Integrating Indicator Reporting to the International Level: Guidelines for Improved Data Interoperability

# Introduction

## Overview

There are several reporting streams for development agendas from the national level to the international level. The SDGs (for the 2030 Agenda), the Sendai Framework Progress Reports (for the Sendai Framework for Disaster Risk Reduction), nationally-determined contributions (for the Paris Agreement), and country-determined indicators for the Beijing Declaration for Equality (supporting women’s empowerment) all represent such reporting streams. In some cases, reporting is voluntary, or done according to agreements on a country-by-country basis. There are generally strong relationships between the various indicators which are reported. The UNDRR indicators have a high degree of overlap with many of the SDG indicators; the nationally-determined contributions (NDCs) supporting the Paris Accords are subsumed in the reporting for the SDGs. Other sets of reported indicators bear a strong relationship to the SDGs, such as the Beijing Declaration's minimum set of gender indicators, and the EU’s recommended indicators that also based on the Beijing Declaration regarding gender issues.

The burden of reporting at the national level is high: lowering the barriers to accessing data coming from the national level is in everyone’s interest. At the same time, there is not a formalization of the relationships between the different reporting requirements. If a detailed picture of the relationships between the various sets of indicators existed, then duplication could be identified and the duplication of effort reduced, lessening the burden on countries.

From the international level, understanding how the various reported indicators are related would enable better decisions to be made when requiring the reporting of new data, and when negotiating the data to be reported by countries. Further, the users of the international data have needs which are not always well understood by those deciding what data is to be collected in support of any particular set of goals/targets.

These issues are well summarized by UNDP’s 2017 report, “Aligning NDCs & SDGs: Lessons Learned and Practical Guidance”

**‘*Important issues and guiding questions:***

1. *Which data is being collected and analyzed to inform the design [roadmaps for the different development agendas], implementation of actions, and measurement of their achievements?*
2. *Where is the information being collected? Where does it sit?*
3. *How is data shared between relevant actors? Are systems already in place to make this happen? Which access restrictions are in place for the data?*
4. *Are end-users able to understand the best way to use this data?’[[1]](#footnote-1)*

In order to address these issues, several different approaches and actions can be recommended. These guidelines are aimed at those who are working at the international level to define what data are to be collected, and those at the national level who must make the decisions about which data are of highest priority, and how best the reporting can be achieved across multiple governmental and non-governmental organizations.

Several approaches are advocated, building on existing initiatives. One leverages the ability of semantic technologies to describe complex sets of inter-relationships between different data structures. While there is a high degree of overlap across the different sets of indicators reported to the international level, it is not easy to see their specifics. An easy way to understand and navigate these overlaps has many applications both for those reporting the indicators from the national level, and those asking for such data to be reported.

Another approach builds on the more familiar SDMX standard, and the related Data Cube Vocabulary published by the W3C which is based on it. While some of the reporting streams to the UN (e.g., the SDG indicators) already utilize these standard approaches, others do not. There is an established technology platform associated with SDMX which could be better utilized if the statistical concepts used in the various reporting frameworks were formalized in line with that standard.

Further, the aligned Data Cube Vocabulary provides an entry point into mainstream Web approaches such as Schema.org which would enable users of the international data to access it using familiar tools such as search engines in a nuanced fashion which is impossible today. This would increase the discoverability and usability of the data in a significant fashion. [Also DCAT Application Profile, or DCAT-AP, for data portals in Europe?]

A third set of recommendations is made to those specifying and agreeing the data reported to the international level, so that it is in a form which is optimally useful to those working "on the ground." The needs of users as perceived by international statistical institutions are not always aligned with the needs of end users at a local and regional levels. This connection is one which could also benefit from having better visibility into how the different concepts comprised by the various reporting frameworks are presented. **[Insert additional section on Feedback & Support for Implementation at Sub-National / Regional levels? ]**

## Audiences

This document presents a set of guidelines for those working on the various reporting chains from the national to the international level, using the SDG Indicators, the UNDRR Indicators, the data needed to support the Paris Agreement, and that associated with the Beijing Declaration for Gender Equality. These are exemplary reporting streams, but the guidelines presented would also apply to other, similar types of reporting.

The guidelines are presented at three levels: a general statement of the goals and the case for their application; a description of the impact on those reporting and receiving the data as reporting is performed and managed; and guidance for those implementing the systems which support these activities at a technical level. [Use term “practitioners”?]

