the non-message parts of social media and from the contents of web pages. Applications that can compile consumer confidence indexes, report on public opinion indicators and analyse mobility and sentiment will be included. In addition to the handbook, the Global Pulse team, a member of the social media task team, has reached an agreement with Twitter on the use of substantial amounts of monthly tweets, which can be shared with the members of the Working Group, possibly through their use in the "sandbox".

D. Big data project inventory

11.   In addition to the Working Group survey on big data projects of 2015, the task team on big data and the Sustainable Development Goals surveyed a variety of institutions on the use of big data specifically with respect to the Goals. The information collected has further augmented the inventory of big data projects. The inventory is a catalogue of big data projects relevant to the production of official statistics, Goal indicators and other types of statistics used for the management and monitoring of public sector programmes and projects. Procedures for the maintenance and updating of the inventory have been standardized, and the mapping of big data projects to the Goals and their targets should further enhance its usefulness for purposes of official statistics. With the establishment of that unique big data inventory, further efforts will be made by the Working Group to promote that catalogue of practices in the use of big data in the official statistical community and in the wider big data community.

E. Training, skills and capacity-building

12.   The more the official statistical community utilizes big data and the new information technology infrastructure in the preparation of official statistics, the

greater the need for new skills and new IT-based methodologies in order to apply big data in multisource statistical production processes. The different sources of big data pose challenges not only in the processing and analysis of the data, but also in the use of IT infrastructure. To overcome those challenges in the transfer of knowledge, the task team on training, skills and capacity-building has focused its work on four main objectives:

(a) Developing methods and tools (including coordination and facilitation) for a baseline needs assessment for big data skills in national statistical systems and an assessment of their institutional readiness to use big data for statistical production;

(b) Providing guidance on the development of a modular training programme based on the outcome of the baseline needs assessment of skills and institutional readiness;

(c) Facilitating the establishment of a global network of training and research institutions for training and capacity-building in the use of big data in statistical production;

(d) Coordinating more advanced training programmes in line with the global survey on big data for official statistics and training courses developed by the various tasks teams of the Working Group.

13. For 2017, the task team has formulated the following deliverables:

(a) An assessment of basic skill needs and institutional readiness through a new round of the global survey on big data covering national statistical agencies;

(b) Training courses to develop the basic foundational skills required and institutional readiness for the use of big data in statistical production processes, in collaboration with national statistical offices, academic institutions and private sector partners;

(c) A global capacity-building network consisting of national and regional training and research institutions and data innovation centres.

III. Outcome of the third Global Conference on Big Data for Official Statistics

14. The third Global Conference took place in Dublin on 30 and 31 August and 1 September 2016 and was attended by some 250 statisticians from around the world. The conference was opened by the Minister of State for Financial Services, e-Government and Public Procurement of Ireland, Eoghan Murphy, who placed the relevance of big data in the context of the 2030 Agenda for Sustainable Development. The conference demonstrated the progress achieved and the outcomes of the big data projects using mobile phone data, social media data and satellite data in a variety of statistical applications. Progress was also shown in the areas of capacity-building, data access and partnerships, quality and methodology and

communicating the value of big data. All presentations have been made available on the website of the conference.[1]

15. The conference covered three main themes, including access and partnerships for big data, capacity-building strategies for the transformation and modernization of official statistics and a multisource and big data statistical production environment for official statistics, including on monitoring with respect to the Sustainable Development Goals.

A. Access and partnerships

16. The high-level panel on access and partnerships stressed that access to mobile phone, social media and other big data through public-private partnerships would bring enormous benefit to both the private partners and the national statistical offices in their service delivery. The partnerships should build on the technological cloud computing infrastructure, the data and data analytics expertise of the private sector and public trust in the independent and professional role of national statistical offices in providing impartial high-quality information. Maintaining public trust was identified as an essential element in the use of big data for official statistics. To that end, full transparency must be observed in the collection, storage and use of mobile phone or social media data, all while acting in accordance with the regulatory and legal frameworks for official statistics and data protection. The benefits include reducing the response burden and costs and improving the timeliness and quality of statistics. However, the official statistical community should recognize that innovative financing mechanisms have to be explored to meet the for-profit interests of the private sector in the partnership. Further discussion of access to proprietary data is contained in section IV of the present report.

