

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

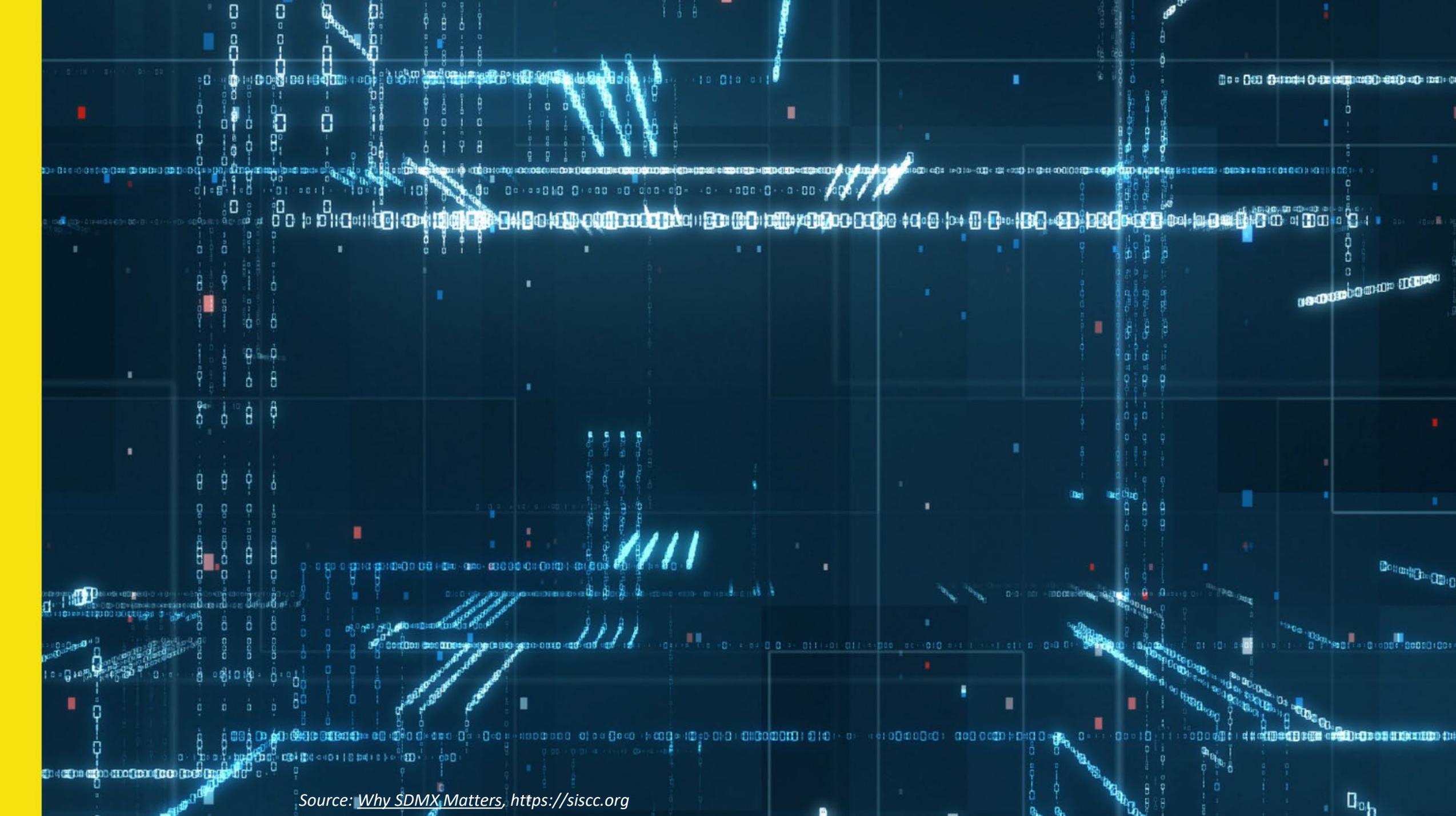
Keywords

#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>

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

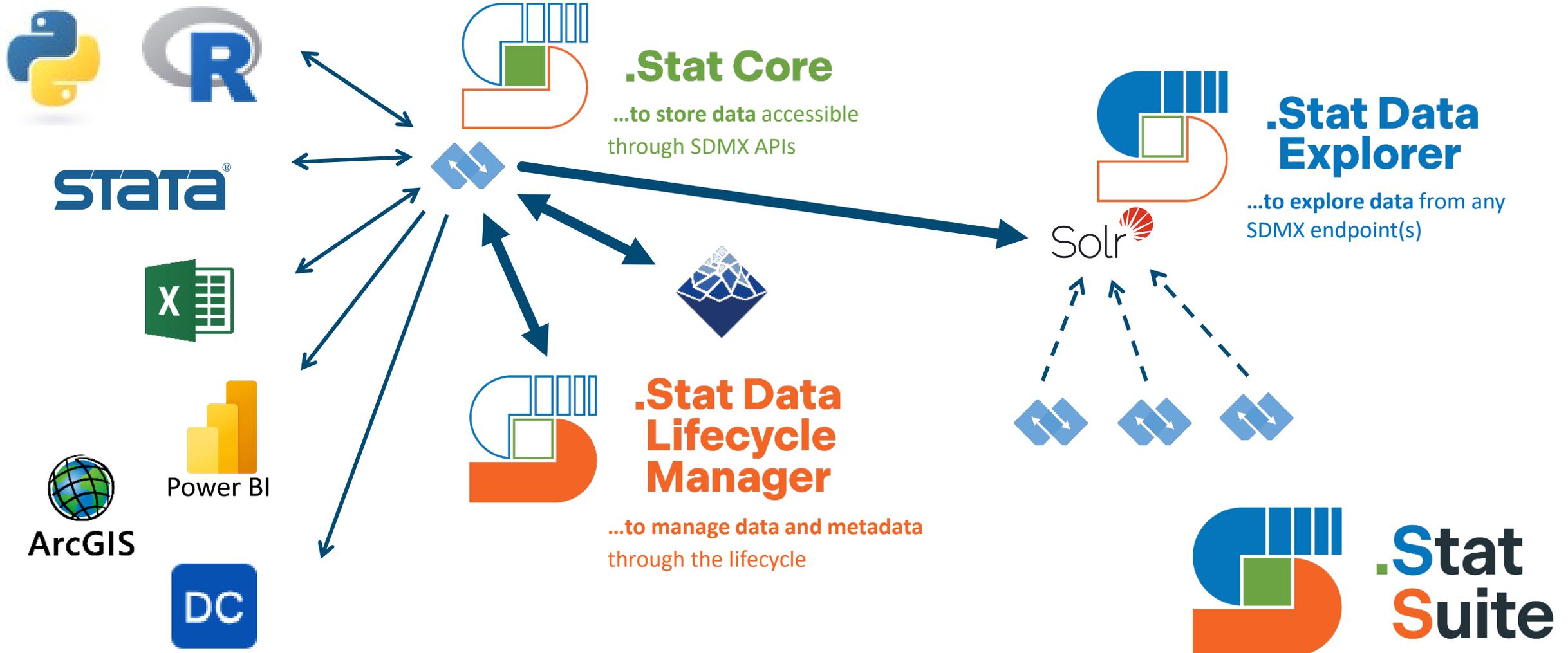
SIS-CC (“the Community”) is a reference open-source community for official statistics, focusing on product excellence and **delivering concrete solutions to common problems through co-investment and co-innovation.**



