Table of Contents

Background

Background

There are two separate problems:

1 A vision of federated Data Hubs for the exchange and dissemination of statistical and geographical information.

2 The coordination and execution of innovation platforms for the development of new indicators (or new ways to produce indicators).

Traditional statistics production lines are independent and should not be mixed with the above issues.

A vision of federated Data Hubs
For the exchange and dissemination of statistical and geographical information

Current state of affairs

- Several countries and agencies have developed platforms related to SDG indicator collection and reporting
- For the most part, these platforms have evolved without enough coordination:
  - mushrooms rather than yiest
- Better coordination across countries and agencies would improve the whole process of collecting and reporting SDG indicators
- Hence the proposal of a vision of a set of two parallel interconnected data hubs

Country Data Hub (CDH)

- A Country Data Hub (CDH) should be a collection of local data from multiple sources organized for dissemination, sharing and storage

- One CDH per country
  - Local producers of SDG data will make it available through the Data Hub.

- SDG information at national or subnational level will ONLY be available through the CDH

- Each country has total sovereign control over its CDH

CDH characteristics

- Country ownership: owned, led and managed by each country

- Operational independence: autonomous of any international organization or agency

- Standards-driven: SDG publication aligned to agreed international standards and protocols

- Service oriented: information should be retrieved by agencies or international organizations using web services (for example, SDMX Pull)

-

Prototypical CDH
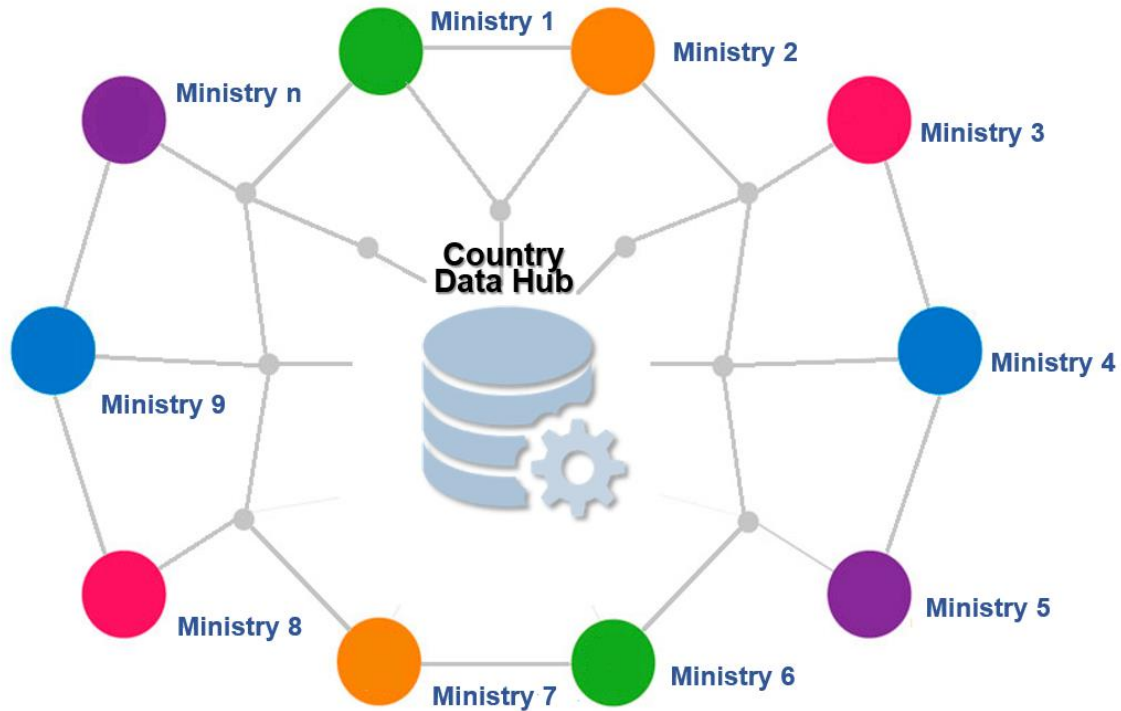
Agency Data Hub (ADH)

- Managed and owned by Agencies and International organizations
- ADH Retrieve Information from CDHs by executing SDMX Software services (web services)
  - Pull mode
- Every ADH selects the set of CDHs needed to retrieve information from
- No country information should be permanently stored in the ADH
- Information has to be referenced to the respective CDH
- ADH can present, consolidate and aggregate information, if and only if, the methodology is public and strictly follows international quality principles

Federated System of SDG Data Hubs

Data flows from countries to international agencies could originate from either **Ministry n** or the correspondent **CDH**

