Economic and Social Council

Distr.: General
22 December 2017

Original: English

Statistical Commission
Forty-ninth session
6 - 9 March 2018
Item 3 (e) 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 2017/228 and past practices, the Secretary-General has the honour to transmit the report of the Global Working Group (GWG) on Big Data for Official Statistics. Over the last four years, the working group identified many initiatives, which aim to make new (Big) data sources, corresponding services and innovative applications accessible, and tried to harness their use in research and capacity building for statistical production processes. This report responds to the request made by the Commission at its last session and reports on the progress made in the articulation and development of a global platform. This platform will be developed by the GWG for global collaboration on trusted data, trusted methods, trusted partners and trusted learning, and is envisaged as a marketplace for sharing and developing trusted data, services and applications. The GWG platform will operate as part of a federated network of autonomous platforms at national, regional and global level with defined interfaces to ensure interoperability and information sharing. The GWG task teams will collaborate on their use cases through the GWG platform, which will be developed and operated under the auspices of the UN Statistical Commission. The Commission is invited to comment on the articulation and development of this GWG platform, on the progress made in 2017 and actions proposed for 2018. Points for discussion are contained in the last section of the report.

---

*   E/CN.3/2018/1.

I.   Introduction

1.   The Statistical Commission created the Global Working Group (GWG) on Big Data for Official Statistics at its 45th session in 2014. In accordance with its terms of reference[1] and Decision 46/101 (iii)[2], the GWG provides strategic vision, direction and the coordination of a global programme on Big Data for official statistics, including for compilation of the SDG indicators for the 2030 Agenda for Sustainable Development.

2.   In Decision 48/105 in 2017, the Commission requested the GWG to develop the business case and proof of concept for a global platform of data, services and applications by applying an incremental and step-by-step approach.

3.   In this report, the GWG responds to the request made by the Commission and presents the progress with the business case and proof of concept for a GWG platform of trusted data, trusted methods, trusted partners and trusted learning supported by core services, which will enable and advance the work of the GWG task teams. Moreover, the GWG reports on its 4th annual management meeting in Bogota on 7 November 2017. This management meeting was attended by senior managers from 20 national statistical agencies and 10 international organizations, complemented by senior representatives of several NGOs and private sector companies[3]. Its agenda covered the progress reports of the task teams, the proof of concept for the GWG platform and the business case for this platform. The meeting agreed the Bogota Declaration.

4.   There is a need for general platform facilities for the larger international community involved in big data for official statistics as well as specific platform facilities for the GWG itself, in particular for its task teams. Therefore, it is expedient to make a distinction between the *global network of platforms* at national, regional and global level serving the larger international community on the one hand, and the *GWG platform*, defined as the platform of and for the GWG itself, on the other. The GWG platform is part of the global network and can be seen as its precursor.

5.   This report will highlight the main elements of the Bogota Declaration that describe the progressive articulation of the business case for the GWG platform based on ongoing practices of the task teams of the GWG. The Bogota Declaration also provides the strategic direction and vision for the GWG by placing the thrust of its work to advance innovation and modernization of national statistical systems through trusted global data collaboratives developed by the task teams under the GWG based on trusted data, trusted methods, trusted partners and trusted learning.

6.   This report describes the Bogota Declaration and the progress made in the articulation of the business case and proof of concept for the GWG platform in Sections II and III. Section IV shows the progress made in the various task teams of the GWG and elaborates on the capacity building prospects in this domain. Section V presents the outcome of the 4th international conference on Big Data in Bogota, Colombia, while Section VI describes the future actions by the GWG to advance its work programme. The report concludes with points for discussion for the consideration of the UN Statistical Commission.

II.   The Bogota Declaration

7.   The Bogota Declaration builds on the recommendations contained in the Secretary General's Data Revolution report "A World that Counts", on (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, 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

---

[1] See E/CN.3/2015/7

[2] See E/CN.3/2015/34

[3] The GWG meeting was attended by a large majority of its current members and by 5 new members, namely Switzerland, Germany, IMF, African Development Bank and UNECA. Saudi Arabia was also welcomed as a new member, even though it was unable to be represented in this meeting. The full updated list of GWG members is available in Annex I.

organizations, the media and academia to promote good practices and principles in data sharing, open data and data rights. The declaration also builds on the recommendations contained in the Cape Town Global Action Plan with respect to applying statistical standards and new data architecture for data sharing, exchange and integration, and to facilitating the use of new technology and new data sources in statistical production processes.