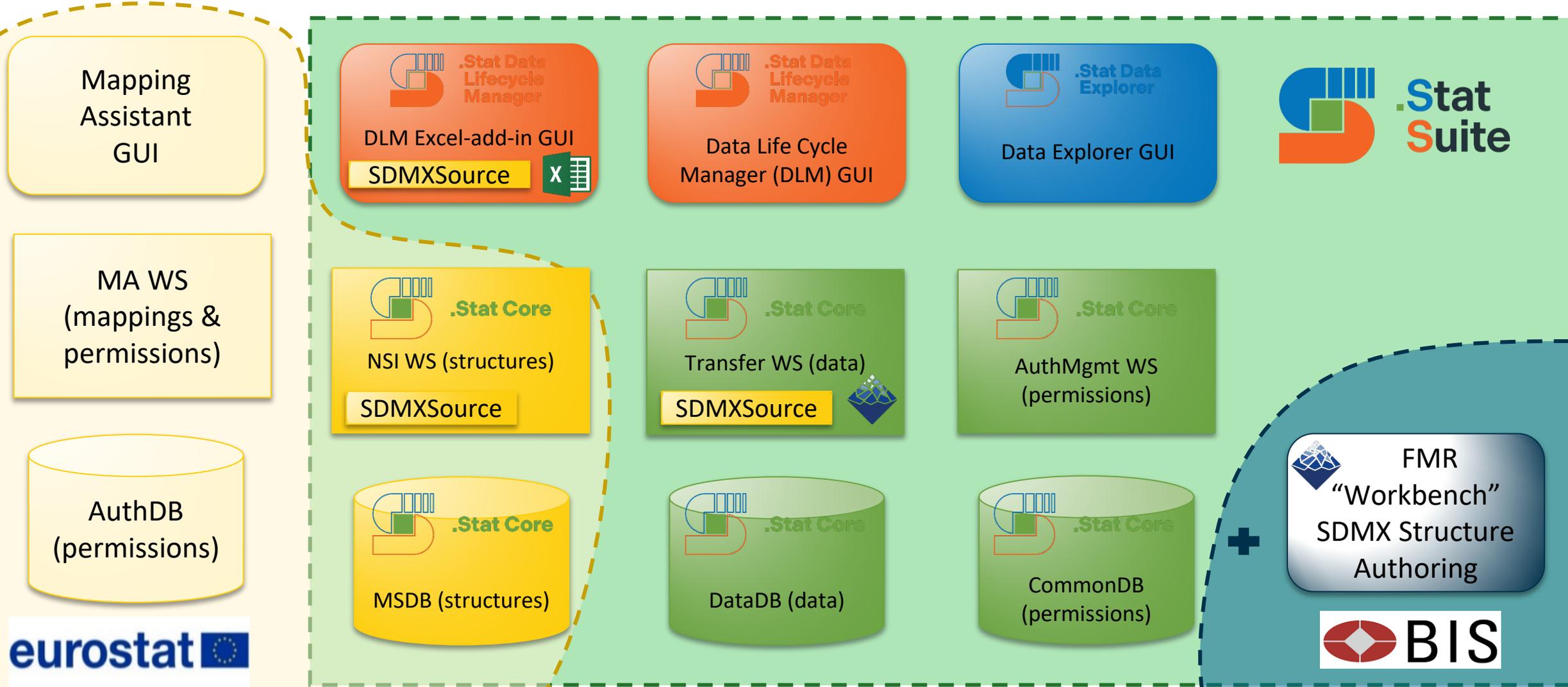
Members

Partners

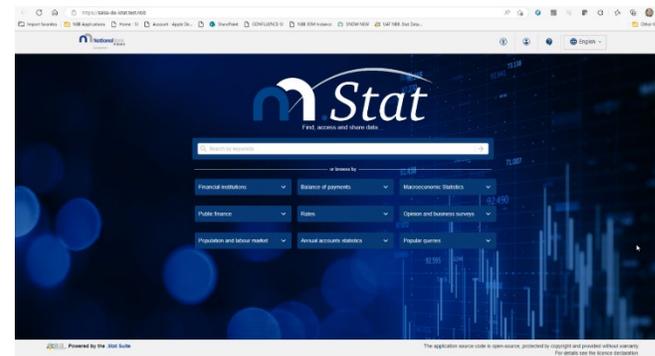
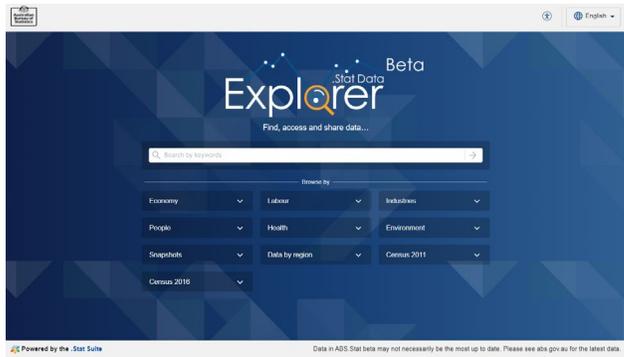
The #OpenSource .Stat Suite



Component-based Architecture → Cooperative “Business Model”



Examples of implementations

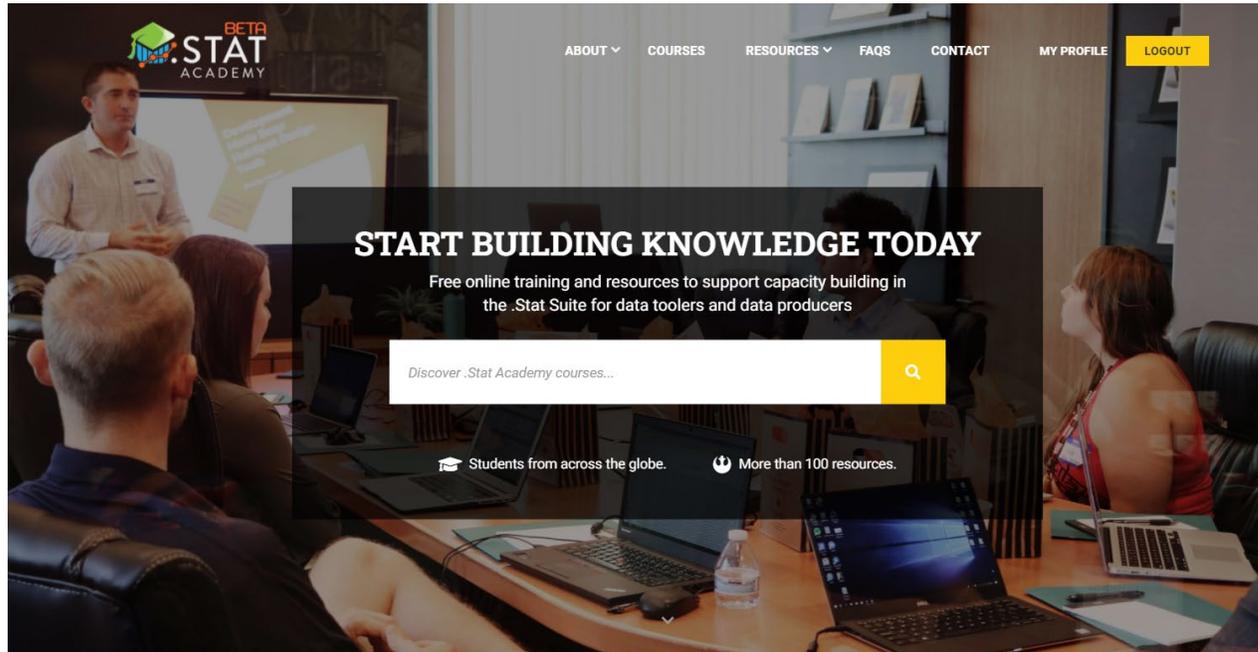




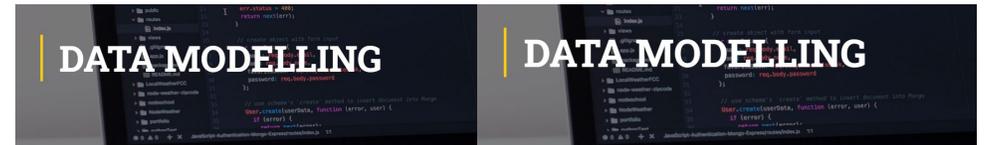
Search, visualise and share Pacific data...

Pacific Data Hub .STAT

The #OpenKnowledge .Stat Academy



<https://academy.siscc.org>



An introduction to SDMX structural model producers

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

An introduction to SDMX for data producers

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

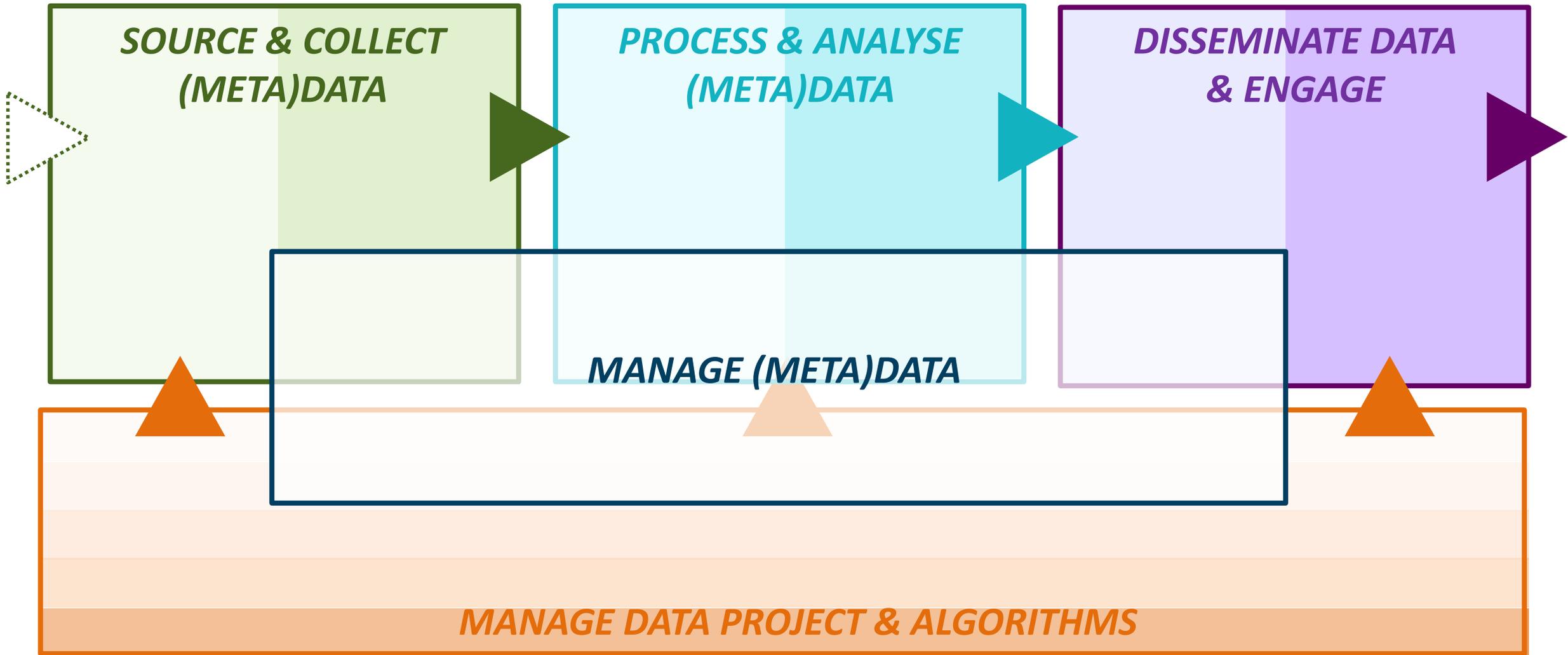
Instructor: Jens Dossé
Categories: Data modelling, SDMX
Review: ★★★★★

