On line, September 15 2022

2022 EXPERT MEETING ON DISSEMINATION AND COMMUNICATION OF STATISTICS

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

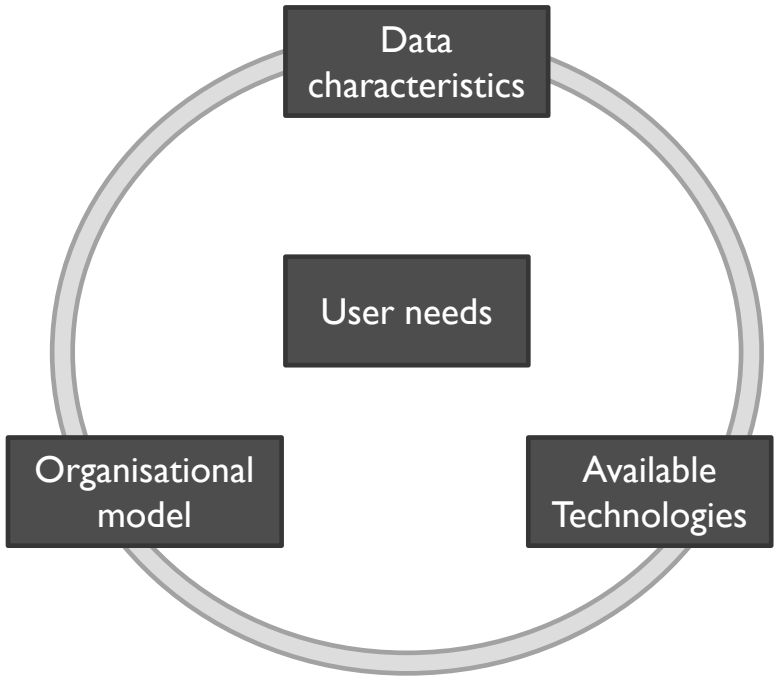
CARLO BOSELLI
Istat | Directorate for Communication

Index

- The choice of a dissemination platform
- Context and strategies in Istat aggregate data dissemination
- The techonological 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: Admistrative data, survey data, micro and aggregate data

Different strategies and processes in the same organization

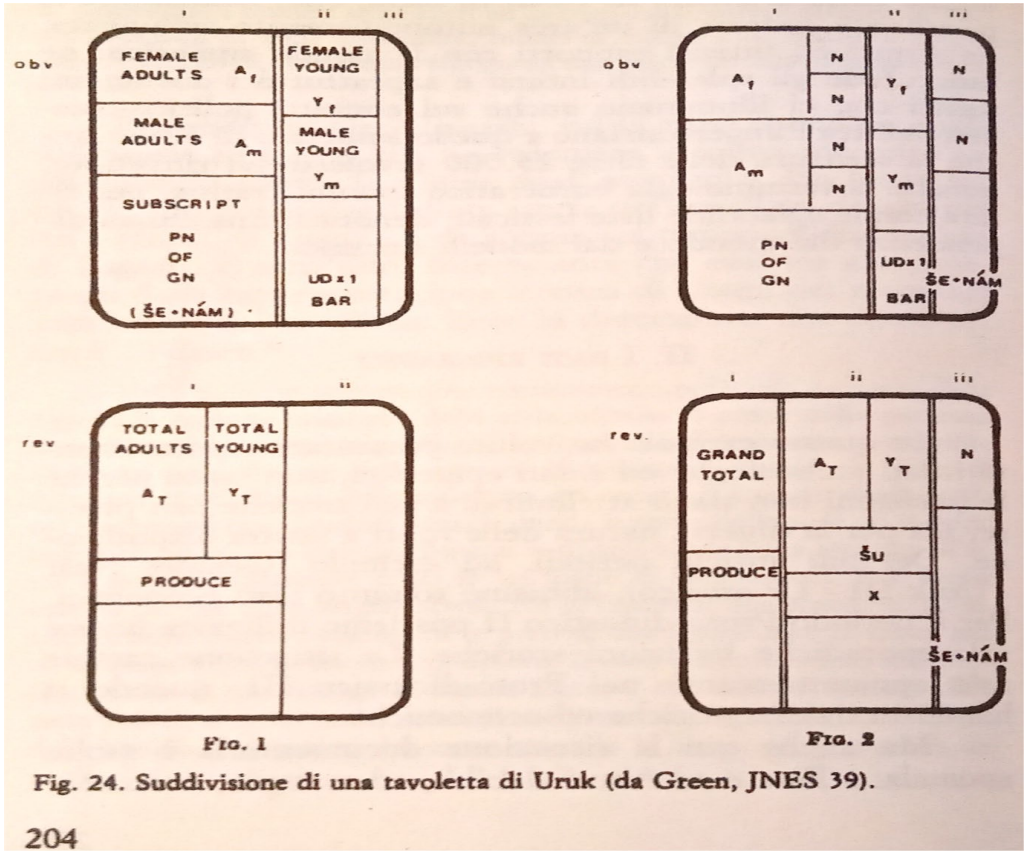
There is no better platform technology ever for data dissemination, but the one having the best balance between different aspects.

