4TH INTERNATIONAL UN CONFERENCE ON BIG DATA FOR OFFICIAL STATISTICS

KEYNOTE

Alexandre Barbosa (NIC.br / Cetic.br)

Innovation and Modernization of National Statistical Systems through "Trusted Data Collaboratives"

Bogotá, Colombia, 8 November 2017

Excellencies, Ministers, Honorable Guests, Distinguished Delegates, dear Colleagues, Ladies and Gentlemen. A very good morning to all of you!

Muy buenos días a todos y a todas. Es un placer y un honor dirigir la palabra a ustedes en esta importante Conferencia de Naciones Unidas que se lleva a cabo por primera vez en Latinoamérica.

I am happy to be here with you today in the beautiful and dynamic city of Bogotá, and I am grateful to both the UNSD and the UN Global Working Group on Big Data for Official Statistics for having invited me to give this address. I also wish to thank the Government of Colombia and, in particular, the Ministry of Information and Communication Technologies and the National Administrative Department of Statistics for hosting this Conference.

It is indeed a great honor to be able to share my thoughts with you and to offer some insights gleaned from Brazil's experience in partnership building for the production of statistics in the context of the Monitoring Framework for the Sustainable Development Goals.

The UN 2030 Agenda for Sustainable Development is a plan of action for people, planet and prosperity and it calls on all countries and stakeholders to act together to rise to the challenge of ensuring that "no one is left behind." The agenda, which comprises 17 Sustainable Development Goals and 169 targets, serves as a framework for countries to take the bold and transformative steps that are urgently needed to shift the world onto a sustainable and resilient path.

To bring about change, we will need to measure and monitor the actions of myriads of stakeholders. Data and statistics production will be crucial for tracking the progress towards the objectives set out in this Agenda.

National statistical systems will be under increasing pressure to produce high-quality data in a timely manner on a wide range of areas – often with limited resources. Innovation and modernization of our national statistical offices is therefore of utmost importance.

When we think of innovation and modernization, technology is usually the first thing that comes to mind. However, although this is a key driver, technology alone is not enough to promote innovation and modernization.

Institutional innovation – a fundamental change enabler for NSOs seeking to reform corporate cultures marked by established beliefs, values and norms – is as important and strategic as

technology. It has the potential to bring about new collaborations through partnerships and data sharing.

This year's Conference addresses important topics related to transforming NSOs and national statistical systems. It also shows the importance that countries – represented here by Ministries, National Statistical Offices, International Organizations and other key stakeholders – are attaching to measurement, and in particular to the use of big data to complement traditional data sources for the production of official statistics.

My focus today is on the importance of multi-source data, multi-stakeholder partnerships and data sharing for the innovation of official statistics.

We are living a veritable data revolution – data is currently produced faster than it can be used and transformed into information that can drive sustainable development. Information in itself is now a key resource, not only for nations but also for organizations; and national statistical agencies will increasingly depend on their capacity to partner with other stakeholders to develop and disseminate information and knowledge to promote development.

A wide range of technologies such as cloud computing, trusted data platforms and open data are crucial for processing large and distributed data. Moreover, standards on open data and metadata documentation, including the Data Documentation Initiative (DDI) and Statistical Data and Metadata eXchange (SDMX) encoding rules for micro and macro-data files are fundamental to data sharing. All these new developments have sparked debate on data-driven innovation and on the value of shared data, as well as on legal issues regarding data access and sharing.

Traditionally, what we see is that NSOs want to maintain control over their data sources, over their working arrangements, data compilation and dissemination processes so as to guarantee quality, independence, impartiality and confidentiality. However, the data revolution requires that NSOs step out of their comfort zone.

These organizations will have to move beyond their own boundaries if they are to establish partnerships that will allow them to shift from using traditional data sources to using alternative data sources and shared data. This is largely due to the fact that the ICT industry and the private sector own very large amounts of relevant data that would likely be unavailable to NSOs if it weren't for agreements and partnerships. This new environment poses numerous challenges for NSOs, requiring actions that enable trusted cross-organization data sharing.

This Conference is an excellent opportunity for all of us to reflect on how we can engage our governments and our institutions in the debate on innovation and modernization of national statistical systems through "trusted data collaboratives." Seen as a process, collaboration between organizations represents new forms of resource sharing that go beyond the traditional public-private partnership model.

I am referring here to collaborations between participants from different sectors, using a multi-stakeholder model to exchange data and to create public value.

I would say that there are three strategic areas for innovation and modernization of NSOs: (i) institutional reform with respect to multi-stakeholder partnerships and the leading role of NSOs; (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.

At the global level, I believe that it is up to the UN to provide guidance on how to access private sector data, to moderate data and technological standards, and to develop capacity building programs, in particular, for less developed statistical systems.