8.   The Bogota Declaration proposes to provide a major thrust for the strategic area of the Cape Town Global Action Plan on innovation and modernization of national statistical systems through trusted global collaboration and supported by a global network of platforms, which is envisaged as a marketplace for sharing and developing core catalogues of services, data, metadata, methods, APIs, IT tools and training materials. The global network operates as a federated network of platforms at national, regional and global level, which ensures interoperability and information sharing among the platforms in the network through agreed and defined interfaces. The GWG platform is home for usage by the GWG task teams and is envisaged initially to meet the requirements for research and development in the use of multi-source data. However, when relevant and legally possible, it will gradually incorporate the use of multi-source data for statistical production, including the production of SDG indicators. The GWG platform is expected to be developed and operated under auspices of the UN Statistical Commission.

9.   Through this declaration the GWG recommends fostering global collaboration, facilitated by the global network of platforms. The GWG task teams can collaborate on their own GWG platform in this federated network and should allow trusted data, trusted methods, trusted services and trusted applications to be shared as a public good, where useful and legally possible. Transparent partnership agreements will need to be developed with private- and public-sector organizations so that GWG platform partners contribute and derive value through a business model which is individually sustainable for all stakeholders and ensures access to a trusted data and technology infrastructure.

10.   The GWG unanimously agreed upon the Bogota Declaration in person on 7 November and further in writing through subsequent exchange of emails. The full text of the Bogota Declaration can be found in Annex II.

III.   The GWG Platform

11.   The Bogota Declaration sets forth the details and drivers of the GWG platform as facilitator for sharing, exchanging and developing trusted data and metadata, trusted methods, trusted partners and trusted learning. This collaboration will gradually develop catalogues and libraries for data, metadata, methods, partners and learning on the GWG platform (see illustration below).

12.   The GWG platform will be part of an interconnected and federated network of platforms, which is based on the best practices of private and public Big Data initiatives. The GWG platform offers technology infrastructure and constitutes a network for data innovation to the official statistical community to facilitate global collaboration on:

(a)   the exchange of ideas and methods for processing, analyzing and visualizing Big Data between official statisticians, data scientists and domain experts from the public and private sectors with a focus on research and development building towards modernized statistical production;

(b)   the sharing and exchange of trusted metadata, methods, services and applications for continuous development and reuse, including sharing trusted data where useful and legally possible. The sharing of sensitive data between trusted partners will be agreed bilaterally, whilst widely applicable open data sources can be generally made available; and

(c)   the development, jointly among the official statistical community, private sector technology companies and other communities, of a trusted data architecture so different types of data can be shared safely and securely.
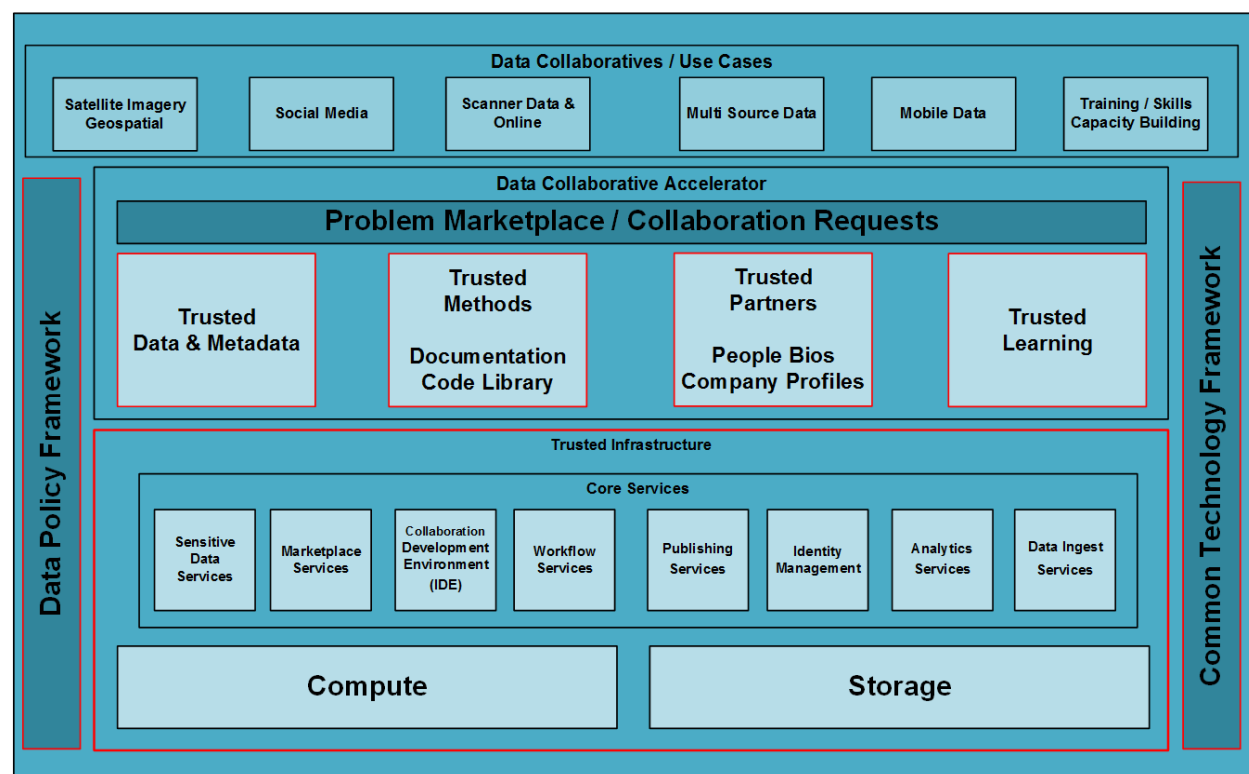
Figure 1: GWG Platform Architecture

