SDG Reference Metadata Exchange

Jan 27, 2021

Abdulla Gozalov and Matthias Reister
United Nations Statistics Division

SDG Reference Metadata

- Word-based metadata reporting template approved by the Interagency and Expert Group on SDG Indicators (IAEG-SDGs) in 2016

- Custom SDG-specific metadata concepts

- SDG Custodian Agencies (International Organizations) provide updates to reference metadata using the template under track changes

- IAEG-SDGs reviews and approves the updates

MSD for Reference Metadata

- First draft Metadata Structure Definition (MSD) for SDG reference metadata developed by UNSD and presented to the SDMX-SDGs Working Group in Sep 2019

- Metadata Concept Scheme based on the original SDG metadata concepts

- Target type: Partial Key
  - Metadata Set is attached to a region of the hypercube defined by the DSD, i.e. a group of related time series
  - MSD depends on dimensions defined in the DSD
  - Metadata Set depends on codes defined in the DSD

Harmonization of metadata concepts

- Raised by country members of the SDMX-SDGs Working Group, many of whom had mapped their metadata to "Single Integrated Metadata Structure" (SIMS)

- The Working Group invested substantial effort to align concepts from original metadata template with global standards, particularly SDMX Glossary and Global Metadata Concept Scheme (both derived from SIMS)

- Harmonization involved mapping original template to harmonized concepts. In some cases, this led to splitting or merging the metadata concepts after careful consideration.

sdmx
Statistical Data and Metadata eXchange

Mapping of IAEG-SDG File Format to SDMX Detailed Concepts (1/2)

IAEG-SDG Headers and Subheaders	Detailed Metadata Concept Name	IAEG-SDG Headers and Subheaders	Detailed Metadata Concept Name
Last updated: [DATE]	0.e. Metadata update	Rationale	4.a. Rationale
Goal	0.a. Goal	Comments and limitations	4.b. Comment and limitations
Target	0.b. Target	Methodology	**4.c. Method of computation**
Indicator	0.c. Indicator	Computation method	
	0.d. Series		4.d. Validation
Institutional information Organization(s)	0.g. International organisations(s) responsible for global monitoring		**4.e. Adjustments**
	1.a. Organisation	Treatment of missing values	**4.f. Treatment of missing values (i) at country level and (ii) at regional level.**
	1.b. Contact person(s)	* At country level:	
	1.c. Contact organisation unit	* At regional or global levels:	
	1.d. Contact person function	Regional aggregates	4.g. Regional aggregations
	1.e. Contact phone	Sources of discrepancies	**6. Comparability / Deviation from international standards**
	1.f. Contact mail	Methods and guidance available to countries for the compilation of the data at the national level	**4.h. Methods and guidance available to countries for the compilation of the data at the national level**
	1.g. Contact emails		
Concepts and definitions Definition Concepts	**2.a. Definition and concepts**		
			4.i. Quality management
	2.b. Unit of measure	Quality assurance	**4.j. Quality assurance**
	2.c. Classifications		**4.k. Quality assessment**

sdmx
Statistical Data and Metadata eXchange

Mapping of IAEG-SDG File Format to SDMX Detailed Concepts (2/2)

IAEG-SDG Headers and Subheaders	Detailed Metadata Concept Name
Data sources Description	**3.a. Data sources**
Collection process	**3.b. Data collection method**
Data availability Disaggregation	**5. Data availability and disaggregation**
Calendar Data collection	**3.c. Data collection calendar**
Data release	**3.d. Data release calendar**
Data providers	**3.e. Data providers**
Data compilers	**3.f. Data compilers**
	3.g. Institutional mandate
References URL References	**7. References and Documentation**
Related indicators	0.f. Related indicators

sdmx
Statistical Data and Metadata eXchange

Metadata pilot timeline

- **Sep 2019**: First MSD proposal, extensive discussion. Metadata sub-group established.
- **Dec 2019**: Draft Metadata Concept Scheme approved.
- **Jan 2020**: Draft MSD v0.1 approved.
- **Apr 2020**: UNSD and WB join forces on metadata pilot. Switched to MS Word template.
- **May 2020**: Draft MSD v0.3 approved to be used in the pilot.
- **Jun 2020**: First version of the web-based authoring tool.
- **Jul 2020**: Metadata import and display at SDG Lab implemented
- **Sep 2020**:Metadata exchange begins
- **Oct 2020**: Metadata API and comparison report implemented
- **Nov 2020**: Final updates to metadata template and comparison report

SDG Metadata Pilot: Challenges

- Lack of tools in support of SDMX metadata authoring and exchange
  - Very few SDMX metadata tools available in comparison to the extensive support for data exchange
- Little experience in the community on metadata exchange and application of the standard
- **Conversion of rich text from a Word document is a major issue**
  - Formulas in particular
  - Workarounds had to be found