Starting point: Build capacity in SDMX data modelling

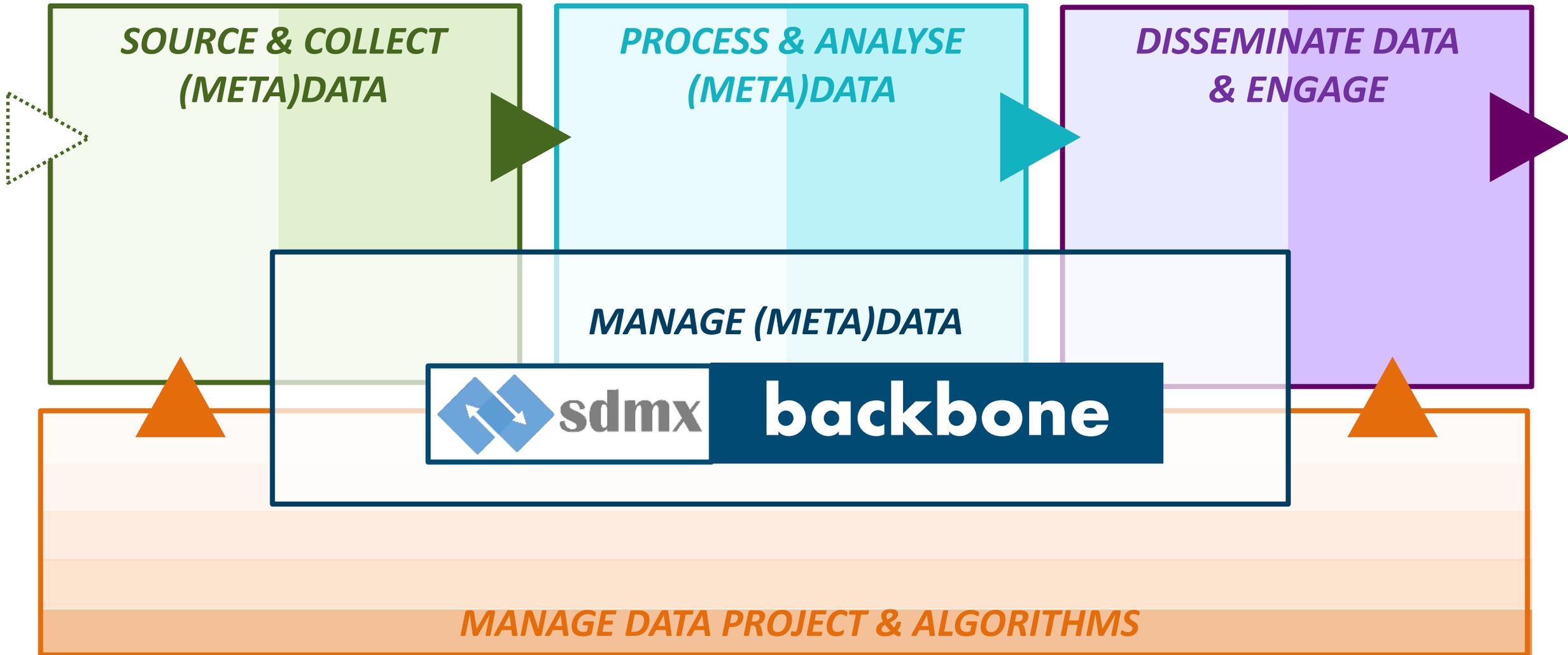
THE #SDMX METADATA-DRIVEN APPROACH



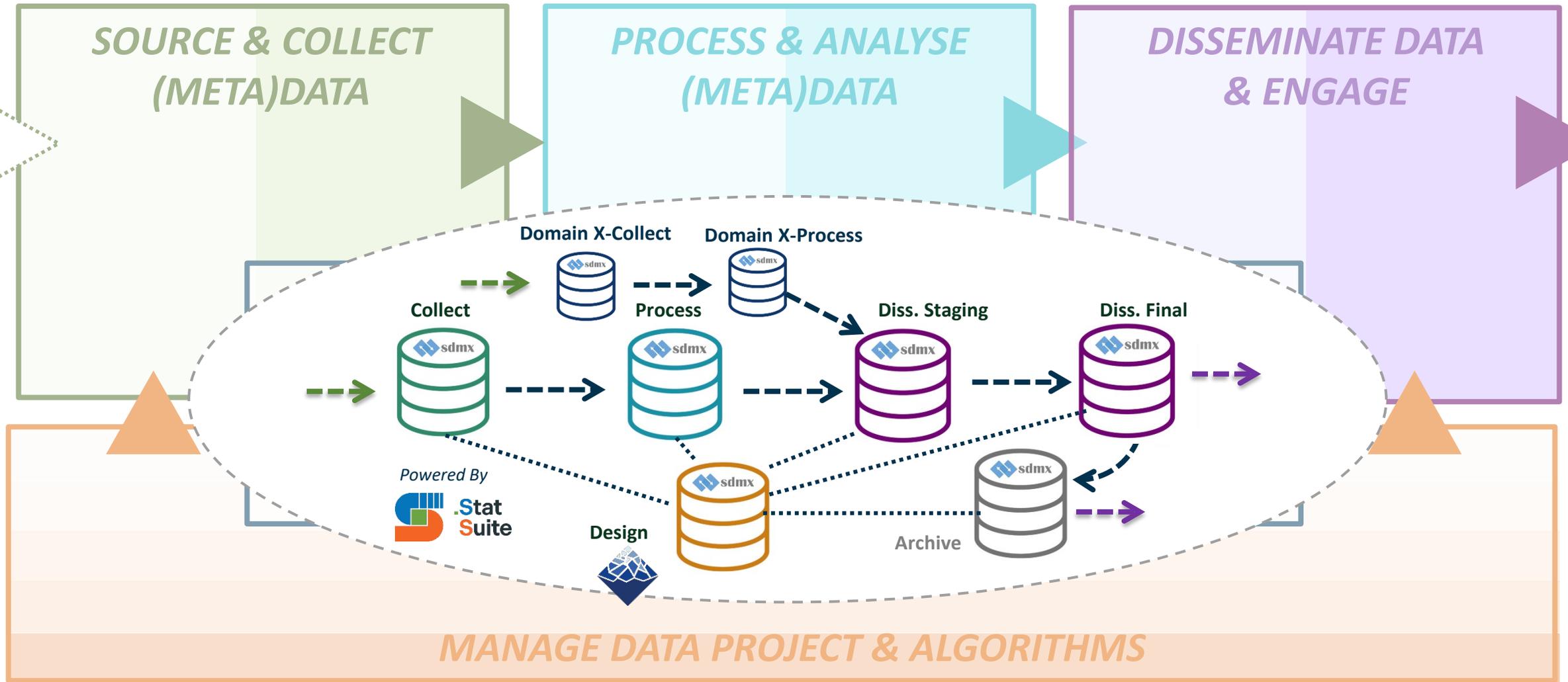
The Data Platform



The #SDMX Backbone

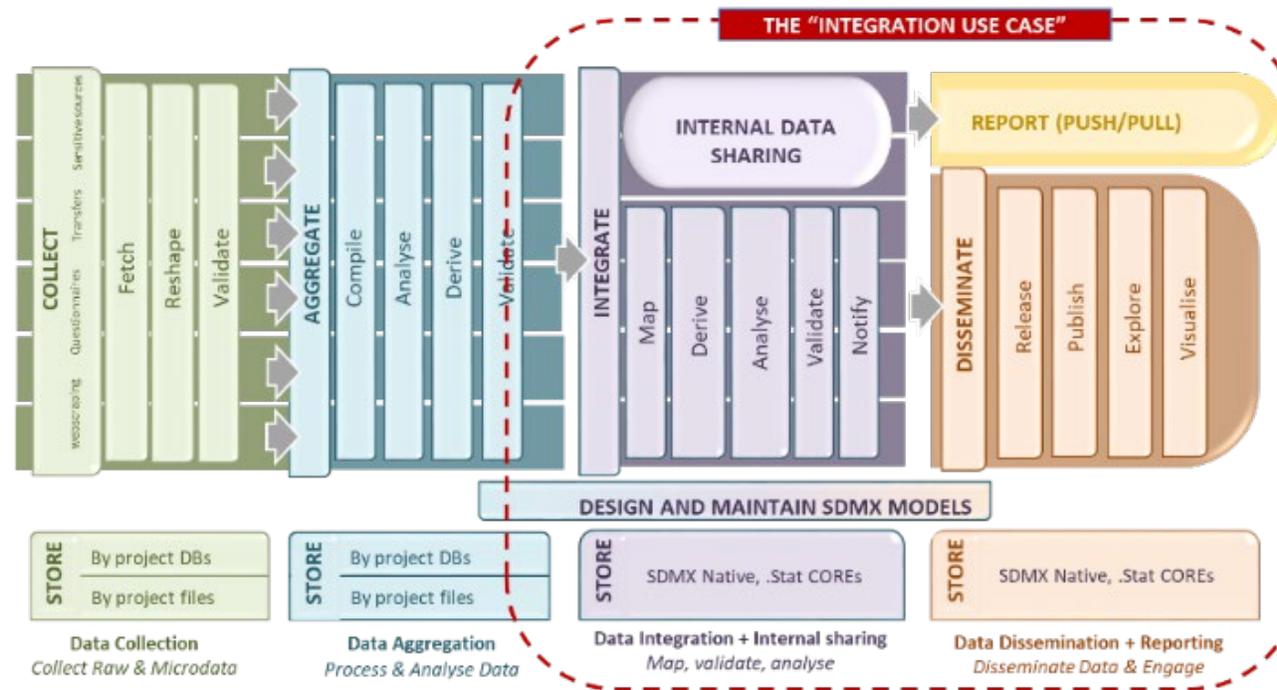


The #SDMX Backbone



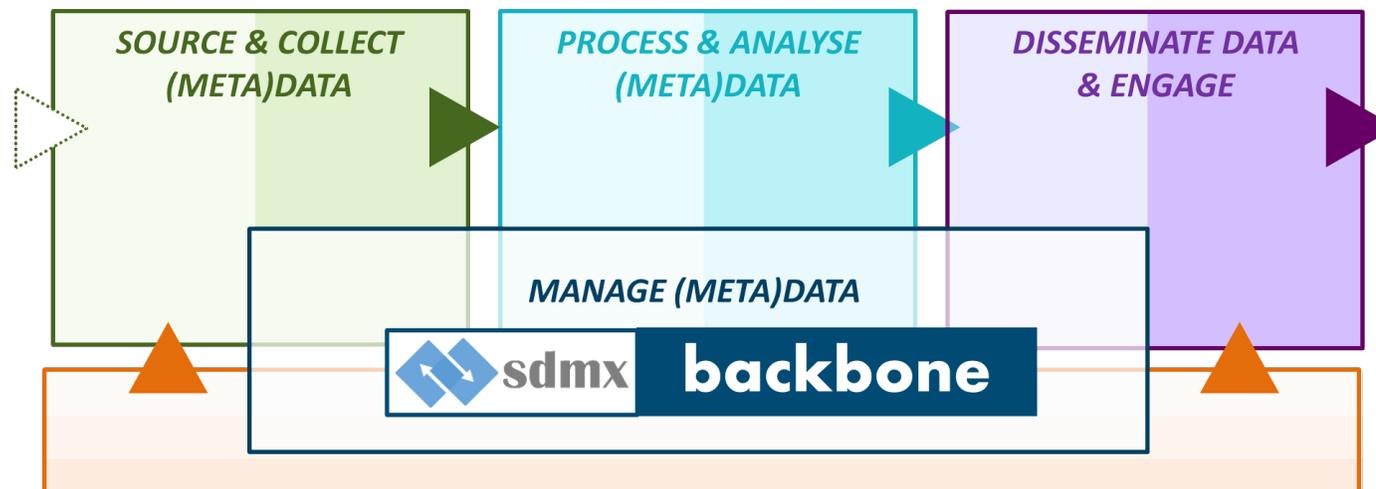
The #SDMX Backbone

The "Data Integration" use case at the National Bank of Belgium



The #SDMX Backbone – Wrap-up

- **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

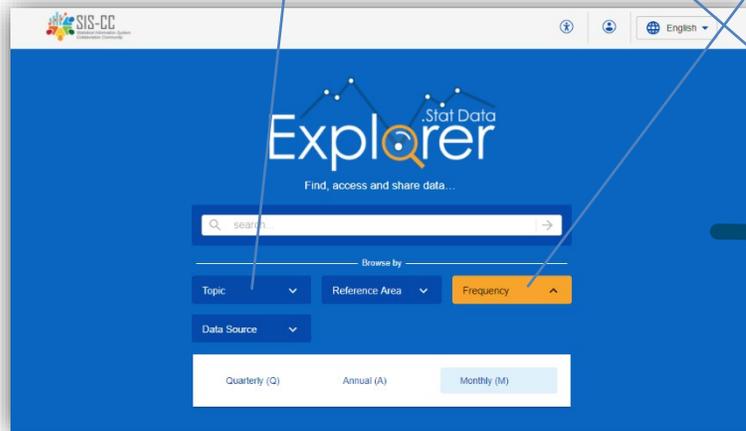


