

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**

Background document  
Available in English only

**Global Platform for Data, Services and Applications**

Prepared by the Global Working Group on Big Data for Official Statistics

# **GLOBAL PLATFORM FOR DATA, SERVICES AND APPLICATIONS**

## **Introduction and background**

The Global Working Group (GWG) on Big Data for Official Statistics is seeking a mandate to explore and develop the business case for a Global Platform for data, services and applications as a public good to be presented to the UNSC in 2018.

The purpose is to develop a business case to specifically explore the possibility of building a public-private partnership that should extend current and future initiatives to make better use of innovative data sources at the national and regional levels. The Global Platform will make traditional and non-traditional data, services and applications more available and accessible to the official statistical community for statistical and research purposes for the public good. It will be based on the fundamental principles for official statistics adding value to data and statistics whilst adhering to strict ethical principles on the appropriate use of data for statistics and research, with suitable standards for privacy, confidentiality and data security. This background document for the consideration of the Statistical Commission at its 48<sup>th</sup> session contains an initial roadmap of proposed Workstreams for 2017 intended to underpin the development of a business case for the establishment of the Global Platform.

The proposal to develop a business case for the Global Platform has been discussed supportively by panels at the third Global Conference on Big Data for Official Statistics in Dublin, Ireland and the series of regional Conferences on a Transformative Agenda for Official Statistics both held in 2016. The proposal is consistent with the objectives of the strategic area of "Innovation and modernization of national statistical systems" in the Global Action Plan as formulated by the High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development. The recent UN World Data Forum devoted a session to exploring the development of the global policy and technology framework for a Global Platform of data, services and applications delivered through public-private partnerships.

## **Objective**

The business case will specify how the establishment of the Global Platform will be organised, governed and financed with the aim of creating the policy and technology infrastructure to facilitate the widespread combination of traditional statistical survey and census data with administrative, geospatial and big data from a variety of sources. This will enable increased data sharing between national and regional statistical systems, offering data and related applications and services for the official statistics community and its users for statistical and research purposes with the view of better serving the public good across the globe.

## **Approach**

The Global Platform will store, offer and deliver data, applications and services, including training and capacity building services, for statistical data producers and data users. It will be delivered through an appropriate cloud-based technology infrastructure.

The technology platform will further the wider availability and access to data, applications and services on a reciprocal basis whereby network partners give and gain access to the shared data, technology and knowledge assets included on the Platform.

To build and retain trust in the Platform and its products and services, it should be operated under an independent governance structure to be provided, initially, by the GWG Committee and required to set out clear and transparent policies.

The Global Platform will learn from the work and expertise of the different GWG task teams whether this is in relation to forming partnerships between public and private sector organisations to gain access to big data sources for official statistics; or through understanding, for example, the kind of social media data which can be collected, how it can be collected and how it can be analyzed and processed into statistics, useful for policy purposes.

The GWG Committee on the Global Platform (GWG GP) will consist of representatives from national statistical offices (Canada, Colombia, Denmark, Mexico, the Netherlands, Philippines, Senegal and the United Kingdom) and international organizations (UNSD and World Bank) under the leadership of ONS, United Kingdom. This Committee will apply the accumulated expertise of its members to provide governance and oversight and consider best practices in the design and implementation of the Global Platform. It will actively pursue involvement of representatives of technology companies and potential donors.

This GWG GP will follow a roadmap of three workstreams in 2017/2018, for which various sub-committees will be created: these are elaborated below along with their expected outputs.

### **Workstream 1: Governance**

This workstream will focus on the overall governance of the Global Platform (GP), developing a business model, financing options and considering the legal structure of the GP including network partner agreements to clarify how partners can contribute to the GP and under what terms.

The GWG GP will retain independent oversight of the GP taking advice from a number of private sector partners who will be part of a GP Advisory Group.

We will explore how the GP could be constituted from a network of partners, including national, regional and international statistical organisations, research institutes and private companies. We will also explore how partners can provide content and assets, and create and maintain the underlying infrastructure, such as cloud infrastructure, data, services and applications for the GP. Network partner's obligations in meeting negotiated contractual, license and service level agreements will be considered, alongside processes through which they can be closely monitored for compliance and performance.

The GWG GP will also oversee a ring-fenced trust fund for the GP to ensure probity in financial dealings where donors or partners are contributing to the development of the GP.

