Use of Standards and Models to Implement Transversal IT Platforms for Official Statistics

Conference on Smart Metadata for Official Statistics (COSMOS 2024)

Paris, France

Prepared by: Juan Muñoz / Silvia Fraustro
Problem Statement

Systems/Data Silos

Transversal Platforms
Proposal

Official Statistics Lifecycle

Enabling

Statistical Interoperability

Supporting

Transversal IT Platforms

Making use of

International Standards and Models
Governance Framework

- Provides a direction
- Promotes and sponsors the change
- Clarifies decision rules and solves discrepancies
- Sets standards, tools, and technologies

Architecture

- Aligns concepts
- Defines components and interactions
- Reveals the path and aligns the required efforts

DATA GOVERNANCE FRAMEWORK FOR STATISTICAL INTEROPERABILITY (DAFI)

THE OPEN GROUP ARCHITECTURE FRAMEWORK STANDARD (TOGAF)
Governance Framework

Organization

Guidelines

Processes

Principles

GAMSO

SDMX Governance Bodies

Modernisation Maturity Model

The Handbook on Management and Organisation of National Statistical systems

DAFI

GSBPM

Internationally Agreed Methodologies

FAIR
GSIM
Sets a common language

TOGAF
Sets an enterprise-wide integration model (Enterprise Architecture)
BUSINESS PROCESSES

1. Specify Needs
2. Design
3. Build
4. Collect
5. Process
6. Analyse
7. Disseminate
8. Evaluate

DATA

GSIM objects

GAMSO Supporting Activities

MMM
Skills and Abilities to Improve Processes

CSDA
Skills and Abilities to Manage Data

Specify
Needs

Design

Build

Collect

Process

Analyse

Disseminate

Evaluate

Inputs

Inputs GSIM objects

Outputs

Outputs GSIM objects

Supporting
Activities
Capacity to share and make use of statistical information among different parties or electronic systems without distortions of its meaning, not needing to communicate to get additional specifications or make ad-hoc adjustments for each specific case.
Domains Integration and FAIR Principles

SDMX Information Model
Artefacts that help to enable interoperability and implement FAIR Principles

Connecting Points:
• Conceptual Definition
• Geographical Reference
• Period of Time

<table>
<thead>
<tr>
<th>Statistical Concept</th>
<th>Description</th>
<th>Concept ID</th>
<th>Version</th>
<th>Representation</th>
<th>Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of Measure</td>
<td>Unit in which the data values are aggregated</td>
<td>UNIT_MEASURE</td>
<td>1.0</td>
<td>CI_UNIT_MEASURE</td>
<td>SDMX</td>
</tr>
<tr>
<td>Time Period</td>
<td>Temporal or point in time to which the observation actually relates</td>
<td>TIME_PERIOD</td>
<td>1.0</td>
<td>Observational Time Period</td>
<td>SDMX</td>
</tr>
<tr>
<td>Frequency of Observation</td>
<td>Time interval at which observations occur over a given time period.</td>
<td>FREQUENCY</td>
<td>1.1</td>
<td>CI_FREQ</td>
<td>SDMX</td>
</tr>
<tr>
<td>Reference Area</td>
<td>Reporting Country in ISO code (The country, or geographical political group of countries that the standard economic phenomenon relates to)</td>
<td>REF_AREA</td>
<td>1.0</td>
<td>CI_AREA</td>
<td>SDMX</td>
</tr>
</tbody>
</table>

SDMX Concept Scheme
Data Management

1. Specify Needs
2. Design
3. Build
4. Collect
5. Process
6. Analyse
7. Disseminate
8. Evaluate

- Creation
- Enabling
- Collection
- Storage / Maintenance
- Integration / Processing
- Analysis
- Synthesis / Representation / Visualization
- Synthetic / Designed Data
- Existent Data
- Inception / Design
- Synthetic / Designed Data
- Existent Data
- Detection of Need / Evaluation
- Specify Needs
- Design
- Build
- Collect
- Process
- Analyse
- Disseminate
- Evaluate

GSBPM Mapped to the Data Lifecycle
Technologies

- CSPA
- Microservices
- Virtualization, Containerization
- APIs: SOAP, RESTful, GraphQL, gRPC, WebSocket, Webhook
- Artificial Intelligence
- Data Science
- Big Data
- Geolocation: GPS, Galileo, Glonass, Beidou, QZSS, GNSS
- CAWI
- CAPI
- CATI
- Cloud Computing: IaaS, PaaS, SaaS, Private, Public, Hybrid
- SOA, Microservices
- SOA, Microservices
- Design
- Collect
- Process
- Analyse
- Disseminate
- Evaluate
- Specify Needs
- Build
Cybersecurity TP

Applications Architecture
- Information Systems for Specific Production Support Purposes

Data Architecture
- Internal and Temporary Databases and Repositories for Production

Technological Architecture
- Common Software and Collaboration Services TP
- Communications and Computing Infrastructure TP

Information Systems TP
- Information Systems for Specific Dissemination Support Purposes

Electronic Data TP
- Electronic Data TP (Main Databases and Repositories)

Auxiliary Databases and Repositories for Dissemination

TP: Transversal Platform
INEGI’s Statistical and Geographic Process Model (MPEG) is an adaptation of GSBPM.

MPEG

Modelo del Proceso Estadístico y Geográfico

INEGI’s Statistical and Geographic Process Model (MPEG) is an adaptation of GSBPM.

INEGI’s Site

DWH

Data Lake

Reports

PTIE

Transversal Platform for Electronic Information

PTIE

Transversal Platform for Electronic Information

Integrated Metadata Database

Metadata Consultation Systems

Metadata Consultation Systems

Metadata Distribution Center (geographic)

Open Data View of IPI Programs

Data Flows

SDMX-DSD

Metadata

Evidence Record Templates

Evidence Registration System

PTRACKING

Standardized Evidences from the Production Process

Specify Needs

Design

Build

Collect

Process

Analyse

Disseminate

Evaluate
Conclusions

• A high-level governance framework is an indispensable condition for achieving the goals of enterprise-wide projects.

• Transversal platforms represent a concrete option that contributes to achieving FAIR principles.

• It is desirable to achieve the objectives of the FAIR principles as they promote features that facilitate better information management and the implementation of better information services to satisfy the user’s needs.

• The existent standards and models from the official statistics community complement the general ones to design specialized environments that support the processes related to the statistics life cycle.
THANK YOU

Juan Muñoz
Juan.Munoz@inegi.org.mx