#SDMX #MetadataDriven Data User Experience

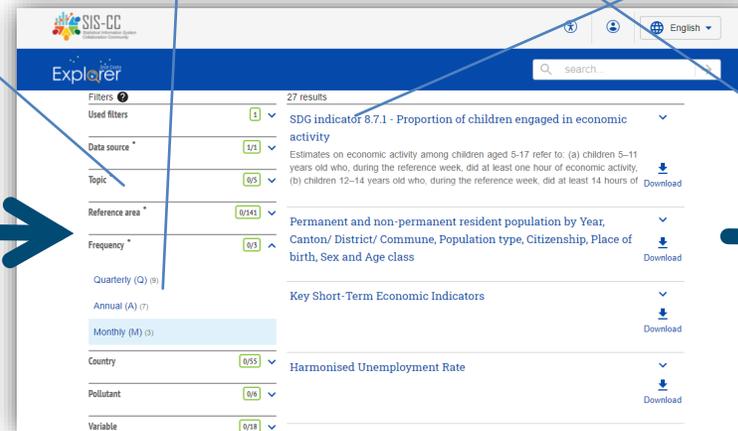

category scheme(s)

 dimensions:
concepts + codelists

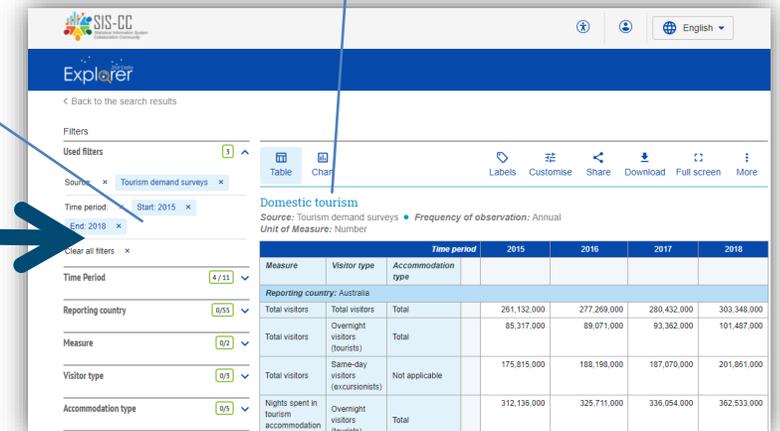
 dataflow
+ presentation annotations



Step 1 Home page
Lexical search or Topics

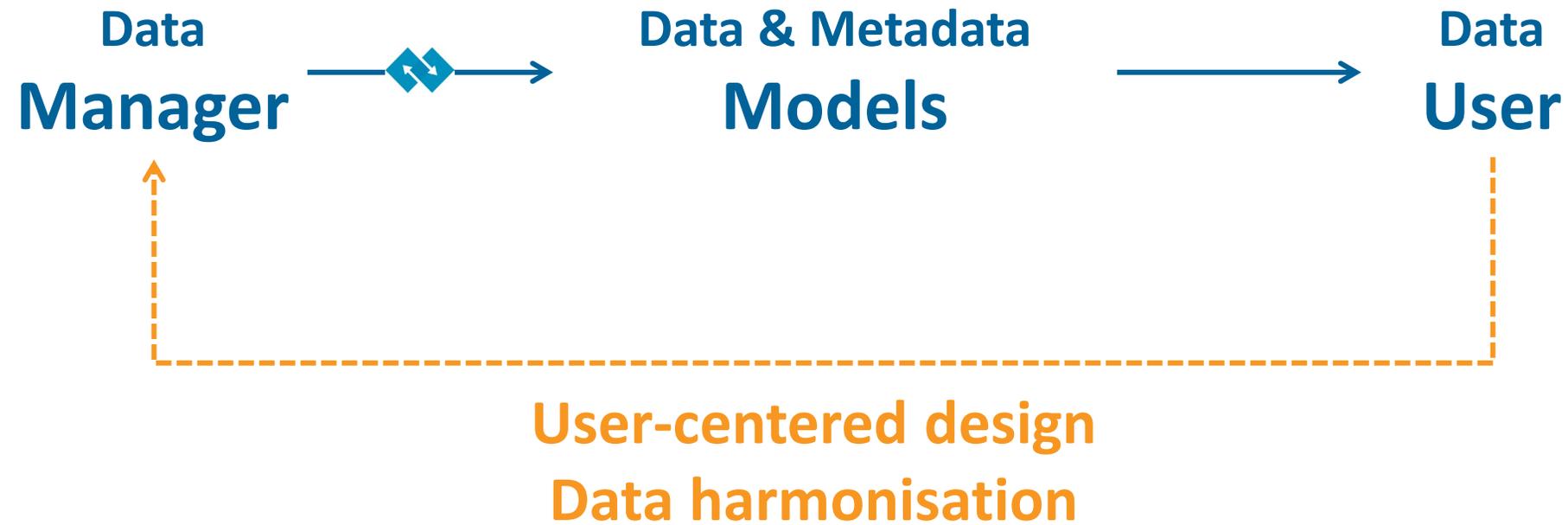


Step 2 Search results
Filtering on dynamic facets



Step 3 Visualisation page
Selecting dimensions / variables

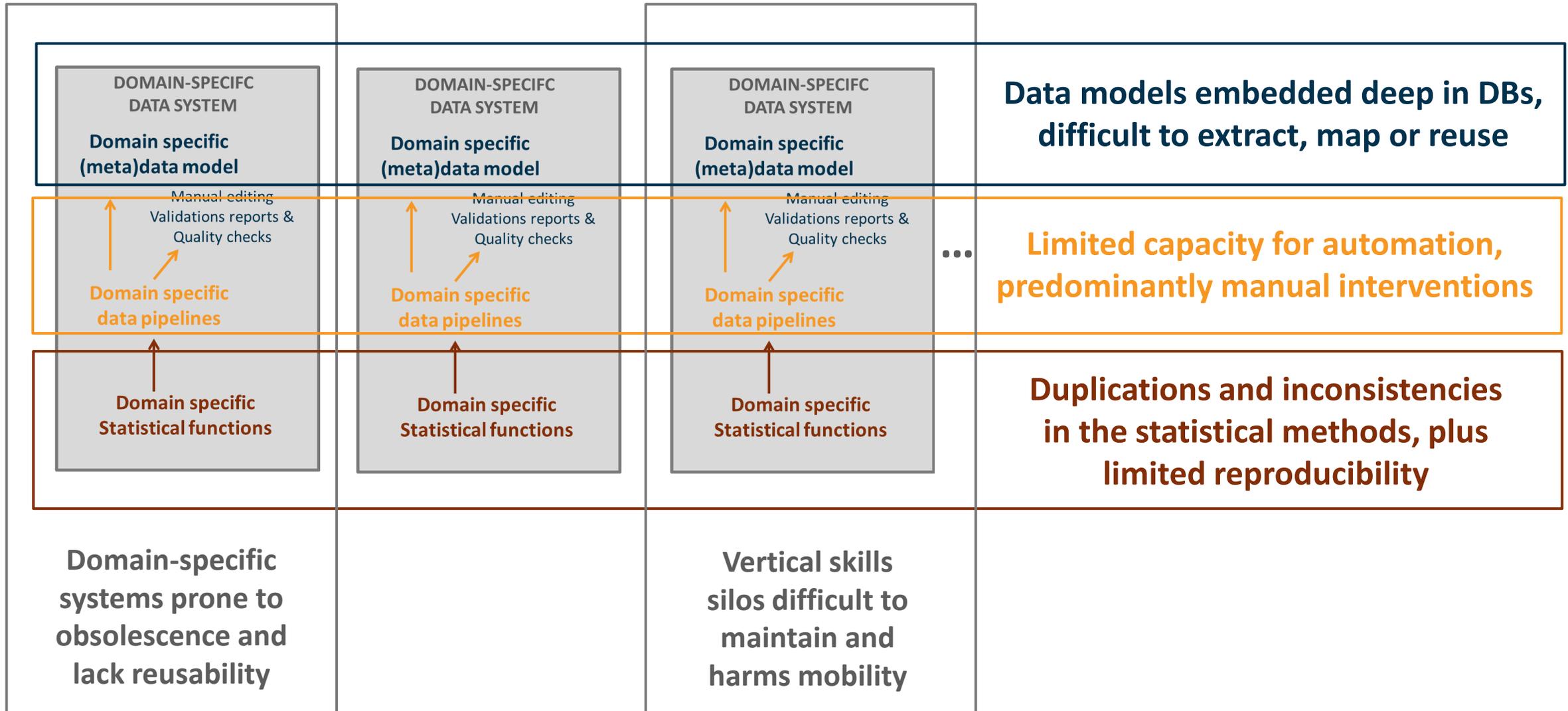
#SDMX #MetadataDriven Data User Experience