13.   In the illustration of Figure 1, the Data Policy Framework of the GWG platform is meant for data governance and information management and consists of a set of rules and regulations under which the GWG platform must operate in storing, processing and disseminating information[4]. Following these rules will reduce the risks associated with inconsistently managed information, which will, in turn, reduce the cost of handling these risks and allow for increased flexibility in a changing technology and information environment over time. The Common Technology Framework is a policy and technology framework meant to deliver data, accessible services and open source applications and associated APIs as a trusted public good to which all GWG platform public and private sector network partners subscribe and adhere to through contractual agreements.

14.   The task teams of the GWG would be the first partners to progressively develop the data and technology architecture of the GWG platform based on their requirements and on real user demands and business cases. Other proofs of concept will include the Trade Data Lake, which will initially consist of a data lake with publicly available data on international and transport data, such as the trade data from the UN Comtrade database, data from the detailed UPS database on parcel post trade, aviation data from ICAO and vessel tracking data. It is envisioned that gradually more sensitive data will be added to the platform, such as administrative data from the Customs administration (for example, ASYCUDA data), which could be processed by trusted users, under strict protocols, directly in the data lake, making use of the latest tools and applications in this Cloud server environment.

15.   The GWG platform is further expected to support capacity building via a library of trusted training materials, methods and software applications and via conducting workshops on modernization of official statistics, the use of alternative data sources (such as Big Data) and the application of new tools, services and analytical techniques.

16.   The development and maintenance of the platform will be undertaken under the auspices and guidance of the Commission, in support of national statistical systems of developed and developing

_____

[4] Some standards are already available at https://github.com/UNGlobalPlatform/standards

countries. The development and implementation of the GWG platform within the federated network of platforms will strengthen modern trusted data usage and information infrastructure for the global community of official statistics.

17.   In further developments of the GWG platform, the underlying standards, methods, models and applications, the GWG will have to make critical decisions on both technology and scientific matters. The GWG has limited expertise in these areas. Therefore, the GWG considered establishing two advisory groups, one on technology matters and one on scientific matters. These advisory groups would guide the GWG on decisions regarding matters of technological development and scientific quality. It is foreseen that these groups will have a core group of advisers, which can be temporarily supplemented with additional advisers, if certain matters so require. In addition, other advisory group could possibly be established on legal matters and on fund raising to expedite the progress made by the Task Teams, including the capacity building, testing and experimentation in pilot countries.

IV.   GWG Task Teams and capacity building

18.   The GWG has active task teams on the use of satellite imagery data, mobile phone data, social media data, and scanner data and a task team on training, skills and capacity building. Progress of each of these task teams is reported in this section. A more detailed overview of their work and their training workshops will be posted on the GWG website.

