

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

SDMX Global Conference, *Empowering Data Communities*

Manama, October 2023

## Keywords

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





A hiker is seen from behind, walking along a dirt trail on a rocky, grassy mountain ridge. The landscape is expansive, with rolling hills and mountains in the distance under a clear blue sky with some light clouds. The text is overlaid on the image in a clean, sans-serif font.

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

**2. The #SDMX #MetadataDriven approach**

**1. Creating value with #SDMX – A #Community journey**

# CREATING VALUE WITH #SDMX A #COMMUNITY JOURNEY

# Who we are



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**.

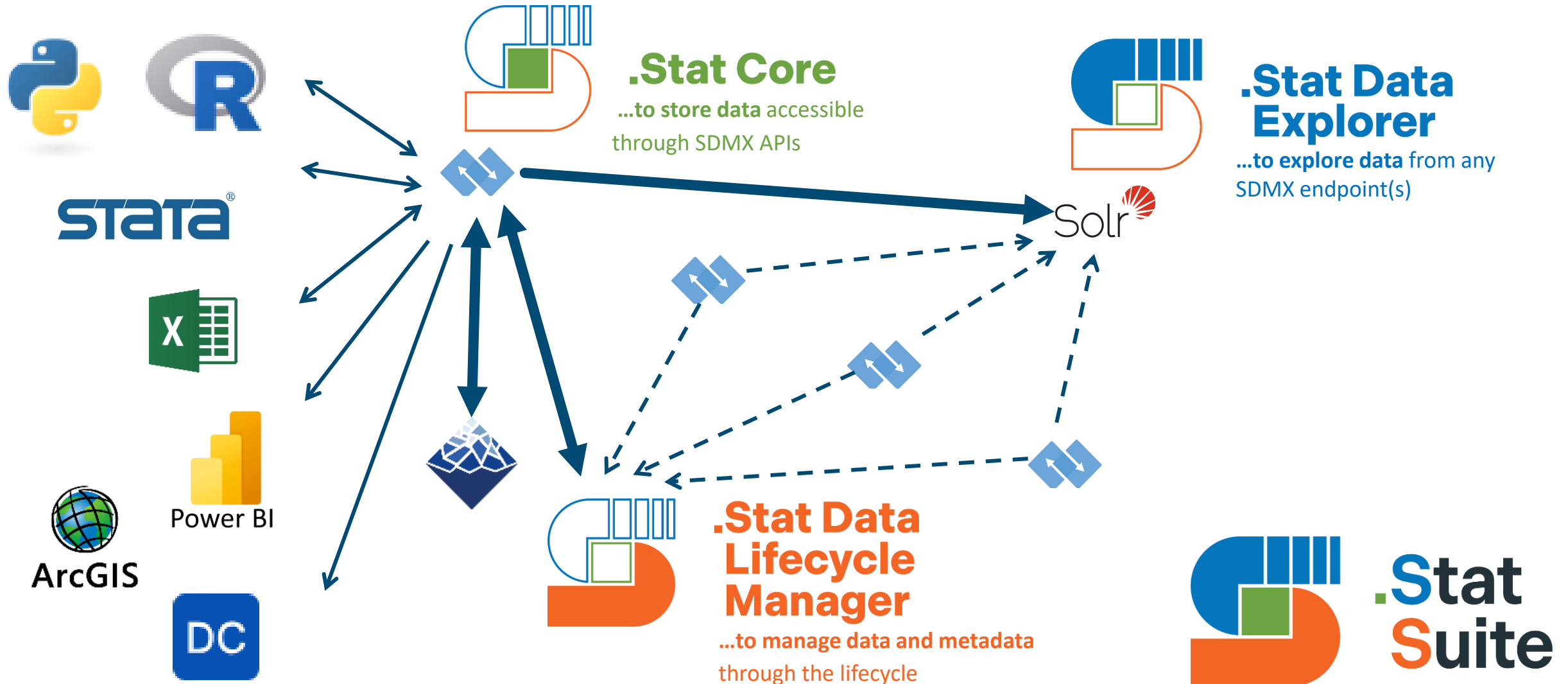


Members

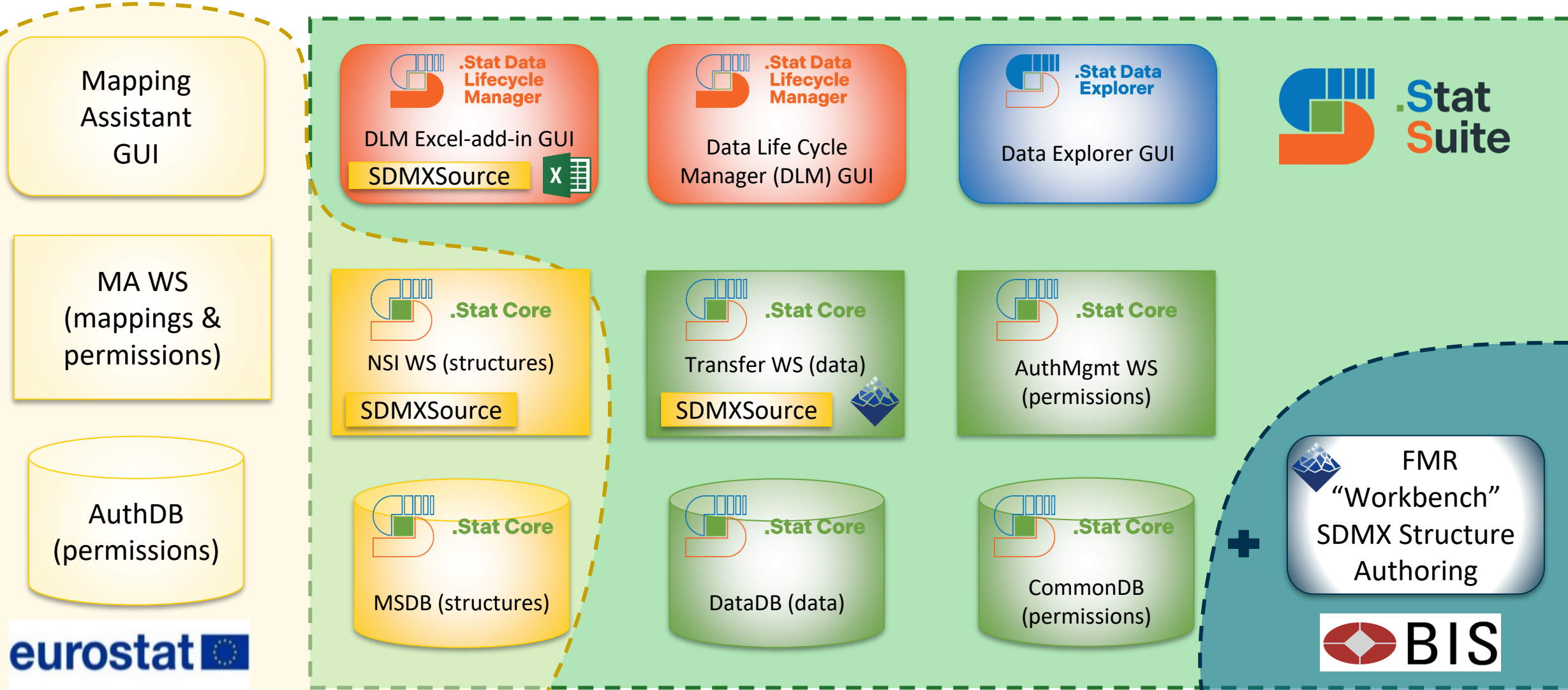


Partners

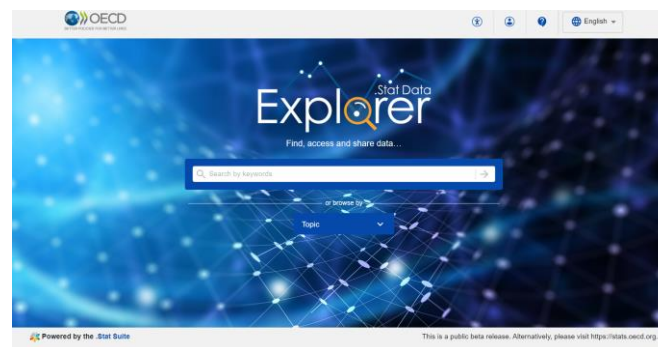
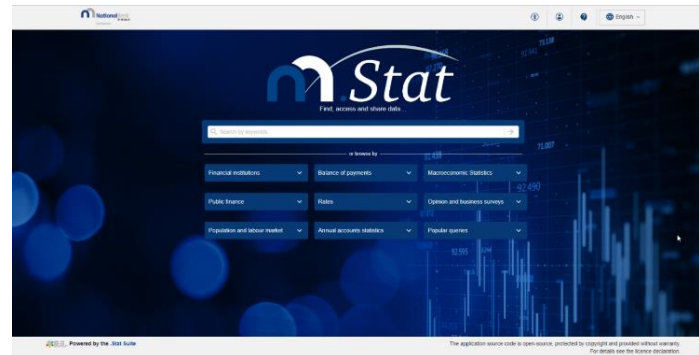
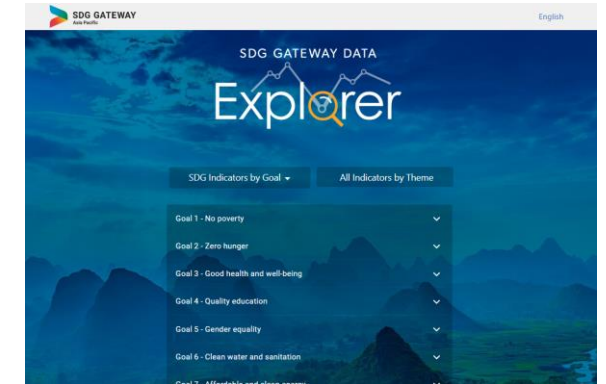
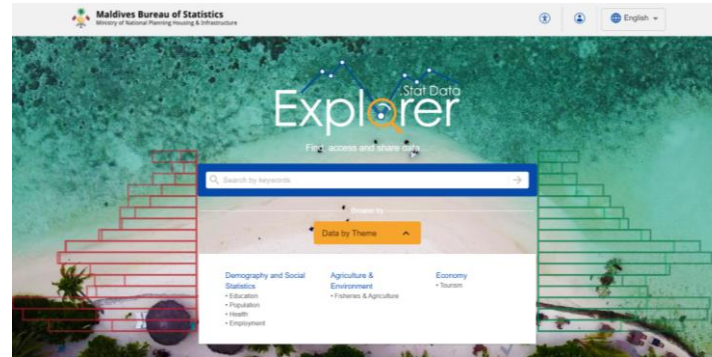
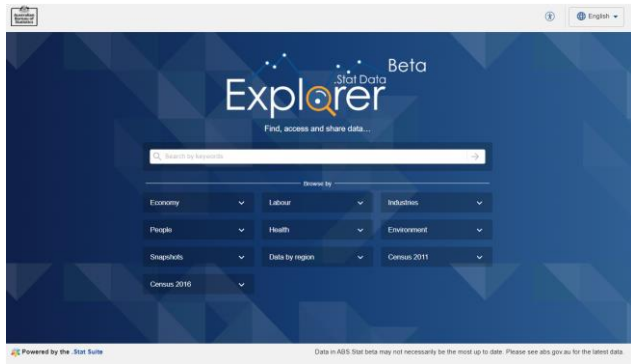
# #SDMX #OpenSource – The .Stat Suite



