

IFC-Bank of Italy Workshop on "Data science in central banking: enhancing the access to and sharing of data"

17-19 October 2023

**Why SDMX matters? A community journey towards
SDMX as an AI-enabler and a data mesh enabler¹**

Eric Anvar,

Organisation for Economic Co-operation and Development
(OECD)

¹ This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Why SDMX matters? A community journey towards SDMX as an AI-enabler and a Data Mesh enabler

IFC Workshop “Data science in central banking: Enhancing the access to and sharing of data”

Rome, October 2023

Keywords

#OpenSource #OpenKnowledge #Community #SDMX #MetadataDriven #AI #DataMesh

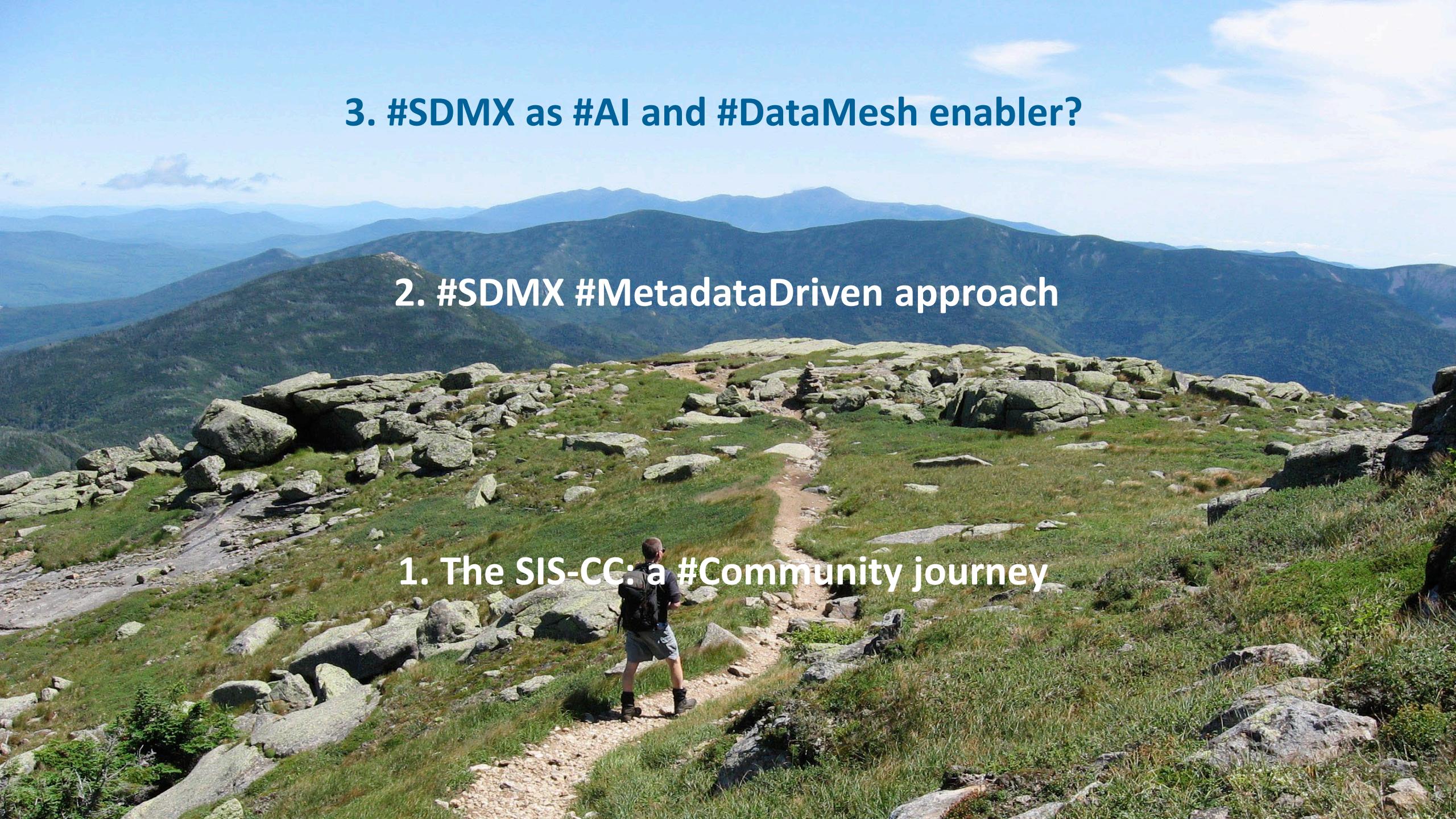




*Perhaps there is more to SDMX than just a tool
to report to international organisations?*



Source: [Why SDMX Matters](https://siscc.org), <https://siscc.org>

A scenic mountain landscape with a hiker on a trail. The foreground shows a rocky, grassy mountain slope with a dirt trail. A hiker is walking away from the camera on the trail. In the background, there are more mountains under a blue sky with some clouds.

3. #SDMX as #AI and #DataMesh enabler?

2. #SDMX #MetadataDriven approach

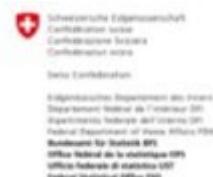
1. The SIS-CC: a #Community journey

THE SIS-CC: A #COMMUNITY JOURNEY

SIS-CC: 12 years, 20 Members and Partners



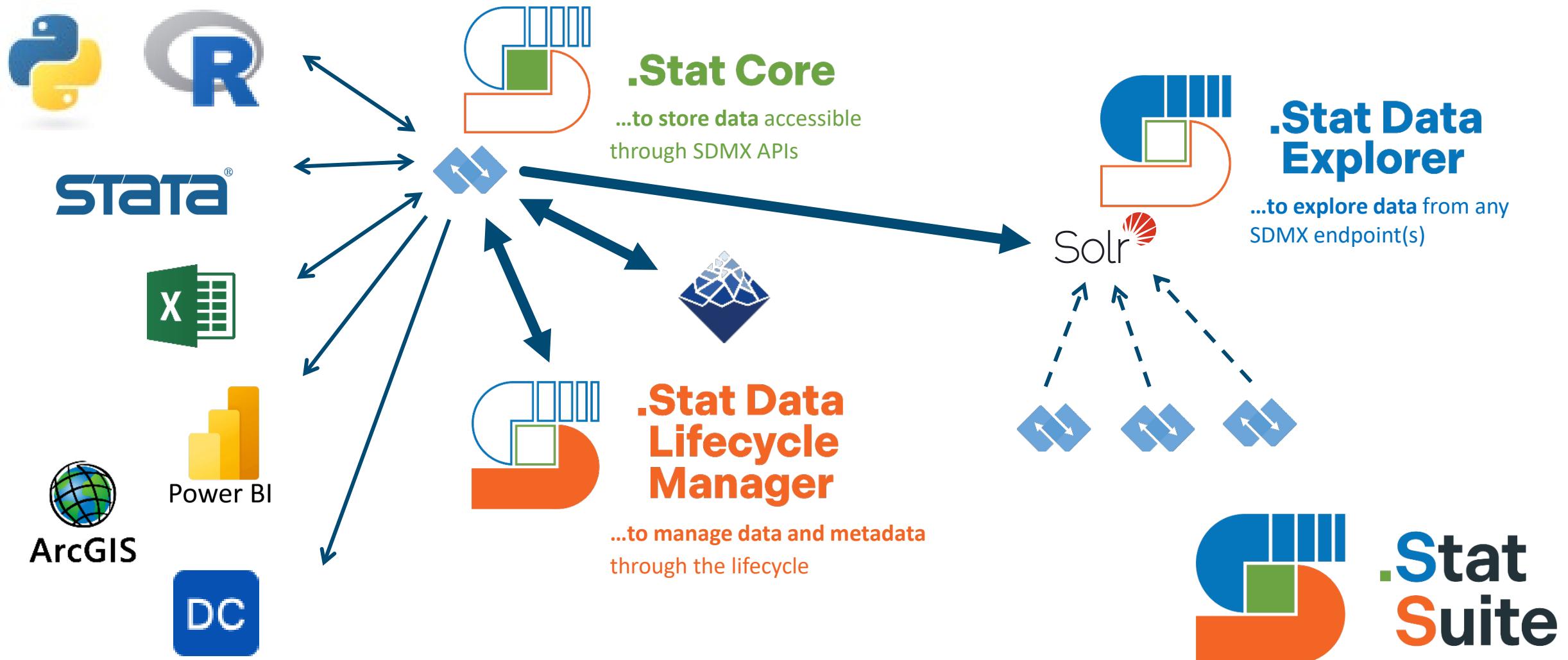
SIS-CC (“the Community”) is a reference open-source community for official statistics, focusing on product excellence and **delivering concrete solutions to common problems through co-investment and co-innovation.**



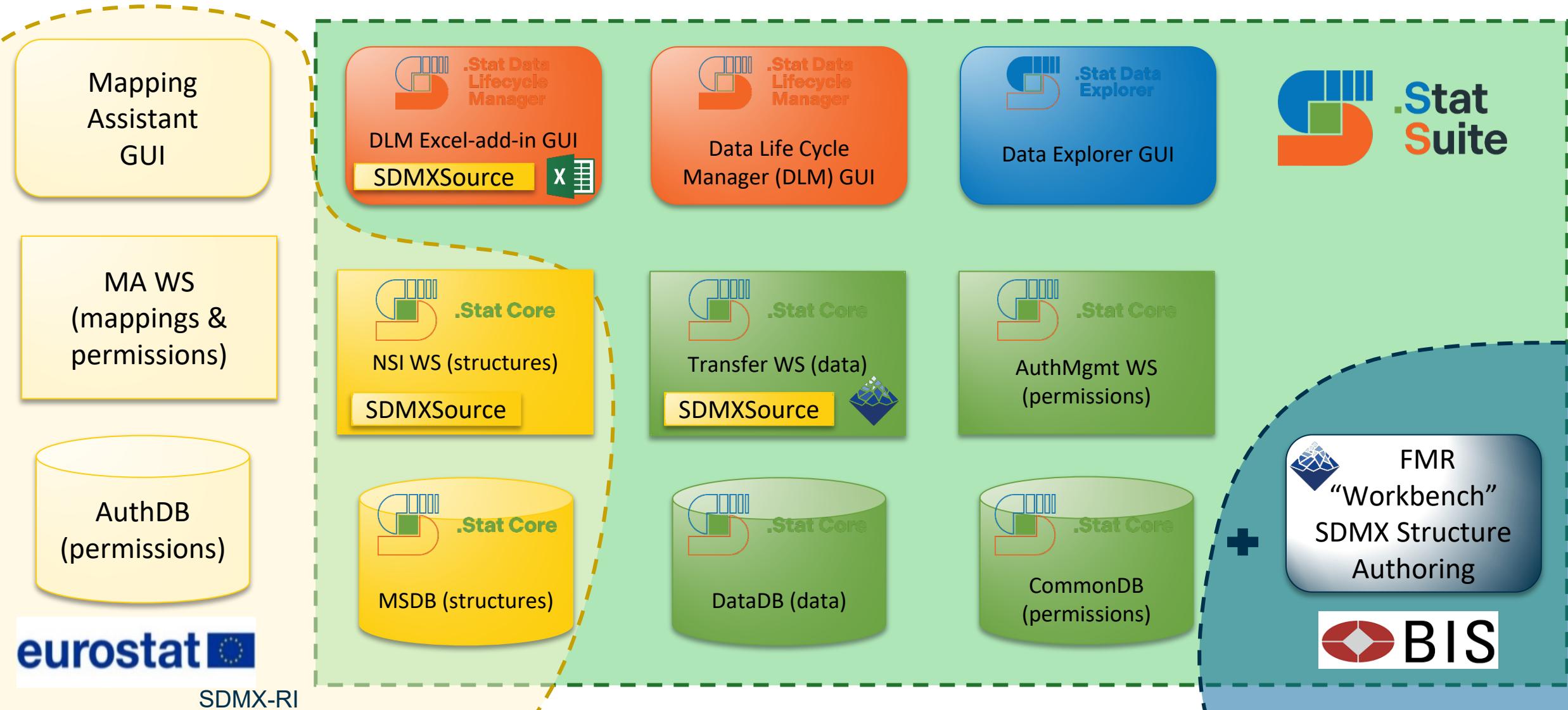
Members



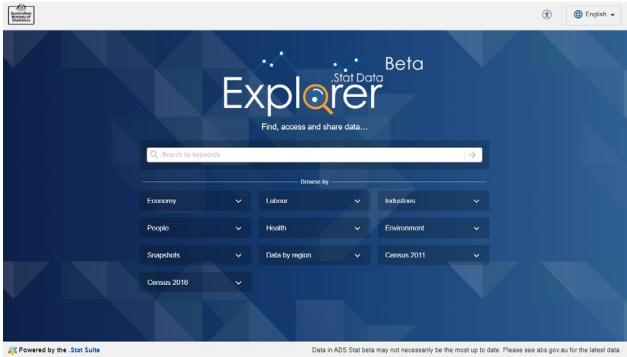
Partners



Component-based Architecture → Cooperative “Business Model”



Examples of implementations



ABS Stat Beta Explorer. Find, access and share data...
Powered by the Stat Suite

Search by keywords or browse by English

Economy, People, Snapshots, Census 2016

Labour, Health, Data by region, Census 2011

Industries, Environment

Find ABS Stat beta may not necessarily be the most up to date. Please see abs.gov.au for the latest data.