A.   The Task Team on Satellite imagery data

19.   This task team on satellite data has completed its Handbook, which contains information about sources of earth observation data, methodologies for producing crop and other statistics using satellite imagery data, outlines of the task team's pilot projects and general guidance for the statistical community exploring the use of earth observation data for statistical purposes. The Handbook is finalised and will be reviewed by the GWG advisory group on scientific matters, when established, before release on the public website of the GWG.

20.   On behalf of the GWG, Professor Kerrie Mengersen of the Queensland University of Technology (QUT) in Brisbane, Australia, a QUT colleague, a scientist from the Australian Science Department in Brisbane and a senior statistician of Statistics Canada developed a training course on the use of earth observation data for agricultural crop statistics. This course was conducted on 6-7 November 2017 in Bogota, prior to the Big Data Conference. The 20 participants on the course came from statistical institutes from around the world.

21.   The workshop gave advice on selecting data sources, on preparing raw satellite images for analysis and on different methods to produce statistical outputs. It also provided case studies on use of earth observation data to monitor key SDG indicators, demonstrations in the use the software package R and hands-on assistance to produce statistical outputs from satellite imagery data. The practical focus of the methods was on crop-type identification and classification based on Landsat 7 and ground reference data. The workshop led to a better understanding of the strengths and limitations in the use of satellite data, the management and manipulation of satellite data and the application of satellite data for crop statistics.

B.   The Task Team on Mobile phone data

22.   This task team has finalized the first full draft of its Handbook on the use of mobile phone data for official statistics, which describes in detail applications, data sources and methods. The Handbook also includes partnership business models between national statistical offices and mobile operators for access to the mobile phone data and concludes with two country cases from France and Indonesia. The Handbook will be reviewed by the GWG advisory group on scientific matters before release on the public website of the GWG.

23.   Prior to the Big Data Conference and on behalf of the GWG two data scientists from Positium, a private sector intermediary, and one statistical methodologist from Eurostat conducted a workshop on the use of mobile phone data for official statistics for about 25 participants from national and international statistical institutes. The workshop used materials based on projects with mobile operators across the EU, Middle East, and Indonesia, resulting in the analysis of mobile data from 11 countries.

24.   The trainers elaborated on the technical details of data preparation, data processing, data modelling using hands-on examples, such as applications to model population movement, migration patterns, mobility, displacement due to diseases, tourism, and real-time crowd mapping. Data related to these areas can be collected very quickly, in some cases in real-time, from the mobile phone operators. Access to the data varies from country to country based on different legal and political frameworks. The workshop concluded with a discussion on possible roadmaps for embedding the use of mobile phone data into the official statistics of the participating countries.

C.   The Task Team on Social media data

25.   The task team on Social Media data is also in the process of completing a Handbook on the use of social media data for opinion and sentiment indicators. It includes descriptions of various social media data sources (such as Twitter and Facebook), explanations of methods and techniques to analyse such data sources and descriptions of applications. Once the Handbook is finalized, it will be reviewed by the GWG advisory group on scientific matters before release on the public website of the GWG.

26.   Prior to the Big Data Conference and on behalf of the GWG, statisticians from INEGI Mexico and DANE Colombia, and a data scientist from Statistics Netherlands conducted a workshop with materials based on projects covering mobility analyses, sentiment analyses and subjective poverty measurement using mostly Twitter and Facebook data. The trainers explained the process of obtaining social media and web scraping data through APIs, the software and programming languages to clean and process the data, and the latest methodologies for data analytics, including machine learning.

27.   Understanding of the scope, nature and coverage of social media data is important in these analyses. Not all social media platforms are equally popular amongst different age groups. Analysts need to have an understanding about how many users there are and their socio-economic profile, i.e. they need to understand the population of their study. The behavior of the users (e.g. how frequently does one post?) and the reach of the posts (whether public or private) can also influence the opinions which are expressed.

D.   Task Team on Scanner data

28.   This is a new task team, which was establish in April 2017. It hit the ground running by making available the software code and statistical methods developed by Australian Bureau of Statistics, Statistics New Zealand and Statistics Netherlands for the use of scanner data in production of CPI. These statistical methods and software code were sharable and well-documented. The methods have been reviewed and released as open-source applications using R. The task team made an explicit point of keeping the Inter-Secretariat Working Group on Price Statistics informed about the progress of its work.

