Progress in Sharing Statistical Data and Metadata using International Standards

The implementation of SDMX in Istat beyond the dissemination and reporting
Francesco Rizzo (rizzo@istat.it)

Keywords: SDMX, data sharing, data reporting, data dissemination, SDMX Reference Infrastructure, SDMX-RI, RDF, Google/DSPL

1. INTRODUCTION

Since 2010, Istat started a multiannual programme, named Stat2015, aligned with the ESS 2020 vision, whose main aim is to modernize its information system and the way to produce statistics for responding efficiently to the new challenges. An important prerequisite for the realization of Stat2015 is the standardization and industrialization of the statistical processes through the adoption and implementation of statistical and technical standards, such as SDMX.

In this context, Istat has been developing and putting in production a set of cross-cutting building blocks (SDMX Istat Framework) allowing to move from the simple use of the standard for reporting data and metadata to International Organizations towards a more strategic perspective, namely for streamlining internal business processes through the harmonization of content, the management of metadata, and the dissemination of data and metadata in several formats and for several platforms.

The main aim of this paper is to illustrate the SDMX Istat Framework and how it responds to the different business needs identified within the Stat2015 multi-annual program.

1.1. Background

Istat has been working on SDMX since 2004. At the beginning, in order to get experience, Istat participated to pilot projects launched by Eurostat within the European Statistical System: SDMX Open Data Interchange (SODI), Demography Rapid Questionnaire, EuroGroups Register, Census Hub. During those pilots, Istat developed a set of software SDMX 2.0 compliant.

From 2009 to 2012, Istat has participated to ESSnets [4][5] on SDMX (phase 1 and 2), contributing to the development of software and guidelines shared with ESS members. Furthermore, Istat actively participated in the design of the SDMX Reference Infrastructure developed by Eurostat.

Starting from 2010, SDMX became part of the strategic multi-annual Istat program, named Stat2015. In this context, Istat has defined an SDMX implementation strategy that can be summarized in Figure 1:

---

1 ISTAT, Italian National Institute of Statistics.
**Collaborations**: it is a fundamental part of the strategy, allowing a reduction of development costs through the sharing of SDMX software and know-how with other organizations. Currently, Istat is part of several communities that, besides other topics, are also focusing on SDMX aspects (OECD SIS – Collaboration Community, Statistical Network, Eurostat SDMX-RI User Group). Furthermore, Istat is running a bilateral collaboration with INEGI-Mexico and is part of the SDMX Technical and Statistical Working Groups.

**Dissemination and Reporting**: the SDMX Single Exit Point (SEP) [3] is already in production. The SEP is used for reporting to International Organization (e.g. Census data and National accounts data to Eurostat; Short Term Economic Statistical data to OECD, SDDS Plus data to IMF), and to disseminate data/metadata from the dissemination data warehouse (I.Stat) in machine-to-machine modality.

**Metadata**: the Istat Unified Metadata System is aimed at managing structural and reference metadata in an integrated way. Structural metadata are handled in an SDMX registry and synchronized with the dissemination/reporting metadata repository, while the reference metadata system (SIDI/Siqual) has been extended in order to be compliant with the Euro SDMX Metadata Structure (ESMS) and the ESS Standard for Quality Report Structure (ESQRS).

**Other formats**: in order to speed up data interoperability, the SDMX SEP web service has been extended to generate other formats, starting from the SDMX-IM: Resource Description Framework (RDF), Google dataset publishing language (DSPL), CSV and JSON.

**End users**: the interaction with the SDMX SEP is facilitated by a series of guidelines, tools and web GUI, accessible from the Istat website [3]. In particular the web GUI connected to the metadata repository allows browsing structural metadata needed to interpret correctly all of the SDMX datasets that can be extracted from the SEP, while the SDMX MS-Excel plug-in can be used to extract SDMX datasets from the SEP.

**National Statistical System**: Istat has been implementing a “distributed data warehouse” accessible through a hub web application. End users can browse the hub to define a
dataset of interest via SDMX structural metadata and retrieve the data directly from the data providers' databases. The whole architecture will be based on an SDMX hub architecture where the central application communicates with remote web services through SDMX messages.

