

3 - 4 LUGLIO 2024

Automazione e Standardizzazione
nell'acquisizione dei dati
in Banca d'Italia:

il provider SDMX dell'ISTAT

Attilio Mattiocco

Consigliere tecnico | Banca d'Italia



Conferenza Nazionale di **Statistica**

**La statistica ufficiale
nel tempo
dell'Intelligenza
Artificiale**

#CNStatistica15

- **SDMX** (Statistical Data and Metadata eXchange) è un’iniziativa internazionale finalizzata a **standardizzare e modernizzare** (“industrializing”) i processi e i meccanismi relativi allo scambio di dati statistici tra le organizzazioni internazionali e gli stati membri.
- Gli Sponsor dello standard SDMX sono: **ECB, BIS, Eurostat, IMF, OECD, UNSD, World Bank, ILO**
- Banca d’Italia e ISTAT hanno collaborato alla creazione dello standard e continuano a collaborare alla sua evoluzione, facendo parte degli organi di governo (**Technical e Statistical Working Group**) e producendo Software Libero (Open Source) che viene messo a disposizione della comunità degli utenti.
- Nel 2014, nell’ambito del **Comitato di Coordinamento tra Banca d’Italia e ISTAT**, il gruppo di lavoro per la “**Razionalizzazione dello scambio dati tra Banca d’Italia e ISTAT**” propone i web services SDMX come strumento d’elezione.
- Nel 2015 compare per la prima volta, nei **Connettori SDMX**, il provider ISTAT (**Single Exit Point**) coerente con **SDMX v2.1**
- Nel giugno 2023 il passaggio alla piattaforma nativa SDMX **IstatData** allinea completamente dati e metadati esposti con il DWH sottostante e potenzia la diffusione dei metadati (*footnotes*).

SDMX Providers

European Central Bank
OECD
EUROSTAT
IMF
International Labour Organization
Bank for International Settlements

UNICEF
UN Data
World Bank

ISTAT
Bundesbank
INSEE
Australian Statistics



SDMX Connectors



Caso d'uso principale: lettura di serie storiche da un provider esterno, da utilizzare **direttamente** in un tool di calcolo (i.e. R, Matlab, SAS, Python, Excel, STATA)

In passato: senza i connettori SDMX



Tutta la filiera può essere automatizzata con

- Meno rischi di **errore**
- Elaborazioni più **rapide**
- **Processo** più chiaro

Oggi: con i connettori SDMX



- Ogni anno circa **400K query SDMX** nei programmi degli utenti, in aumento
- **Efficienza, robustezza di processo, semplicità di utilizzo**
- **Qualità del dato**, certezza delle informazioni (metadati, riproducibilità)
- L'universo SDMX si pone come un **database aperto e flessibile** di dati statistici **certificati**, dotato di una **lingua franca** che lo rende utilizzabile come (e in alcuni casi più di) un database gestito in-house
- **Necessità di una governance chiara, coerente e continua** e di una **strategia di diffusione** verso gli utenti (e.g. la recente migrazione a IstatData)
 - Gli utenti dei dati devono essere considerati come **parte integrante del processo di diffusione**
 - Tutti i **cambiamenti** devono essere comunicati per tempo
 - Il supporto al «**change**» implica delle azioni specifiche (e.g. tool di conversione automatico delle query)
 - Si instaura una **relazione «simbiotica»** tra utente e produttore dei dati

SDMX + AI

- La forte strutturazione e la ricchezza di metadati contribuiscono allo sforzo di **trasparenza e affidabilità** dei risultati
- **L'universo dei dati strutturati SDMX crea il substrato ideale per l'AI**
- Possiamo immaginare un «**Data Copilot**» in grado di guidare l'utente nella ricerca e acquisizione di dati, anche in modalità «conversazionale» (IMF - StatGPT)
- Un elemento distintivo rispetto all'utilizzo generico di fonti internet sta nella possibilità di **referenziare le fonti con precisione** e quindi di «certificare» il dato

VTL

- Il **Validation and Transformation Language** è un linguaggio sviluppato nell'ambito della comunità SDMX per la standardizzazione della manipolazione dei dati (validazioni e trasformazioni)
- Grazie al VTL si crea l'opportunità di un ulteriore passo in avanti: creare nuovi dati, a partire da quelli esistenti, con la **certezza del processo con cui sono costruiti**
- Anche in questo ambito ISTAT e Banca d'Italia sono stati tra i primi a sperimentare: dei quattro engine VTL oggi esistenti, due sono stati **sviluppati dai nostri istituti**.

3 - 4 LUGLIO 2024

grazie
per l'attenzione

Attilio Mattiocco

Consigliere tecnico | Banca d'Italia

<https://github.com/amattioc/SDMX>



Conferenza Nazionale di **Statistica**

**La statistica ufficiale
nel tempo
dell'Intelligenza
Artificiale**

#CNStatistica15