

SDG METADATA TRANSLATION PROJECT

Using SDMX to facilitate computer-assisted
translation of reference metadata

Jennifer Park (Consultant, World Bank) Steven (Brock) Fanning (CODE), Abdulla
Gozalov and Harumi Shibata Salazar (both UNSD)



WORLD BANK GROUP



**United
Nations**

DESA
Statistics Division

**SDG Metadata
Translation Team**

October, 2023
Global Meeting on Statistical
Data and Metadata Exchange
(SDMX)

Why translate statistical metadata?



- ✓ Supports statistical capacity of national statistical offices
- ✓ Engagement with local public improves data relevance
- ✓ Facilitates user feedback to improve data accuracy

In the case of SDGs, translations were requested by *countries*.

What are the challenges?



- Structural and reference metadata can be large and complex.
- In the case of the SDGs:
 - 259 reference metadata files to date (each about 6 pages)
 - Technical language used, often with graphics and equations
 - Reference metadata are routinely updated
- Traditional translation methods are costly.
- Versions for source and translations can be difficult to manage.

What is computer-assisted translation?



- Software used to assist a human translator in the translation process from one language to another.
- An automated but not automatic process; human expert is key
- A variety of online translation services are available.

How to coordinate files, processes, and people?



- Online translation APIs can be used for translation of words or phrases, but organizing
 - the files
 - the process
 - and the translators

takes additional effort

- An easy-to-access and easy-to-use solution was needed
 - *Enter SDMX—with our methods to make more user-friendly*