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



Structure of Uruk tablets for the dissemination of agricultural data. (M.W. Green, Journal of Near Eastern Studies 39 - Animal Husbandry at Uruk in the Archaic Period), in «I Sumeri» - Giovanni Pettinato - Bompiani

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.

The techonological solution: Data Browser, Meta Data Manager

Open source platform developed by Istat and based on two components:

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

Population and private households - regions, provinces and large cities	Piemonte	Valle d'Aosta / Vallée d'Aoste
Liguria	Lombardia	Provincia Autonoma Bolzano / Bozen
Provincia Autonoma Trento	Veneto	Friuli-Venezia Giulia
Emilia-Romagna	Toscana	Umbria
Marche	Lazio	Abruzzo
Molise	Campania	Puglia
Basilicata	Calabria	Sicilia
Sardegna	Time series	Download files (all municipalities)

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

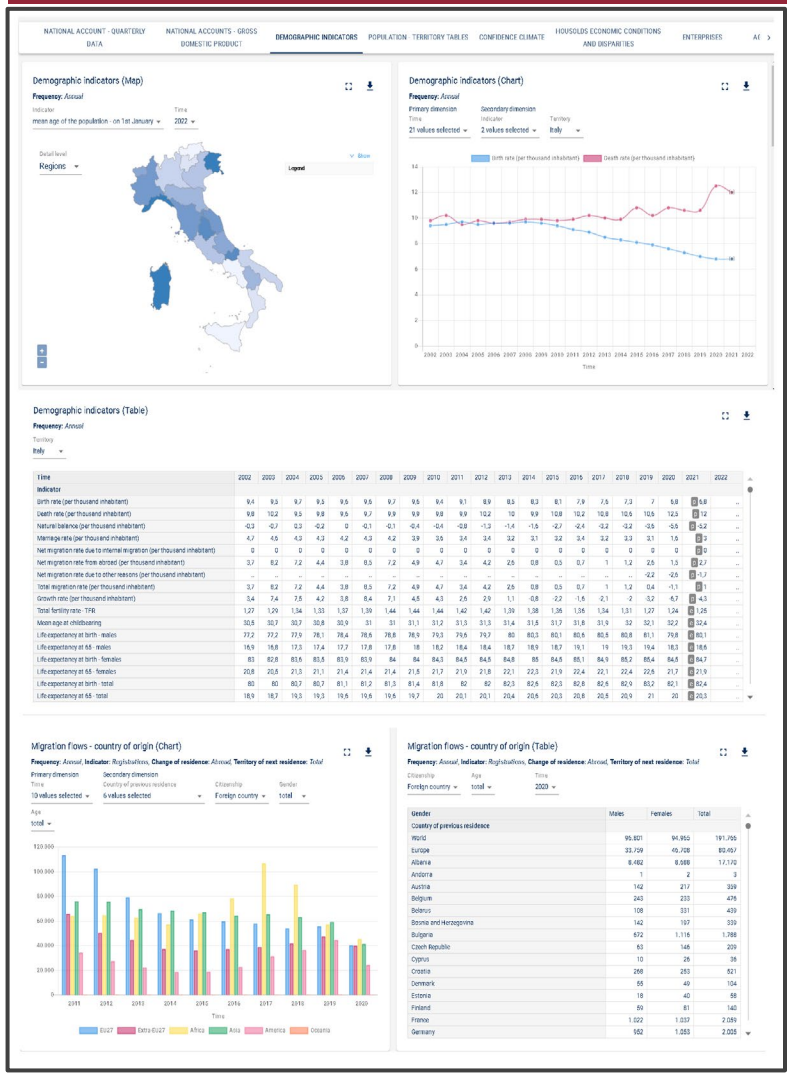
Data browser

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

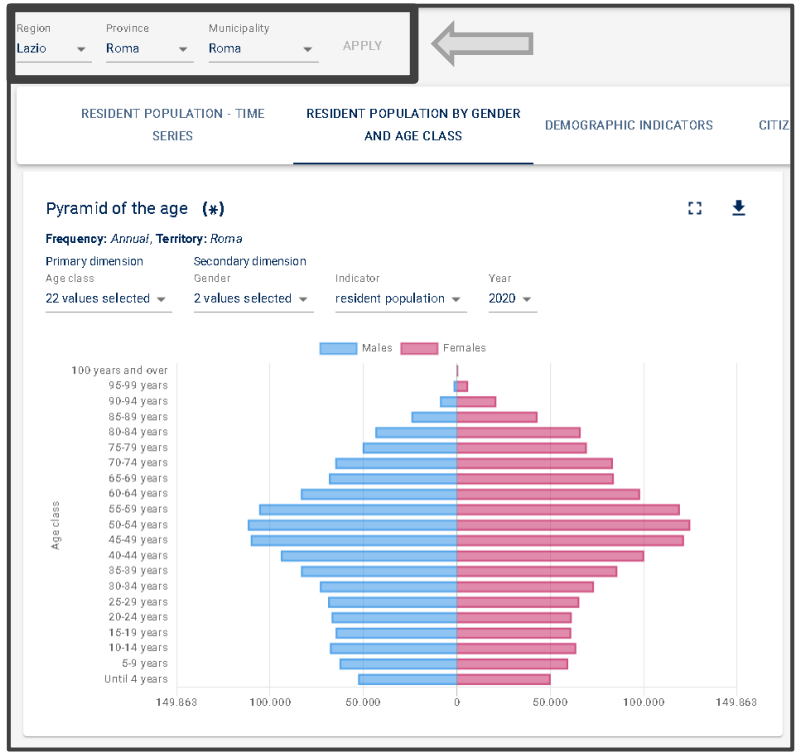
Meta Data Manager

Open source, modular, native based on an international standard (SDMX)

DATA BROWSER: Data comparison and synthetic prospects (Dashboard)