The legal structure and operating policy framework for the GP will be specified through this work. It may vary according to the source and nature of the information being handled taking into account the following dimensions:

Long term Business Model: The aim of the GP is that it should be available to as many users as possible, as a public good. To achieve this, the workstream will develop the proposal for a long term business model which is sustainable for all network partners. This is a non-trivial challenge which needs careful consideration, particularly in the context of the global legislative and regulatory frameworks across which the GP will need to operate. A study of the overlaps and gaps across different legislative and regulatory frameworks will be required to identify where synergies and blockers exist.

Financing Options: This workstream will consider and set out transparently a proposed business model which will serve the interests of all network partners, to ensure the success and long term sustainability of the GP.

Legal Compliance: The GP information management and governance framework must comply with applicable international, regional and national laws, regulations and policies. In order to do so, the platform must ensure that the information is managed (collected, processed and disseminated) according to the best legal practices, following applicable rules and regulations and ensuring that appropriate internal control mechanisms are in place, thereby guaranteeing legal compliance.

Products and services and customer network: Building on the work of existing and new GWG Task Teams (TT), both the product and services and the customers network will be explored in terms of data, software applications and capacity building modules organized around the use of both the traditional and non-traditional data sources. This exploration and discovery of the match of products and services and customers will be undertaken through collaboration between private and public sector partners.

Network Partner Agreements: This workstream will further the creation of policy framework for the establishment of legally binding and verifiable partner agreements for the GP to share, use and maintain the deployed assets of the technology platform, data, services and applications to help identify options for a sustainable business model.

Accountability: A senior responsible owner should be appointed who has the overall responsibility for the management and governance of the GP. He/she will oversee the work program to deliver the platform, monitor progress and put in place an independent audit regime to make sure the platform is meeting its legal obligations.

**This workstream is expected to draft and seek endorsement by the Statistical Commission of a governance and operating framework that will govern the Global Platform. It will also seek endorsement for partner agreements, resource mobilization and a long term business model.**

**Workstream 2: Data policy framework for data governance and information management including: ethics, privacy, confidentiality and security and technical infrastructure.**

The data policy framework for data governance and information management is a set of rules and regulations under which the GP must operate in storing, processing and disseminating information. 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.

Consistent with the UN guidance note on Big Data for the Achievement of the 2030 Sustainable Development Goals (SDGs) Agenda, we will establish appropriate governance structures and mechanisms to ensure compliance with the law, protection of privacy and confidentiality and the ethical use of data. The range of expertise we will employ will include global leaders in data governance and security (including those from both national statistical institutes and the private sector) and we will use the latest technologies available to ensure the identities of individuals and indeed groups of individuals (e.g. religious, ethnic and other vulnerable groups etc) are appropriately protected. We will be transparent in our approach and seek opportunities to communicate in a clear and accessible way so that the public are aware of how their data is being used and how this use translates in to wider public and societal benefits.

This work stream will develop and propose principles and policies for the GP to cover the ethical use of data and the methods and procedures for the collection, processing, storage and presentation of data taking full account of data privacy, confidentiality and security issues. The GP must be operated and managed by independent professionals who can

make judgments, comments or decisions strictly within their areas of expertise, independent from third-party interests or political pressures.

The GP will implement a technical infrastructure based on a common policy and technology framework adhering to the data policy framework to deliver data, accessible services and open source applications and associated APIs as a trusted public good to which all GP public and private sector network partners subscribe and adhere to through contractual agreements.

The technology requirements for the Global Platform will be progressively specified on the basis of the work of the present and future task teams of the GWG and the use of a variety of different data sources. For example: satellite data, mobile phone data, social media data, household and business statistics, administrative data and other data and applications related to the measurement of the Sustainable Development Goals (SDGs) or other public good tasks. The number and scope of GWG task teams is expected to grow into other areas of official statistics such as merchandise trade data, population and housing census, national accounts, environmental accounting, price statistics, etc. The work of these future GWG task teams is expected to be undertaken in close collaboration with established subject-matter related expert and working groups and committees functioning under the Statistical Commission.

It is anticipated that various GWG task teams will deliver their expected outputs in the course of 2017. A partner agreement will be established for each task team whereby the members of each task team will appoint a lead, who will collaborate with private sector partners, using Agile development practices, to deliver technology infrastructure, expertise and data feeds to support the task team in preparing the expected outputs. Agile management will support the iterative, incremental design and build of the new products and services of the GP in a highly flexible and interactive manner.

