United Nations | Department of Economic and Social Affairs
Statistics

Overview of recent work on SDG Data Interoperability

Background

- 2030 Agenda and the need for high-quality, timely, and reliable data, disaggregated and supplemented with necessary contextual information.
- National Statistical Systems: challenges not only in producing statistics to fill data gaps, but also in integrating and making existing SDG-related data and information accessible to decision makers in a meaningful way.

Data interoperability for the SDGs

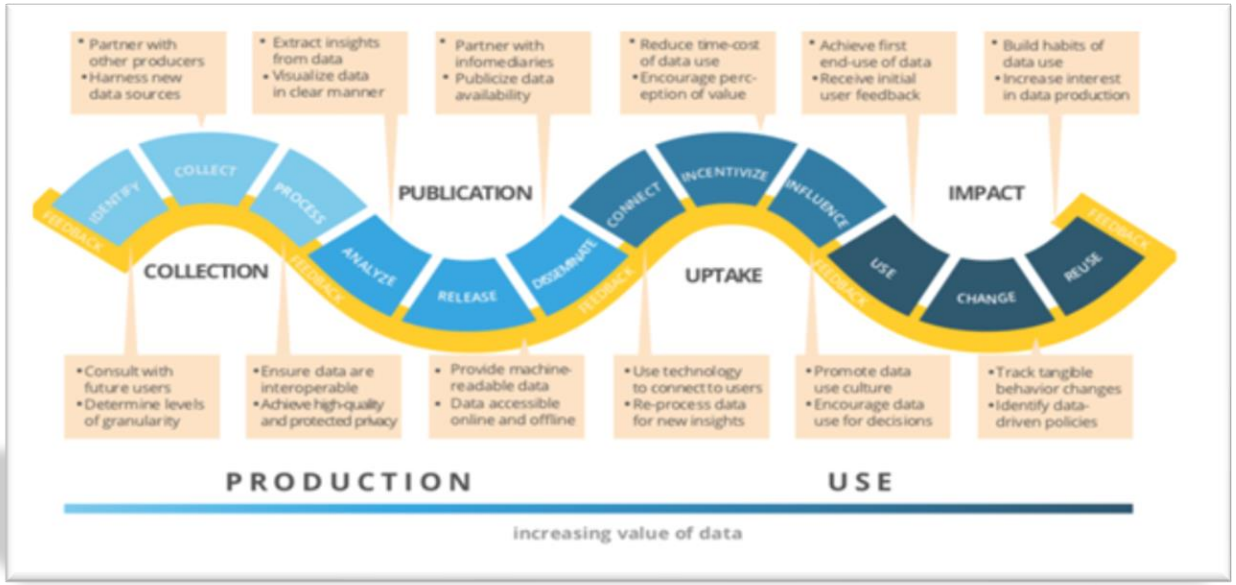
- There are many unrealized opportunities to extract value from data that already exists to meet information needs of the 2030 Agenda

- Investing time and resources in the development and deployment of data interoperability solutions will help us make better use of the data that currently sits in sectoral and institutional silos to implement and monitor the SDGs

Interoperability and Open Data

- Many organizations are now adopting open data policies that authorize and facilitate the reuse of their data assets
- Open data requires data to be interoperable not only from a technical perspective, but also from a legal and institutional perspective

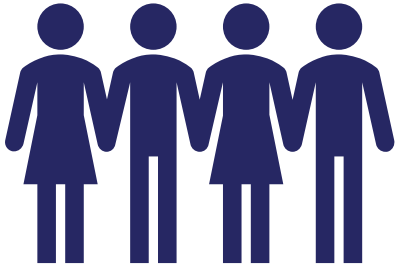
Interoperability in the data value chain



- Planning for interoperability should take place early in the data life cycle
- Interoperability considerations should inform every step in the data value chain

Source: Open Data Watch

Global Partnership for Sustainable Development Data



Collaborative on data interoperability

- **First UN World Data Forum, January 2017:**
  Jointly convened by UN Statistics Division and Global Partnership for Sustainable Development Data

- **Second UN World Data Forum, October 2018:**
  Launch of Data Interoperability Guide

Vision

- Integrating and joining up data from multiple sources, and across systems, to realize the data revolution for sustainable development, where more and better data is open, accessible, and used to the fullest extent possible in improving people's lives

Getting the **governance** and **institutional framework** right

Designing **data structures** with users in mind

Standardizing the **data content**

Providing **standard interfaces** to access and work with data

Disseminating **linked open data** for knowledge creation

Pathway to data interoperability

Data management and governance

- Very often interoperability is less a technology problem and more a data management and governance issue
- To be effective, data needs **oversight** and **accountability** across its lifecycle.
- **Institutional frameworks** create the environment where data, technology, and business processes **fit** with each other

Data and metadata models

- Interoperability is highly dependent on data and metadata **modelling decisions and practices**
- Need to ensuring that systems are designed with interoperability in mind from the outset
- The focus should be on producing **simple, self-contained** datasets that are **easy to understand and manipulate** by users and client applications.

Standard classifications and vocabularies

- Classification systems shape the way data is collected, processed, analyzed and shared with users
- They allow data producers to **express the meaning of data without ambiguities**
- They enable users to **find and link** related pieces of information, from the unit record level to the dataset level, across different information systems.

