



sdmx.io – where tools and learning resources come together to address official statistics use cases

Douglas Araujo and Brian Buffett

Bank for International Settlements

Rome, 18-20 February 2025

Context

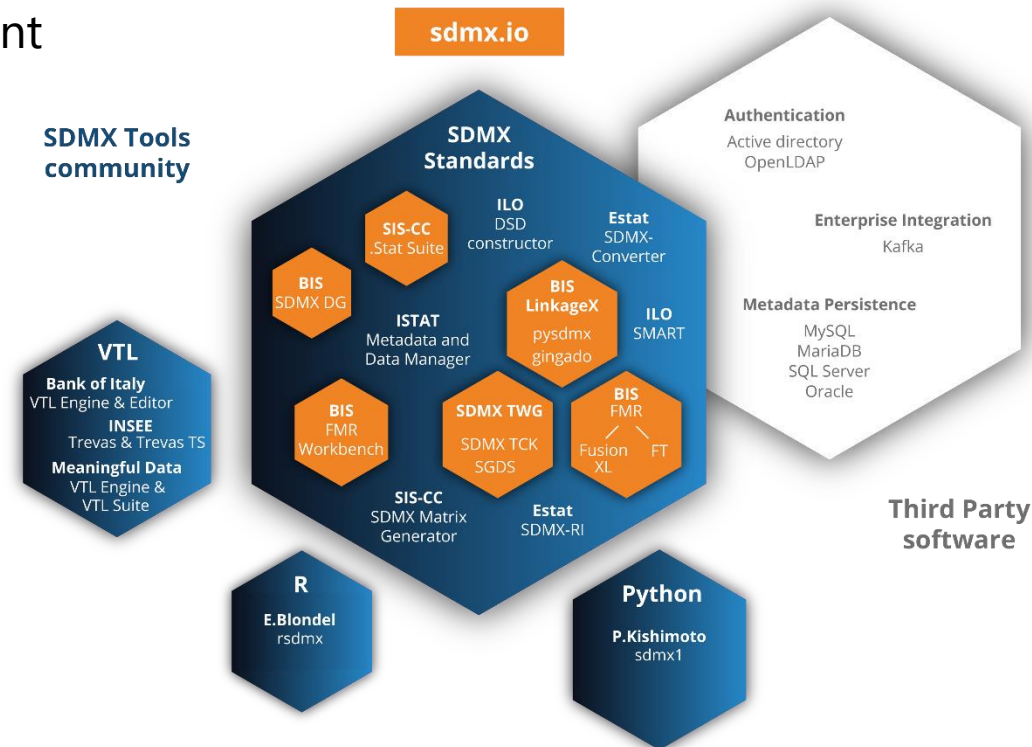
- **Increasing demands for data:** pressure for more agile data organisations, processes, and software
 - AI and advanced analytics only add to the demand
- **SDMX** (**S**tatistical **D**ata and **M**etadata **eX**change) standard:
 - information model and technical architecture that support robust metadata-driven processes...
 - ... like **data analysis** for one-off use (eg, research) or for recurrent use (eg, an internal dashboard)
- **SDMX** is reliable and trustable:
 - an ISO standard with a **well-defined governance**
 - Sponsored by BIS, ECB, EUROSTAT, ILO, IMF, OECD, UN, WB
 - Used routinely by various authorities to produce and disseminate **official statistics**
- But for most data scientists in economics, **SDMX is still unknown and has high barriers to entry**

SDMX – Only for Data Producers?

- **Rich Metadata Support:** Metadata which describes datasets in detail, including variables, units, methodologies, and data sources. This helps data scientists:
 - Quickly assess data relevance.
 - Understand data structure without diving deep into raw files.
 - Ensure consistency across datasets for time series or comparative analyses.
- **Improved Data Discovery:** Metadata enables efficient searching and filtering of datasets based on specific criteria like time periods, geographic regions, or indicators.
- **Standardized Data Models:** Common formats for sharing statistical data, making it easier to integrate datasets from different sources (e.g., BIS, ECB, IMF, central banks).
- **Automation & API Integration:** APIs allow automated data retrieval and updates, which is crucial for dashboards or monitoring systems.
- **Interoperability:** Facilitates seamless data integration across various tools and platforms commonly used in data science (Python, R, BI tools).

Enter **SDMX*io***

- Goal is to facilitate the use of SDMX-based solutions
 - throughout the data lifecycle (eg, production -> dissemination -> use)
 - and for organisations of all sizes and stage of development
- To achieve this goal, **sdmx.io** is an ecosystem of resources:
 - open-source tools,
 - pre-configured containerised environments,
 - learning material (tutorials, webinars),
 - and more



What about data science?


- SDMX is typically used by authorities responsible for producing and disseminating statistics
- But it has ample potential for use in data science as well (Araujo, 2023):
 - Standard definitions of data and metadata modelling can help users find relevant data...
 - ... from reliable sources ...
 - ... in a programmatic and reproducible way.
- **sdmx.io** will increasingly also cater to this target user group of end-users / data scientists
- Today, we are introducing a new series **sdmx4ds** (SDMX for Data Science)

sdmx4ds: a repository of data science work using SDMX

- Data scientists can share their work that uses SDMX (and potentially other data sources)
 - Citation is encouraged
- Goal is to publish Jupyter Notebooks in python, R, Julia showcasing:
 - an interesting analysis that uses data downloaded via SDMX
 - How to download data with SDMX? [gingado](#) `sdmx1`
- Submission process:
 - All public, via GitHub
 - Submissions should be short (circa 3-5 pages of PDF)
 - A template offers guidance on structure

Live demo of the submission process

Next steps

- Call for papers – coming out soon. Keep an eye on sdmx.io
- Publication
 - Lightly-edited Jupyter notebooks / Rmarkdown / quarto file in GitHub repo
- Easy for other users to fork and adapt to their own cases
 -  button allows for quick replication and customisation by readers



Thank You!