[Are researchers an audience? OBO interface allows for many useful links. Mention.]

## Domain Coverage

The various reporting streams discussed here are fundamentally cross-domain: the Sendai Framework for Disaster Risk Reduction and the 2030 Agenda for Sustainable Development are examples of global policy agendas which have extremely broad data requirements. Depending on the specific focus, they may also have very detailed requirements in terms of the needed data (eg, the Beijing Declaration's focus on gender). This document is thus broadly relevant across all the various domains which be covered by these types of policy agendas at an international level, and those involved in their local and national implementation. [ADD “Three Pillars” (“cross-pillar”) idea to appeal to suited UN wonks?]

# Policy Initiatives: Examples

These guidelines use as examples a small set of well-known policy initiatives which are related in different and inter-connected ways with the reporting of data to the international level. These are briefly described here.

[For each, cover data requirements, reporting agreements/arrangements, reuse, and characterize the state of play for implementation. Enabling mechanism for localization (excluding exaggerated claims of capacity-building – focus on data) support for coordinators at the national level in light of the recent re-org of development programs within the UN.]

# Policy Initiatives: Examples

These guidelines use as examples a small set of well-known policy initiatives which are related in different and inter-connected ways with the reporting of data to the international level. These are briefly described here.

[For each, cover data requirements, reporting agreements/arrangements, reuse, and characterize the state of play for implementation. Enabling mechanism for localization (excluding exaggerated claims of capacity-building – focus on data) support for coordinators at the national level in light of the recent re-org of development programs within the UN.]

**Overview of selected development agendas and their reporting:**

## The 2030 Agenda for Sustainable Development (SDGs)

The 2030 Agenda, with its 17 Sustainable Development Goals (SDGs) and 169 associated targets, marks the first time that the international community committed to integrated targets for sustainable development that are time-specific and universal.

Member states’ national reporting and dissemination platforms are the key mechanisms for the implementation and ongoing review of the SDGs. Whilst all UN member states agreed on the 17 SDGs and 169 targets, and the UN Statistical Commission was delegated the process of determining indicators to allow measurement of whether the SDGs will ultimately be considered a success. The Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs) established baseline data and developed the *Global Indicator Framework* of 232 indicators for the SDGs that are nationally-relevant and human rights-sensitive.

It is understood that not all of the 232 indicators of the *Global Indicator Framework* are applicable to all national and sub-national contexts. Annual SDG progress reports on the *Global Indicator Framework,* combined with data produced by member states’ national statistical systems (also reporting on regional information), thereby serve to inform the High-level Political Forum for ongoing review and follow-up (Paragraph 83, *2030 Agenda for Sustainable Development*).

## The Sendai Framework for Disaster Risk Reduction (Sendai)

The Sendai Framework (2015-2030) sets four global targets

7 global targets

38 indicators

4

## The Paris Agreement under the UN Framework Convention on Climate Change (‘Paris Agreement’) –

2°C

Reporting mechanisms and indicators are self-selected by countries through their nationally-determined contributions (NDCs.

# Goals and Benefits

# Approaches and Guidance

[TO DO: Enumerate specific guidelines according to the ideas summarized in each section below]

## Semantics Tools for Understanding and Using the Data

### Description of the Indicator Content Graph (ICG)

### High-Level Guidelines

The main benefit of connecting to the ICG is to gain visibility into the entire network of indicator reporting at a national and UN level, so that reporting can be better managed and coordinated at the national level, with coherent support from the UN through increased visibility into the data requirements being established by the relevant UN agencies. If one reporting stream for a particular policy agenda can see at a detailed level how it relates to the reporting streams for the others, then it becomes easier to make good decisions about how data collection is prioritized, and what data are already being reported to facilitate reuse. The end result is a lower reporting burden, and the potential for better management tools at the national level.

### User-Level Guidelines

User guidelines will focus on how the terms and concept definitions needed for describing the data to be compiled to support a specific reporting chain. Once described, they can be mapped to the agreed ICG model (BIS and EUROVOC). These activities require an understanding of both the indicators and their disaggregations [Reference the mapping Steve and Massimo are working on – what is this?]

[NOTE: How do workflows leverage available online tools for working with BIS/EUROVOC and (soon-to-be-released) SDG-KOS tools. It should be possible for business users to find and reuse material that already exists in the library.]