2. METHODS

One of the main constraints in developing the SDMX Istat Framework has been the optimization of the development costs and searching the suitable funds. Istat decided two strategic lines:

- harness the opportunity of Eurostat grants;
- reuse and extend the SDMX Reference Infrastructure (SDMX-RI) [2] developed by Eurostat.

The SDMX-RI “ultimate version” is based on the SDMX Common APIs, an open source project fostered by the SDMX Secretariat and implemented (SdmxSource) by Eurostat and Metadata Technology Ltd.

2.1. Financing the development

In December 2013, Istat signed with Eurostat a grant agreement on “Horizontal and vertical integration: implementing technical and statistical standards in ESS”, with a duration of 24 months (extended till August 2015).

The grant foresees 24 deliverables split into 4 working packages:

WP1: coordination;

WP2: enhancement of the SDMX Reference Infrastructure, and its integration within the Istat information system;

WP3: enhancement of the Istat metadata management system;

WP4: contribution to the development of the SDMX standard and implementation of capacity building action.

7 units from 2 different Directorates are involved in the grant, for a total of 778 man/days. Furthermore, the grant is used to support over 1000 man/days from a subcontractor.

2.2. The overall SDMX architecture in Istat

Within Istat, an SDMX architecture based on SDMX-RI and complemented by the SDMX Istat Framework building blocks is already in place.

The main aim of this architecture is to allow end-user applications to browse and query data stored in the dissemination data warehouse I.Stat. To this purpose, the suitable SDMX structural metadata (Data Structure Definitions, Code lists, Concept schemes, Category schemes, etc.) have been developed and mapped against the data stored in I.stat.

The SEP web service gives access to end-user applications in machine-to-machine modality, using SOAP and/or REST protocols, and sending and receiving SDMX messages. From the SEP web service, it is possible to extract data in other formats (RDF, Google/DSPL, JSON).

The same architecture is used for reporting to International Organizations in both push and pull mode. For example, datasets from National Accounts, Job Vacancy Survey and Labour Cost Indicator are extracted by domain managers and sent to Eurostat through the eDamis channel (Eurostat's Single Entry Point). At the contrary, census data and
Economic indicators are collected in pull mode by Eurostat (Census Hub), OECD (STES) and IMF (SDDS Plus).

Figure 2. SDMX overall architecture in Istat

2.3. The SDMX Istat Framework building blocks

The SDMX Istat Framework is a set of pick-and-choose building blocks allowing a statistical office to handle data and metadata that has to be shared with other organizations or end users. It acts as a series of add-ons of the SDMX-RI in order to cover most of the aspects identified within the Istat SDMX strategy.

The following component diagram depicts the view of the SDMX Istat Framework:
The SDMX Istat Framework building blocks are the ones with brown colour, while the orange blocks identify extensions of the SDMX-RI components developed by Istat;

**Builder** - it allows creating an SDMX compliant database.

**Loader** - it allows loading into the SDMX compliant database csv or SDMX files.

**On-the-fly Web Service** - it is responsible for exposing the data and structural metadata using a Web Service interface that follows the guidelines of the SDMX v2.0 standard for Web Services and the Web Service Guidelines for SDMX v2.1 standard [6]. Data can be extracted directly from the SDMX compliant database created through the Builder or from any dissemination database mapped using the Mapping Assistant. Furthermore a Data Collector Application can negotiate other format beyond SDMX-ML such as RDF, Google/DSPL and JSON.

**Metadata Web GUI** – it provides a Web graphical user interface that helps domain managers in designing SDMX artefacts, and end-users to browse and query the structural metadata repository (extended Mapping Store database).

**RSS Generator** – it is responsible for generating a feed entry on the event of new data arriving from the Loader. The generated feed is able to trigger Data Collector Applications providing them the suitable SDMX queries.

3. **Conclusions**

International standards, such as SDMX, can help the modernization processes within the statistical Organizations, but efficient tools should be available in order to speed the implementations. The SDMX Reference Infrastructure could be considered a very good starting point that can be easily integrated with existing information systems and extended using the SDMX Common APIs: in this context the SDMX Istat Framework represents a real use case.

**References**