LU STAT STATISTIQUES.lu Explorer Trouver, accéder à et partager des données...
Powered by the Stat Suite

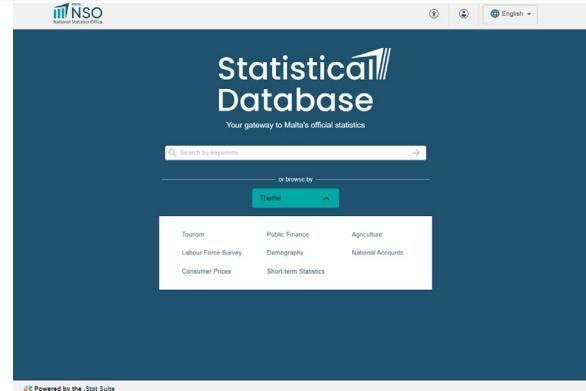
Search by keywords or browse by English

Thèmes

Alimenté par la Stat Suite Le code source de l'application est open-source, protégé par le droit d'auteur et fourni sans garantie. Pour plus de détails, consultez la déclaration de licence.



SDG GATEWAY SDG GATEWAY DATA Explorer SDG Indicators by Goal All Indicators by Theme
Goal 1 - No poverty Goal 2 - Zero hunger Goal 3 - Good health and well-being Goal 4 - Quality education Goal 5 - Gender equality Goal 6 - Clean water and sanitation Goal 7 - Affordable and clean energy
Powered by the Stat Suite

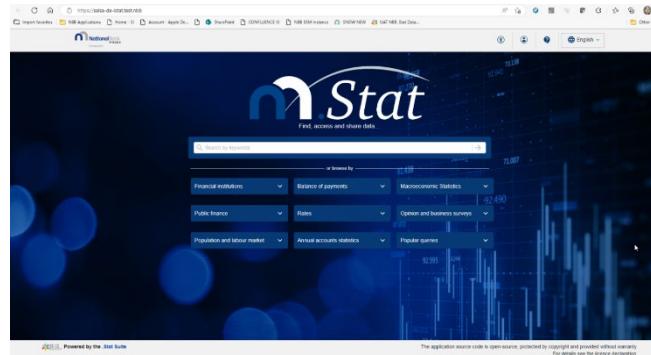


NSO Statistical Database Your gateway to Malta's official statistics
Powered by the Stat Suite

Search by keywords or browse by English

Theme

Tourism, Labour Force Survey, Demography, National Accounts, Consumer Prices, Short-term Statistics



Stat Find, access and share data...
Powered by the Stat Suite

Search by keywords or browse by English

Financial institutions, Public finance, Population and labour market, Balance of payments, Rates, Annual accounts statistics, Macroeconomic Statistics, Opinion and business surveys, Popular queries

The application source code is open-source, protected by copyright and provided without warranty. For details see the licence declaration.



PACIFIC DATA HUB Explorer Search, visualise and share Pacific data...
Powered by the Stat Suite

Search by keywords or browse by English

Topic, Pacific Island Countries and territories, Development indicators

Pacific Data Hub

PDH stat Data Explorer JS 12.0.0 | Core 7.2.0



Statistics Sharing Hub Explorer
Powered by the Stat Suite

Search by keywords or browse by English

Topic, Country, Region, Sector, Data Type, Data Source, Data Type, Data Source

This is a public beta release. Alternatively, please visit <https://stats.oecd.org>.



OECD Explorer Find, access and share data...
Powered by the Stat Suite

Search by keywords or browse by English

Topic

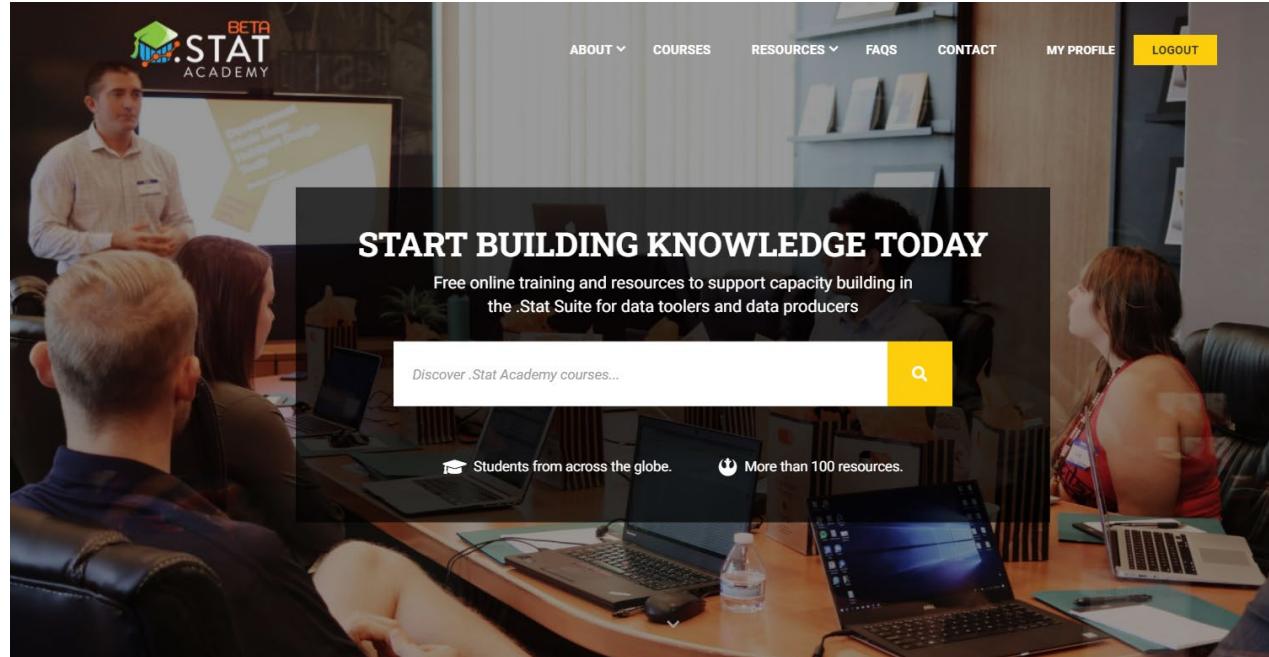
This is a public beta release. Alternatively, please visit <https://stats.oecd.org>.



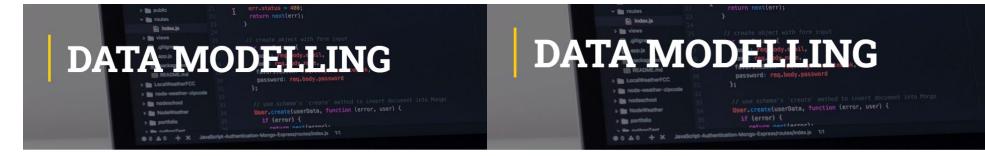
Search, visualise and share Pacific data...

Pacific Data Hub .STAT

The #OpenKnowledge .Stat Academy



<https://academy.siscc.org>



An introduction to SDMX structural mod
producers

Instructor David Barraclough Categories Data modelling, SDMX Review ★★★★☆



An introduction to SDMX for data produc

Instructor David Barraclough Categories Data modelling, SDMX Review ★★★★★



SDMX vocabulary for data managers

Instructor Jens Dossé Categories Data modelling, SDMX Review ★★★★☆

SDMX vocabulary for beginners

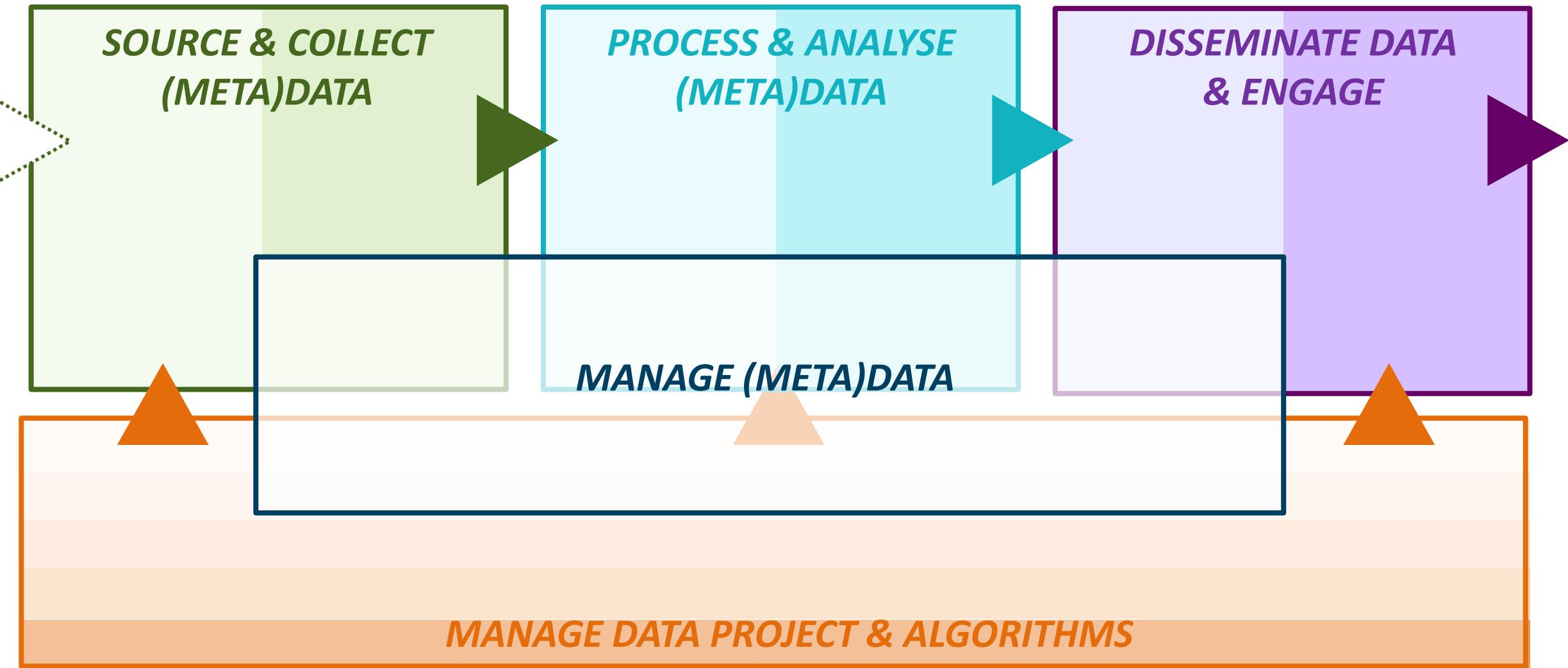
Instructor Jens Dossé Categories Data modelling, SDMX Review ★★★★★