Establish a systematic User Research



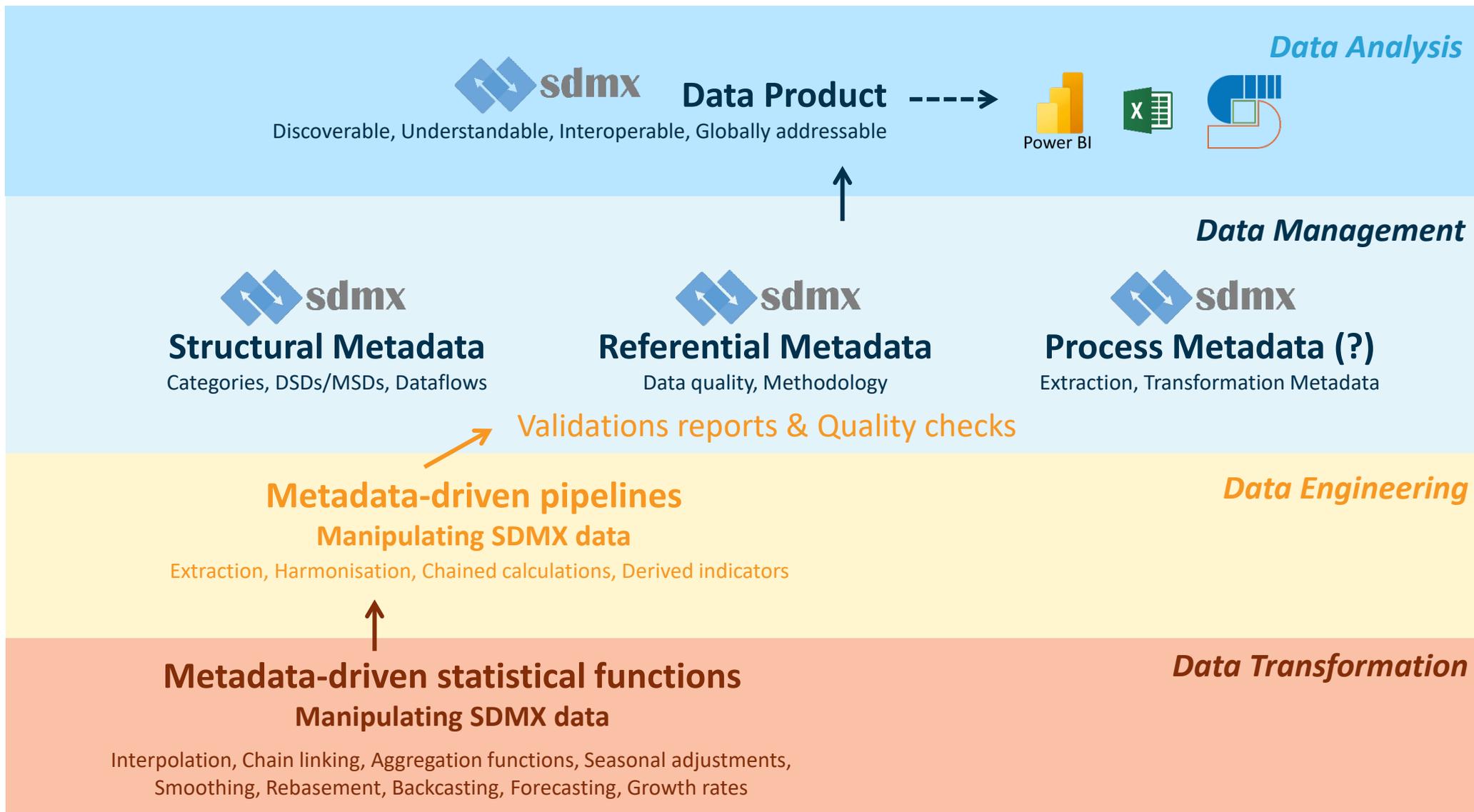
The Data Platform is fragmented



#SDMX #MetadataDriven Data Manager Experience

MANAGE DATA
& METADATA

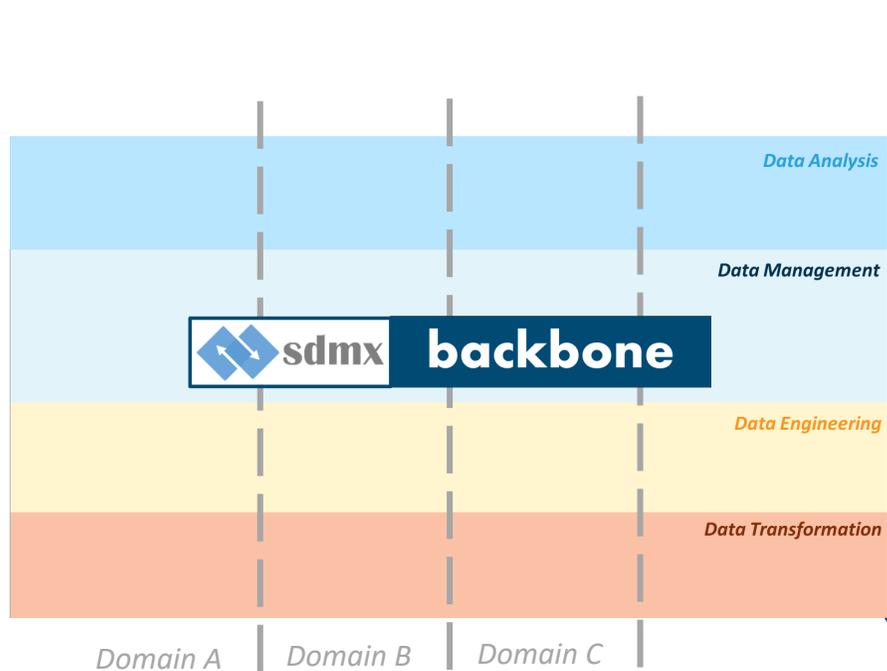
MANAGE DATA PROJECT
& ALGORITHMS



(?) feature that exists partially / peacemeal in SDMX and that ought to be specified

#SDMX #MetadataDriven Data Manager Experience

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



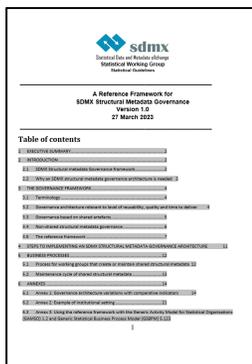
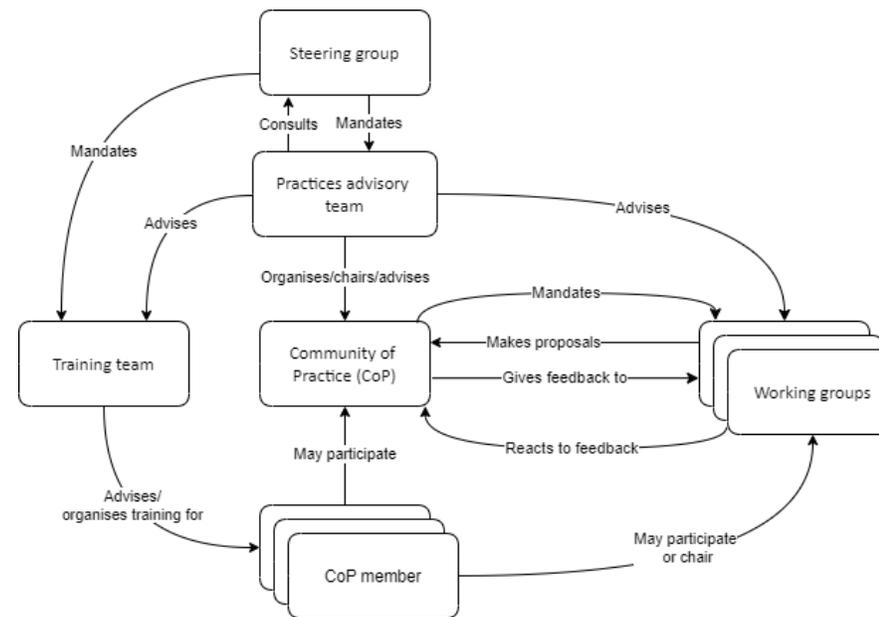
- Decoupled roles and specializations
- Data mgt by non coding-experts
- Facilitation of staff mobility
- Possibility of data tasks mutualization
- Facilitation of platform migration
- Domain specific coding language
- Domain specific interfaces
- Capitalize on shared methods / reuse
- Capitalize on common infrastructure



Structure your Data governance

#SDMX enables Data Governance

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



sdmx
Statistical Information System
Collaboration Community

A Reference Framework for
SDMX Structural Metadata Governance
Version 1.0
27 March 2023

Table of contents

- 1. Introduction
- 2. Objectives
- 3. Scope
- 4. The Reference Framework
- 5. The Reference Framework
- 6. The Reference Framework
- 7. The Reference Framework
- 8. The Reference Framework
- 9. The Reference Framework
- 10. The Reference Framework
- 11. The Reference Framework
- 12. The Reference Framework
- 13. The Reference Framework
- 14. The Reference Framework
- 15. The Reference Framework
- 16. The Reference Framework
- 17. The Reference Framework
- 18. The Reference Framework
- 19. The Reference Framework
- 20. The Reference Framework



sdmx governance

- 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 backbone

**SDMX metadata drives
Data User experience
Data Manager experience**



Interested in developing an #SDMX Backbone & Governance?

Contact me – eric.anvar@oecd.org

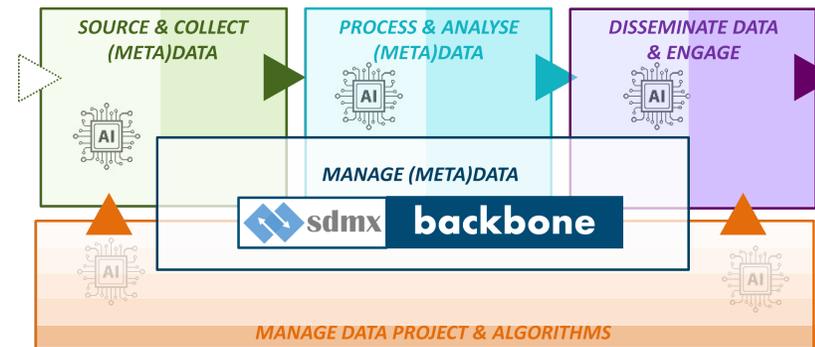
#SDMX #AI
SDMX AS AN AI ENABLER?



