

First Meeting of the Friends of the Chair Group on Economic Statistics, 28-30 May 2019 New York

Session: The institutional transformation of NSOs – New methods, services and roles (Moderator Mariana Kotzeva)

Short Note on New data Sources and Methods including Big Data, AI, and Machine Learning

Prepared by UNSD

This short note will focus on the steps taken by the United Nations Statistical Commission through the Global Working Group (GWG) on Big Data for Official Statistics to leverage new data, methods and services through an UN Global Platform on Big Data in support of statistical agencies at the national, regional and global level to work collaboratively in advancing the use of new methods and methods in the production of official statistics.

The United Nations Statistical Commission created the Global Working Group on Big Data for Official Statistics (GWG) at its 45th session in 2014. The GWG provides strategic vision, direction and the coordination of a global program on Big Data for official statistics, including for compilation of the SDG indicators for the 2030 Agenda for Sustainable Development.

The GWG saw a need for a collaborative environment for the global statistical community to work together (including working with other stakeholder groups) on data innovation and modernization of methods and services. The UN Global Platform (UNGP) is a global Cloud based platform which has marketplace capability to enable collaboration. The capabilities include digital services to support the development, sharing and publishing of trusted data, trusted methods and trusted learning by trusted partners. The platform currently contains alpha services such as access to Alibaba Cloud, Amazon Web Services, Google Cloud Platform and Microsoft's Azure cloud, open data from Automatic Identification System (AIS) and Satellite Imagery (Landsat, Sentinel, Planet and Airbus) combined with services for code collaboration, methods publishing and Earth observation and location data analysis. Progressively, the services are extended, most recently with the developers service, operations service, helpdesk services and market place services. Users of the platform can search, build, deploy and consume data, algorithms and statistical methods and can further develop methods using the main programming languages used by the community (R, Python, Java and Scala).

The UN Global Platform allows the statistical community to work together with its public and private sector partners on new methods and data sources. Examples are the compilation of agricultural crop statistics using satellite data, the measurement of human mobility (migration and tourism statistics) using mobile phone data, the calculation of Consumer Price Indices (CPIs) using global data sets of scanner and other transaction data, and the use of privacy preserving techniques to allow for secure and trusted data sharing. Many more projects can be initiated on the platform involving not only the statistical community and related public sector bodies, but also stakeholders from private sector, academia and civil society.

The more activity the platform orchestrates the more it can learn and the better it can evolve. The development and maintenance of the UNGP are performed under the auspices and guidance of the Statistical Commission. For the UNGP, the rate of innovation is not dependent on the physical size of the

platform but on the size of the ecosystem of the core stakeholder group (the global statistical system) and all their partners. All members of the ecosystem can provide information on improvement, quality control, reliability and price sensitivity. In other words, for a platform organization, the bigger the network of users becomes the more innovative, efficient and customer-focused it can become.

Resources for the initial development of the UNGP are being provided by the partner NSIs (such as UK Office for National Statistics (ONS), Canada, Denmark, the Netherlands and Poland) and the UNSD, collaborating via the GWG. Some of the technical partners in the private sector working with the UNGP have also directly contributed resources to assist, for example, free access to data and tools, and free credits for cloud services.

The GWG is of the view that an independent UN Global Platform Organization will need to be established to deliver the UNGP. The organization must be able to generate and accept funding, plan and deliver the UNGP as it evolves and undertakes operation and commercial activities such as hiring staff, contract negotiation and UNGP Platform evolution. UN mechanisms, led by the GWG, will define and ensure the overall governance for a UNGP Organization which will develop trusted partner relationships and design and operate the UNGP. In any option for the UNGP Organization, there should be a Governance Board which would consist of employees from UNSD/DESA, national statistical institutes and several independent board members. The Governance Board would decide on the data policies, ethics, security and assurance policies and principles. The Board would regularly report to the Statistical Commission.

The GWG considers public trust an essential aspect of the further development of the UNGP. Data providers and individual data owners need reassurance about where their data is stored and who is accessing it for what purposes. A wide variety of legal and regulatory frameworks are in place across different geographies; and social and cultural attitudes to data vary enormously across the world. For these and other reasons, it is envisaged that there will be several regional research hubs in place as part of the overall UNGP Organization. Regional hubs will be able to operate on the Global Platform under a well-defined data policy framework. Linking the regional hubs in a federated system will place the global data collaborative community of official statistics at the heart of modern trusted data usage and information technology, which will offer both developed and developing countries opportunities to realize the benefits of multisource data.

While considerable progress has been made in the deployment of the UNGP, with an on-line community of 200+ people in 14 time zones, a collective strategic push/call for action must be provided/made by the global statistical community to:

- Formalize by March 2020, the governance and business model with a federated network of global, regional and national data hubs that offers the organizational and technological flexibility and interoperability to progressively scale the number of dedicated national and regional Big Data institutions.
- Seek commitments, on a voluntary basis, from national, regional and international statistical agencies to establish Big Data and Data Science institutes/centers that collaborate on the UNGP.

UNSD will shortly sign a Memorandum of Intent with the NBS China for an UN-NBS Big Data Center in Hangzhou.

- Mobilize the collective strength and knowledge of the global statistical system through dedicated project/task teams working collaboratively on promising Big Data and microdata linking projects to advance the skill development, applications and services to produce official statistics.
- Make a global call for the intensification of training and capacity building in data science by national, regional and global statistical institutions through on-line training courses
- Strengthen private sector partnerships for global access to and collaboration in technology, services and data.
- Establish a dedicated network of data scientists from national, regional and international statistical agencies, academia, private sectors, government departments, and international agencies.
- Demonstrate the data science capability of the national, regional and global statistical system in the production of official statistics
- Communicate the success stories in the use of big data and micro data linking to produce official statistics and SDG indicators.