# #SDMX enable a cooperative “Business Model”



# Examples of implementations



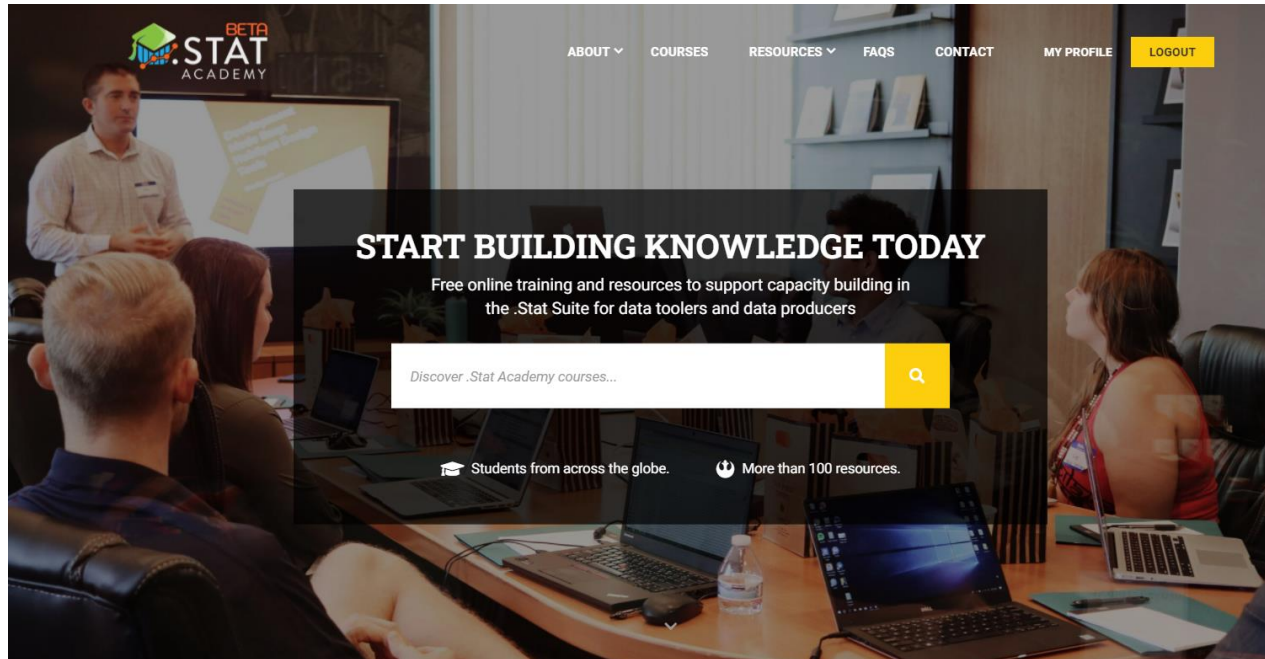




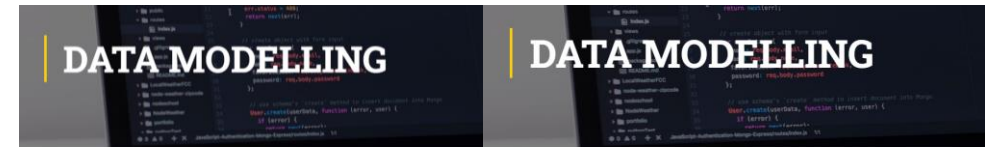
Search, visualise and share Pacific data...