**Nowcasting
Federated Learning**



**Outlier detection
Clustering**

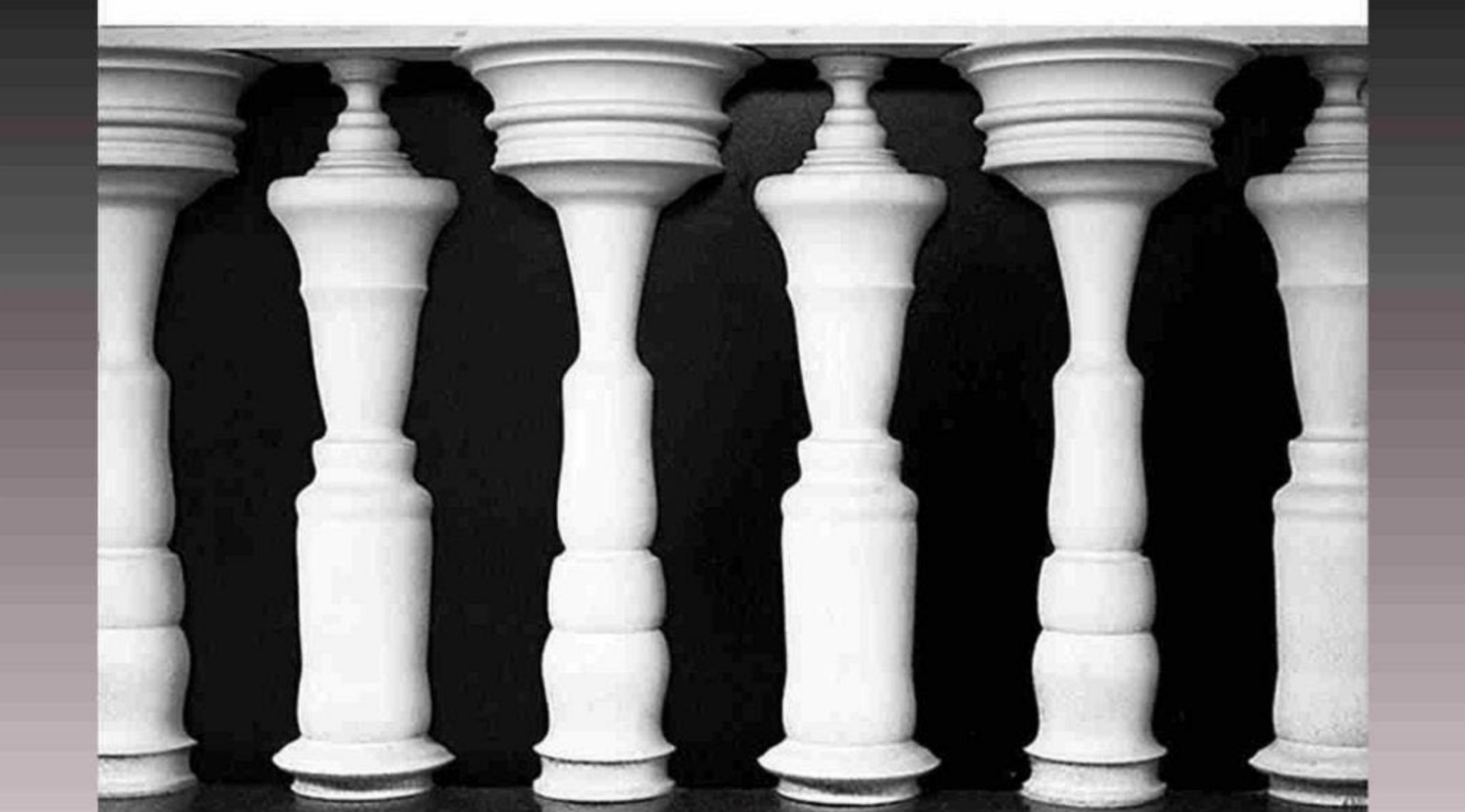


**Natural Language
Conversational
Data Access**



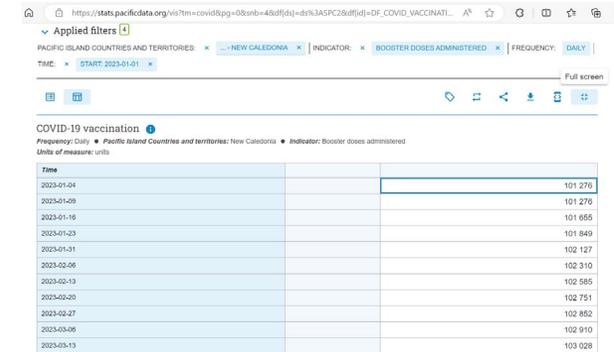
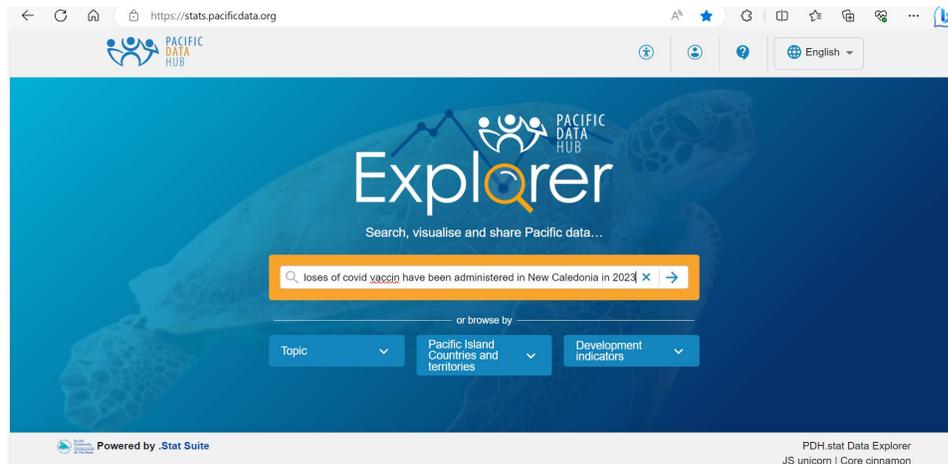
**Code generation
Data Modelling**

The cognitive gap



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

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



Applied filters

PACIFIC ISLAND COUNTRIES AND TERRITORIES: NEW CALEDONIA | INDICATOR: BOOSTER DOSES ADMINISTERED | FREQUENCY: DAILY

TIME: START: 2023-01-01

Full screen

COVID-19 vaccination

Frequency: Daily • Pacific Island Countries and territories: New Caledonia • Indicator: Booster doses administered

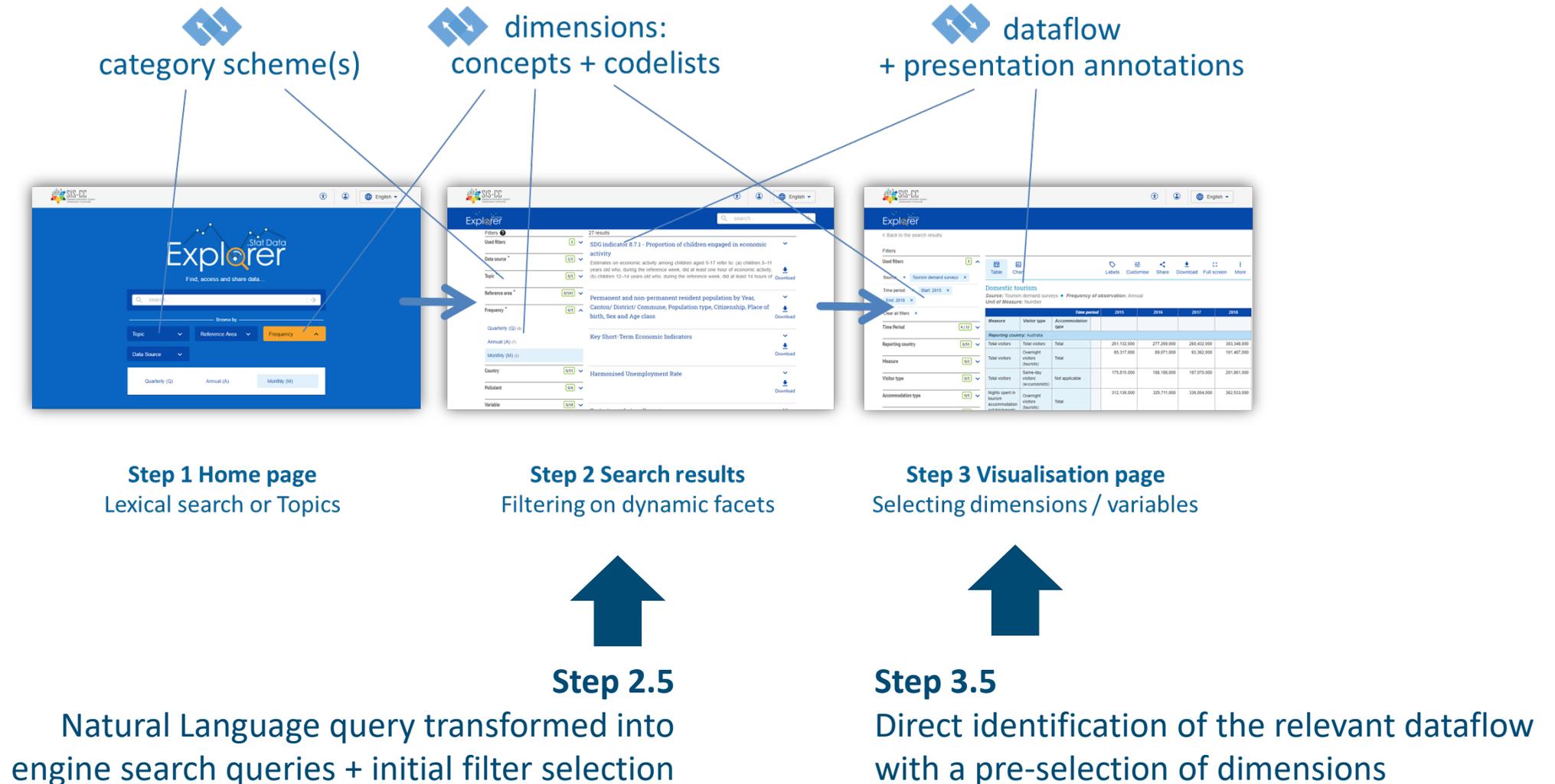
Units of measure: units

Time	Value
2023-01-04	101 276
2023-01-09	101 276
2023-01-16	101 685
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