29.   Members of this task team have also started drafting a guide to explain and demonstrate a broad set of the methods for using scanner data in the calculation of consumer price indices (CPI). The availability of the guide and open source software will allow other statistical offices to experiment and test scanner data for potential use in their statistical production process, along with web scraped and survey data.

30.   On behalf of the GWG, statisticians from the Australian Bureau of Statistics, Statistics Belgium, Statistics Canada and Statistics Denmark conducted a training workshop on scanner and on-line data for the compilation of the CPI in Bogota the evening prior to the Big Data Conference. The workshop provided information on how to access online data, select data sources, prepare raw data for use and described the status of scanner data implementation in different countries. The workshop also contained a live demonstration using the CPI software written in R and provided hands-on assistance in scraping online data from websites. This led to a better understanding of the strength and limitations of the use of scanner data and online data and the strengths and weaknesses of several index methodologies for the compilation of CPI.

E.   Task Team on Training, Skills and Capacity-building

31.   This GWG task team organized its work around developing methods and tools to assess the needs for Big Data skills in national statistical systems, coordinating GWG training courses, developing a modular training curriculum and cooperating within the global network of training programmes on the use Big Data for official statistics. The task team worked to enrol representatives from developing countries in the European Statistical Training Program (ESTP) to ensure they were properly represented.

32.   Experiencing high demand from European statistical offices for ESTP courses related to the use of Big Data for official statistics, the GWG task team accelerated the development of its own training courses with the support of the existing GWG task teams. The success of the four training workshops organized by the task teams of the GWG in Bogota in November 2017 demonstrates that the GWG is able to offer a trusted learning program with the support of the member countries of the task teams. It is expected that a trusted learning curriculum will be further developed and announced for 2018 in collaboration with members of the task teams and the GWG management group.

V.   Outcome of the 4th International UN Conference on Big Data for Official Statistics

33.   The fourth International UN Conference on Big Data for Official Statistics was held in Bogota on 8-10 November 2017 and was attended on average by about 280 statisticians, data scientists and IT specialists from around the world with 350 participants during the first day of the Conference. The Conference was organized by the GWG in close cooperation with DANE, the national statistical office of Colombia. It took place at the Ministry of Information, Communication and Technology in Bogota, Colombia.

34.   The overarching theme of the conference was "trusted data collaboratives" to share, exchange and develop data, services and application for the modernization of national statistical systems. In his keynote speech[5] Alexandre Barbosa, Director of the Regional Center for Studies on the Development of the Information Society in Brazil, clearly laid out three strategic areas for the innovation and modernization of national statistical systems: (i) institutional reform with respect to multi-stakeholder partnerships and the leading role of the national statistical offices; (ii) new standard setting to open data, data security and interoperability; and (iii) the use of new Big Data sources and corresponding technologies, including data storage, data integration as well as machine learning and artificial intelligence. He recognised that the use of Big Data has human, technical and technological challenges. However, the major impediment is institutional in nature. He highlighted good practices in addressing these challenges by establishing multi-stakeholder "trusted data collaboratives" through – for example – national data centers and data science campuses for research and development and statistical production. At the international level, a UN-led GWG platform would complement the emerging network of national and regional platforms for the innovation and modernisation of official statistics.

35.   The Conference programme approached its main theme starting from the need for innovation and modernization of the national statistical systems making a strong link to the Cape Town Action plan regarding the monitoring of progress on the 2030 Agenda for Sustainable Development through the use of new data and innovative technology. New collaborative initiatives were presented in establishing national data centres and science campuses and in creating city data centres in selected cities and municipalities for the purposes of harmonization and value creation in the national statistical system. Subsequently, panel sessions discussed the broadening of partners in the existing global collaboratives led by the existing tasks teams in the use of earth observation data, scanner data and mobile phone data. Each of these panels expressed the role of the statistical community to set standards for trusted data, methods, partners and learning in the use of these data sources. These standards should be accompanied by enabling access and use of these standards through a GWG platform. A dedicated panel session on standards, consisting of representatives of the private and public sector, confirmed the need for defined principles and standards on encryption, data security and data protection for secure information sharing, which are subject to continuous technological update.