This Conference is also an opportunity for key stakeholders to advance the discussion on how to create public value from data. We should keep in mind that all stakeholders can be considered relevant data sources and that local institutional arrangements and global partnerships in support of national efforts will play a key role in this scenario.

As you will see, my talk comprises four topics:

I. A brief look at the Cape Town Global Action Plan for Sustainable Development Data, in which I highlight two of its six strategic areas.

   On the one hand, I will address the challenges of bringing innovation and modernization to national statistical systems in the new digital era. And, on the other hand, I will underscore the need to establish multi-stakeholder partnerships.

II. I will also look at the concepts of "designed data", "organic data", and "trusted data collaboratives". I will discuss the importance of "trusted data collaboratives" for producing reliable official statistics and I will also discuss the need for a paradigm shift – from the development of institutional arrangements towards the creation of multi-stakeholder partnerships.
   I will argue that innovation and modernization of national statistical systems must rely on "trusted data collaboratives." According to the Global Working Group, the large amount of data produced by private and commercial providers is a potential threat for the production of official statistics, since traditional data sources may end up making not-so-timely statistics obsolete. The bright side here is that, once potential barriers are overcome, thanks to trusted data sharing, private big data sources will offer enormous value to NSOs.

III. I will present a proposed layered-model approach to describe the ecosystem of statistics in this new environment.

IV. And, finally, I will share some insights from the Brazilian experience in building partnerships for the production of statistics, highlighting the importance of ICTs for achieving the development goals and the importance of ICT statistics for SDGs.

Hopefully, once I have finished, you will see how all the pieces of the puzzle fit together.

Let me start off by introducing the objectives for the two strategic areas I highlighted earlier: innovation and modernization, and multi-stakeholder partnerships (two first bullets on this slide). The Cape Town Global Action Plan has established three key objectives and actions for the innovation and modernization of national statistical systems:

1. To modernize governance of institutional frameworks;

2. To modernize statistical standards and facilitate data integration and data exchange; and

3. To facilitate the application of new technologies and new data sources.

Great emphasis is put on revising the fundamental principles of official statistics and on the need to embrace open data initiatives as well as to remove barriers to the use of new data sources.

The action plan also highlights the need to establish multi-stakeholder partnerships. This is aimed at developing and strengthening partnerships among different stakeholders involved in the production and use of data for sustainable development.

Having highlighted two important areas of the Plan, I will now introduce the concept of designed data and organic data – also known as big data – and will look at "trusted data collaboratives."

Assuming that countries still face major gaps in the production of official and timely data for SDG monitoring, and that big data has the potential to complement traditional data sources, we can conclude that the so-called data revolution offers many opportunities, while imposing the need to improve statistical system capacity so as to better respond to new requirements related to SDG measurement. It seems unlikely that NSOs will be able to take on this challenge alone.

Let's take a quick look at what happened over the last decade: digital technologies triggered a digital revolution, and new data sources emerged. These sources include data from the telecom and Internet providers, data from social media companies (Twitter and Facebook, for instance); data from mobile and tracking devices; and private-sector transaction data, such as credit cards and banking transactions. Society is assembling data on a massive scale, and this is commonly called "big data," but according to Robert Grove, former director of the US Census Bureau, it might be better labelled as "organic data."

Needless to say, national statistical agencies are facing an increasing demand for high-quality, timely, reliable and disaggregated data to support the 2030 Agenda. Policymakers will need more and more disaggregated data on multiple sociodemographic variables such as income, gender, age, race, disability, as well as on variables that are relevant to national contexts. National statistical agencies will have to find a means of combining different data sources so as to produce reliable and policy-relevant data.

Robert Grove is right in saying that "the combination of designed data with organic data is the ticket to the future." Hence, big data will play a central role in transforming statistics production, not only because it offers relevant input, but also because it requires that we rethink data governance.

Partnerships built on multiple-actor engagement will require "trusted data collaboratives" which is technically a secure platform for exchanging protected and open data.

Data is often sensitive; therefore, data-sharing agreements have to be well formulated and supported by a secure technological infrastructure. The innovation we can achieve within NSOs is twofold: at the technological level, tools, methods and processes are of utmost importance; but institutional innovation (including partnerships, governance, legislation, privacy and ethics agreements, and stakeholders' engagement) is an even more powerful driver for modernization of national statistical systems.

The third point that I will now bring to your attention relates to data and statistics production ecosystem.

In this slide, you can see how the data and statistics production ecosystem can be represented in a layered model and how the notion of "trusted data collaboratives" relates to both data users and data producers.

The layers in the middle represent data providers from the traditional national statistical systems and also from a wide range of potential new data sources from private sector. The layer on the top represents data users and their data needs. The bottom layer represents the new possibility of data sharing through partnerships supported by "trusted data collaboratives" platforms. The dots represent technical, legal and institutional frameworks and the arrows represent negotiations that are required to bring innovation to NSOs.