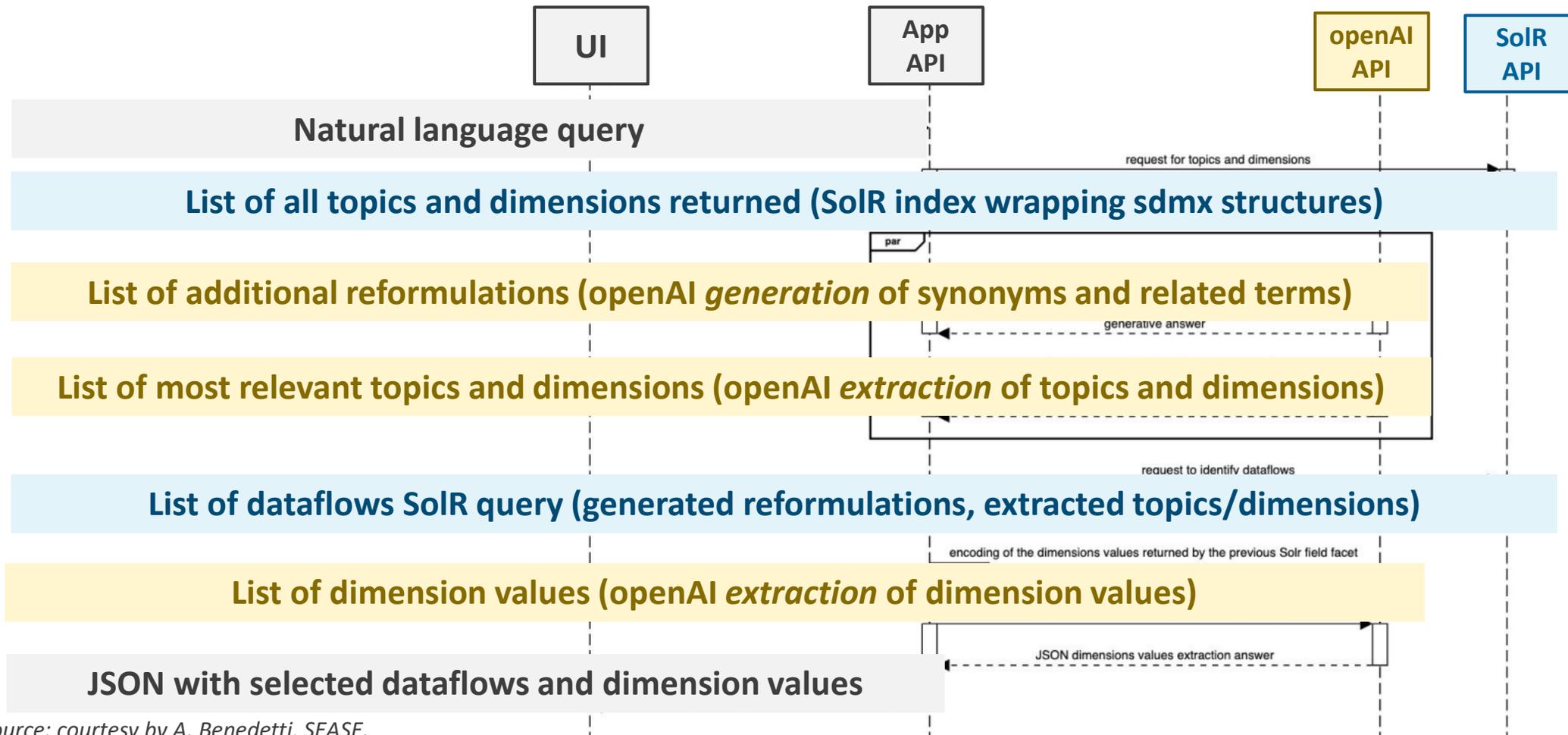
Natural Language query transformed into search queries
+ Pre-selection of dimensions on target dataflow(s)

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



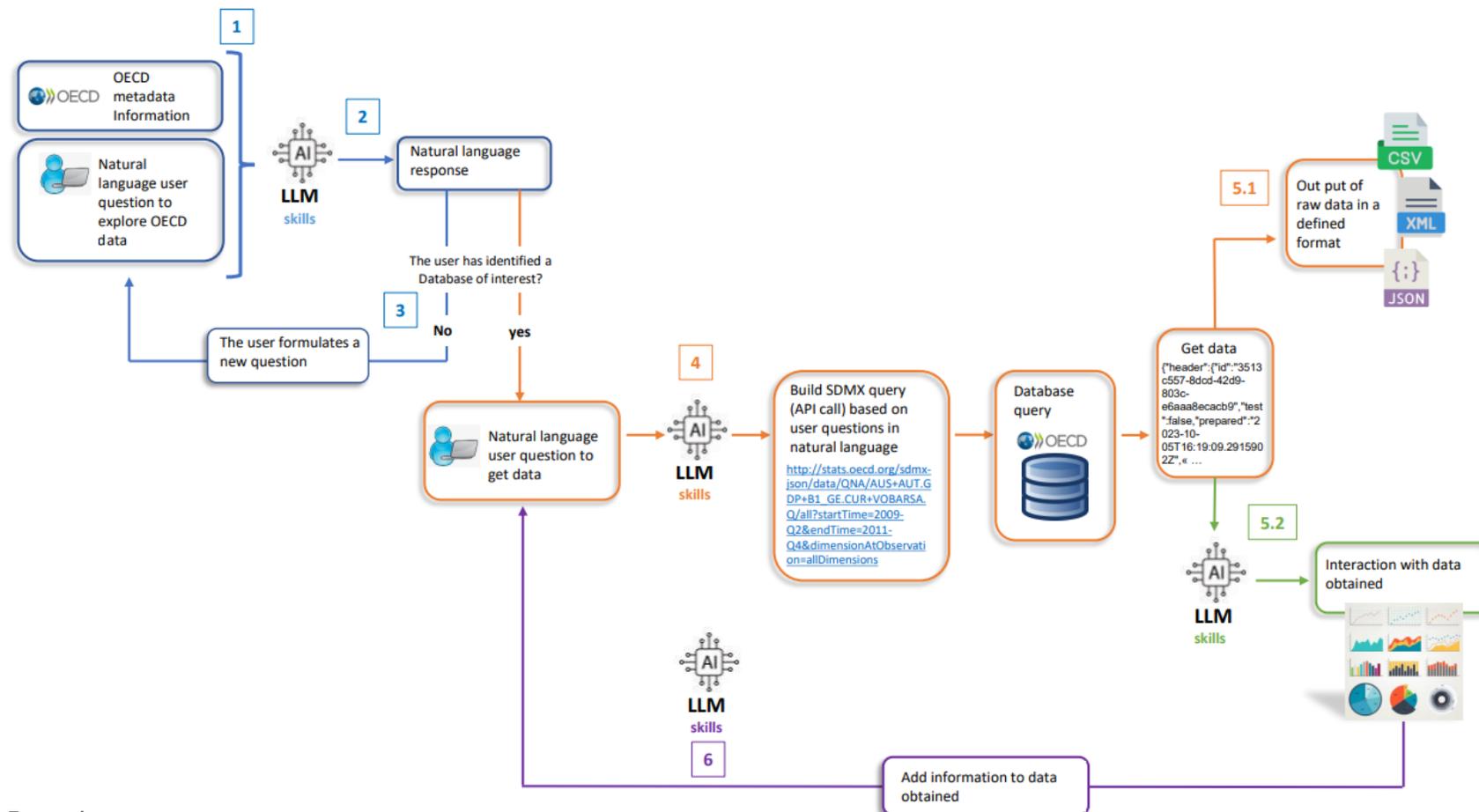
Augmenting the Data Explorer with #SDMX #AI (I)

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

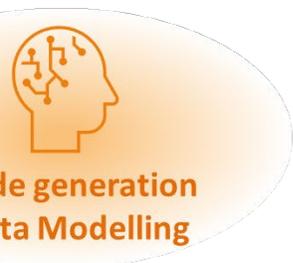
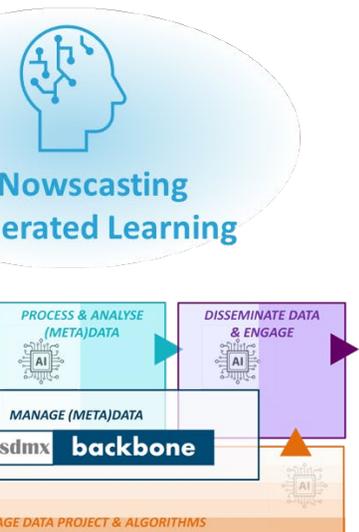


Replacing the Data Explorer with #SDMX #AI (II)

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



Creating #SDMX #AI open and reusable assets



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

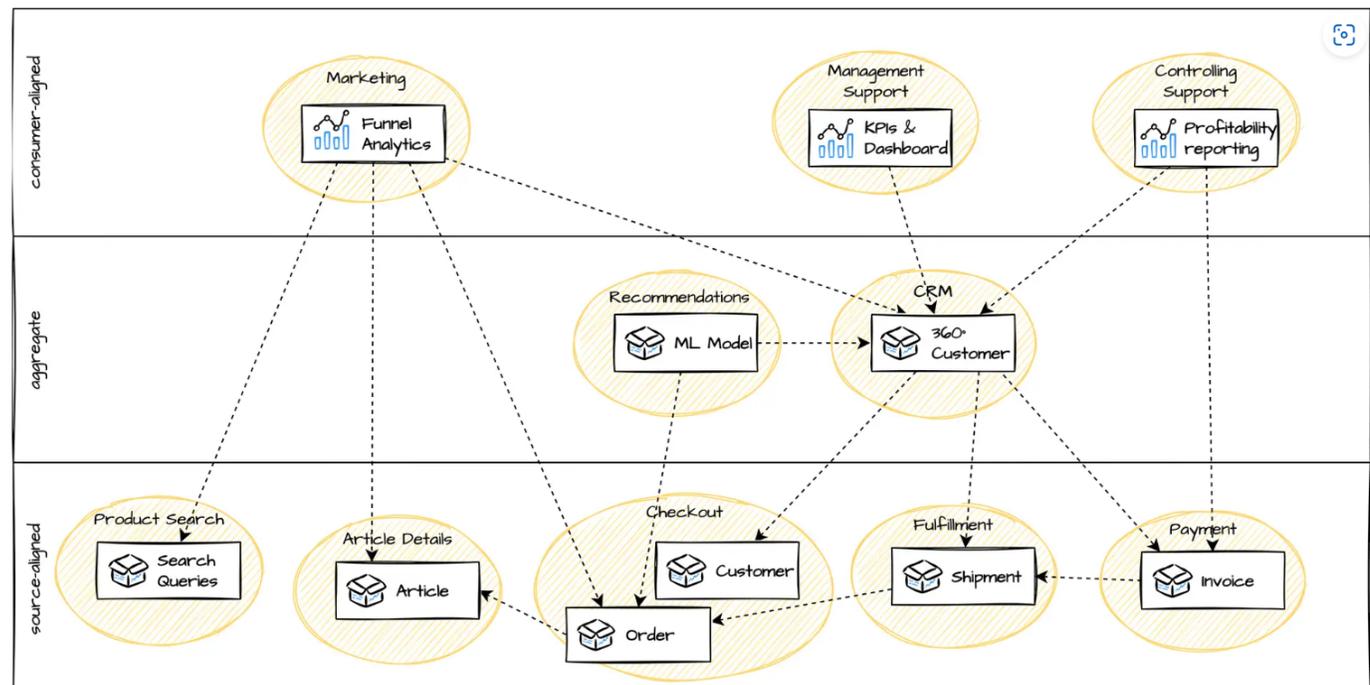


Zhamak Dehghani

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.