B. Capacity-building for innovation in official statistics

17. Business-as-usual in delivering statistical services to national and international clients is no longer tenable given the need to manage the down-side risks posed by external initiatives, competition from other providers of statistics and budget cuts, all while reaping the benefits of the changing information and communications technology and data landscapes through rapid technological development and the data revolution.

18. The panels on capacity-building indicated that a multifaceted transformation of national statistical systems is needed to meet the new challenges and to reap benefits from the new opportunities. Existing capacity-building programmes should be broadened and possibly be focused on transforming the technology architecture and the workforce, exploiting more data sources and redirecting products and services. Regarding the transformation of the technology architecture, the panels indicated that it should facilitate the shift from physical information technology equipment on-site towards the introduction of a cloud-computing environment along with the adoption of a common services and application architecture for data collection, registers, metadata and data management, analysis and dissemination.

_____

[1] Available from http://unstats.un.org/unsd/bigdata/conferences/2016/default.asp.

19. Such investment in the transformation of the technology architecture should be accompanied by capacity-building programmes that support the progressive diversification of the new skill sets of the staffs of the national statistical systems, ranging from data scientists and data engineers using new multisource data and modern technology, to lawyers strengthening the legal environment, to managers leading the change in corporate culture with a continuously improving quality standard. Those new capabilities should allow for the adoption of a standardized corporate business architecture that is flexible and adaptable to emerging demands and should be process-based rather than product-based, with an increasing use of administrative and big data sources (and away from primary data sources) for multiple statistical outputs. In addition, our dissemination and communication strategy should be upgraded and made adaptable in order to target different segments of users by applying a diverse set of data dissemination techniques, including mobile device applications and data visualization of key findings.

20. It was proposed that the regional statistical training programmes should increasingly broaden their capacity-building programmes for both senior managers of national statistical systems with respect to institutional transformation and modernization and for the technical and information technology staff of those systems with respect to data engineering, data analysis and data science capabilities. In order to build capacity through those new training programmes, regional statistical training institutes could adopt regional and global cloud computing platforms with multiple-source and big data processing capabilities. As a corollary, the creation of an international network of regional training and research institutes connected through regional and global platforms was proposed that would provide institutional, technological and data infrastructure for continuous learning of new statistical methods, new management skills and new information technology techniques. Regional networks linked through a global platform would complement the existing national training institutes and newly created national data and research centres.

21. The panel on smart investments in innovation of official statistics indicated that the rapidly changing technological and data landscapes require a rethinking of the support of the donor community to the national, regional and international statistical systems. Initial considerations in that fast-changing environment include focusing on smart investments in small projects that pilot, test and subsequently scale innovative methods and technologies. Such emerging projects and programmes require new collaborative and strategic public-private partnerships between the official statistical community, academia, technology companies and proprietary data owners. Furthermore, a shift seems necessary to handle the recurring costs of having continuous access to private big data sources and modern cloud-based technology infrastructure. That requires a change in donor disbursements from the upfront funding of structural investments in national statistical systems to making a series of investments over time in common scalable regional and global platforms for access to data, services and applications through strategic partnerships.

C. Big data and the indicators for the Sustainable Development Goals

22. Members of the panels on big data and the Sustainable Development Goals also underlined the need for a mixed-mode and multisource approach to meeting the demands of the Goal indicator framework. A challenge is posed not only by the sheer number of indicators, but also by the requirement to deliver on the indicators at a granular level and "to leave no one behind". Administrative and big data sources resources will be needed in order to provide such a granular breakdown. With the emerging demands for access to technology, data and applications in support of major statistical programmes, promising developments have taken place in Senegal, Cabo Verde and South Africa with respect to the creation of national and regional technology and data centres. Overall, the panellists of the third Global Conference supported the proposed global platform as a positive development for international and regional cooperation in sharing data, services and applications, which will help to meet the challenges posed by the monitoring of the Goal indicators.

23. Additional information on the outcomes of the third Global Conference is provided in a background document to the present report.

IV. Recommendations for access to big data as an integral part of a quality framework for official statistics

24. The fifth principle of the Fundamental Principles of Official Statistics concerns access to data. It states that "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". That principle could be interpreted as including access to big data.

