Food and Agriculture Organization
of the United Nations

SDMX-based Modernization at FAO

Data Modelling and Automation for Multi-Domain Systems

Luc Warde, *Data Analyst, Statistics Division*
Daniele Mansillo, *Data Engineer, IT division*

Context and Challenges

FAO is undertaking a comprehensive data migration and harmonization project to implement an SDMX native dissemination platform using .Stat Suite

Multi-Domain Environment

Managing diverse statistical domains in a hybrid SDMX and non-SDMX ecosystem

Domain Complexity

High dimensionality, long classifications, and lightweight flows across different statistical domains

Core Requirements

1. Consistency
2. Automation
3. Traceability

Compliance with FAO's Statistics and Data Quality Assurance Framework (SDQAF)

Our Dual Approach

1 Advanced SDMX Modelling

2 Automated SDMX Pipeline

Integration principle: Modelling defines the structural rules; the pipeline operationalizes them in a reliable way to ensure that updates are consistent and automatic.

Modelling Principles

Centralized system with modular designs adapted to specific domain requirements

SDG Indicators

One complex DSD supporting multiple data flows with comprehensive indicator coverage

Fisheries & Aquaculture

Long classifications and high dimensionality to capture detailed sector information

Derived Domains

Lightweight DSDs with focused flows for specialized statistical areas

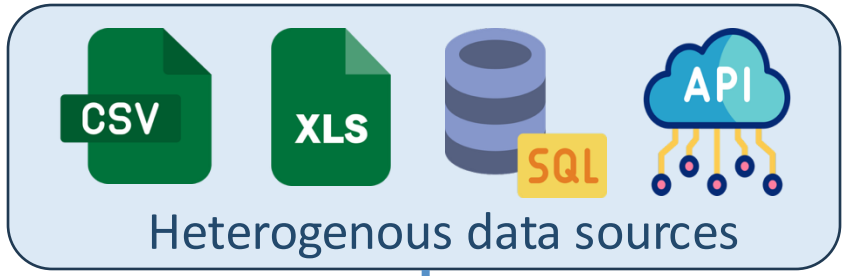
FAO dissemination standards applied in modelling: metadata, unit of measure, and observation status

Structures work both

clear labels

consistent codes

Automated Pipeline Architecture



SDMX Metadata Registry

1. Initial data preparation

Convert source data into SDMX-csv format with minimal manipulation

2. SDMX and non-SDMX Transformations

Apply Mapping using Structure & Representation Maps; use custom rules when necessary

3. Legacy mapping

Map to Legacy Dissemination structure in SDMX-csv

4. New Dissemination Mapping

Mapping from Legacy Dissemination structure to new Dissemination structure

SDMX 3.0 Innovations

- New Representation Maps features:
  - Fixed source and target
  - Time representations mapping
  - Regular expression and substring mappings
- Series constraints
- Constraints (planned): enforce structure, filter by validity period (DSD Constraints).

Pos	Source		Target
#	TIMEPOINTYEARS	TIMEPOINTMONTHS	String
1	(.*)	7013	\1
2	(.*)	7021	\1
3	(.*)	[2-4] (.*)	\1-\2

Example of Representation Map with regular expressions and substring mappings

ⓘ Key principle: Modelling anticipates future needs; the pipeline implements current capabilities while staying ready for evolution.

Where SDMX Meets the Pipeline

01

Reference Model Foundation

DSDs and codelists provide the structural blueprint for all pipeline mappings

02

Transformation Rules

Representation mapping tools convert modelling rules into actual data transformations

03

Metadata Harmonization and Pattern Reuse

Ensures consistent, comparable outputs across all statistical domains

04

Process Repository and Coordination

SDMX serves as both data and process repository, enabling cooperative maintenance (unlike code-oriented systems)

Practical benefits & Next Steps

1 For internal data management

→ Less manual work
→ Timely updates
→ Coherence and consistency across domains
→ Traceability

2 For international reporting

→ Alignment with global standards
→ Interoperability

3 For end users

→ Comparability and coherence
→ Accessibility to updated statistics
→ Clarity on exploring data and large classifications
→ Reinforce trust

© FAO / Cristiano Minichiello

Coming Next: Expanding SDMX 3.0 adoption with advanced constraints, generalized filters, and enhanced hierarchical codelists.

شكرا

謝謝

Merci

Thank You

Благодарю

¡Muchas Gracias!