SDMX can facilitate CAT translation

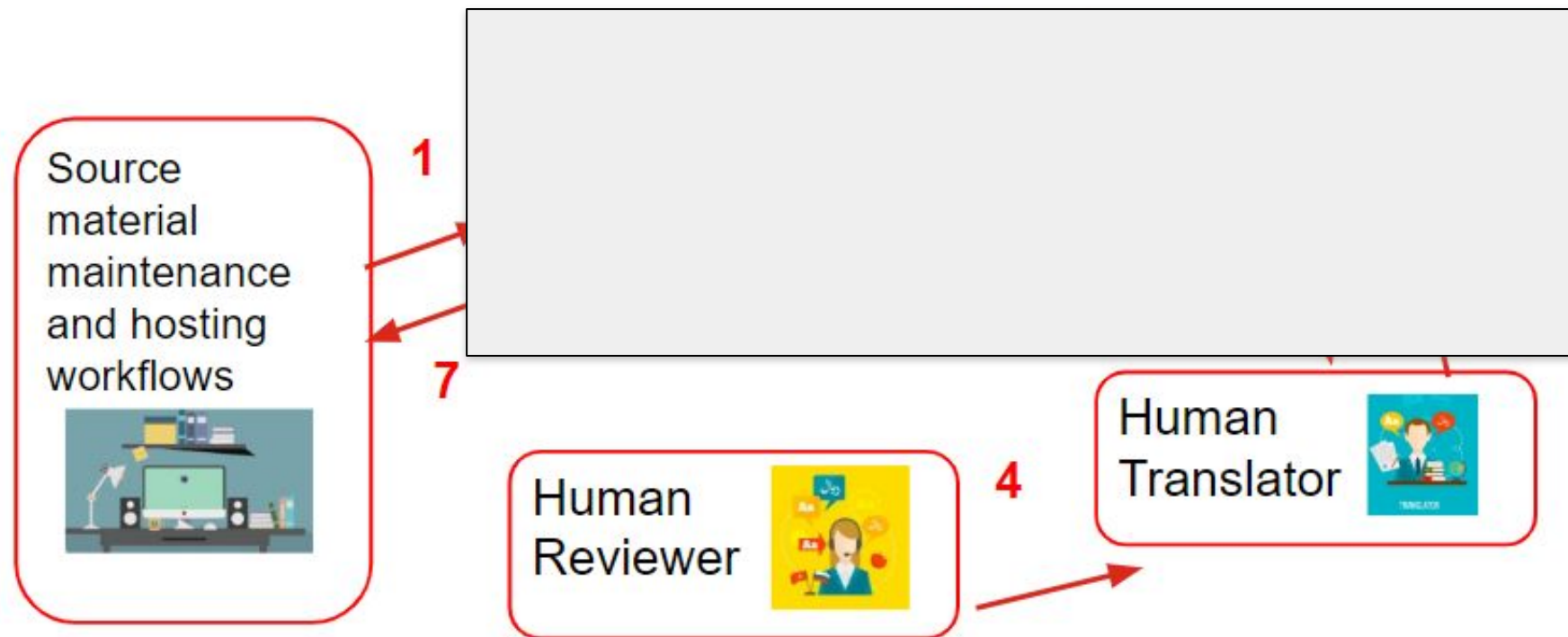


- SDMX's rigorously defined structures simplify metadata translation and organize the process.
- Data structures can be managed as a unit and at its components.
 - code lists can be translated independently of each other
- Reference metadata sets also benefit from uniform structure.
 - text strings associated with each concept can be updated (from English source) and translated independently

Process: Enabled by SDMX!



Hidden Technical Flows=User-Friendly!





Role: Interface for translators

1. Web based
2. Translator friendly
3. Version control integration
4. Effortless use of third-party machine-translation services





Role: Ensures accuracy and internal consistency of computer translation

1. Should be selected for subject knowledge area
2. Reviews translation options offered by software
3. Selects best fit and edits as necessary
4. Can provide preferred translation into Weblate dictionary





Role: Ensures sustainable quality control of translations

1. Identifies priority target languages
2. Selects appropriate human translators
3. Develops and manages quality control process



Applying the CAT Solution:

1a. Pilot results with reference metadata



- With support from the ECASTAT Trust Fund, the World Bank initiated a pilot study to evaluate the feasibility, reliability, and cost of using CAT software to translate SDG reference metadata. ROSSTAT and UNECE provided a human expert.
- Results showed the method was low cost, easy to use, and reduced translation time by 35%. The project expanded to translate all tier 1 metadata into Russian.
- Statistics Canada, and ECLAC and IADB collaborated with the World Bank to produce French and Spanish translations.
- Human translators found the method easy to use (little to no training needed). The software is free and results can be piped.

Applying the CAT Solution:

1b. Switching from Word to SDMX API



- In the beginning of our project, we used Word files containing reference metadata and converted them through Excel into machine-readable form. This was very labor-intensive.
- In collaboration with UNSD, we developed an SDMX authoring tool, which greatly improves accuracy and time required.

Applying the CAT Solution:

1b. Switching from Word to SDMX API



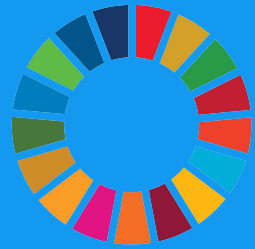
- The metadata authoring is kept flexible by collaborative use of a library to convert proprietary rich-text (Word) into open rich-text (HTML)
 - To **transition from historical processes**, the authoring tool allows database managers to upload prior Word documents containing reference metadata into the HTML format (**easy!**)
 - The authoring tool was recently updated to **apply controls** on version dates, improving data quality

Applying the CAT Solution:

1c. Improved process through harvesting



- With support from FAO, the project now routinely harvests all tier 1 and 2 machine-readable reference metadata from the SDG Data Lab's SDMX Reference Metadata API.
- The method sustainably manages precise translation updates.
- We are currently exploring interest in other target languages with UN Regional Commissions and national statistical offices
 - ✓ **French** reference metadata have been completed
 - **Spanish** reference metadata have been prepared for tier 1, but not yet tier 2 and quarterly updates
 - **Russian** reference metadata translation paused in 2022.



SDG METADATA TRANSLATION PROJECT

Using SDMX to facilitate computer-assisted
translation of reference metadata

Jennifer Park (Consultant, World Bank) Steven (Brock) Fanning (CODE), Abdulla
Gozalov and Harumi Shibata Salazar (both UNSD)



WORLD BANK GROUP



**United
Nations**

DESA
Statistics Division

**SDG Metadata
Translation Team**

October, 2023
Global Meeting on Statistical
Data and Metadata Exchange
(SDMX)