**Pacific Data Hub .STAT**

# #SDMX #OpenKnowledge – The .Stat Academy



<https://academy.siscc.org>

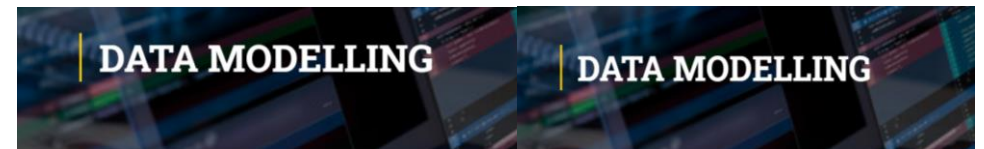


An introduction to SDMX structural mod  
producers

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

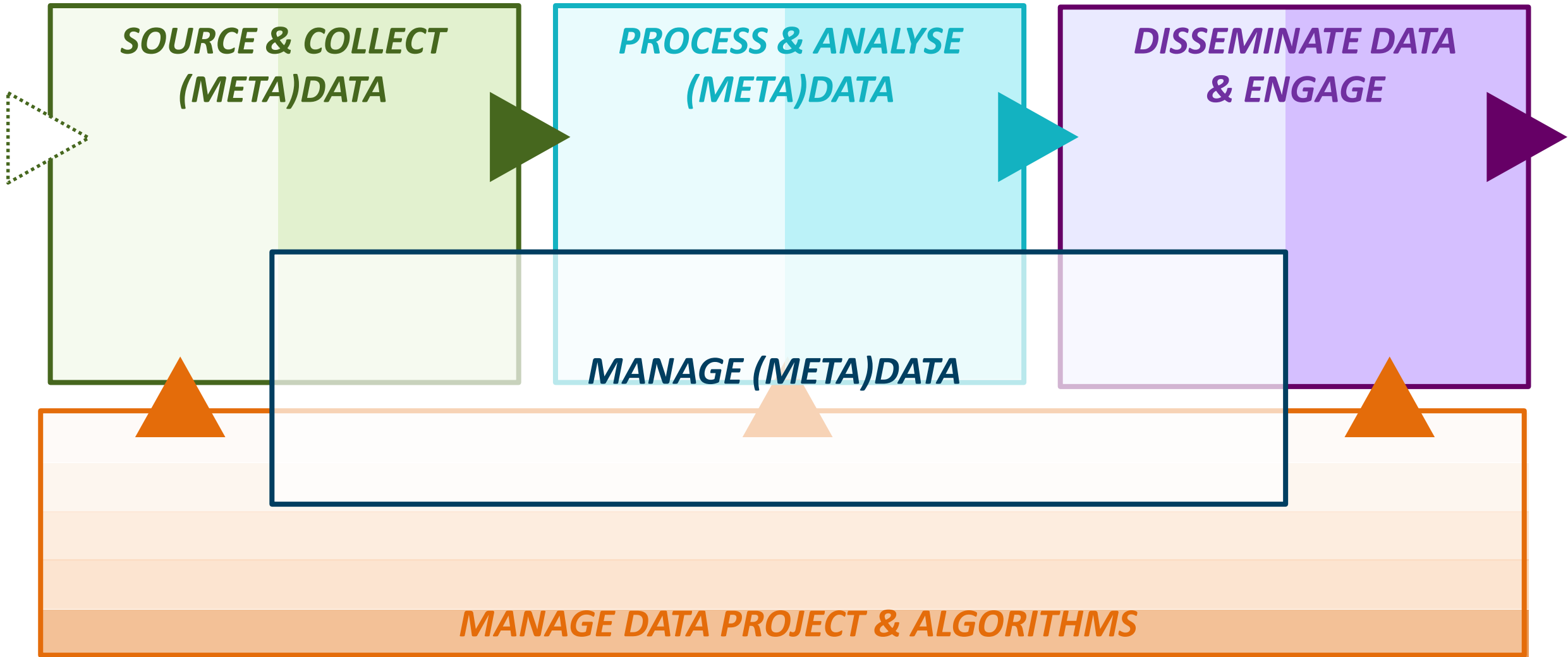
Instructor: Jens Dossé  
Categories: Data modelling, SDMX  
Review: ★★★★★