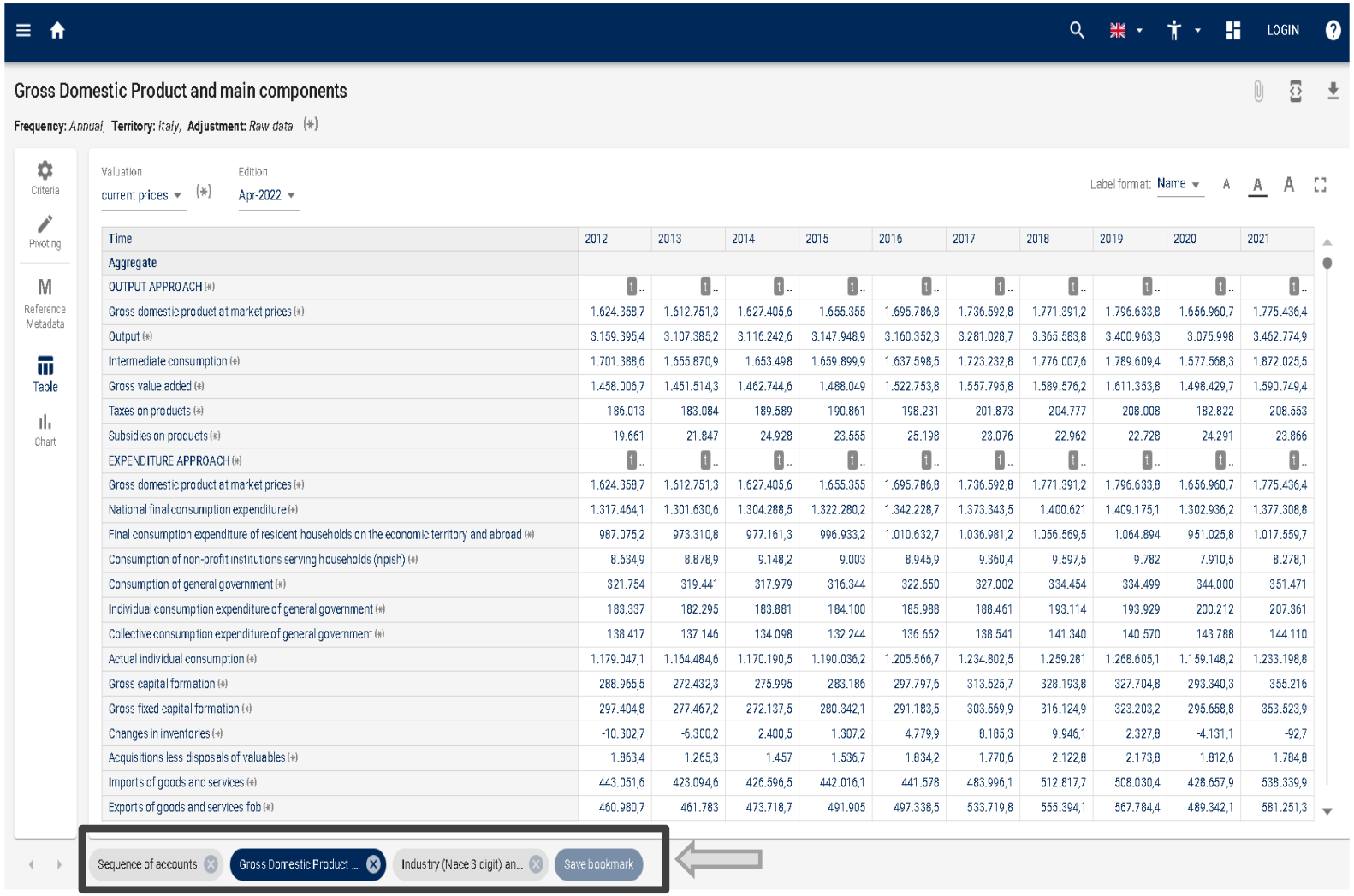


- Combination of maps, charts tables and explanatory text
- Objects created by combining different dataflow from different datasets and different nodes
- Interactive selection
- Full screen enlargement for each object
- Data export in different format
- Image export for charts and maps



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)