Starting point: Build capacity in SDMX data modelling

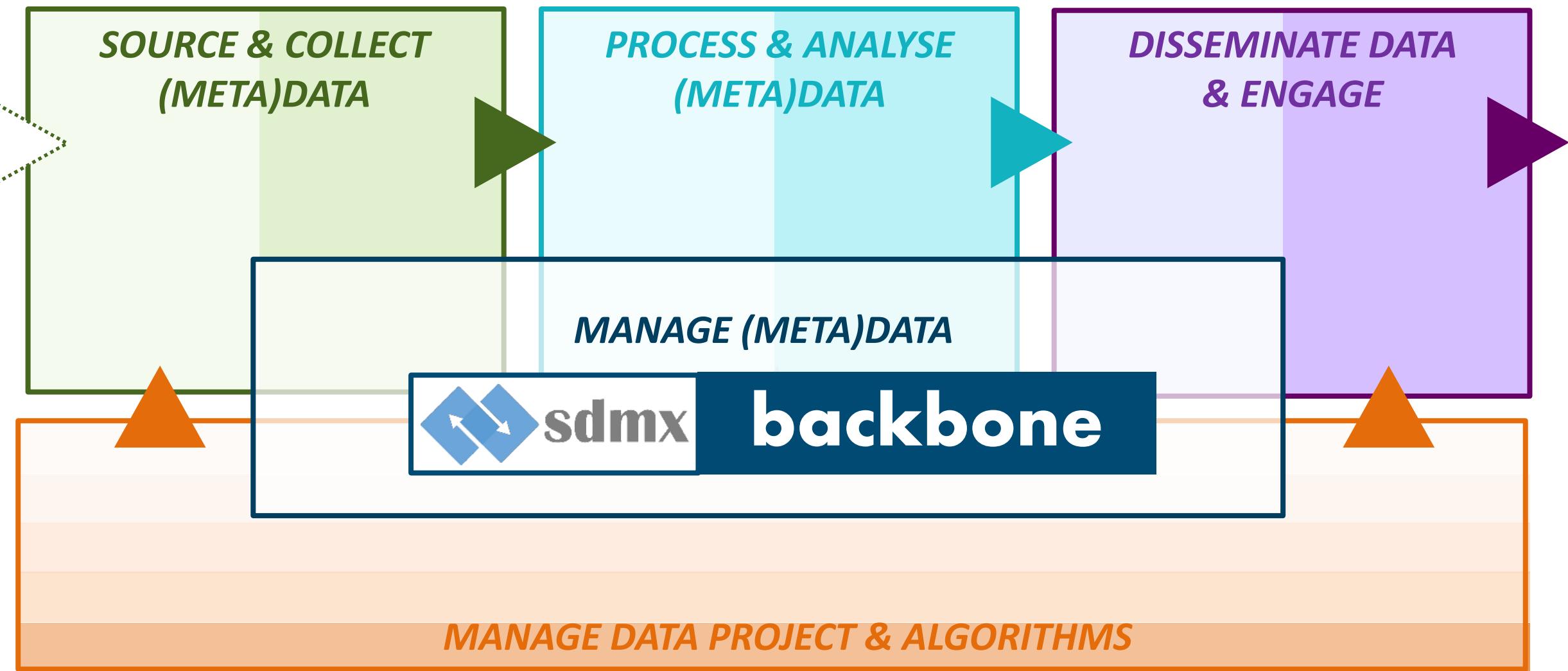


THE #SDMX METADATA-DRIVEN APPROACH

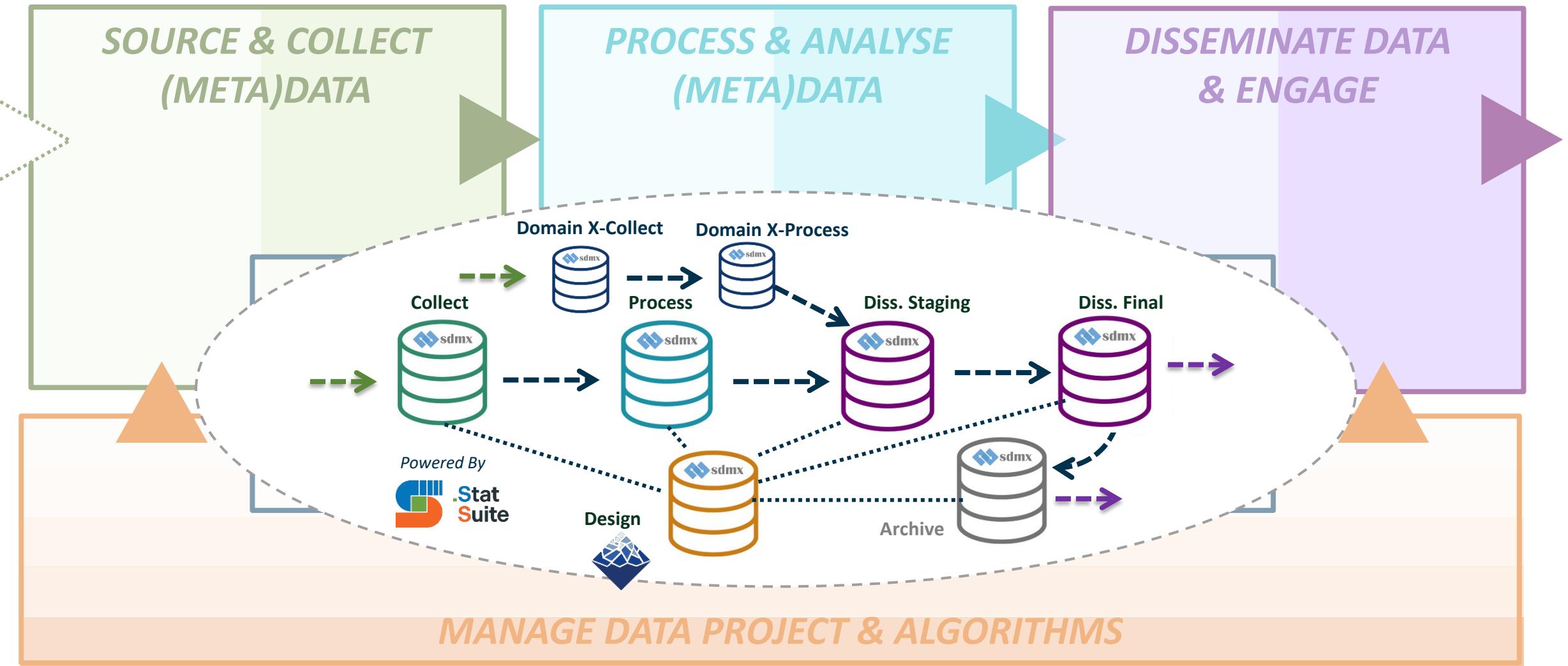




The #SDMX Backbone

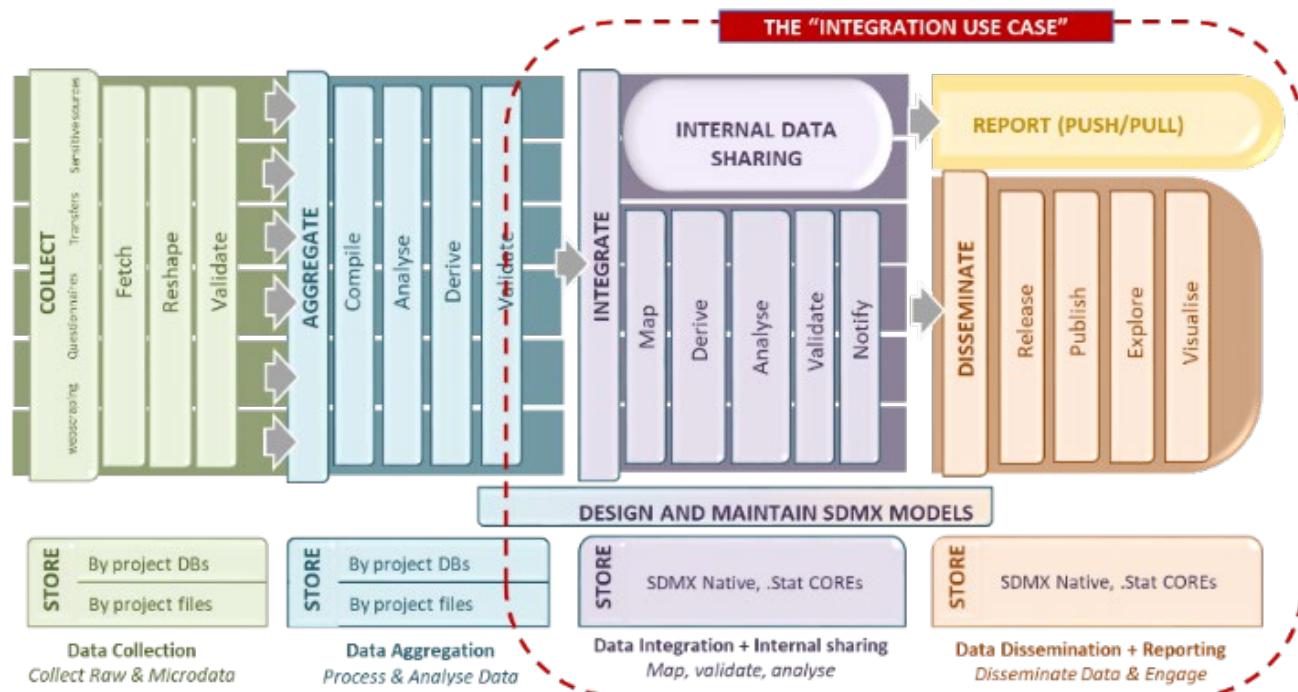


The #SDMX Backbone

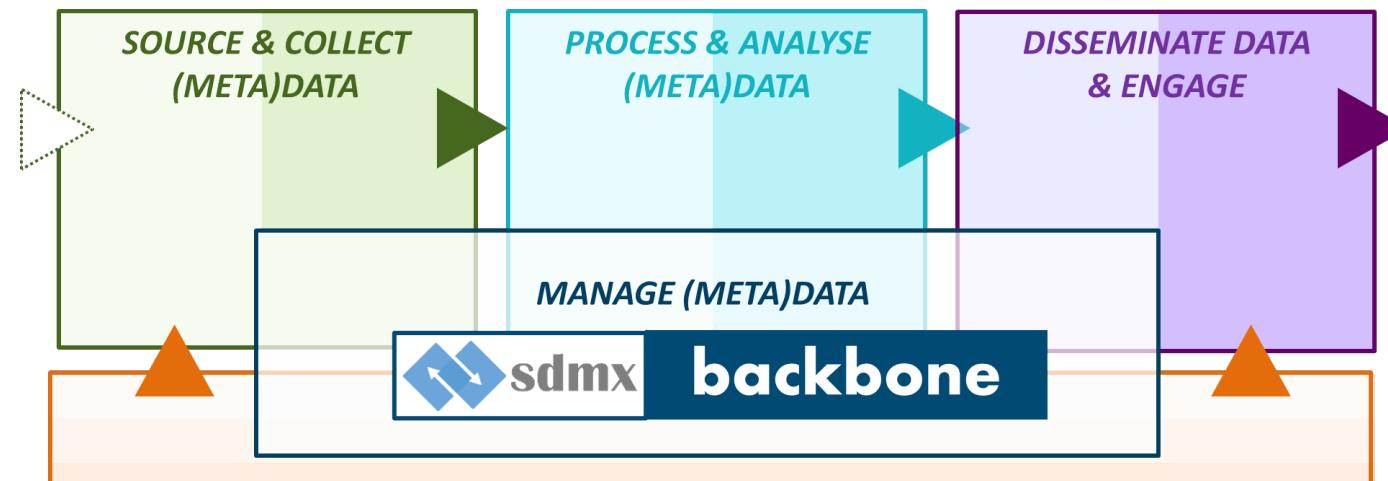


The #SDMX Backbone

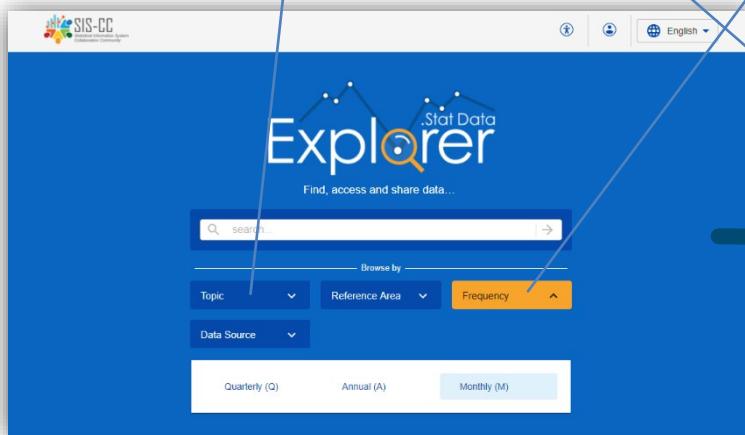
The “Data Integration” use case at the National Bank of Belgium



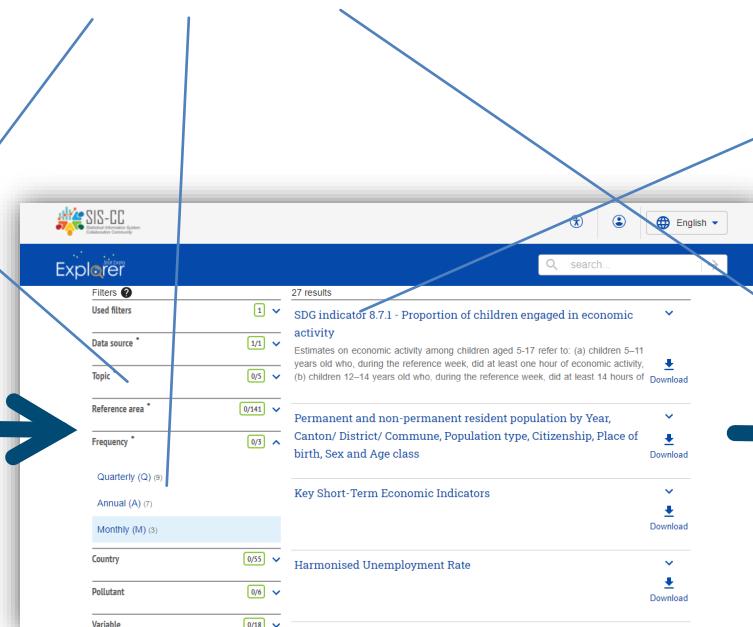
- sdmx data storage becomes a commodity: on premise and over the cloud, and across the entire data lifecycle and business domains.