*Starting point: Build capacity in SDMX data modelling*

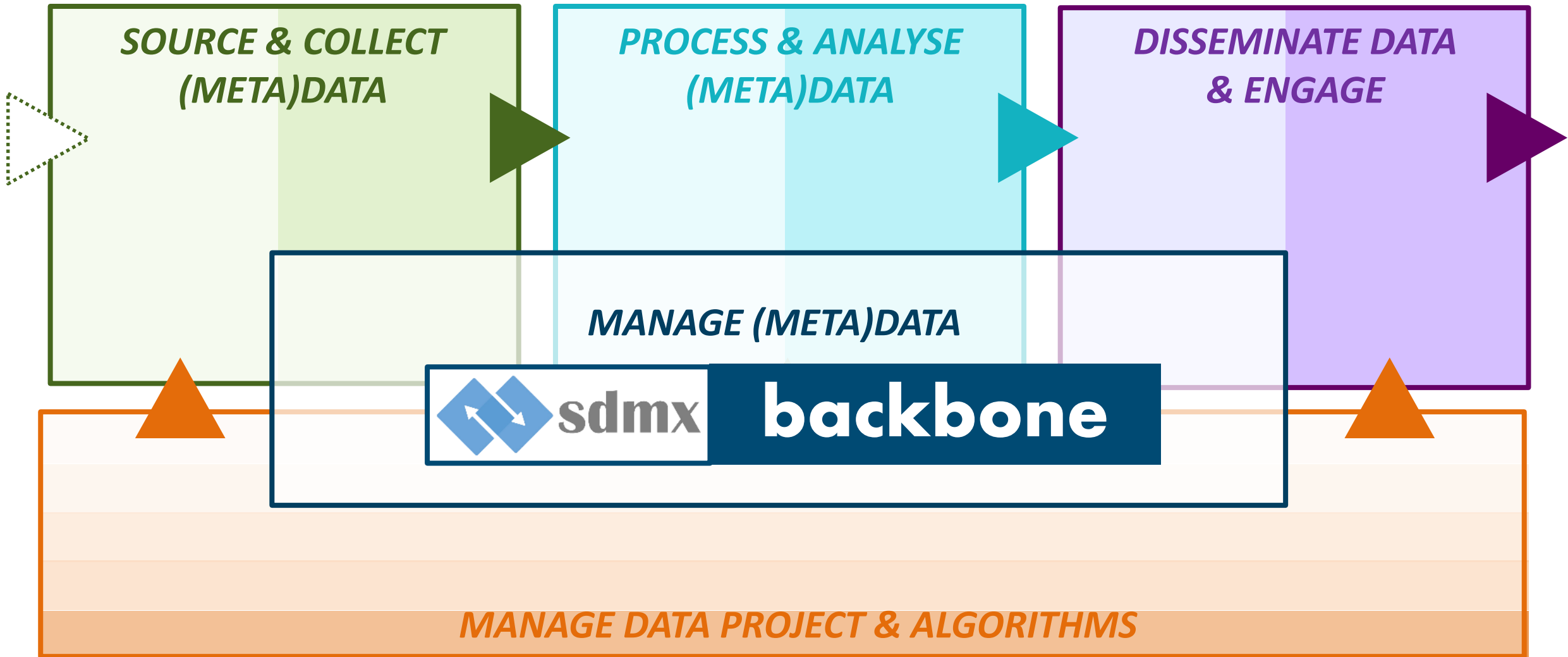
# THE #SDMX METADATA-DRIVEN APPROACH