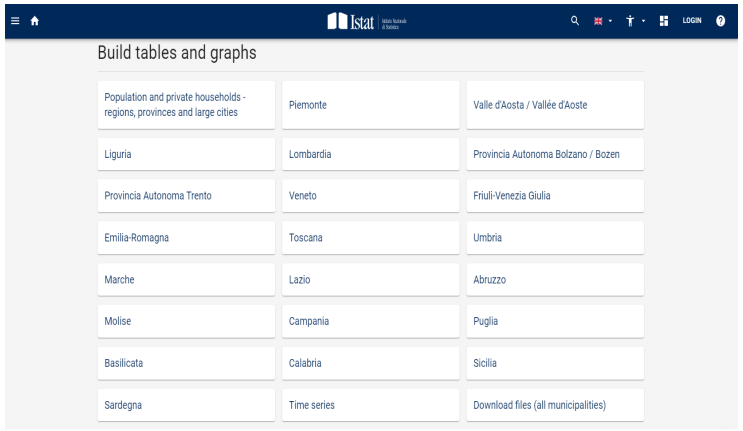


Time	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Aggregate										
OUTPUT APPROACH (*)	..	..	..	..	..	..	..	..	..	..
Gross domestic product at market prices (*)	1.624.358,7	1.612.751,3	1.627.405,6	1.655.355	1.695.786,8	1.736.592,8	1.771.391,2	1.796.633,8	1.656.960,7	1.775.436,4
Output (*)	3.159.395,4	3.107.385,2	3.116.242,6	3.147.948,9	3.160.352,3	3.281.028,7	3.365.583,8	3.400.963,3	3.075.998	3.462.774,9
Intermediate consumption (*)	1.701.388,6	1.655.870,9	1.653.498	1.659.899,9	1.637.598,5	1.723.232,8	1.776.007,6	1.789.609,4	1.577.568,3	1.872.025,5
Gross value added (*)	1.458.006,7	1.451.514,3	1.462.744,6	1.488.049	1.522.753,8	1.557.795,8	1.589.576,2	1.611.353,8	1.498.429,7	1.590.749,4
Taxes on products (*)	186.013	183.084	189.589	190.861	198.231	201.873	204.777	208.008	182.822	208.553
Subsidies on products (*)	19.661	21.847	24.928	23.555	25.198	23.076	22.962	22.728	24.291	23.866
EXPENDITURE APPROACH (*)	..	..	..	..	..	..	..	..	..	..
Gross domestic product at market prices (*)	1.624.358,7	1.612.751,3	1.627.405,6	1.655.355	1.695.786,8	1.736.592,8	1.771.391,2	1.796.633,8	1.656.960,7	1.775.436,4
National final consumption expenditure (*)	1.317.464,1	1.301.630,6	1.304.288,5	1.322.280,2	1.342.228,7	1.373.343,5	1.400.621	1.409.175,1	1.302.936,2	1.377.308,8
Final consumption expenditure of resident households on the economic territory and abroad (*)	987.075,2	973.310,8	977.161,3	996.933,2	1.010.632,7	1.036.981,2	1.056.569,5	1.064.894	951.025,8	1.017.559,7
Consumption of non-profit institutions serving households (npish) (*)	8.634,9	8.878,9	9.148,2	9.003	8.945,9	9.360,4	9.597,5	9.782	7.910,5	8.278,1
Consumption of general government (*)	321.754	319.441	317.979	316.344	322.650	327.002	334.454	334.499	344.000	351.471
Individual consumption expenditure of general government (*)	183.337	182.295	183.881	184.100	185.988	188.461	193.114	193.929	200.212	207.361
Collective consumption expenditure of general government (*)	138.417	137.146	134.098	132.244	136.662	138.541	141.340	140.570	143.788	144.110
Actual individual consumption (*)	1.179.047,1	1.164.484,6	1.170.190,5	1.190.036,2	1.205.566,7	1.234.802,5	1.259.281	1.268.605,1	1.159.148,2	1.233.198,8
Gross capital formation (*)	288.965,5	272.432,3	275.995	283.186	297.797,6	313.525,7	328.193,8	327.704,8	293.340,3	355.216
Gross fixed capital formation (*)	297.404,8	277.467,2	272.137,5	280.342,1	291.183,5	303.569,9	316.124,9	323.203,2	295.658,8	353.523,9
Changes in inventories (*)	-10.302,7	-6.300,2	2.400,5	1.307,2	4.779,9	8.185,3	9.946,1	2.327,8	-4.131,1	-92,7
Acquisitions less disposals of valuables (*)	1.863,4	1.265,3	1.457	1.536,7	1.834,2	1.770,6	2.122,8	2.173,8	1.812,6	1.784,8
Imports of goods and services (*)	443.051,6	423.094,6	426.596,5	442.016,1	441.578	483.996,1	512.817,7	508.030,4	428.657,9	538.339,9
Exports of goods and services fob (*)	460.980,7	461.783	473.718,7	491.905	497.338,5	533.719,8	555.394,1	567.784,4	489.342,1	581.251,3