Standard classifications and vocabularies

- The use of **customized** classifications and vocabularies is sometimes unavoidable
  - **legacy** data management systems,
  - **specific need**s of domain of application
- Often it is necessary to "**standardize after the fact**," mapping "local" terminology to standard vocabularies and taxonomies

Standard classifications and vocabularies

- To meet the needs of a continuously changing data ecosystem, classifications and vocabularies need to **adapt over time** and be **continuously "mapped" to each other** by establishing associations of correspondence between their elements.
- Moreover, they need to be publicly available and accessible in **open, machine-readable** formats, such as CSV, JSON or RDF

Open data formats and standard interfaces

- Interoperability is not only about standardized data production, but also about standardized “**data logistics**” (Walsh and Pollock)
- There is need for common “pathways” to **get data from providers to users** in a fast, convenient, effective, and efficient manner.
- System interfaces should prioritize interoperability and flexibility over specificity and optimization.

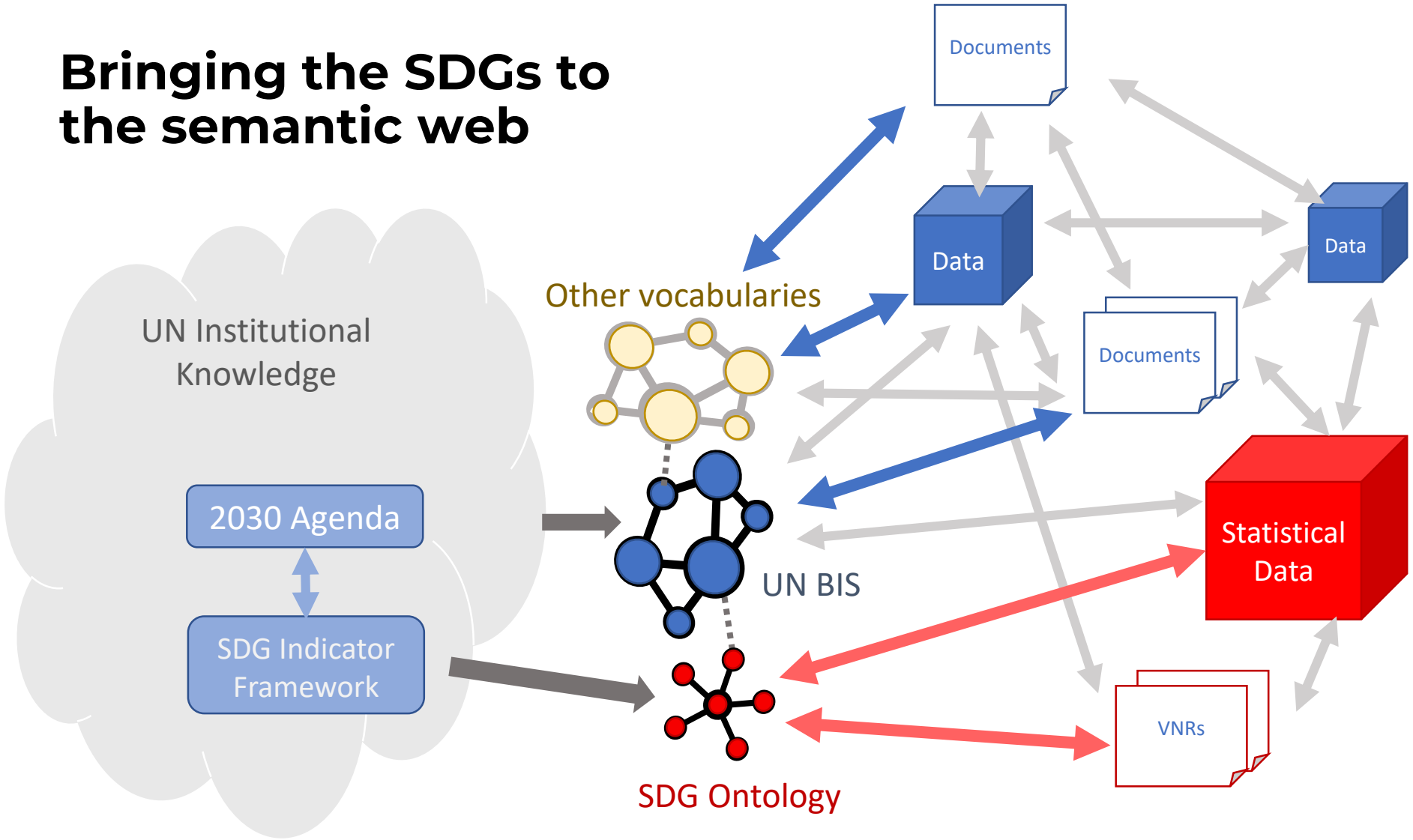
Open data formats and standard interfaces

- A first step is to make the open data available through **bulk downloads in openly documented, non-proprietary formats** (such as CSV, JSON, XML, and GeoJSON, etc)
- Application Programming Interfaces (APIs) provide **machine-to-machine access to data services** that are the building blocks for users applications

Linked open data

- Facilitate integration of data and other information resources on the web
- Enable analysis across disparate and previously isolated datasets
- Need to formulate best practices for:
    1. Publishing SDG data as Linked Open Data
    2. Develop applications that combine and analyze multiple SDG-related datasets and documents from diverse sources over the Web

Bringing the SDGs to the semantic web



Final thoughts

- The adoption of standardized metadata schemas and practices is highly dependent on the social context
  - How easy/difficult they are to understand and use by a wide community of data practitioners
  - How difficult/costly it is to transition out of legacy metadata systems
  - Degree of openness and transparency of their developers and curators

M.A. SICILIA (2014)

United Nations | Department of Economic and Social Affairs
Statistics

Thank you.

Questions or inquiries?
gonzalezmorales@un.org