- sdmx APIs enable universal data interoperability: microservice orientation, component-based architecture, cooperative business model
- sdmx semantics enables “metadata-driven” data user / manager experience: structural & referential metadata, presentation & process metadata



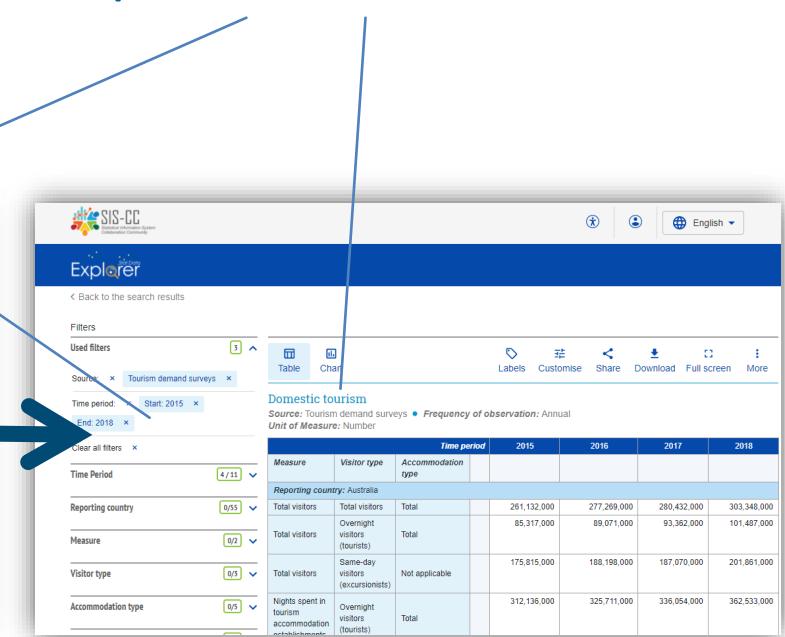
category scheme(s)



dimensions:
concepts + codelists



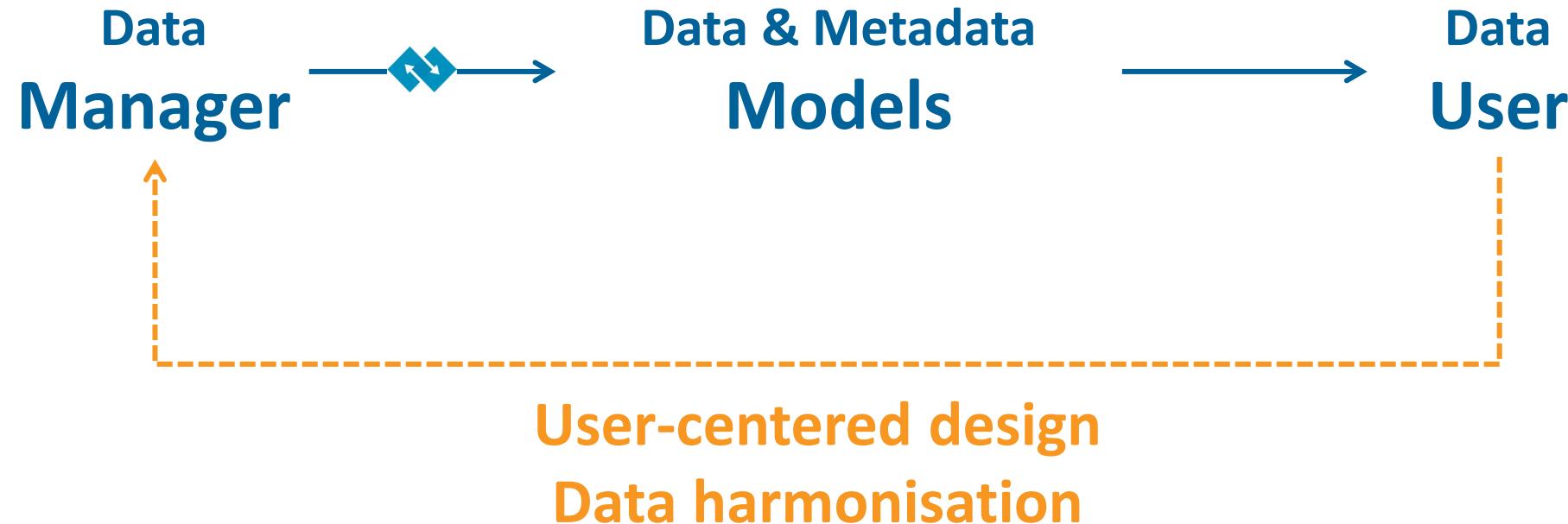
dataflow
+ presentation annotations



Step 1 Home page
Lexical search or Topics

Step 2 Search results
Filtering on dynamic facets

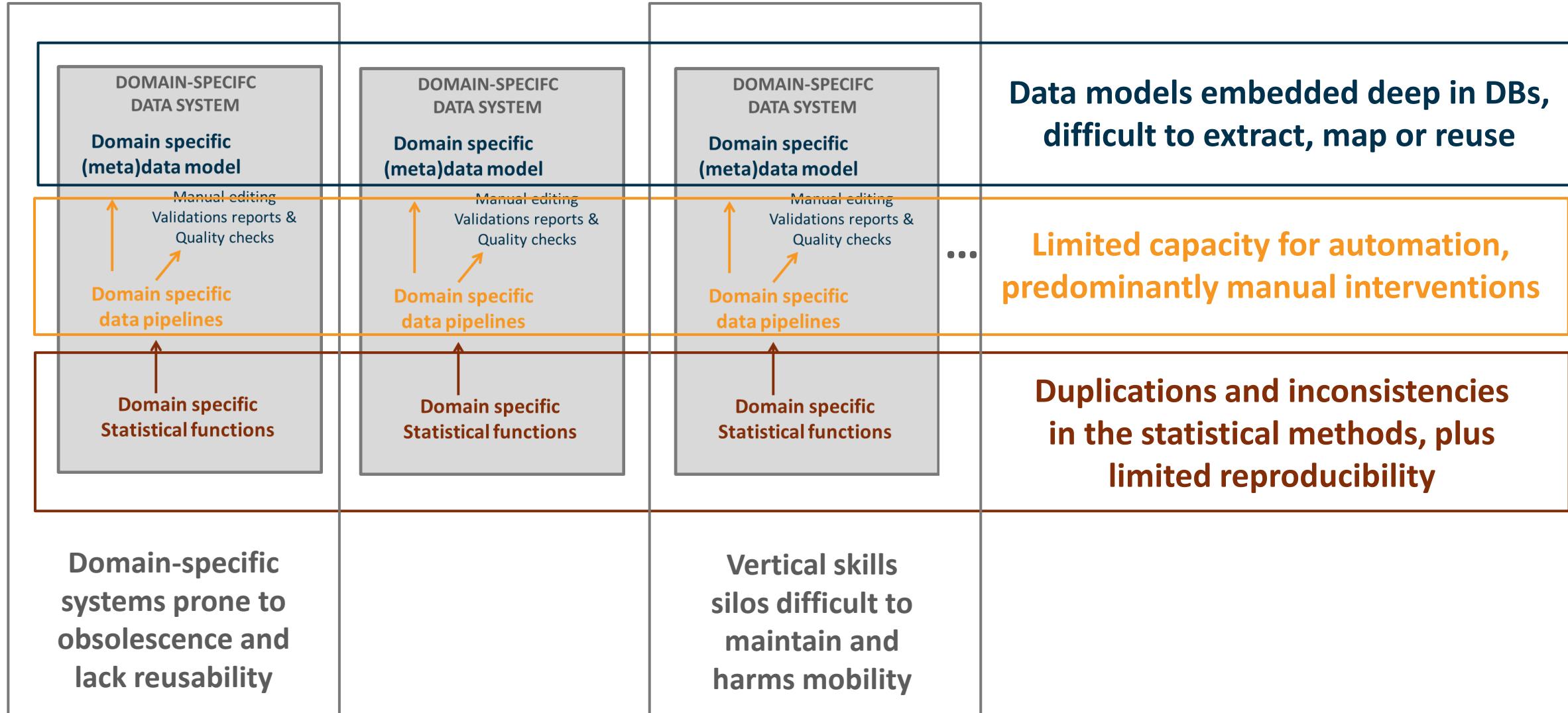
Step 3 Visualisation page
Selecting dimensions / variables