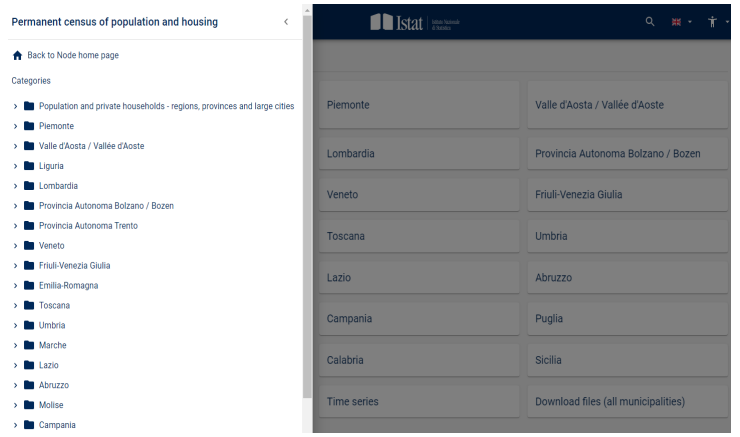
- Save your query as a bookmark
- Switch your researches selecting a bookmark below the query (switch like in the Excel sheets)
- Each bookmark can be related to different dataflows belonging to the same or different datasets
- It is also possibile to save bookmarks from dataflows belonging to datasets of different nodes from other institutions