25. The Working Group is concerned that access to big data means in many cases access to data held by private proprietary data owners with commercial business interests. That fact makes access to big data different from access to traditional data sources and raises concerns regarding the agreements that statistical agencies are allowed to make with those data owners, all while maintaining independence, assuring privacy and confidentiality and producing high-quality statistics. The issues to be considered include the equal distribution of the burden across data owners, the cost and effort required to provide data, the role of data in the value propositions of businesses, the achieving of a fair balance between public and business interests and the operational arrangements between statistical agencies and data providers. Such agreements may need to cover transparency of data provision, provide clear descriptions of metadata and include a public disclaimer on the use of the data to ensure both the quality of official statistics and the public trust.

26. In 2015, the Working Group drafted a set of principles for access to big data sources for official statistics,[2] with the objective of helping statistical organizations to obtain access to new data sources, while striking a balance between the legitimate

---

[2] Available from http://unstats.un.org/unsd/trade/events/2015/abudhabi/default.asp.

interests of organizations, often in the private sector, that hold big data, and the need of the public to have better official statistics. The access principles are linked to the Fundamental Principles of Official Statistics and appeal to the sense of social responsibility of the data providers. The Working Group consulted with stakeholders on the content of the access principles on a number of occasions in 2015 and generally received positive feedback.

27.   At its meeting in Dublin in August 2016, the Working Group discussed the next steps to be taken so that the principles are available to and acknowledged by the global statistical community. First, it was agreed that, since "principles" may be too strong a term to gain acceptance by any intergovernmental body (it took 20 years to adopt the Fundamental Principles), relabelling them as "recommendations" would be preferable. In addition, further consultations of stakeholders would be carried out through a targeted survey. Most importantly, it was agreed that the recommendations on access should be considered by the Commission within an overall framework consisting of the Fundamental Principles of Official Statistics, existing codes of practice for official statistics and national quality assurance frameworks.

28.   A report on the Fundamental Principles of Official Statistics (E/CN.3/2017/9) submitted to the Commission at its forty-eighth session, requested that the Commission consider establishing a "friends of the chair" group, which would be mandated, among other things, to revisit the question of how the Fundamental Principles should be interpreted or extended to the private sector, especially in the context of the use of proprietary data. The "friends of the chair" group may want to take into consideration the concerns of the Working Group with regard to access to proprietary data. Additional information on access to proprietary data is provided in a background document to the present report.

V.   Global platform for data, services and applications

29.   In November 2014, the Secretary-General transmitted the report of the Independent Expert Advisory Group on a Data Revolution for Sustainable Development, entitled "A world that counts: mobilizing the data revolution for sustainable development", which contained recommendations on the following: (a) technology, innovation and analysis to establish a network of data innovation networks for leveraging and sharing data and data research; (b) capacity-building and resources related to capacity-building and technology transfer, data literacy and resource mobilization through innovative financing mechanisms in partnership with the private sector; and (c) governance and leadership related to partnerships and coordination between governments, the private sector, non-governmental organizations, the media and academia to promote good practices and principles in data sharing, open data and data rights.

30.   Over the past three years, the Working Group has identified many promising public and private big data initiatives at the national and regional levels, of both public and private agencies, which aim to make data, services and applications accessible and to accelerate their synergies for research and capacity-building. The work of the Working Group task teams in the areas of satellite data, mobile phone data and social media data is to make those applications accessible through

handbooks accompanied by training materials and courses. Additional task teams are being proposed and established in promising areas with demonstrated uses in official statistics, such as price index compilation using scanner data and web scraping. However, a consolidated effort to create a platform resembling a network of data innovation networks has not yet taken shape.

31. Following its decision at a meeting on 29 August 2016 in Dublin, the Working Group proposed the creation of a global platform for data, services and applications. Building on the best practices of public and private big data initiatives, and offering technology infrastructure and a network for data innovation to the official statistical community, the global platform could address the need to have global hubs for the following:

(a) The exchange of ideas and methods for processing, analysing and visualizing big data between official statisticians, data scientists and domain experts from the public and private sectors;

(b) The storage of big data, related processing, analytical and visualization methodology and services and applications for continuous development and reuse;

(c) The demonstration of the value of big data in improving decision-making through the use of official statistics in pilot studies and case studies;

(d) Resources for training materials and workshops on big data for capacity-building.

