DATA BROWSER AND METADATA MANAGER: THE NEW DISSEMINATION PLATFORM FOR ISTAT AGGREGATE DATA
Index

- The choice of a dissemination platform
- Context and strategies in Istat aggregate data dissemination
- The technological solution: Data Browser, Meta Data Manager
- Data Browser features
- Meta Data Manager features
The choice of a dissemination platform: Multidisciplinary approach

Different factors to consider before choosing a dissemination platform:

- Pre-existing complex system:
  - Different kinds of data: Administrative data, survey data, micro and aggregate data
  - Different strategies and processes in the same organization
  - There is no better platform technology ever for data dissemination, but the one having the best balance between different aspects.
  - Technology must adapt to the objectives, the characteristics of the data and processes

Multidisciplinary approach:

- Experts in business organization, statisticians, data modeling, IT
The choice of a dissemination platform: not only a technological issue

URUK III 4500 years ago, tablets for the dissemination of agricultural data:

- A long form and a short form of the tablets structure
- Subdivision of animals by sex, by age, the animals born from the precedents, the related production of butter and cheese, those responsible for registration and administrative information

Some aspects of data dissemination are a kind of archetype and involve the users needs, the nature of the data, the organisational system, and available technology:

- Metadatation, data modelling, data comparison and linkage
- Standardization and stability of the dissemination process

Consider context, strategies and technological solutions
**Context and strategies**

Specific context in Istat for a large-scale dissemination of aggregate data:

- Aggregate data from a large number of surveys under Eurostat Regulation whose results all need to be disseminated. In some cases aggregate data from a combination of administrative and survey data.

- A large number of processes already existing of data aggregation and dissemination

- Data dissemination structure is almost stable over time

- Data from different surveys that can be integrated only partially for specific domains of analysis
Context and strategies

Strategies:

1) Allow users to have a homogeneous environment and reduce access costs

2) Gain flexibility in comparing pre-structured aggregate data

3) Synthetic data representation solutions without modifying data structure and underlying processes

4) High data store capacity per single data cube

5) Migrate to a standard (SDMX) in order to:
   - take advantage of technological modularity (changes or improvement in each component of the platform) without changes in the data modeling and preserving data dissemination processes in the future;
   - allow machine to machine access using API, and avoid misalignments between what is stored in DB and what is exposed via webservice.
Open source platform developed by Istat and based on two components:

1) **Data Browser (Front end):** to allow external users to browse data

2) **Meta e Data Manager (Back end):** to create SDMX data structure, data cubes, data mapping, data loading and data flows (for Istat internal users)

Front and back end technologies are based on the SDMX standard.

Starting from 2020, Permanent Census of Population and Housing data have been disseminated on the new platform: https://esploradati.censimentopopolazione.istat.it
Different dissemination systems towards a new single platform

Istat aggregate data dissemination systems

- I.Stat (800,000 users per year)
- Permanent census
- COEWEB (data on foreign trade)
- Other data bases storing aggregate data

New single platform

Unique environment for basic and advanced users

New features

Data browser

Improved performance in data loading, interactive dashboards, bookmarks, enhancement of the data search and data selection at territorial level

Meta Data Manager

Open source, modular, native based on an international standard (SDMX)
DATA BROWSER: Data comparison and synthetic prospects (Dashboard)

- Combination of maps, charts, tables and explanatory text
- Objects created by combining different dataflow from different datasets and different nodes
- Interactive selection
- Full screen enlargement for each object
- Data export in different format
- Image export for charts and maps

Territorial data selection system at regional, provincial and municipality level: all the dashboard data automatically updated for a specific territory.
DATA BROWSER: Data comparison and synthetic prospects (Bookmarks)

- Save your query as a bookmark
- Switch your researches selecting a bookmark below the query (switch like in the Excel sheets)
- Each bookmark can be related to different dataflows belonging to the same or different datasets
- It is also possible to save bookmarks from dataflows belonging to datasets of different nodes from other institutions
The thematic navigation is possible both through a horizontal navigation system based on thematic buttons and also through a vertical thematic tree.

Combination of text search and thematic filter: the textual search system returns a list of data flows (queries) that can also be filtered using the thematic tree.
DATA BROWSER: Other features

- Caching system for fast data loading
- Export data and metadata (also data structure) in different formats (SDMX, CSV, EXCEL)
- Export the query in Excel:
  - current table visualization on the screen
  - the entire dataflow: in this case entire dataflow is split into different Excel sheets
- Management of metadata (notes) for each item, and referential metadata at the dataflow level
DATA MANAGER: Migration process to a standard

Full migration of data and metadata (notes and referential metadata)

- Independence from a specific technological framework
- Full consistency between internal metadata and metadata exposed with SDMX webservice

DATA MANAGER: Migration process to a standard

Independence from a specific technological framework
Full consistency between internal metadata and metadata exposed with SDMX webservice
DATA MANAGER: webservice and single exit point

- Architecture natively based on an internationally recognized data transfer standard (SDMX)
- High data store capacity per single datacube
- Modular infrastructure with access via webservice
- New technological framework is developed by ISTAT and completely Open Source

Meta Data Manager Tool SDMX

OUTPUT

IT reference colleagues for technological aspects:
Francesco Rizzo (rizzo@istat.it)
Alessio Cardacino (alcardac@istat.it)
Simone Coccia (sicoccia@istat.it)
Thank you for your attention

CARLO BOSELLI | carlo.boselli@istat.it