DATA BROWSER: Data flow search system

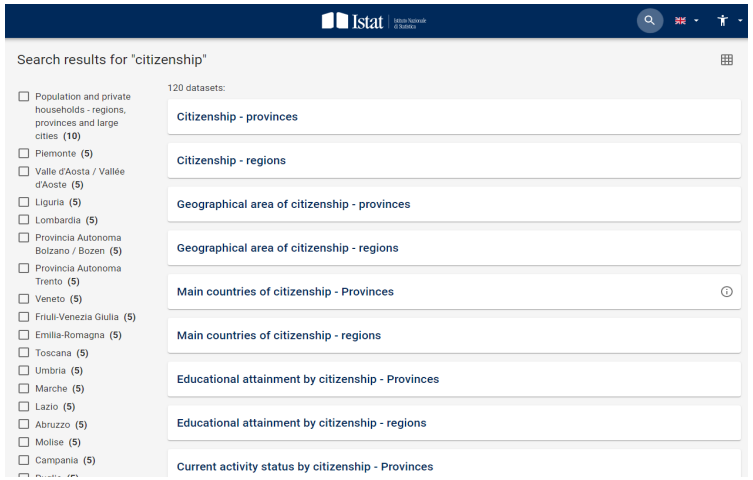
Thematic horizontal buttons



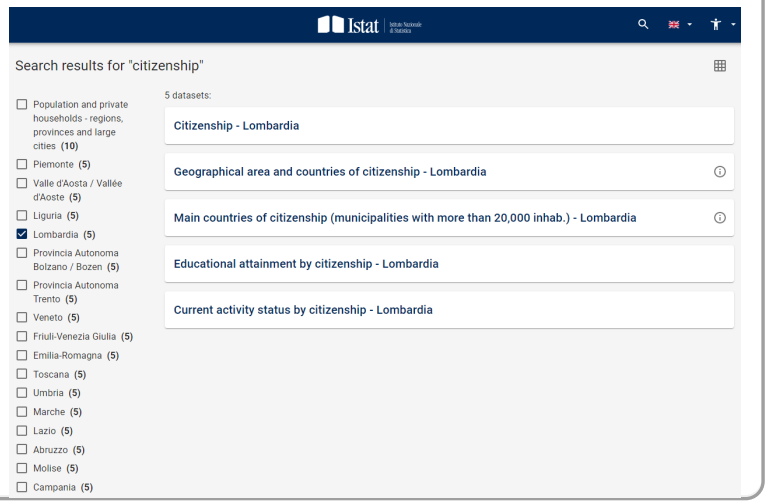
Thematic vertical tree



Text search