Establish a systematic User Research

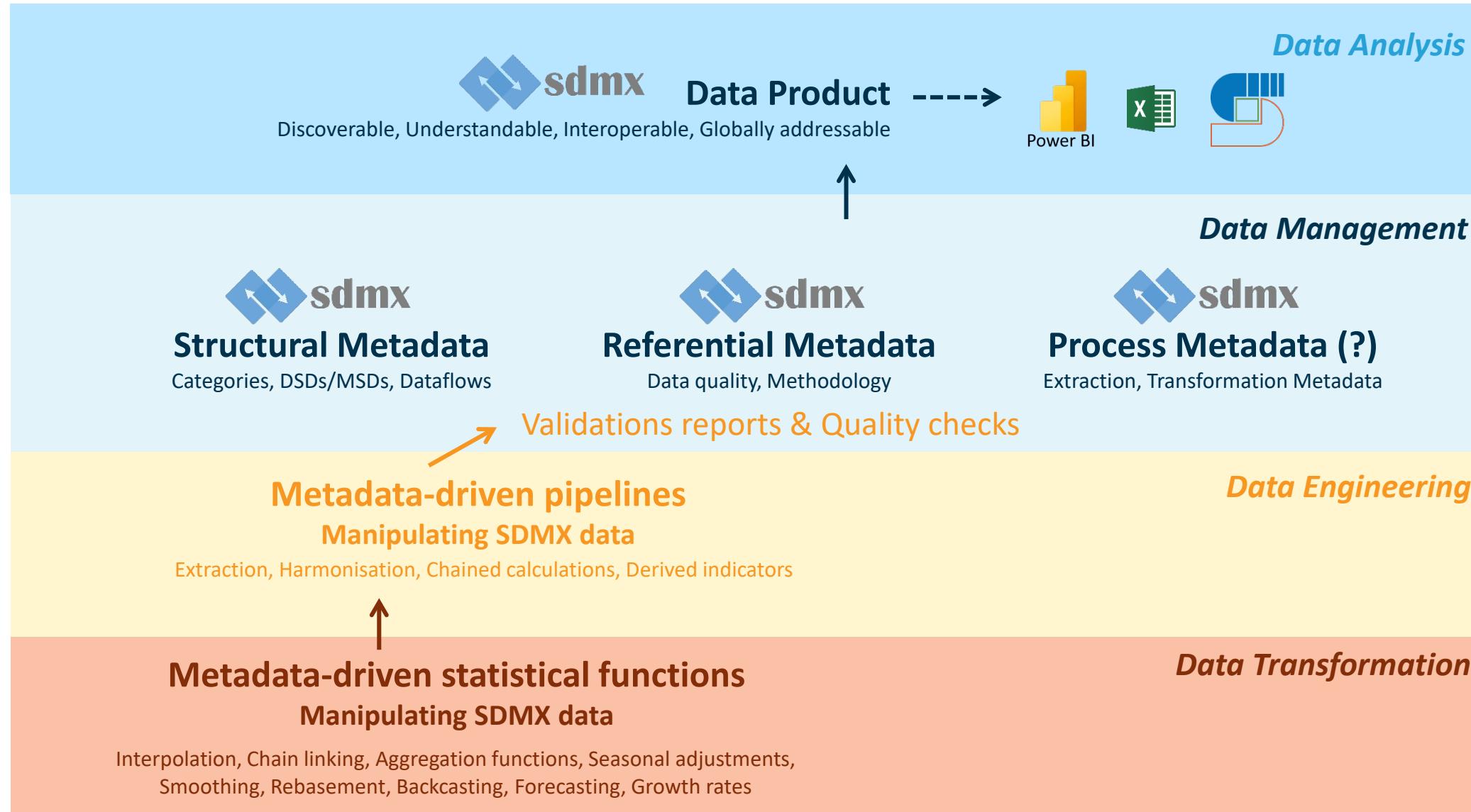


The Data Platform is fragmented

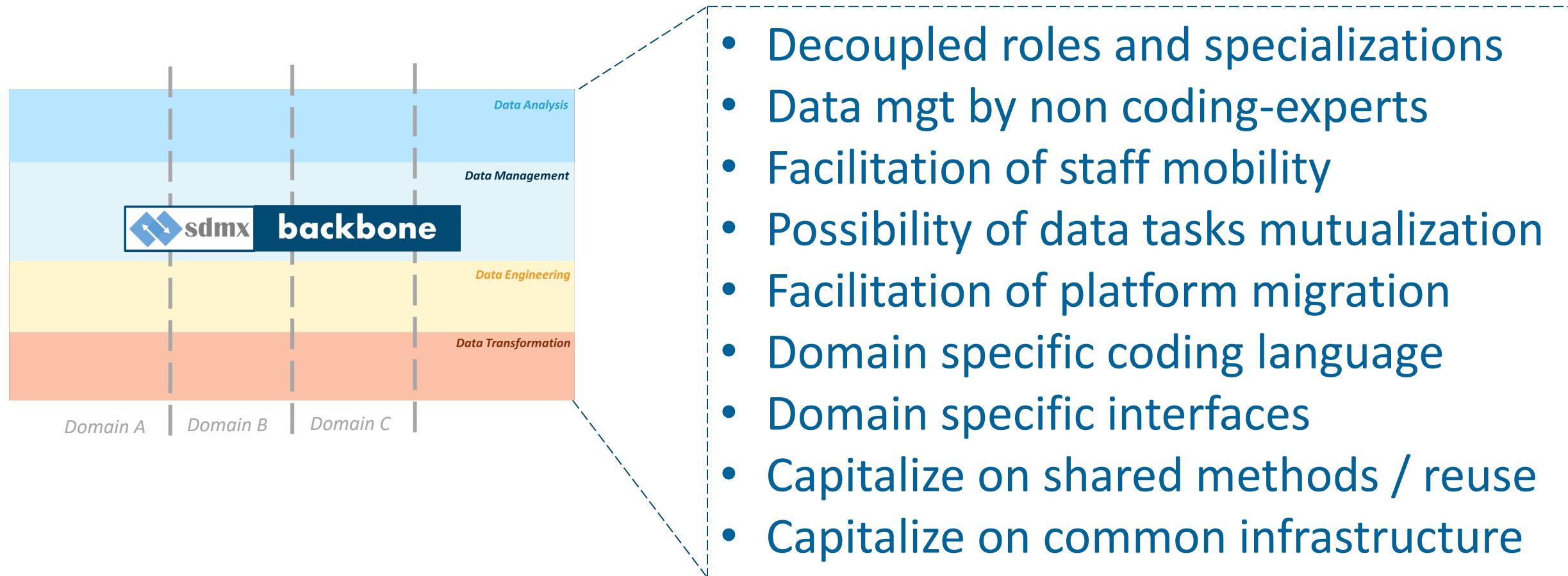


MANAGE DATA & METADATA

MANAGE DATA PROJECT & ALGORITHMS



The #SDMX #MetaDataDriven approach combines (centrally managed) data infrastructure with the possibility of (decentralized) data operations, creating efficiencies at both ends

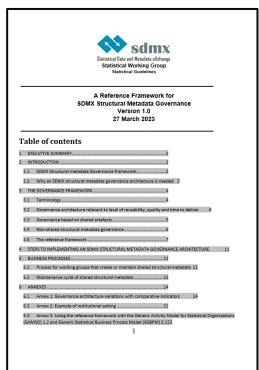
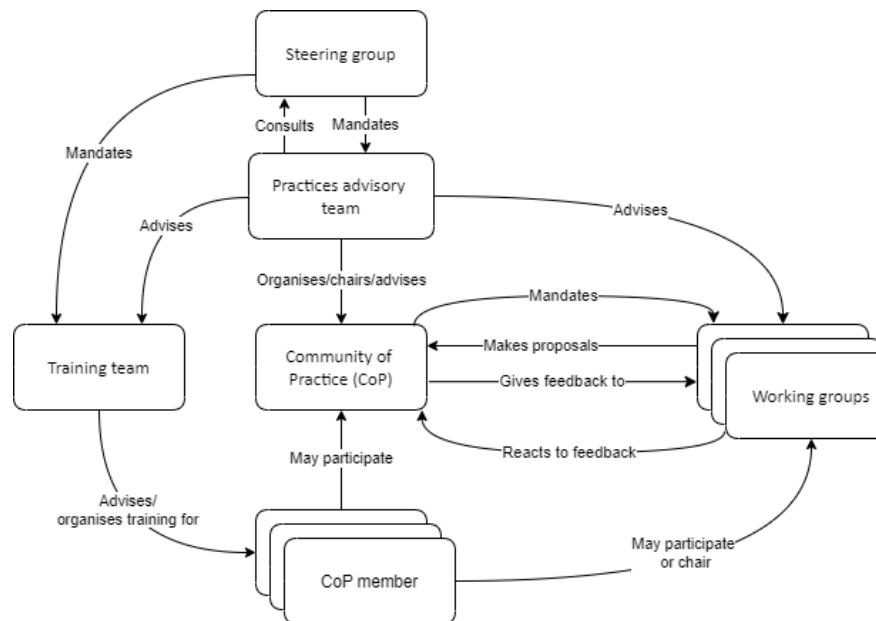


Structure your Data governance



Photo by [UX Indonesia](#) on [Unsplash](#)

SDMX provides the tools (conceptual and technical) to operate your governance and progressively harmonise data and metadata models





- Systematic user research – enabled by SDMX
- Structured data governance – enabled by SDMX
- SDMX data storage becomes a commodity
- SDMX APIs enable universal data interoperability
- SDMX semantics enables “metadata-drivenness”



SDMX metadata drives
Data User experience
Data Manager experience



Interested in developing an #SDMX Backbone & Governance?

Contact me – eric.anvar@oecd.org

#SDMX #AI SDMX AS AN AI ENABLER?

