Istat | Istituto Nazionale di Statistica

"SDMX Global Conference 2023"

Rome, Italy

29 September 2025 – 03 October 2025

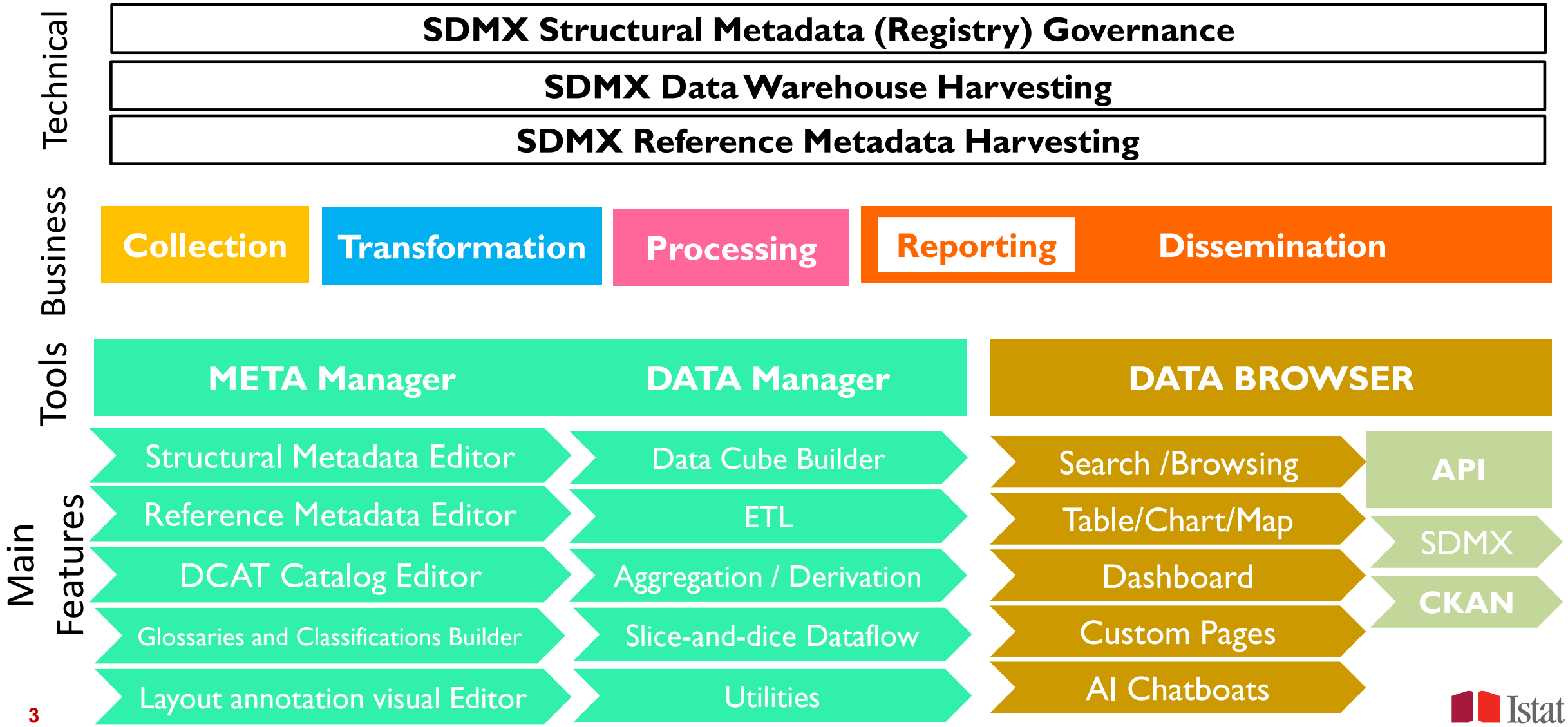
Bridging complexity and capacity: the strategic evolution of the Istat Toolkit to accelerate SDMX adoption and facilitating implementations

Francesco Rizzo (SDMX dissemination platform – project manager), Carlo Boselli (Dissemination database initiative - coordinator), Alessio Cardacino (SDMX toolkit development initiative – coordinator)

Inside the presentation

- Business and technical processes supported by the SDMX Toolkit
- The Toolkit's evolution: simplifying while expanding capabilities, scalability, and interoperability
- Integrating AI to reduce access barriers for Data Users
- Chatbots in the Data Browser Web Application
- Conclusions