Getting it from the CDH would make information pulls **easier** and **standardized**
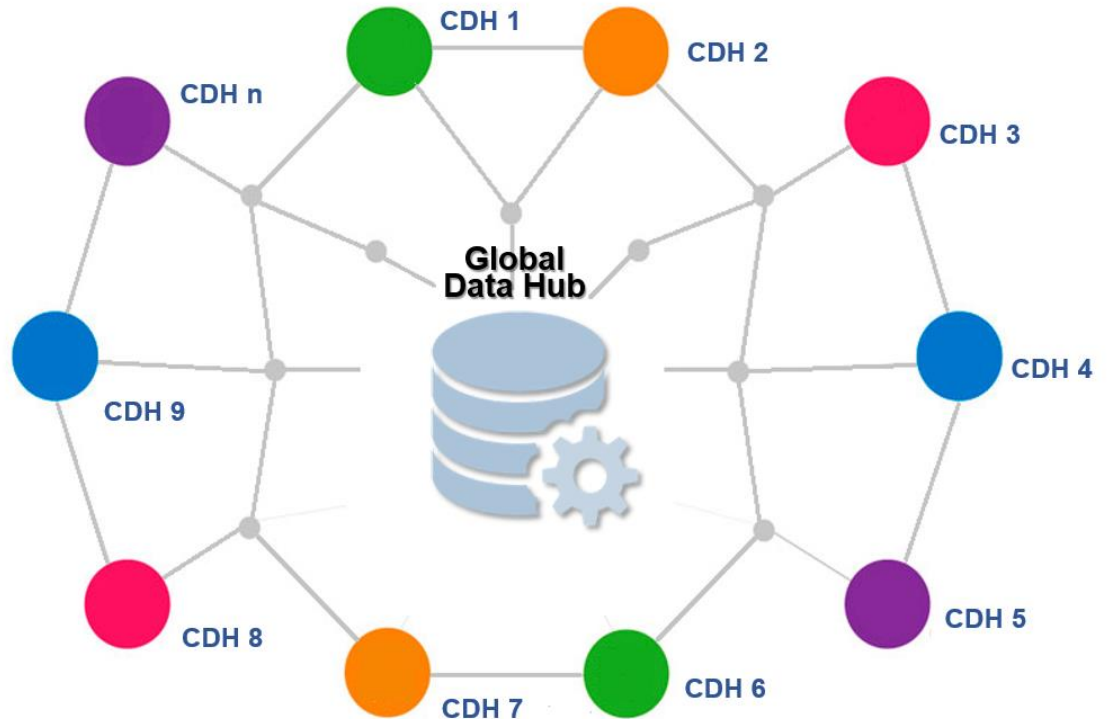
Next steps

- Develop the business architecture of the Federated System of Country and Agency Data Hubs:
  - Objective
  - Governance structure
  - Roles and responsibilities of participants
  - Expected performance
  - Controversy resolution
- Functioning and specifications of each CDH and ADH
- Technical development and interfaces between hubs to make an interoperable system

The coordination and execution of innovation platforms

For the development of new indicators (or new ways to produce indicators)

The first challenge

How to measure SDGs in an integrated, standardized and consistent manner?

Natural response: Use GSBPM!

It was created to:

- Modernize statistical production processes
- Share methods and components
- Integrate data and metadata standards
- Harmonize statistical computing infrastructures
- Provide a framework for process quality assessment and improvement

However, while it requires to use international methodological standards, it was intended to standardize production processes and not methodological standards

How to measure SDGs in an integrated, standardized and consistent manner?

1. GSBPM limits innovation by prioritizing the use of standards to produce information while discouraging the experimentation with new data sources and/or technologies

2. On the other hand, it is a general framework that does not provide guidelines when there are no national or international standards

   - Each country then can do as they see fit (subject to their production capacities)
   - Different methodologies arise and it is impossible to compare
   - Huge international effort is needed to reconcile methods

To address the first challenge, we need to solve this coordination dilemma for the innovation realm effectively

How to measure SDGs in an integrated, standardized and consistent manner?

The best approach is to complement GSBPM with an independent R&D process framework that will provide the international standards for production (tier III) and take advantage of new opportunities in the data ecosystem for the improvement of current standards (tier I and II).

What do we need?

1. Generic Research and Development (R&D) Process.
2. Common working framework for R&D teams.

1. Generic R&D process

Definition: Describes and defines the business processes needed to produce a **methodology** that satisfies an information need

It provides a standard framework and harmonized terminology to help statistical organizations to modernize or create methodologies for the production of official statistics, as well as to share methods and components

This process's output is meant to facilitate the implementation of a GSBPM model whenever a methodology has been deemed suitable for official statistics

2. International working framework for R&D teams

In order for the R&D process to be **trusted**, **usable** (possible for countries to adopt methods seamlessly) and **improvable**, its development should:

1. <u>Be replicable</u>. Output of R&D teams needs to include code, computational environment and training/test data.
2. <u>Documented</u> under minimum common guidelines:
   a. Have a common directory structure for all projects.
   b. Share a set of common FAQs (which indicator, who is working on it, where is the code, what are the data sources, what is the contact, and so on).
   c. Have a common way in which other countries can actively provide feedback while methods are being developed.
   d. Have a common way to make work, progress and documentation publicly available.
3. <u>Be agile</u>. Specify metrics of success/failure in each team from the get-go and use agile methodologies to minimize time allotted to bad investments.

Next steps

Each country needs to establish investment priorities based on their national and international agendas, methodological capacity and data availability.

In this first proposal our guiding principles for the R&D process are:

1. Focus on the production of information, particularly tier III indicators
2. Take research that has proven useful in academia or industry and apply it to the context of official statistics
3. Create international teams by identifying common priorities and multidisciplinary capacities.
4. Invest in technology only within the needs of R&D teams, without competing with off-the-shelf technologies (explore the landscape and choose from it!)
5. Each team should follow the generic process for R&D and the international working framework; inform it iteratively so as to improve and mature it

Vision of a federated system of SDG data hubs and innovation platforms

**Julio Santaella,
President of INEGI, Mexico**

March 9, 2018