I would like to recall the potential of ICTs for promoting sustainable development. Many international organizations acknowledge the substantial contribution of ICTs in achieving the 2030 Agenda. ICTs can potentially support all 17 SDGs, integrating and accelerating the three pillars of sustainable development: economic growth, social inclusion and environmental sustainability.

→ ICTs are so important to the SDGs that this UN Conference is being hosted by the Ministry of ICT here in Colombia.

Let's now move to my last point. I will discuss the importance of a multi-stakeholder model for promoting institutional innovation in the production of statistics. I will illustrate this by taking a look at ICT statistics production in Brazil.

On the one hand, institutional innovation in Brazil is linked to engaging stakeholders from government such as ministries and the NSO, as well as international organizations, academia and civil society in the production of ICT-related statistics. Multi-stakeholder engagement has played a key role in establishing the legitimacy of the CETIC, which is the Regional Center for Studies on the Development of the Information Society that produces ICT-statistics in Brazil. And, on the other hand, innovation is linked to transparency and collaboration.

CETIC promotes microdata sharing through stakeholder agreements, and establishes partnerships for capacity building in the use of survey data and big data sources for the production of statistics. The Center is also a channel for the dissemination of international frameworks and methodologies for countries in Latin America and Portuguese-speaking countries of Africa.

CETIC is fully funded by the Brazilian Network Information Center linked to the Brazilian Internet Steering Committee. These two organizations represent Internet governance structure in the country and rely upon a multi-stakeholder governance model based on principles of multilateralism, transparency and democracy.

The solid cooperation and partnership established over the years between CETIC and the Brazilian NSO, as well as with ITU, UNESCO and UNECLAC has been central to the exchange of data and best practices in the production of internationally comparable data.

Now, I would like to very briefly mention that the Brazilian government created the National Commission for Sustainable Development Goals so as to advance the implementation of the global agenda proposed by the UN in Brazil, thus giving transparency to the implementation process of the 2030 Agenda. This Commission comprises government representatives and members of civil society and serves as a mechanism for articulation, mobilization and dialogue with key stakeholders.

According to our NSO, the Commission is an opportunity for partnership creation and development – partnerships that are developed and geared towards data production for SDGs, and relies on the capacity of potential partners to participate in national statistical system in the country. The network of stakeholders created by the Commission is also expected to evaluate the use of administrative and big data sources, and to rethink the regulatory framework in the country.

I will now highlight three important international initiatives related to the use of big data source to produce ICT statistics:

1. ITU and UN are leading the Task Force on ICT for SDGs with the objective of proposing of a list of ICT indicators by topic that could be used to measure ICT availability and use in sectors relevant to the SDGs.

2. The UNECLAC Big Data Pilot Project on measuring the digital economy is intended to improve measurement capabilities in the digital economy using big data sources.

3. The ITU Big Data Pilot Project to explores ways of using big data sources from the ICT industry to improve and complement existing statistics and methodologies in measuring the information society.

Finally, the four topics that I have addressed today express my view on the importance of big data sources as a transformative tool for official statistics and on why their use has become a real concern for national statistical offices.

If we are to leverage the potential of big data to improve the production of official statistics, we will need to accelerate the creation of "trusted data collaboratives," in which stakeholders can exchange data for the public and social good.

There is, however, little consensus on best practices for data sharing and on the establishment of "trusted data collaboratives" platforms.

Private companies and governments still have limited knowledge about how to maximize the benefits of data sharing, while minimizing its associated risks, which are usually related to confidentiality and privacy. We must think about how to share data in a trusted manner, within data governance frameworks and a proper legal environment so as to safeguard the data being shared.

I will conclude by saying that the biggest problems with big data are not of a technical nature, but are rather related to governance and established institutions. Efforts by Member States and NSOs to produce statistics for monitoring SDGs should be directed at establishing "trusted data collaboratives," by engaging stakeholders in the national statistical system, for example through

national data centers and data campuses, which can collaborate at the international level through, for instance, a UN-led global data center.

I am confident that these topics will be taken forward in the various panels of this Conference over the next three days and I hope that they will be also reflected in the Bogota Declaration. Together, these topics define an ambitious multi-disciplinary and multi-partner program for the Global Working Group. In this regard, after this Conference we may want to change the working group's name to the UN Global Working Group on "trusted data collaboratives" and innovative technologies.

Nuestra misión como formuladores de políticas, expertos técnicos, representantes del sector privado, académico y de la sociedad civil es continuar el debate sobre cómo implementar lineamientos para compartir datos confiables y cómo mejorar la calidad en las diversas fases de recolección, procesamiento y diseminación de datos.

Our mission, as policymakers, technical experts, representatives from the private sector, academia and civil society is to continue the debate on how to implement guidelines for trusted data sharing and how to improve quality in the various phases of collecting, processing and disseminating data.

Thank you for your attention.

Muchas gracias por su atención.