SDG pilot metadata exchange components

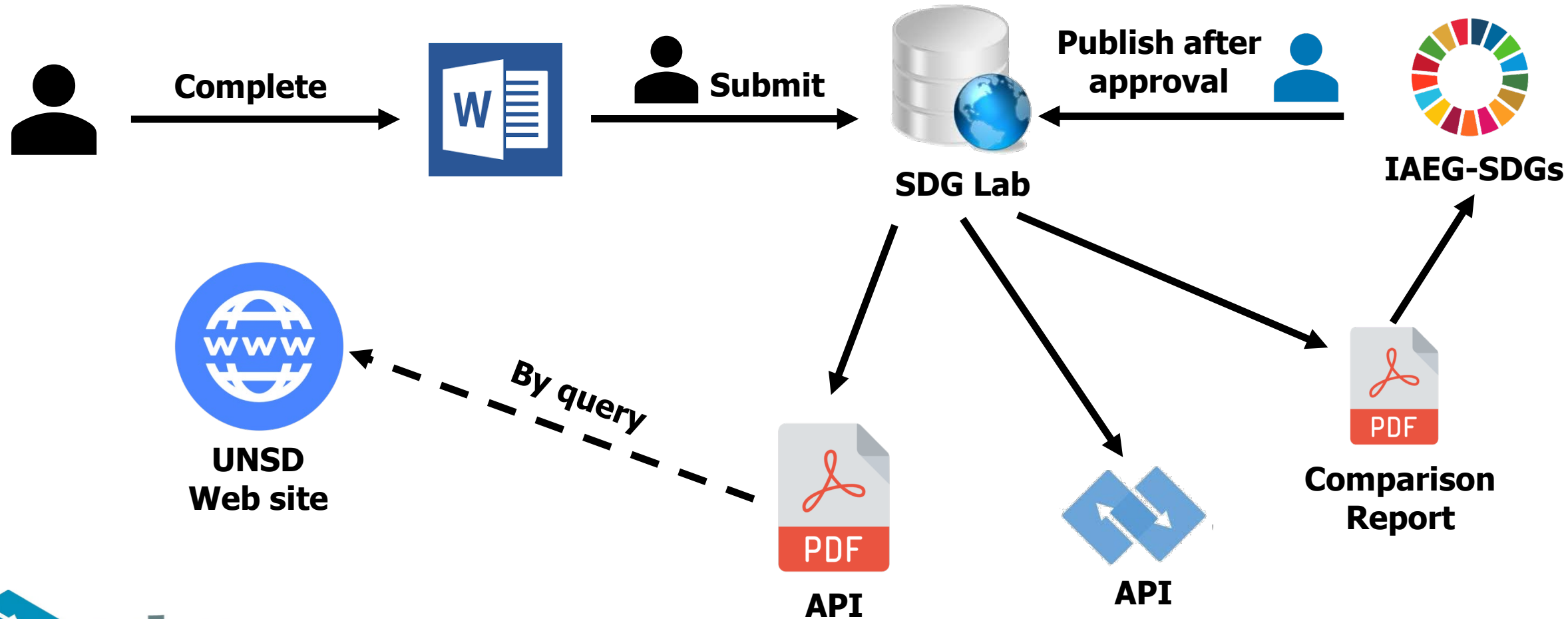| Metadata Template |
| --- |
| MS Word-based |
| Drop-downs for target selection, multiple selection possible |
| Locked uneditable parts |
| Multi-lingual support |
| DSD import to update target drop-downs |

Authoring Tool
Mammoth Library for Word to HTML conversion
Parses template and generate SDMX-ML metadata set(s) as well as PDF report for validation

SDG Lab
Part of the SDG Global Platform
Implements data and metadata exchange
Receives SDMX-ML messages and inserts into the database
Implements API for SDMX Metadata

sdmx
Statistical Data and Metadata eXchange

Planned improvements to metadata exchange

- Template
  - Restructure to improve user friendliness
  - Add concept codes for improved validation, processing, translation
  - Import metadata set
  - Import MSD?
- Authoring tool
  - Resolve formatting issues
  - API for metadata submission to facilitate comparisons
  - API for PDF metadata
- SDG Lab
  - Integrate with Authoring Tool
- MSD
  - Implement vintages

Metadata exchange: Vision

Lessons Learned

- Rich-text, complex content SDMX metadata exchange feasible
- Conversion of rich text still a challenge but can be overcome
  - Maintaining user-friendliness and facilitating uptake
- Simplification of the standard highly desirable
  - Improvements coming in SDMX 3.0
- Poor tool support at the moment
  - Even metadata set validation is a major issue
- Harmonization of metadata concepts requires an effort but pays off
- SDMX is a good way of metadata provision in a machine-readable format
  - Facilitating activities including the World Bank SDG Metadata Translation Project

THANK YOU