Business and Technical Processes supported by the SDMX Toolkit



Istat

The Toolkit's evolution: simplifying while expanding capabilities, scalability, and interoperability

Easy SDMX

- Excel files for creating DSDs, Data Cubes, Data Flows, Mappings, and loading data

Extract/Converting

- Excel
- .PX
- CSPro
- Excel2MSD
- Excel2Json
- *Parquet*

- SDMX Pull Requestor

Transform, Validate, Transcoding

- Editor to create formulas to be applied as transformation rules during data ingestion

Processing

- Schedulable actions for data aggregation and derivation

Mapping

- Structural and semantic mapping between heterogeneous data structures, along with the integration of legacy databases

Istat

Scalability and high-performance as primary conditions

- Decoupling M2M access from Web GUI access
- Scheduling large-scale data loading or data transfers across different environments
- Caching of data and metadata to improve response time in the Data Browser

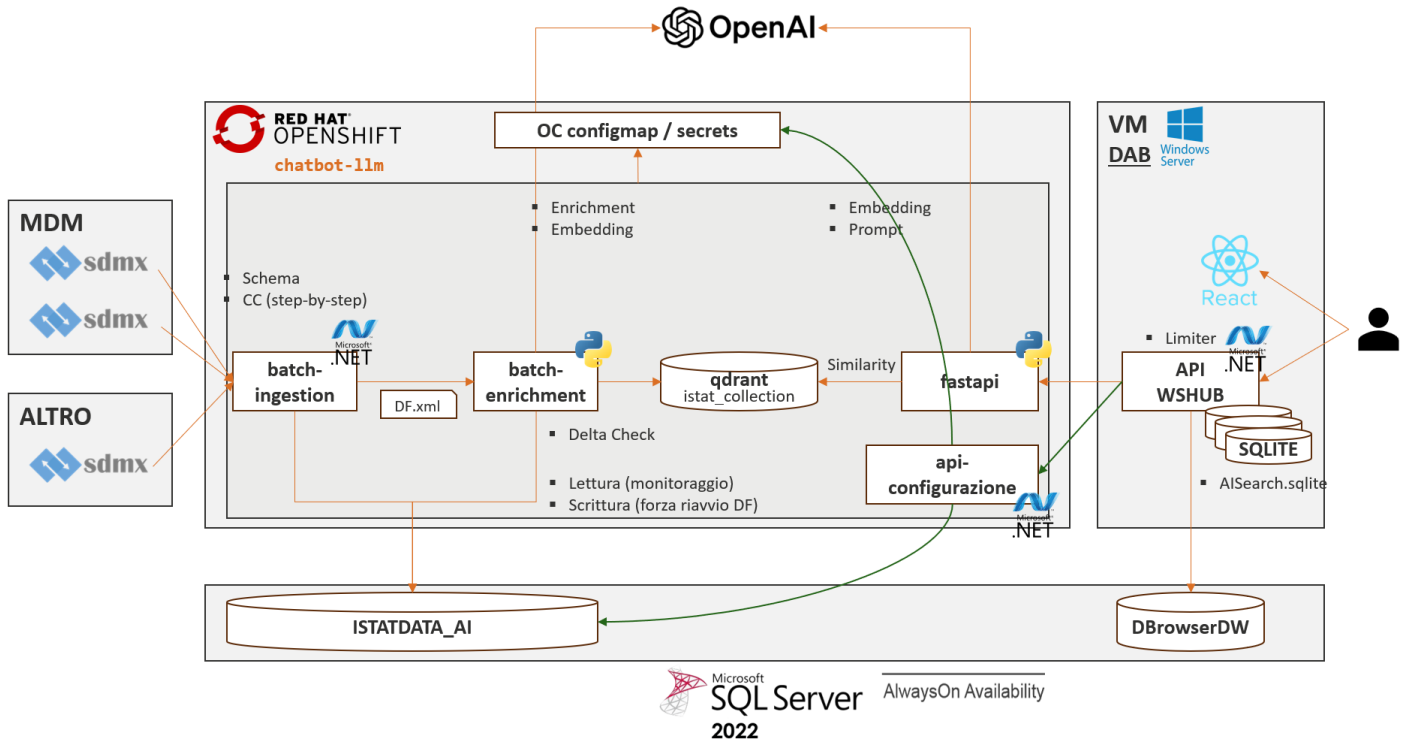
Tests in Istat's dissemination environment used