Text search and thematic filter



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

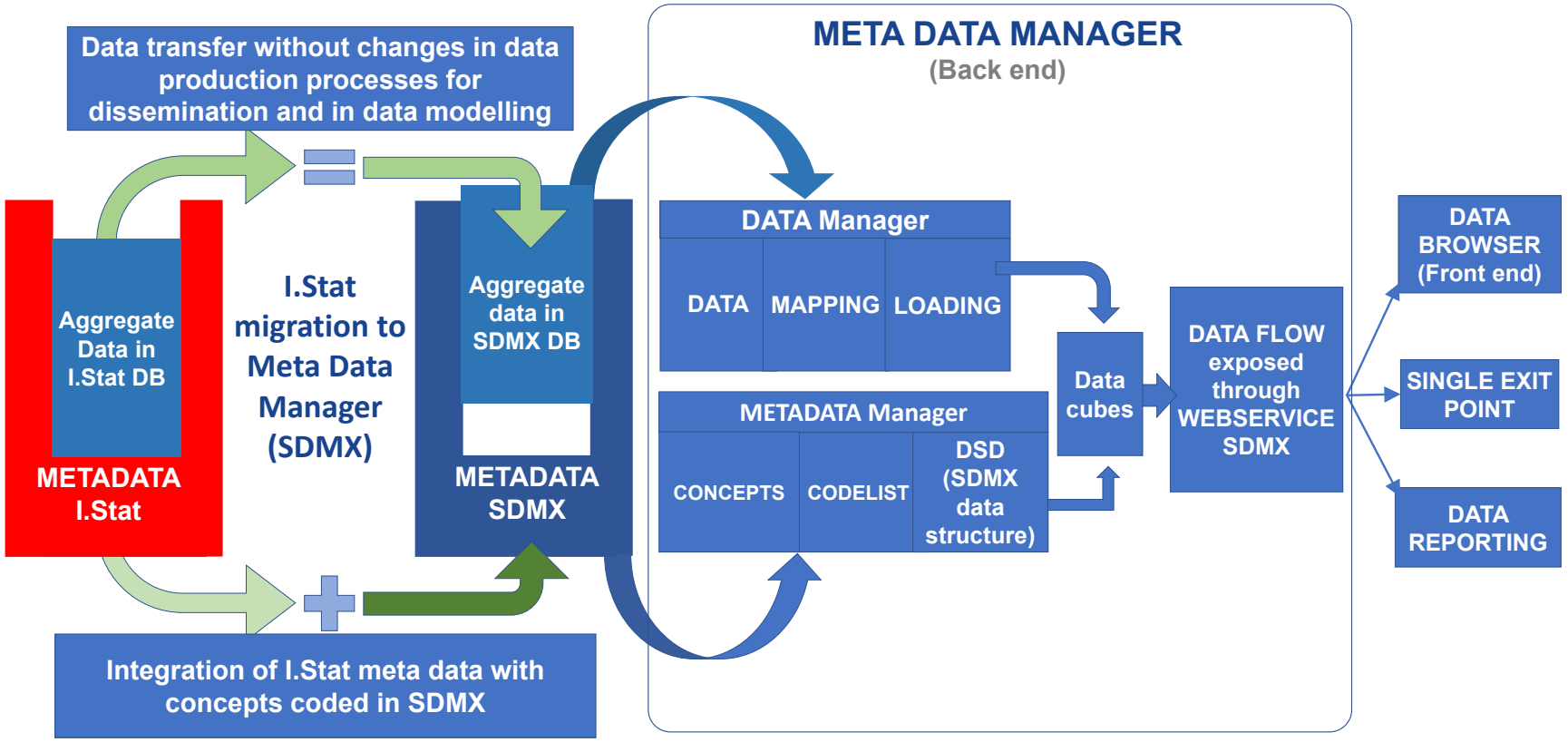
  - the entire dataflow: in this case entire dataflow is split into different Excel sheets