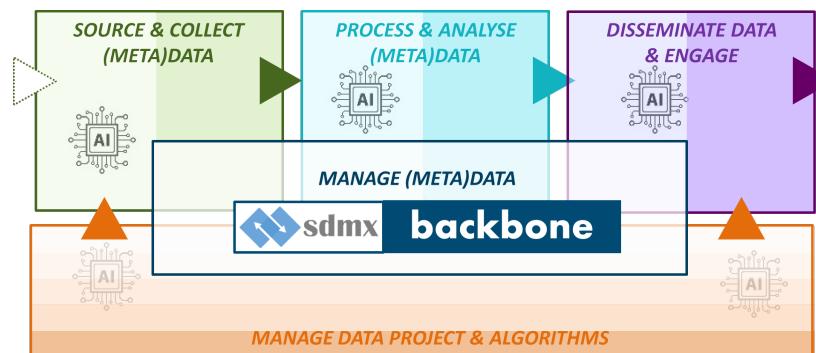
Opportunities #SDMX #AI



Nowcasting Federated Learning



Outlier detection
Clustering

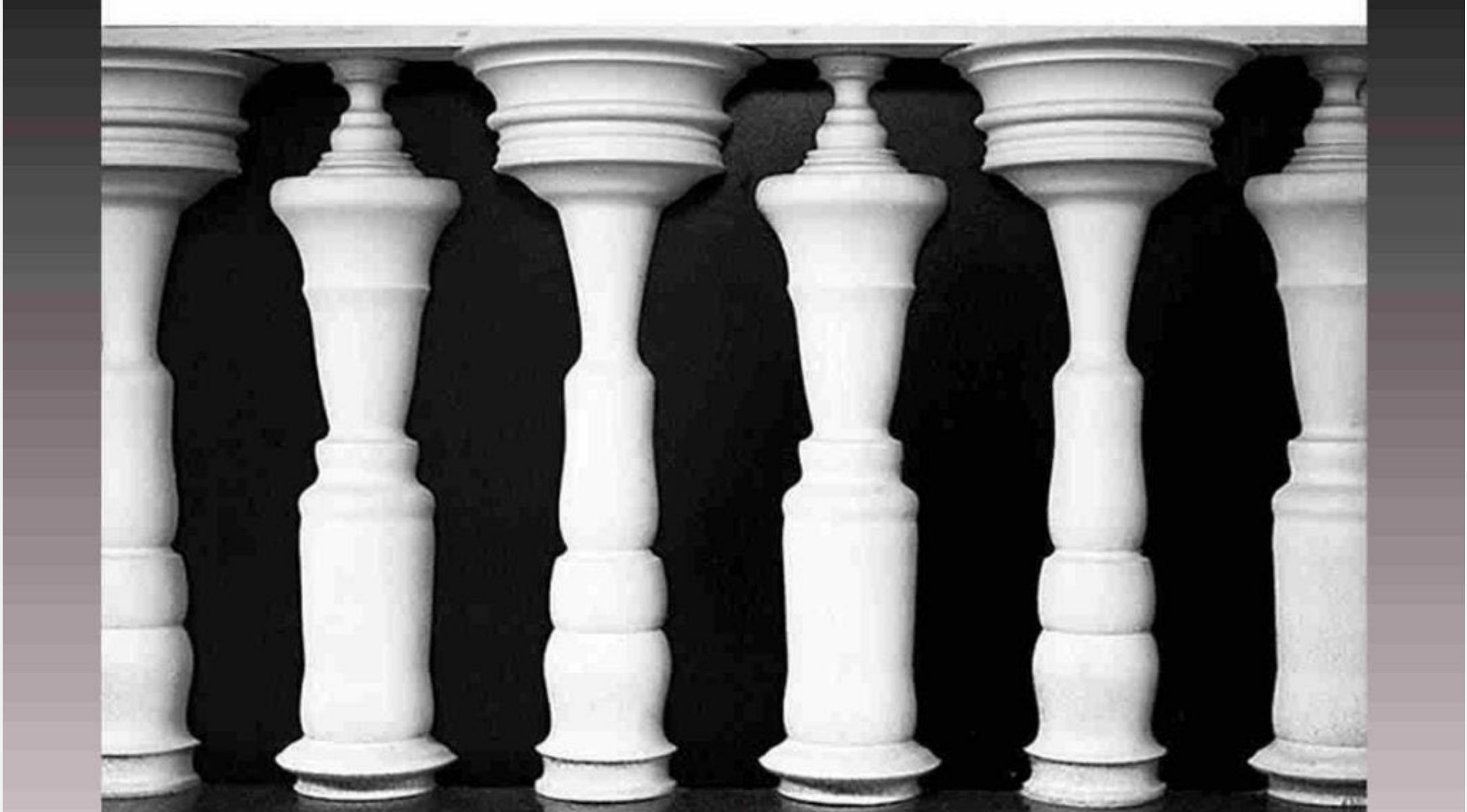


Code generation
Data Modelling



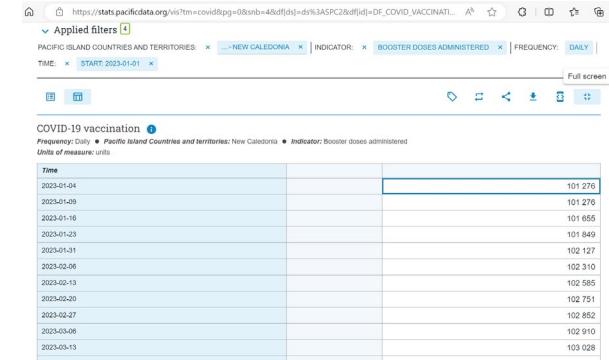
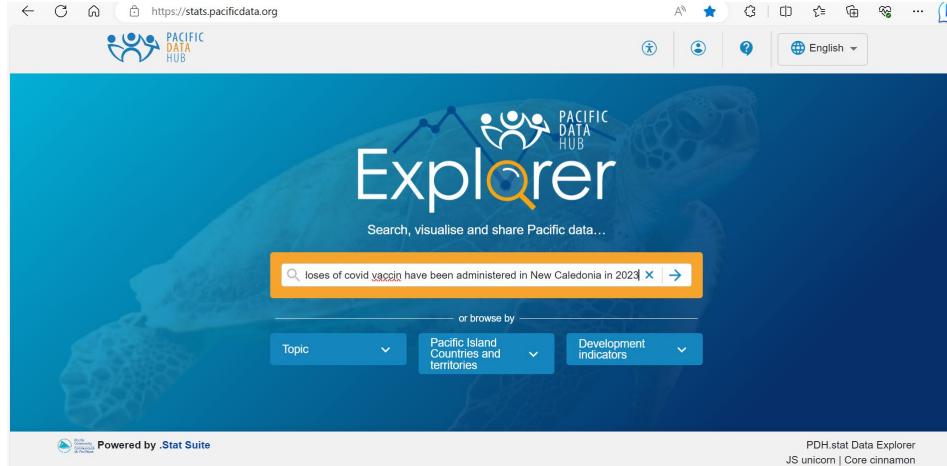
Natural Language
Conversational
Data Access

The cognitive gap



Augmenting the Data Explorer with #SDMX #AI (I)

How many booster doses of covid vaccines administered in New Caledonia in 2023?

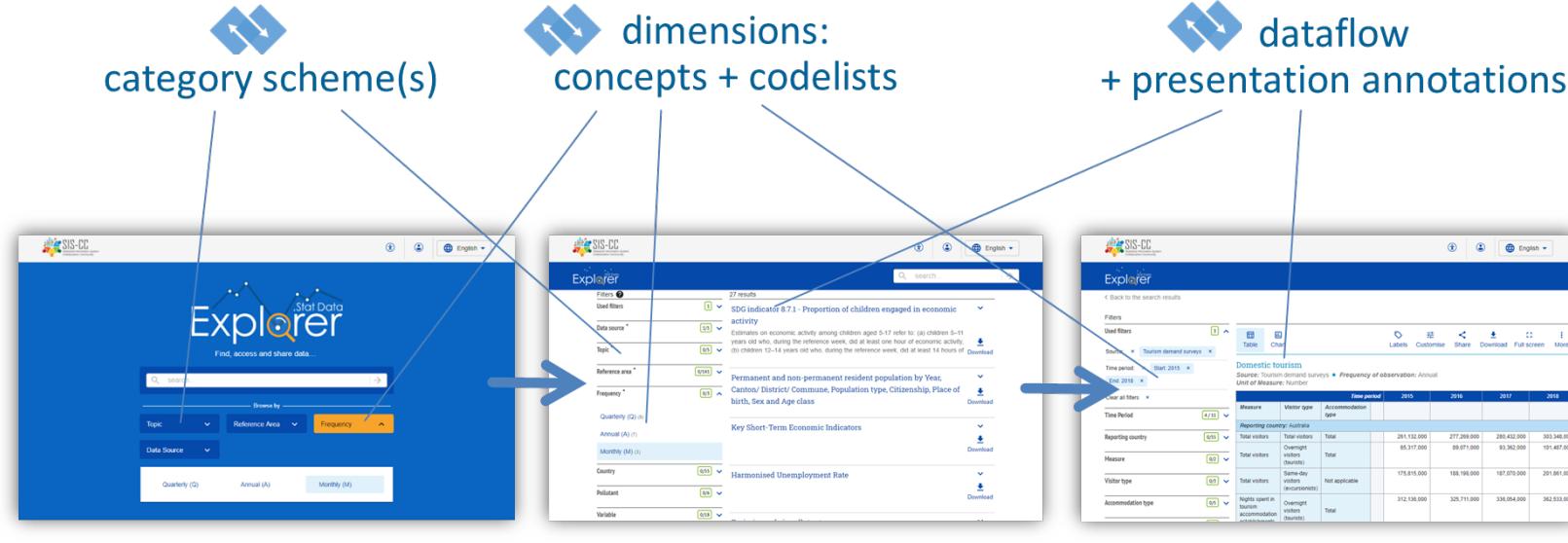


Time	Value
2023-01-04	101 276
2023-01-09	101 276
2023-01-16	101 655
2023-01-23	101 849
2023-01-31	102 127
2023-02-06	102 310
2023-02-13	102 585
2023-02-20	102 751
2023-02-27	102 852
2023-03-06	102 910
2023-03-13	103 028

Step 2.5 or 3.5

Natural Language query transformed into search queries
+ Pre-selection of dimensions on target dataflow(s)

Augmenting the Data Explorer with #SDMX #AI (I)



Natural Language query transformed into
engine search queries + initial filter selection

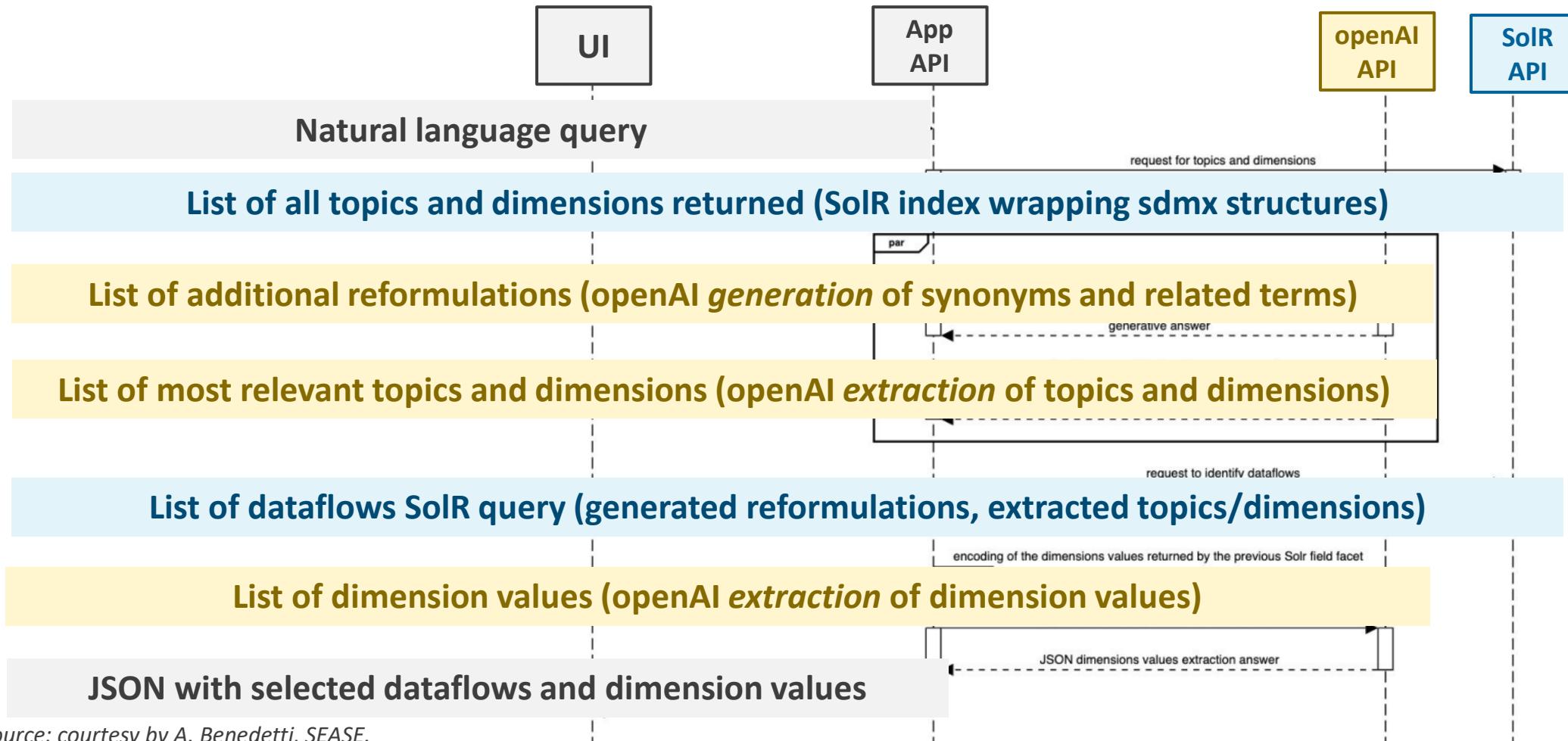
Step 2.5

Step 3.5

Direct identification of the relevant dataflow
with a pre-selection of dimensions