36.   In the closing session of the Conference, a Proof of Concept of the envisaged data and technology infrastructure of the GWG platform was presented by the Office of National Statistics (ONS) of the United Kingdom (UK). This data and technology model is built on the concept of trusted data, trusted methods, trusted partners and trusted learning to connect and create value for users and stakeholders and is supported by large sets of agreed core services, which are shareable within the international collaboration. Specific examples were shown, supported by video demonstrations covering some of the platforms core supporting services relating to sensitive data services, software development, publishing services, analytics services

---

[5] See Conference website at https://unstats.un.org/unsd/bigdata/conferences/2017/default.asp

and data ingest services. All of these proposed core services were based on open source software and are therefore freely available.

37.   The Conference concluded by sharing views on the Bogota Declaration and with the expression of commitments to collaborate on the GWG platform through the existing and through new task teams.

VI.   Proposed actions of the GWG

38.   Over the last year the GWG had already created a Committee dedicated to the articulation and development of the GWG platform. Collaboration on the GWG platform is organized through the GWG task teams as the global data collaboratives. Task teams have a known reporting mechanism within the GWG and new data collaboratives will be created as new task teams under the GWG. . Short term actions for the GWG are the completion of proof of concepts for the GWG platform, the upgrading of the GWG website to provide a better vehicle for communication and the advancement of the GWG capacity building program.

A.   Adapting the GWG governance structure

39.   The GWG has agreed to have two co-chairs because of the increased work load, in general, and because of the required specialized knowledge in building the GWG platform, in particular. It was agreed that Statistics Denmark and ONS, UK, will be the two co-chairs of the GWG. UNSD as secretariat will keep the co-chairs informed about all activities of the GWG.

40.   Statistics Denmark will oversee the GWG work programme of existing, expanding and setting up new task teams. The current management team consists mainly of the GWG task team leaders, led by Statistics Denmark as GWG chair. This team is now renamed the GWG Committee on Global Data Collaboratives, given that the task teams are functioning as global data collaboratives. Statistics Denmark will be chair this GWG Committee. The existing GWG Committee on the Global Platform will continue its work under leadership of ONS, UK. This Committee will gradually develop the business case, proof of concept and communication for the GWG platform, with a focus on its data and technology framework across a federated network of platforms. On a regular basis UNSD will inform the full GWG membership about the developments led by both GWG Committees.

41.   Two advisory groups are being considered: An Advisory Group on scientific matters chaired by Statistics Denmark, and an Advisory Group on technology matters chaired by ONS, UK. Members of the advisory groups would be invited on personal basis, determined by their accomplishments and experience in their respective fields. Appropriate Terms of Reference and the composition of the members of the groups would need to be determined. The Advisory Group on scientific matters would, as one of its first tasks, review the existing handbooks prepared by the task teams on satellite imagery data, mobile phone data and social media data, whereas the Advisory Group on technology matters would advise on the set up of the GWG platform and the interoperability of the platforms in the federated network.

42.   For the progressive detailing of the business case for the GWG platform based on the experiences of the GWG Task Teams, the Task Teams will be asked to elaborate on the legal considerations of their data collaboration like access and use of privately owned (big) data, collaboration with private sector, data integration and the like. Also, the Task Teams will be asked to elaborate on the financial implications of the operation and maintenance of their data collaboration, including capacity building, testing and experimentation in pilot countries. Depending on the progress made by the Task Teams, specific advisory groups may be established for legal and funding matters.

B.   Broadening the tasks of existing GWG Task Teams and establishing new ones

43.   Existing task teams will be requested to broaden the scope of their work by explicitly (a) including the sharing of proprietary data and methods, (b) developing training workshops for participants in developing countries, and (c) expanding their partnerships with the private sector. Moreover, tasks teams are expected to start using the GWG platform and its core services to share information. In light of the new work programme, the tasks teams will be asked to review their terms of reference in early 2018 and report back to the GWG Committee on global data collaboratives.

44.   New task teams will be created, notably a task team on the Trade Data Lake led by UNSD, a task team on multi-source data led by Statistics Denmark and a task team on data security and encryption led by ONS, UK. The GWG will continue to explore the creation of other teams on thematic issues such as economic and financial statistics, the digital economy and ICT, climate change, working with trusted partners. A critical aspect of establishing the new task teams will be to determine, for each, the chair and committed team members who will make their participation part of the regular program of work within their organisation and who are willing to work using Agile development approaches with a dedicated approach.