### Technical Guidelines

Technical users will need to understand how mappings against the BIS and EURVOC translate into the integration with the SDG-IO, and how this connects to the OBO Foundry upper model. The business-level definitions will be expressed using an agreed profile of SKOS, and they will need to understand how this is done.

Additional guidance should be provided for those policy agenda areas which wish to elaborate their own ontologies, in relation to the others found within the ICG.

[- Include examples]

## Standards and Models for Sharing and Exchange (The SDMX and Data Cube Vocabulary)

### Description of the SDMX Reporting for Indicators (SRI) Approach

### High-Level Guidelines

The infrastructure for SDG Indicator reporting is based on the SDMX architecture, some reporters to the Global Database and some reporting countries do so using SDMX formats. Because the SDGs are so broad-reaching and have similar data requirements to reporting streams for other policy agendas, the existing infrastructure can in some cases be leveraged at the national level. The SDMX architecture is not specific to a single reporting stream (e.g., the SDG Indicators) but is configured to fit each reporting stream using a Data Structure Definition (DSD).

This has several benefits: duplication of effort can be lessened, because many of the same tools and technical staff can be employed across reporting streams. This framework can be leveraged to provide a direct integration with common Web tools for finding and using data (search engines, etc.). This facilitates the dissemination of data to those outside the UN institutions, at both the UN and national levels. While potentially an expensive and complicated undertaking, the use of a shared approach based on alignments with the SDMX architecture can help reduce costs and maximize the utility of such publication. [QUERY: Can these tools be built and then shared at the national level? This is the discussion with Dan Brickley and Doug Fils regarding Data Cube/Schema.org, DCAT, JSON-LD, etc.]

### User-Level Guidelines

The creation of a Data Structure Definition describing the data involved in any particular reporting stream is an important activity for using any SDMX technology. The DSD is based on the terms defined in the activites related to the ICG described above and uses the same understanding of the relevant indictators and their disaggregation. Further, any classifications or codelists needed for describing the data must be comprehensively understood.

A common DSD (at least one) must exist for the entire community supporting any given policy agenda reporting stream. National customizations are also supported, and guidelines from the UNSD should be followed for these.

Developing DSDs and distributing/disseminating them are important areas of focus, and guidance is provided for how these can be documented etc. [TO DO: Leverage existing best practice for the SDGs and other UN SDMX implementations.]

When DSDs exist, data must be reporting according to the structures they describe. There is a range of tools for doing this, including some strongly SDMX-based tools (.STAT, SDMX RI-based tools) and some other approaches based on Excel/CSV (etc.) [These need to be summarized, and then addressed in a coherent way. Meet with the UNSD SDMX training guy.]

The Working Group on Open Data is emphasizing the sharing of UN data with users outside the UN community through the adoption of open data policies (and relevent technologies). SDMX is well-aligned with many of these approaches (the W3C Data Cube Vocabulary, DCAT, Schema.org, etc.) and can be directly leveraged using a straightforward approach. This approach is outlined…. [TO DO: Figure out how to interface with this!]

### Technical Guidelines

At a technical level, there are many sources of advice for the implementation of SDMX, and specifically the way in which SDMX has been used to support the SDG Indicators. The focus here is on the data-related aspects of that architecture, but the metadata reporting parts of SDMX are also used within the SDG reporting stream. [Specify the sources des cribbing best practice as relevant, based on what is promoted by the UNSD to countries.]

The SDMX Information Model was the basis of the W3C Data Cube Vocabulary (an RDF specification for describing data and data structures) and it in turn is well-integrated with DCAT/Schema.org, and some related RDF Vocabularies (e.g., JSON-LD). To leverage these alignments, the transformations and relationships between them need to be implemented in transforms, and the basic approach to publishing data and structural descriptions in SPARQL endpoints (etc.) must be understood. [TO DO: Provide guidance as appropriate here, based on discussions with Dan Brickley, Doug Fils, Simon C., others.]

## Feedback and Support for Localization/National and Local Implementation

### Description of the Implementer’s Feedback and Support (??????)

### High-Level Guidelines

### User-Level Guidelines

### Technical Guidelines

1. “Aligning NDCs & SDGs: Lessons Learned and Practical Guidance.” United Nations Development Programme. December 2017 <[www.undp.org/content/undp/en/home/librarypage/climate-and-disaster-resilience-/ndcs-and-sdgs.html](http://www.undp.org/content/undp/en/home/librarypage/climate-and-disaster-resilience-/ndcs-and-sdgs.html)> [↑](#footnote-ref-1)