32. The Working Group also proposed that the platform should be developed under the auspices of the Commission, which means that the Commission would decide on the business model for the platform, the road map for its development and the implementation of the platform. Under the Commission's oversight, it would be a platform by and for the official statistical community serving primarily the national statistical systems of both developed and developing countries. As indicated above, the proposal for the global platform has been moderated and supported by the various panels of the third Global Conference and the series of regional Conferences on a Transformative Agenda for Official Statistics, held in 2016. The proposal is also in line with the objectives in the strategic area of innovation and modernization of national statistical systems, contained in the draft global action plan for sustainable development data, as formulated by the High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development (see E/CN.3/2017/3) and in line with the outcome documents of the regional conferences (see E/CN.3/2017/5).

Tentative business model of the global platform

33. With the aim of assisting the Commission in thinking about the structure and content of the global platform, the following paragraphs sketch out a tentative business model for the platform. They provide one scenario for the building of the platform. The Working Group has created a committee that will explore other possible scenarios for the platform.

34. The global platform could provide scale and scalability for access to and capacity-building in the use of big data and its integration with administrative sources, geospatial information and traditional survey and census data and in the use

of the related services, applications and infrastructure. In terms of products and services, the global platform could: (a) facilitate access to and use of information technology infrastructure and proprietary data sources; (b) host specific global projects such as those on global enterprise registers and bilateral or multilateral data asymmetries; (c) store and provide access to new open source applications for the use of multisource data sets for more timely and flexible production processes for official statistics; and (d) provide training and capacity-building services in the use of new data sources for statistical production. In order to be a truly global platform for official statistics, it was envisaged that the data access could also include access to trusted (micro) datasets from official statistical sources and access to verified and curated data provided by private and public partner organizations. Services of the global platform could include cloud storage and cloud computing, technical statistical services (including stand-alone services or those that act as building blocks for developing applications), and related capacity-building services.

35.   As an example of how the provision of services and applications could work, a project could be initiated to transfer the methodology and provide access to the data and technology infrastructure of satellite imagery data for use in the calculation of agricultural statistics in either a developed or developing country environment. Such a project would require having access to satellite data for training and testing purposes, preparing the data in a secure and appropriate information technology environment, testing the methodology, compiling the agricultural statistics and training staff to carry out related operations on a routine basis. The global platform, as a network of partners that includes statistical offices, research institutes and private companies, through its technology infrastructure, would be able to offer and deliver those services, including the services relating to training and capacity-building.

36.   Consideration also needs to be given to the business institutional environment. To work out the details of the business case for the global platform and to prepare a road map for its implementation, the Working Group has created a committee consisting of representatives of national statistical offices, including those which have recently established national data centres. The committee will apply the accumulated expertise of its members and best practices in the design and implementation of the global platform. In addition, separate dialogues will be held with private data and technology companies and international and regional agencies. In articulating the business case and road map for the implementation of the platform, the committee will consider issues such as the deployment of technological infrastructure, access to the proprietary data of private data owners, the sharing of existing and development of new big data applications, the offering of capacity-building services based on public and private sector courses, governance and financing arrangements, including the positions and roles of the Commission and the Working Group, and links to national, regional and global initiatives such as the national data centres, the "sandbox" of the Economic Commission for Europe, the African regional centres and the Global Pulse labs. An initial road map with milestones and a timetable, including for the testing of the beta version, is presented in a background document to the present report.

37.   In conclusion, the Working Group proposes to establish a global platform for data, services and applications under the auspices of the Commission. The platform

would be a network of partners, including statistical offices, research institutes and private companies, with the appropriate technology infrastructure to offer and deliver data, applications and services to the official statistical community, including services for training and capacity-building. The platform would be gradually built according to the road map described in the background document.

VI. Points for discussion

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

(a) **The progress made by the task teams of the United Nations Global Working Group on Big Data for Official Statistics;**

(b) **The outcome of the third Global Conference on Big Data for Official Statistics;**

(c) **The recommendation of the Working Group to take into account access to proprietary data in the deliberations on the proposed "friends of the chair group" on the Fundamental Principles for Official Statistics;**

(d) **The proposal of the Working Group to establish a global platform for data, services and applications servicing the community of official statistics that would be progressively developed under the auspices of the Commission.**

_____