C.   Advancing the trusted GWG platform

45.   In 2018-19 the GWG Committee on the Global Platform will work, together with the Advisory Group on Technology, to develop a prototype platform whose use will be piloted by the existing GWG Task Teams for initial testing. The GWG platform will be built on a flexible cloud-based infrastructure initially to allow (public) data, services and applications to be shared as a public good where useful and legally possible.

46.   As soon as is feasible this GWG platform will be opened up to other 'trusted partners' to facilitate the development of trusted data collaboratives. The initial principles and processes to set out the requirements of 'trusted partners' will be agreed by the GWG Committee.

47.   Adopting Agile development principles the GWG platform will be iterated to develop and deliver a marketplace which will offer trusted partners the ability to a) collaborate between themselves to innovate and explore new data sources; b) test and share different code, statistical and data science methods; c) participate in learning and capability building of benefit to their own organisations; d) choose between a variety of core services, compute and storage solutions (including federating their own compute and storage facilities as part of the platform; and d) collaborate to agree which data sources, code and methods are considered 'trusted' i.e. tested and agreed robust enough for general purpose use in the delivery of statistics and indicators.

48.   Working with the Advisory Group on Technology, the GWG Committee on the Global Platform will seek to develop the necessary technical solutions to extend the GWG platform to allow differentiated access to data of different types and classes, by user and use-case by 2019-2020. This will enable data providers to share more sensitive data, not just public data, in a way where they can be confident about which users have access to their data and for what purposes. This Proof of Concept of differentiated access to data and data sharing will be undertaken in close consultation with countries taking into account their legislative frameworks to allow them to participate in the global network of platforms at a level which suits their local situation.

D.   Developing the GWG work programme on trusted learning

49.   The four training workshops organized in Bogota demonstrated that the GWG is able to offer a course curriculum on the use of new data source for the compilation of official statistics. The courses were conducted with strong support from the members of the GWG task teams. It is expected that this capacity building program will be further developed in 2018 with a number of courses to be conducted at regional level.

50.   There is tentative agreement for a regional training workshop on the use of satellite imagery data for crop and related statistics in Asia and the Pacific region in the second quarter of 2018. QUT would take the lead supported by UNSD. A regional training workshop on the use of mobile phone data for official statistics will very likely take place around the 3rd quarter of 2018. Positium would again take the lead supported by UNSD. The task team on Training, Skills and Capacity Building will confirm these and further activities in due course.

E.   Fifth International Conference on Big Data for Official Statistics

51.   It is proposed that the fifth International Conference on Big Data for Official Statistics will be organized by the GWG in the fall of 2019 in Africa. In the interim, the GWG will organize one or more sessions on topics related to its work at the UN World Data Forum in October 2018 in Dubai.

F.   Renaming of the GWG

52.   Further discussion will take place on renaming and rebranding of the GWG as well as its task teams. Alexandre Barbosa suggested in his keynote speech that the GWG should be renamed to bring it more in line with the full scope of activities which the GWG is undertaking. The new name should reflect the two streams of work of the GWG, namely the global data collaboratives and the global platform. Furthermore, the task teams could be renamed by theme or statistical domain rather than by data source, as is currently the case.

VI.   Points for discussion

53.   The Commission is invited to express its views on:

(i)   the progress made by the GWG on Big Data for Official Statistics and its task teams;

(ii)   the outcome of the 4th International Conference on Big Data for Official Statistics;

(iii)   the proposal for the GWG platform as a global collaborative for trusted data and metadata, trusted methods, trusted partners and trusted learning servicing the community of official statistics, which would be progressively developed under the auspices of the Commission.

ANNEX I

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

Countries

Australia
Bangladesh
Brazil
Cameroon
China
Colombia
Denmark
Egypt
Germany
Indonesia
Ireland
Italy
Mexico
Morocco
Netherlands
Oman
Pakistan
Philippines
Republic of Korea
Saudi Arabia
Switzerland
United Arab Emirates
United Kingdom
United Republic of Tanzania
United States of America

Organisations

African Development Bank
CARICOM
Eurostat
GCCSTAT
International Telecommunication Union
International Monetary Fund
Organization for Economic Cooperation and Development
United Nations Economic and Social Commission for Asia and the Pacific
United Nations Statistical Institute for Asia and the Pacific
United Nations Economic Commission for Europe
United Nations Economic Commission for Africa
United Nations Statistics Division
United Nations Global Pulse
Universal Postal Union
World Bank

ANNEX II

Bogota Declaration

Taking into consideration the recommendations contained in the Secretary-General's 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",* including:

(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.*

Recalling the adoption of the Cape Town Global Action Plan for Sustainable Development Data (see E/CN.3/2017/3) and the related outcome documents of the regional conferences on the Transformative Agenda (see E/CN.3/2017/5) by the Statistical Commission at its 48th session in March 2017.