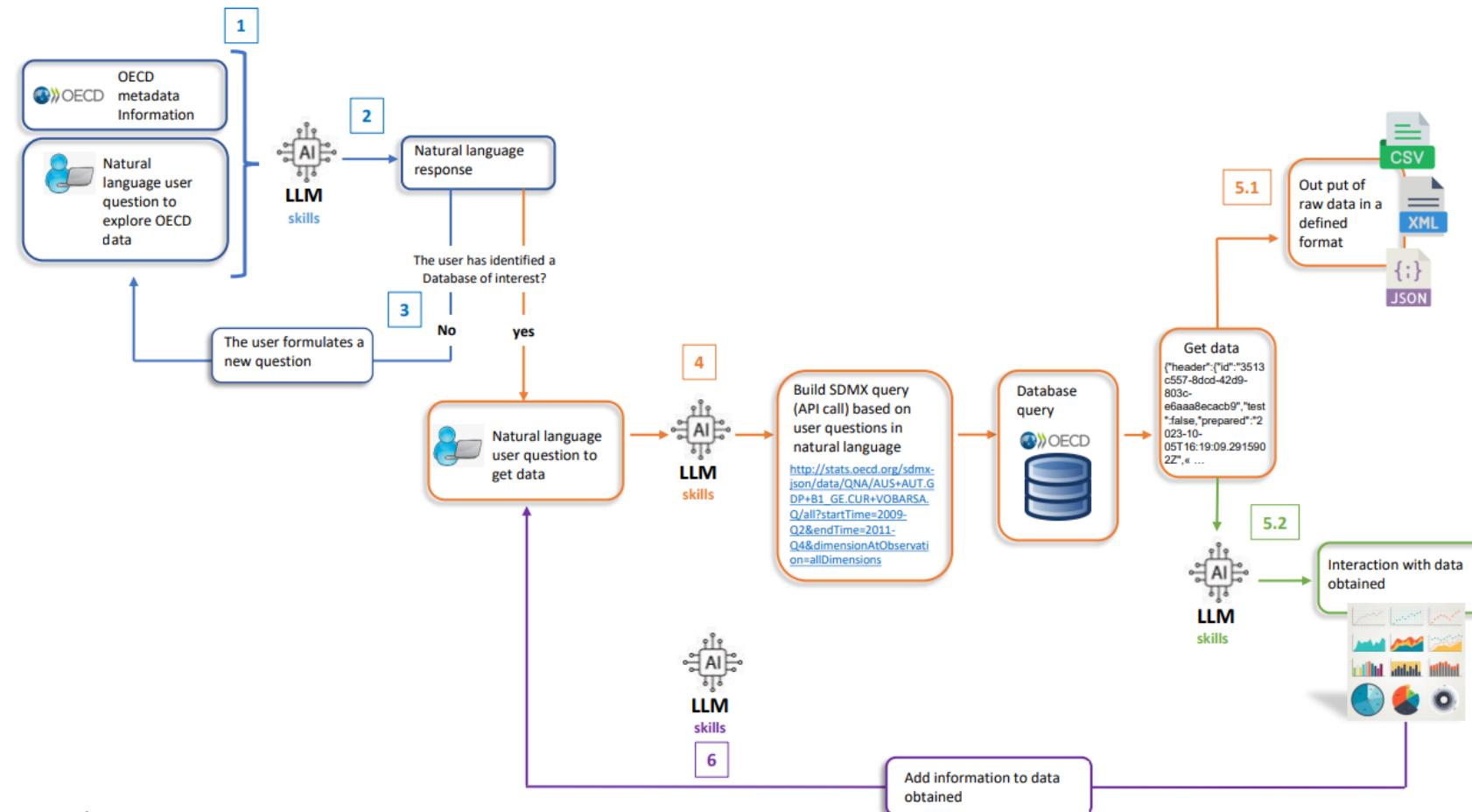
Augmenting the Data Explorer with #SDMX #AI (I)

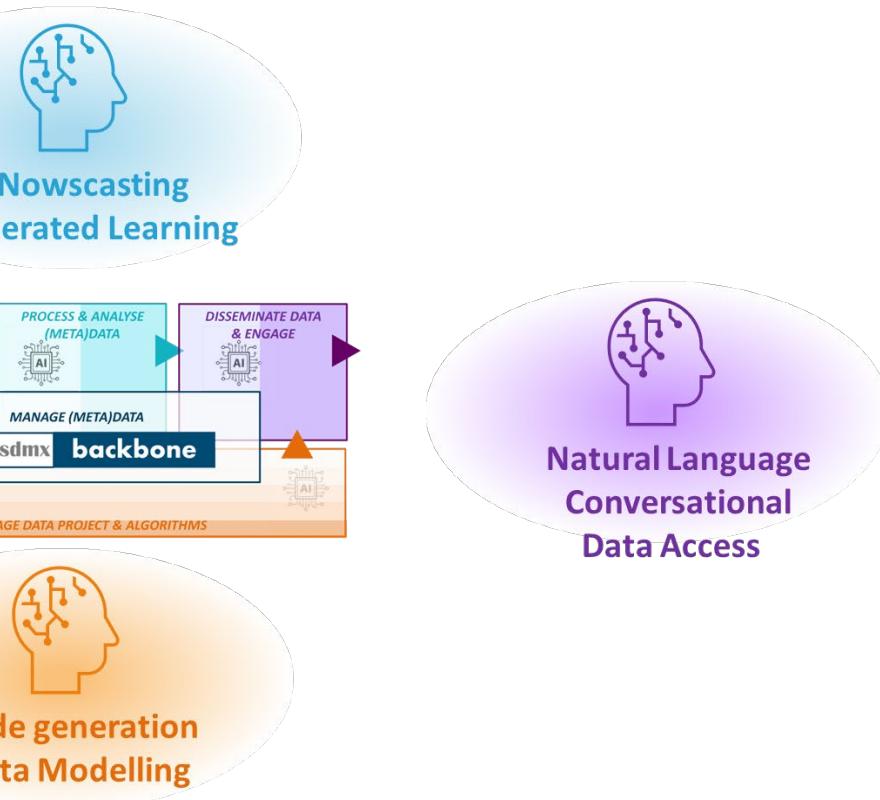
Proof of Concept ongoing, mixing Generative AI & SolR techniques.



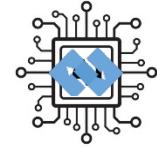
Replacing the Data Explorer with #SDMX #AI (II)

The StatsBot complements / replaces the Data Explorer with a conversational data assistant





Natural language capable Data Explorer (I)



Universal StatsBot to access any SDMX source through a conversation (II)



Capacity building – Prompt engineering practice in SDMX context



Specialized, "SDMX-literate" open source LLM



Interested in co-investing in #SDMX #AI open assets?

Contact me – eric.anvar@oecd.org

#SDMX #DATAMESH SDMX AS A DATA MESH ENABLER?



O'REILLY®
Data Mesh
Delivering Data-Driven Value at Scale

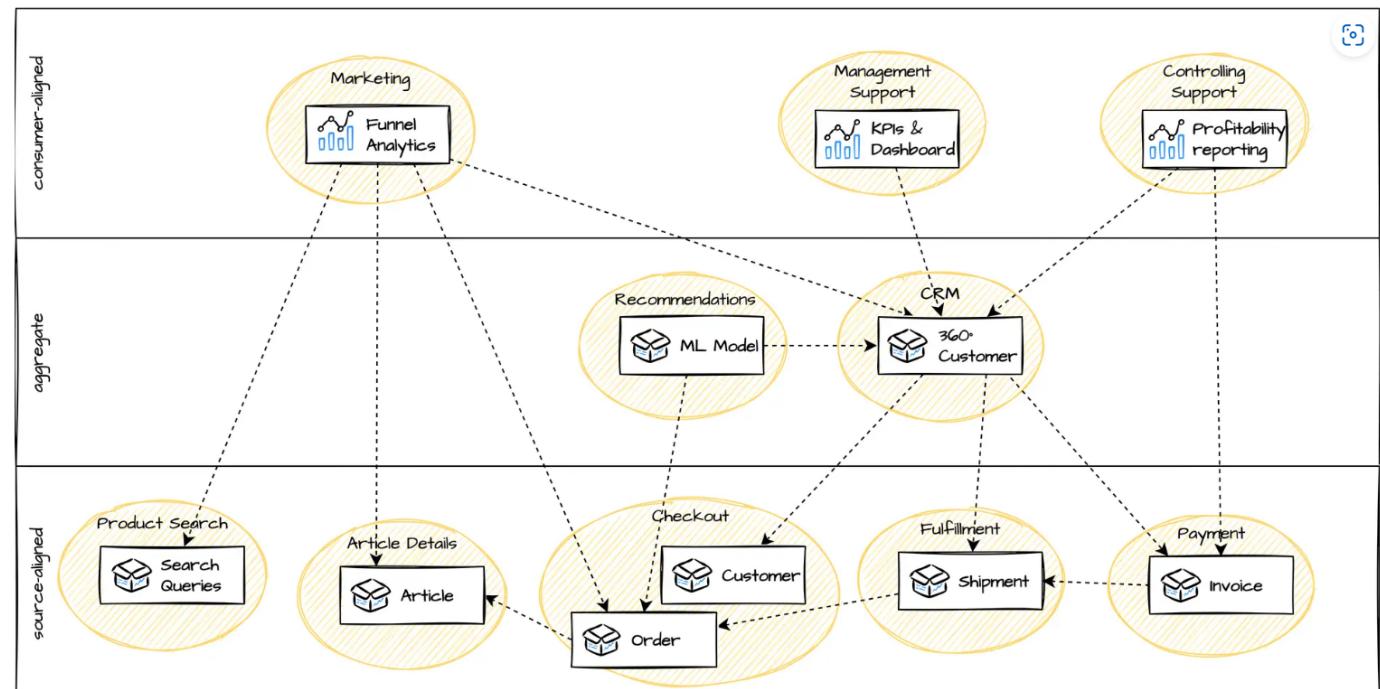


Zhamak Dehghani

What is a #DataMesh?

Address the problem with the centralized organization behind the DWH/Data Lake approach to analytical data: centralized expertise often leads to bottleneck and failure.