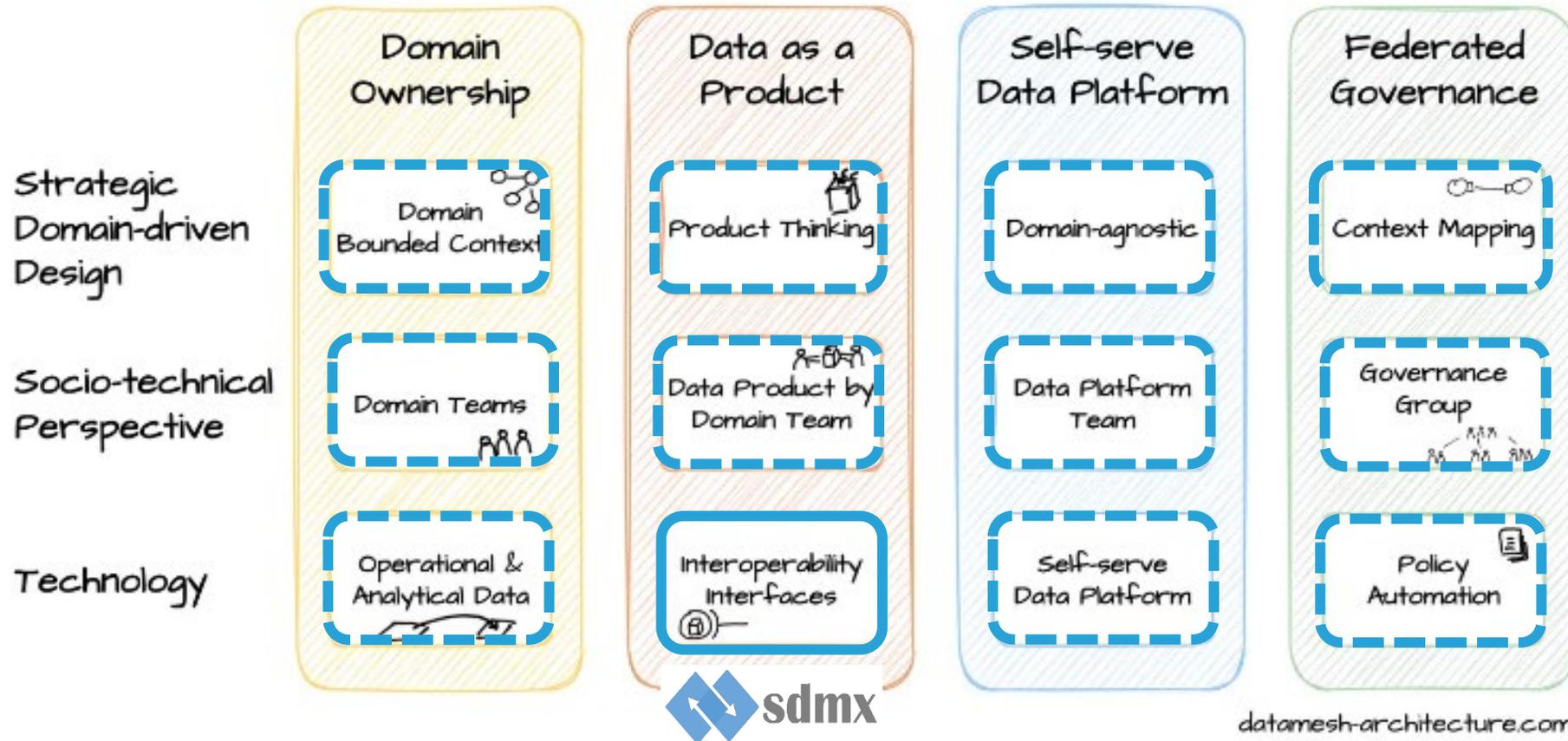


#DataMesh requires common data semantics

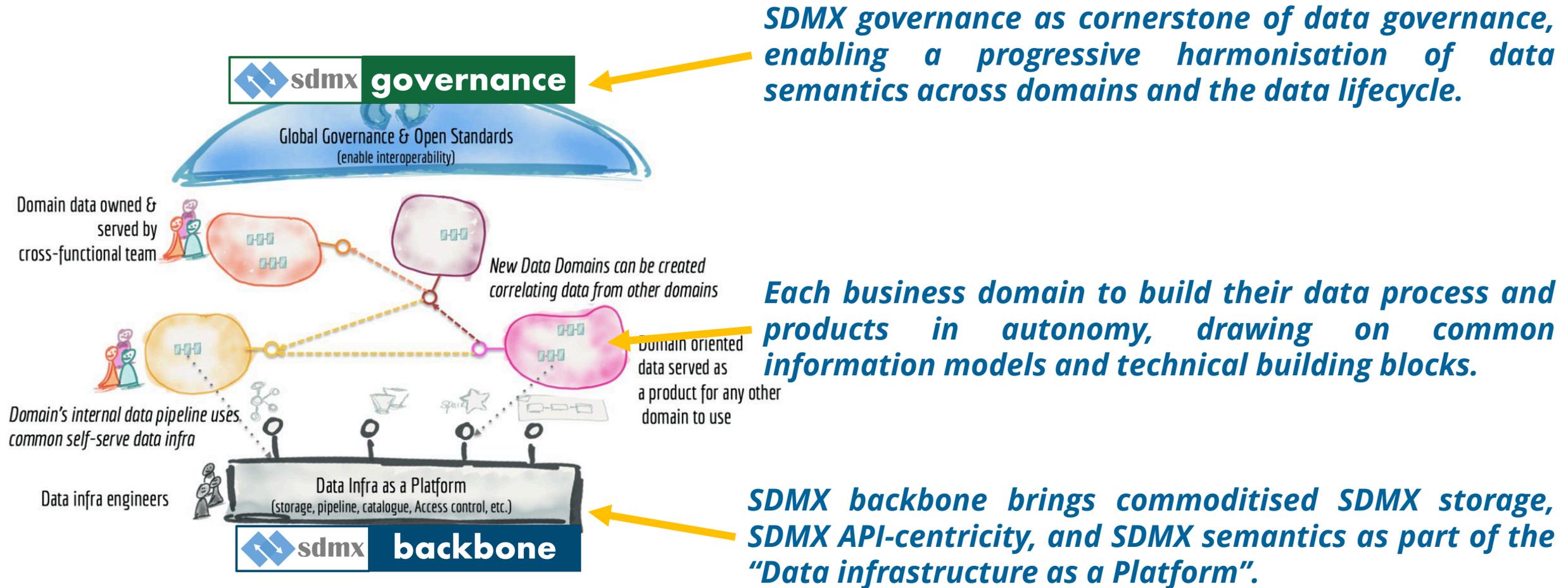
*“Data mesh attempts to **strike a balance between team autonomy and interteam interoperability and collaboration**. It gives domain teams autonomy to take control of their local decision making, for example, **choosing the best data model for their data products**, while it uses computational governance 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

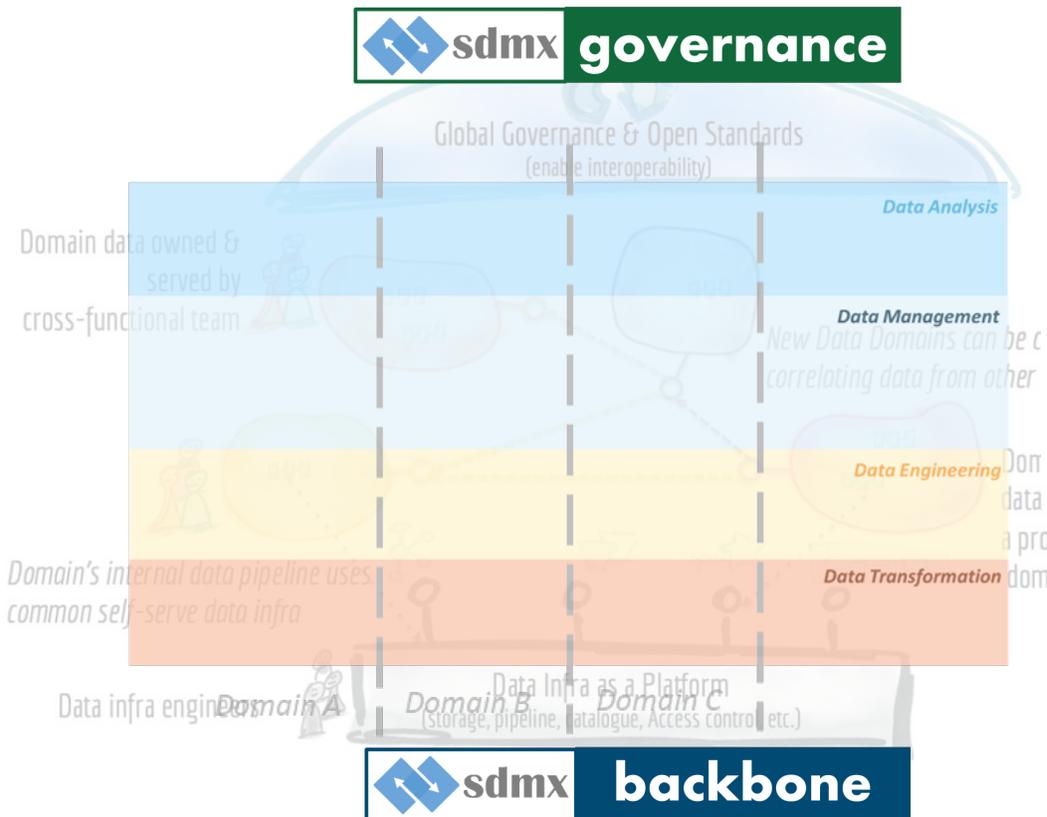
#SDMX as a #DataMesh enabler?



#SDMX as a #DataMesh enabler?



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 & experiemnt



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



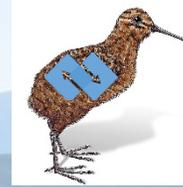
- .Stat Suite – [High level product overview and flight planner](#)
- .Stat Suite – [Kanban Board](#)
- .Stat Suite – [Changelog](#)
- .Stat Suite – [Extensive documentation](#)
- .Stat Suite – [Code repository](#)

.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

