A new dissemination tool for the Italian statistical System

**Keywords:** Sistan, hub, dissemination, datawarehouse, SDMX

# Introduction

The National Statistical System (*Sistema Statistico Nazionale* ‒ Sistan) is a network linking more than 3 thousand statistical offices located in Ministries and other central institutions, Regions, Provinces, Municipalities and other territorial authorities, such as Prefectures and Chambers of commerce, and selected private organisations.

Among the Sistan bodies, only 13 institutions are Other Statistical Authorities producing European statistics (ONAs), therefore having a specific role and responsibilities in the wider European statistical system. Nevertheless, a significant part of the others regularly produces relevant data at the national level, included in the National statistical program.

According to legislation, Istat, as the central node of Sistan, has a clear responsibility to direct, coordinate, promote and provide technical assistance and training to Sistan offices. It includes supporting their activities and data dissemination.

In order to fulfil this task, as recommended during last round of peer reviews [1], Istat has been working on the creation of the “Public statistics Hub”. Such project consists in the construction of a national DataWarehouse (DW) for the dissemination of macrodata produced by the national statistical system, in order to improve coordination, integration and quality in the Sistan data dissemination. An important objective of the project is also to create a unique entry point where users can find data in a user-friendly, harmonized format. A pilot tool was developed to highlight the system potentialities and possible drawbacks. Based on that experience, and in the light of the new Istat role as provider of strategic services within the Italian Public Administration framework [2] the project is now moving to a wider, more operational phase.

# The “Sistan.Hub” project

The project aims at reinforcing cooperation between Istat and the other Sistan bodies in order to improve data dissemination and spread knowledge and usage of international standards such as the ISO IS-17369 Statistical/ Data and Metadata eXchange standard (SDMX). [3]

Final users will also benefit of the new DW as it offers the possibility to find and download data through a unique access point and a single navigation environment. In fact, one of the issues more often highlighted by users is the heterogeneity of data presentation in different dissemination DW. For non-professional users, it can be also challenging to find the data of interest in the vastness of the web, often in different formats and with scarce if any metadata.

An important feature of the project is the territorial point of view: an ad hoc area of the DW will be devoted to Regions, so that data of specific local interest can be easily found.

In principle, participation in the project is on a voluntary basis, through a simple procedure published on the Sistan web site (www.sistan.it), but in general, it is strongly recommended to bodies producing data with a national relevance. In the territorial area, data produced only for a specific region can be disseminated, given that they fulfil the quality criteria for official statistics.

The final objective is to build a federation of the DWs managed by the Sistan bodies, gradually and progressively integrating them in a unique framework.

# Methods

The Hub architecture allows to disseminate datasets stored in distributed databases (peripheral nodes) from a unique central point (Hub node). A node is a webservice that can be registered into the application to retrieve information by it. The machine-to-machine communication by web service occurs on the basis of SDMX exchange protocol.

The dissemination of data is carried out through every node of the system: it has the effect of entering each individual entity within a network, making its data navigable through the Hub. In such design, the central body is not a real node, but it is a tool that seeks congruence between the data held by different bodies. Its main purpose is to provide an integrated dissemination system gateway through which interconnected and integrated data from different organizations can be shared.

The end-user will browse, select and visualize high quality integrated statistical information always using the same Graphical Interface. Data are not stored centrally but fetched, in pull modality, from the data producers’ databases upon request of the data users. Datasets of different data producers can be categorized in the same theme tree.

From a technical point of view, the portal is based on the SDMX Reference Infrastructure, a set of tools developed by Eurostat. The advantages of using such architecture are as follows:

* The data and metadata contained in the "local" databases are neither modified nor duplicated but only mapped on the SDMX Data Structure Definitions (DSDs): so, the warehouse of aggregate data integrates and synchronizes components and activities so far “disconnected”;
* Data relating to different statistical domains and coming from different databases can be disseminated through a single end-point in a transparent way for end users;
* The infrastructure allows a harmonization of metadata, as well as a harmonization of the workflow (some functions are performed by statisticians, others by IT staff).

The idea is to take advantage of the SDMX Information Model in order to increase the standardisation of statistical metatada among different statistical domains and in the various phases of the statistical process (starting from classifications and concepts by creating, collecting, editing, managing and sharing SDMX Codelists and ConceptSchemes),

The design of a multipurpose web application has been developed, in a twofold perspective: the “MetaManager” feature of the application serves the purpose of managing (import, create, edit, delete, export in SDMX-ML 2.0, SDMX-ML 2.1, csv in case of ItemSchemes, etc.) SDMX structural metadata (Codelists, ConceptSchemes, CategorySchemes, DSDs). The use of a second (strictly combined) feature, the “DataManager”, is particularly indicated for dissemination purposes because through the creation of the SDMX DSDs is possible to define the multidimensional structure of the hyper-cubes that can be physically created inside a relational SDMX- compliant database for storing data to be disseminated, through the definition and publication of SDMX Dataflows linked to specific views (queries) on the data of the hypercubes.

In particular, the structural metadata repository can actually be considered a real SDMX Registry that can be used, precisely with a view to harmonizing and standardizing the statistical contents, as a reference repository for the creation of an information architecture that is "metadata-driven". Furthermore, in order to make the information (variables, official classifications, operational code lists) on structural metadata stored in the registry fully usable also for end users, a web interface is being created (currently available only in a prototype version and usable only within the institute) for the metadata navigation that describes metadata in an understandable way for non-specialized users, with advanced search and correlation functions provided.

Istat considers this approach strategic, so it is intended to overcome the issues that have held back the project. Therefore, apart from consultancy and training, Istat is offering the adhering entities a suitable open source toolkit, which allows facilitating the industrialization of the following sub-processes:

* metadata management;
* data exchange (push and pull) between organizations;
* dissemination of data / metadata (GUI and machine-actionable);
* creation of data dissemination and updating databases;
* transformation into other formats (RDF, DCAT);
* implementation of scalable architectures.

# References

1. <https://ec.europa.eu/eurostat/documents/64157/4372828/2015-IT-improvement-actions/a1b18bf0-56de-4984-a8ea-6e188979f4dd>
2. https://forumpa2020.eventifpa.it/it/event-details/?id=9461
3. <https://sdmx.org/>