- Databases with over 10 billion data points
- Data cubes with more than 100 million data points

Integrating AI to reduce access barriers for Data Users

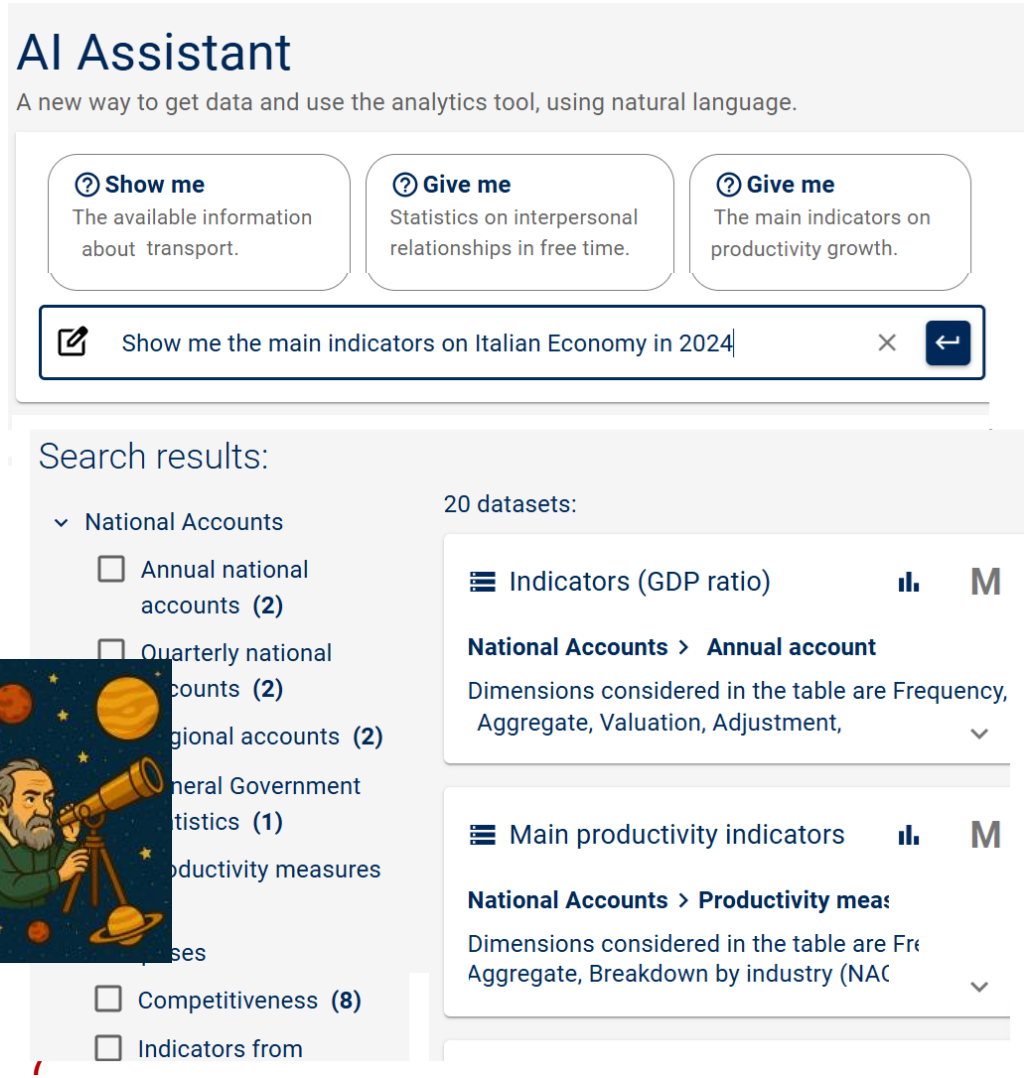
AI backbone component

- Connects directly to SDMX APIs, retrieving both structural and content-related metadata essential for constructing relevant and context-aware "*data products*"
- Accepts requests from a *Chatbot*
- Interfaces with *OpenAI* services to interpret and generate natural language responses