Ethics: The GP should be governed by strict ethical principles designed to retain the confidence of providers of information, end-users and the public. This will be achieved through the sound use of transparent and regularly updated principles and guidance on practice. The Platform should minimise, wherever possible, the burden on providers of information and must demonstrate its cost effectiveness. This is in line with the Fundamental Principles of Official Statistics which lay out requirements for professional standards, integrity and ethics.

Privacy and Confidentiality: Protection of privacy and confidentiality is an essential aspect of gaining and retaining trust in the procedures for data collection, processing, storage and access. One of the key considerations in creating and maintaining the GP is that it has to operate under strict access and security protocols. Effective legal protections against unauthorized disclosure need to be put in place to ensure confidence in the existence and use of the GP.

Data Security: The information management and governance framework must ensure appropriate protection of information against any type of breach, corruption or loss of sensitive information. System security should be maintained and regularly updated in order to protect sensitive/personal data. Security testing such as penetration testing must be carried out to safeguard data against cyber attack.

Data Access: The GP will also develop principles and standards for access to data by network partners and other users. For example; Open Data can be shared and used by all parties whereas access to sensitive data will be limited to those who need to use it based on an accreditation process and an understanding of their needs to access these data. All access to data will be fully monitored.

Data Integrity: The data being used and stored must have a reasonable and suitable guarantee of authenticity and reliability. Authenticity can be achieved by ensuring the source being used is reputable and can be verified. Reliability is ensured by having standard policies in place to process the acquired data. For example, data may be of high quality when acquired from a source, but may lose its quality if sub-optimal processing methods are applied.

Data Discovery and Availability: The data policy framework must ensure that information can be accessed in a timely, accurate and efficient manner. An efficient technology infrastructure is critical to meet this requirement. To enable data discovery and availability the data policy framework should set out suitable metadata standards and practices and the system requirements for high performance data access, analysis and dissemination.

Data Retention: The information being processed must be retained for appropriate time, taking into account the legal, regulatory, fiscal, operational, risk and historical requirements. For example, source data used for official statistics may need to be legally retained for future verification. With the expected increase in the use of multiple data sources, including big data, the data retention requirements need to be agreed in consultation with data owners and data users. Careful consideration must be given to the retention of personal data.

**This workstream is expected to draft and seek endorsement by the Statistical Commission of a data policy framework for governance and information management including ethics, privacy, confidentiality and security, which will shape the workings of the Global Platform. It will also seek endorsement for a common technology infrastructure to enable its delivery.**

### **Workstream 3: Communications, stakeholder engagement and best practice sharing.**

Following the deliverables of the existing GWG task teams, each of the present and future GWG task teams will deliver three outputs: a) the methodology applied in the software application, b) the open source software application and its documentation, and c) the hands-on capacity building programme for the use of the software application, including the trainer data.

To underpin capacity building and best practice sharing, it is essential that the outputs of the GWG task teams are made widely available as exemplars of best practice; the overall progress towards delivering the GP should also be widely communicated to encourage new partners and users to join in this initiative.

Transparency: The processes and activities of the GP must be documented in a transparent manner and made openly available to relevant staff and interested parties. It is both legally and ethically vital to keep the subjects (individuals and entities about whom the data is collected) informed about what data is being processed, how and why.

**This workstream is expected seek endorsement by the Statistical Commission of a framework for communications and best practice sharing for the Global Platform.**

#### **Working methods**

#### **Governance**

The GWG Committee on Global Platform for Data, Service and Applications (GWG GP Committee) will undertake virtual meetings (monthly) and physical meetings (quarterly). It will be chaired by the UK's Office for National Statistics Deputy National Statistician and Director General for Data Capability, Heather Savory. The Terms of Reference and membership of the GWG GP Committee is established and can be found [here](#).

GWG GP sub committees will be established for the various workstreams for which ToR and memberships will be established. A primary and secondary lead will be identified for each workstream to carry out preliminary work and establish the delivery programme for the Global Platform.

A Global Platform Advisory Group will also be established: including a limited number of representatives from the national and international statistical community, private sector, academia, and open data community.

The discovery phase for the Global Platform will be carried out during 2017.

The GWG GP Committee will report to the 49<sup>th</sup> session of the UN Statistical Commission through GWG on Big Data.