Highlighting that the Cape Town Global Action Plan for Sustainable Development Data calls upon the global statistical community to take action on the strategic area of modernizing and strengthening the national statistical systems with a focus on modernizing the governance and institutional framework; on applying statistical standards and new data architecture for data sharing, exchange and integration; and on facilitating the use of new technology and new data sources in statistical production processes.

The Global Working Group at the 4th Global Conference on Big Data for Official Statistics in Bogota

Proposes:

- to provide a major thrust for the strategic area of the Cape Town Action Plan on innovation and modernization by advancing global data collaboratives, facilitated by a trusted federated global platform initially for research and development in the discovery, access and use of data, statistical methodology, software applications and capacity building for the production of statistics and indicators. These partnerships will innovate and help modernize official statistics and their use of new data sources, including Big Data. It will enable data driven transformation in the production of specific statistics or SDG indicators for better decision making.

- to progressively invest in research and development via task teams of the Global Working Group whose main objective is the innovation of current statistical production processes and the creation of new ways of compiling SDG indicators; define a framework for the evaluation of the task teams' work centered on the quality of the information produced as well as their cost effectiveness and scale successful projects by transforming them into data products for global consumption;

- to progressively develop the data and technology architecture of the global platform based on the requirements of the work programs of the task teams of the Global Working Group, and underpinned by real user demands and business cases;

- that this work program for trusted data, services and applications is undertaken under the auspices and guidance of the United Nations Statistical Commission, in support of the global community of official statistics by putting the national statistical systems of developed and developing countries at its heart;

- that the global platform should build on the best practices of private and public Big Data initiatives, offers technology infrastructure and a network for data innovation to the official statistical community and addresses the need for an interconnected and federated network to facilitate:

    (a) the exchange of ideas and methods for processing, analyzing and visualizing Big Data among official statisticians, data scientists and domain experts from the public and private sectors with a focus on research and development building towards modernized statistical production

    (b) the sharing and exchange of trusted metadata, methods, services and applications for continuous development and reuse, including sharing trusted data where useful and legally possible. The sharing of sensitive data between trusted partners will be agreed bilaterally, whilst widely applicable open data sources can be generally made available

    (c) the development, jointly among the official statistical community, private sector technology companies and other communities, of a trusted data architecture so different types of data can be shared safely and securely;

- to support capacity building via a library of trusted training materials and a catalogue of trusted guidance material, methods and software applications and via conducting workshops on Big Data and new analytical techniques.

Underlines:
- that the implementation of the global data collaborative as a federated system will place the community of official statistics at the heart of modern trusted data usage and information technology.
- that it will offer both developed and developing countries opportunities to realize the benefits of multi-source data, including Big Data, administrative data, census data and survey data, to better understand economic, environmental and societal changes without investment in expensive technologies.
- that the global collaborative for trusted data, services and applications will benefit all parties involved via synergies in sharing methods and data, creating a global culture of best practice and capability sharing.
- that the sharing of knowledge and capacity building in human resources in the discovery, access and use of multi-source data is a shared responsibility of national and international statistical community and should be scaled in existing statistical capacity building programs.

Recommends:
- global collaboration, facilitated by the global platform with the potential to accommodate many different types of trusted data, trusted services and trusted applications, which should

    (a) make it easy for all nations to gain value by participating in the global network;

    (b) deliver a marketplace and a flexible cloud-based technology infrastructure to allow trusted data, methods, services and applications to be shared as a public good where useful and legally possible; and

    (c) develop transparent partnership agreements with private- and public-sector organizations so that network partners contribute and derive value through a business model which is individually sustainable for all stakeholders and ensures access to trusted data.

United Nations Global Working Group assembled for the 4th International Conference on Big Data for Official Statistics, Bogota, Colombia, November 2017.