- Management of metadata (notes) for each item, and referential metadata at the dataflow level

DATA MANAGER: Migration process to a standard

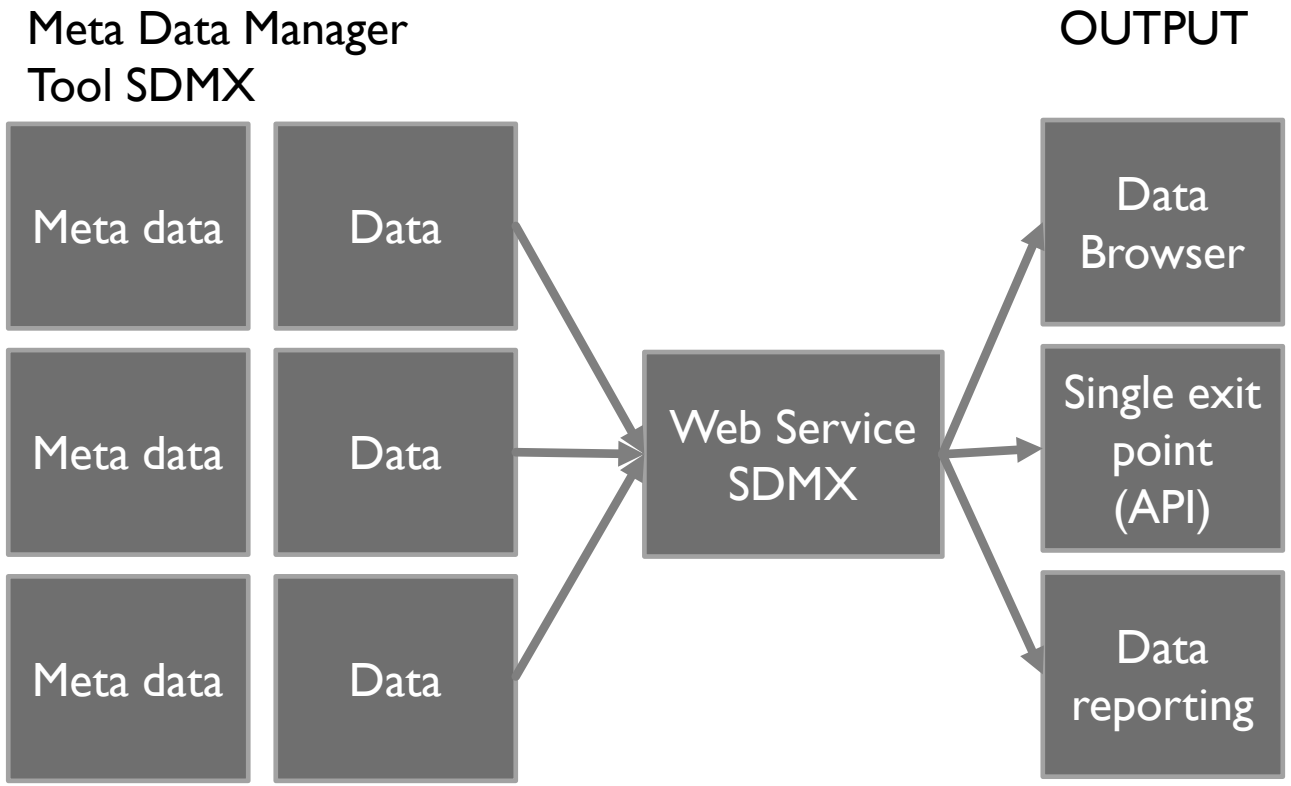
Full migration of data and metadata (notes and referential metadata)

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



DATA MANAGER: 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



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

Istat | Istituto Nazionale di Statistica