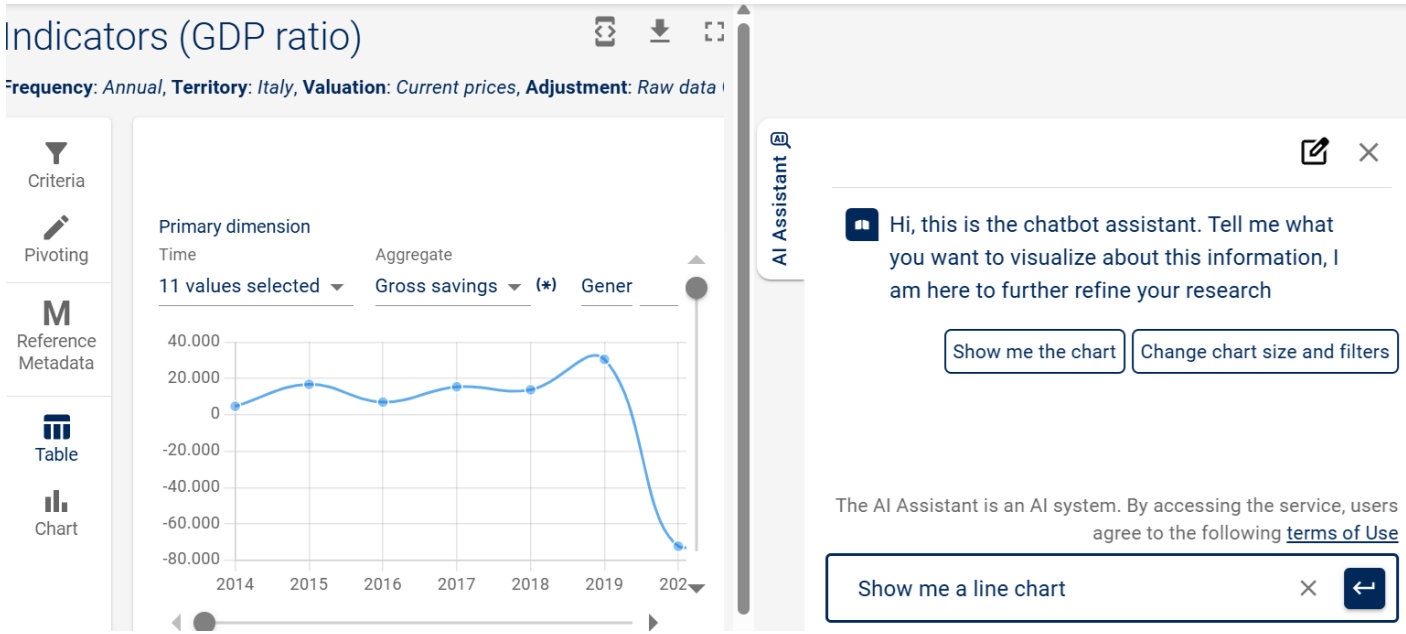


Chatbots in the Data Browser Web Application

Search AI assistant



Visualization AI assistant



Conclusions

- The toolkit hides much of the complexity of the SDMX standard, providing an intuitive environment where users can work through visual interfaces, guided configurations, and pre-built components without needing to master the SDMX technical specifications. This dramatically lowers the entry barrier, especially for institutions with limited SDMX expertise or IT support
- The new strategy embraces emerging capabilities in data engineering, data architecture, and data analysis. By aligning with these disciplines, the toolkit supports now more robust data pipelines, clearer metadata governance, and the ability to design interoperable, scalable infrastructures
- The integration of AI technologies enables data users to search and customize dataset visualizations while interacting with a “human-like” virtual expert assistant

Thanks

- SDMX Istat toolkit download: https://sdmxistattoolkit.github.io/index.html
- Corporate dissemination data warehouse (IstatData): https://esploradati.istat.it/databrowser/#/en
- Permanent census of population and housing: https://esploradati.istat.it/databrowser/#/en/censpop
- Agriculture Census 2020: https://esploradati.istat.it/databrowser/#/en/censimentoagricoltura
- External trade dissemination system (CoeWeb): https://esploradati.istat.it/coeweb/databrowser/#/en

Francesco Rizzo | rizzo@istat.it
Carlo Boselli | carlo.boselli@istat.it
Alessio Cardacino | alessio.cardacino@istat.it

Istat | Istituto Nazionale di Statistica