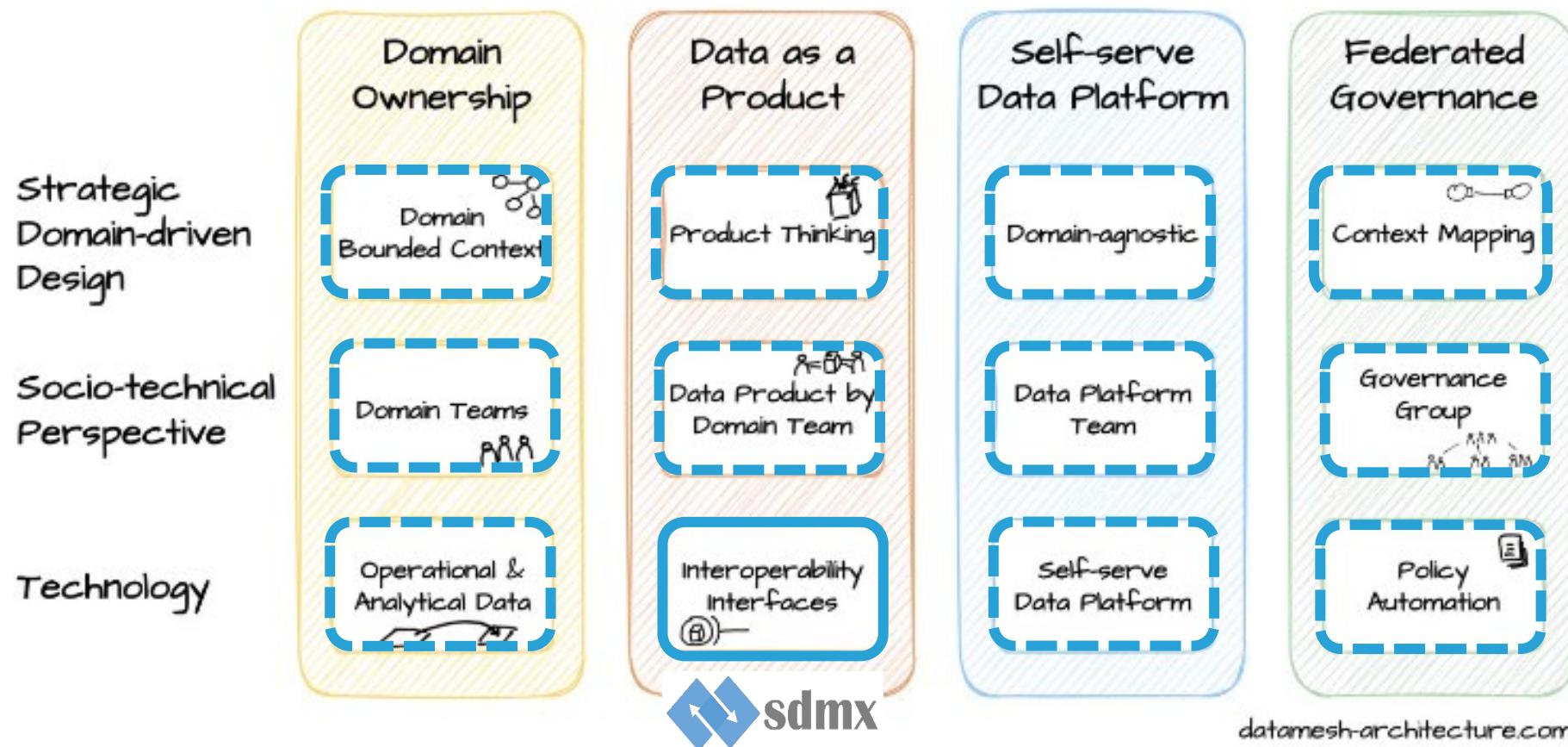
The #DataMesh is the socio-technical paradigm to switch from a purely centralized model to a model that enables domain-oriented decentralization of analytical data operations.



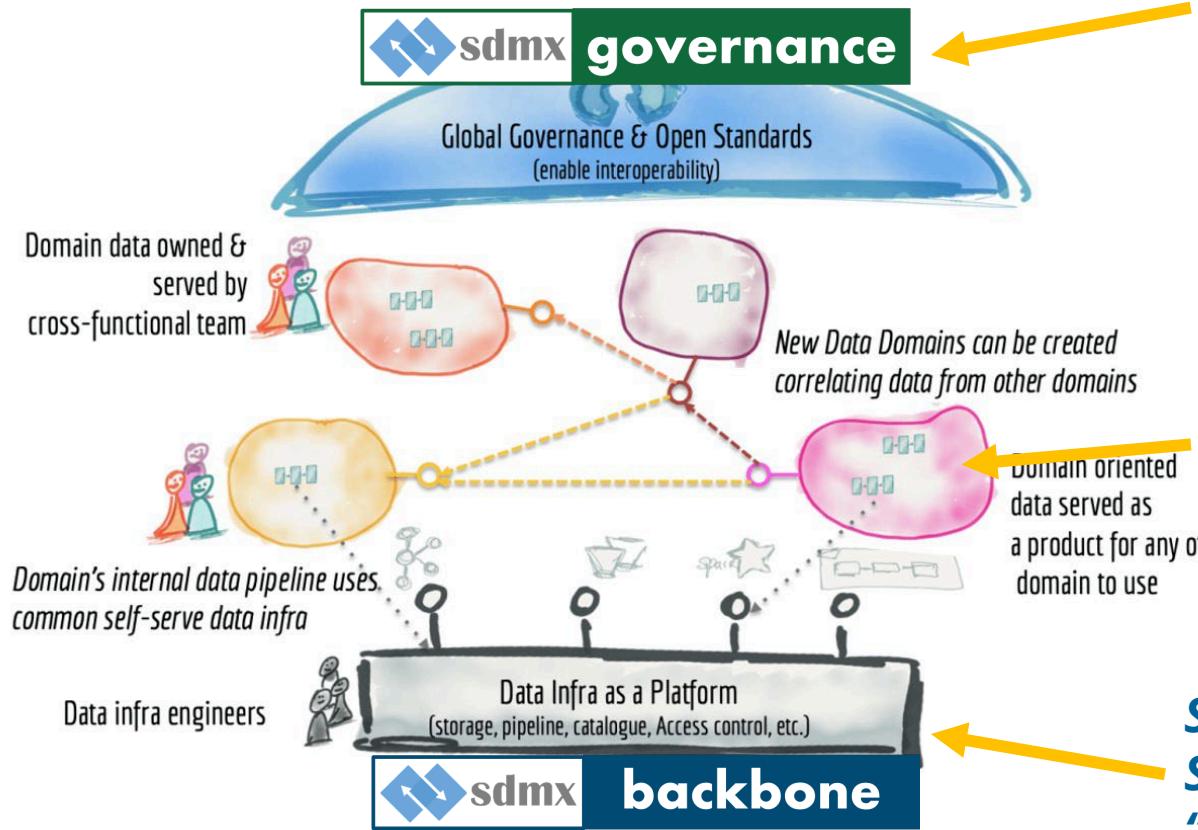
*“Data mesh attempts to **strike a balance between team autonomy and interteam interoperability and collaboration**. It gives domain teams autonomy to take control of their local decision making, for example, **choosing the best data model for their data products**, while it uses computational governance policies to impose a consistent experience across all data products, for example, **standardising the data modelling language that all domains utilize**. Data mesh gives domain teams autonomy to build and maintain their data products, while it places a **domain-agnostic data platform in place for teams to do so in a consistent and cost-effective way**.”*

Zhamak Dehghani, Data Mesh

#SDMX as a #DataMesh enabler?



#SDMX as a #DataMesh enabler?

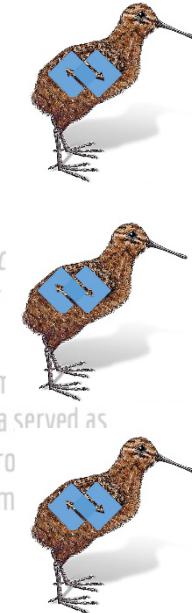
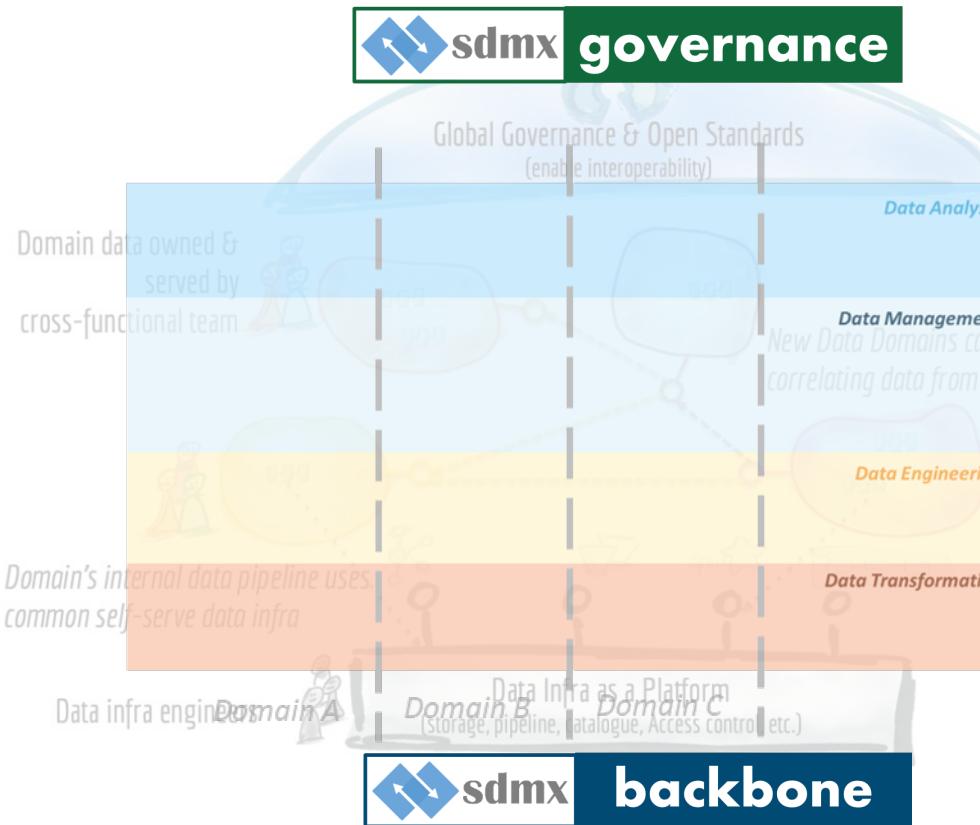


SDMX governance as cornerstone of data governance, enabling a progressive harmonisation of data semantics across domains and the data lifecycle.

Each business domain to build their data process and products in autonomy, drawing on common information models and technical building blocks.

SDMX backbone brings commoditised SDMX storage, SDMX API-centricity, and SDMX semantics as part of the "Data infrastructure as a Platform".

Creating #SDMX #DataMesh open and reusable assets



Develop the architecture vision & business case for #DataMesh in Official Statistics

Adapt and augment #SDMX (and #VTL?) with the #DataMesh paradigm

(data product, data contract, DDD, computational governance...)

Develop implementation guidelines for the #SDMX #DataMesh & experience



Interested in developing the #SDMX #DataMesh framework?

Contact me – eric.anvar@oecd.org

SIS-CCCommunity



- [SIS-CC Community](#)
- [SIS-CC 2020-25 Strategy](#)
- [Why SDMX Matters](#)

SDMX.IO Ecosystem

- [The SDMX.IO ecosystem](#)



SDMX Community



- [SDMX Community](#)
- [SDMX 2021-25 Strategy](#)
- [The business case for SDMX](#)
- [SDMX 3.0 packages under review](#)
- [SDMX Guideline: Modelling statistical domains and exchange frameworks in SDMX](#)
- [SDMX Guidelines: A Reference Framework for SDMX Structural Metadata Governance](#)

.Stat Suite, open source platform



- .Stat Suite – [High level product overview and flight planner](#)
- .Stat Suite – [Kanban Board](#)
- .Stat Suite – [Changelog](#)
- .Stat Suite – [Extensive documentation](#)
- .Stat Suite – [Code repository](#)

.Stat Academy, open knowledge platform



- .Stat Academy – Learning Paths for [Data Producer](#); [Data Tooler](#)
- .Stat Academy – [Courses](#)
- .Stat Academy – [Webinars](#)
- .Stat Academy – [Become an instructor](#)

3. #SDMX as #AI and #DataMesh enabler?



2. #SDMX #MetadataDriven approach



1. The SIS-CC: a #Community journey