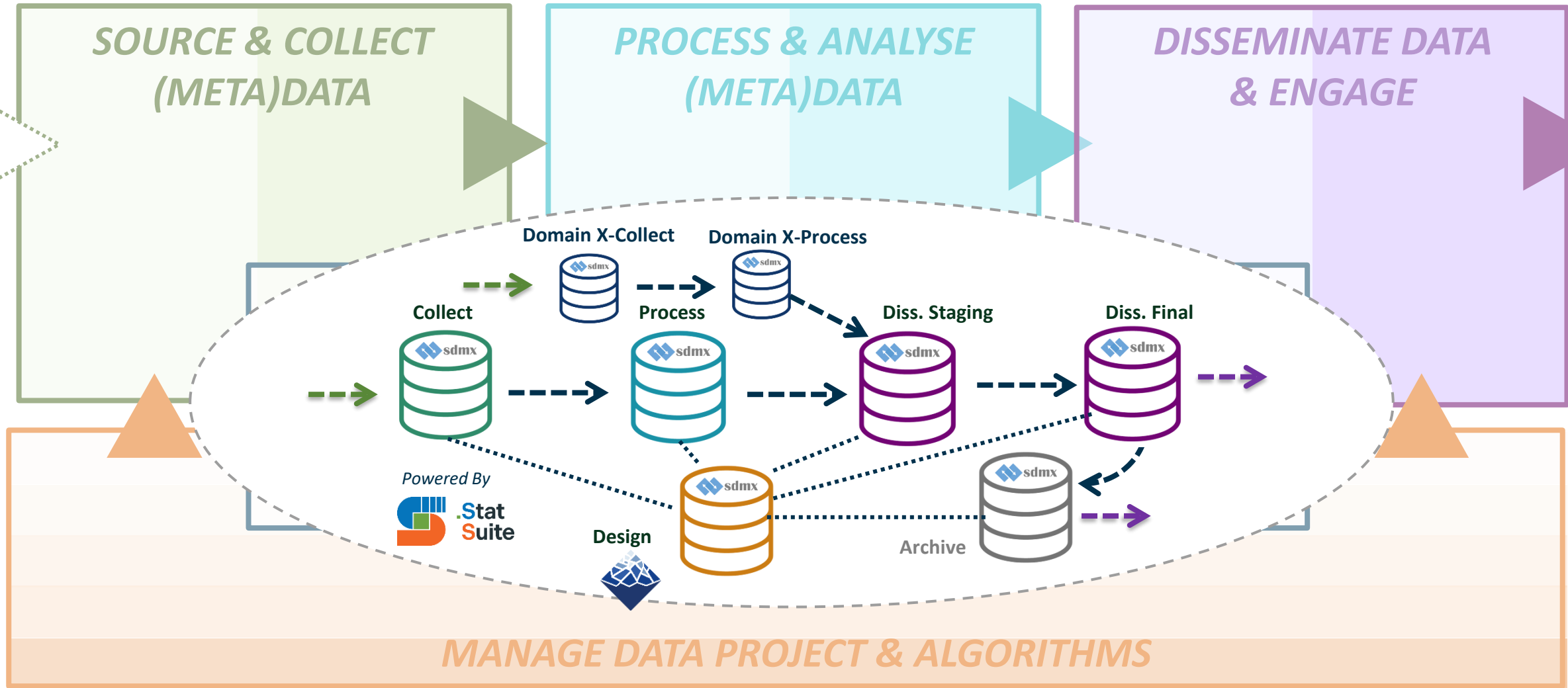
# The Data Platform



# The #SDMX Backbone

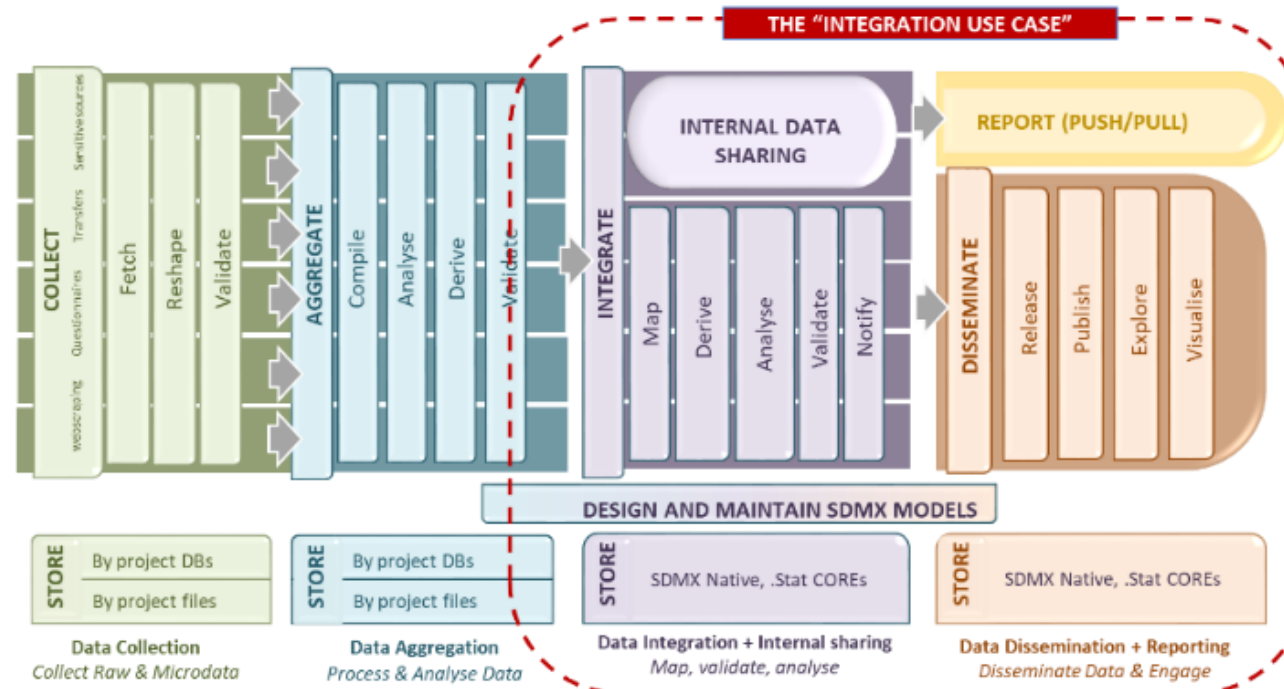


# The #SDMX Backbone




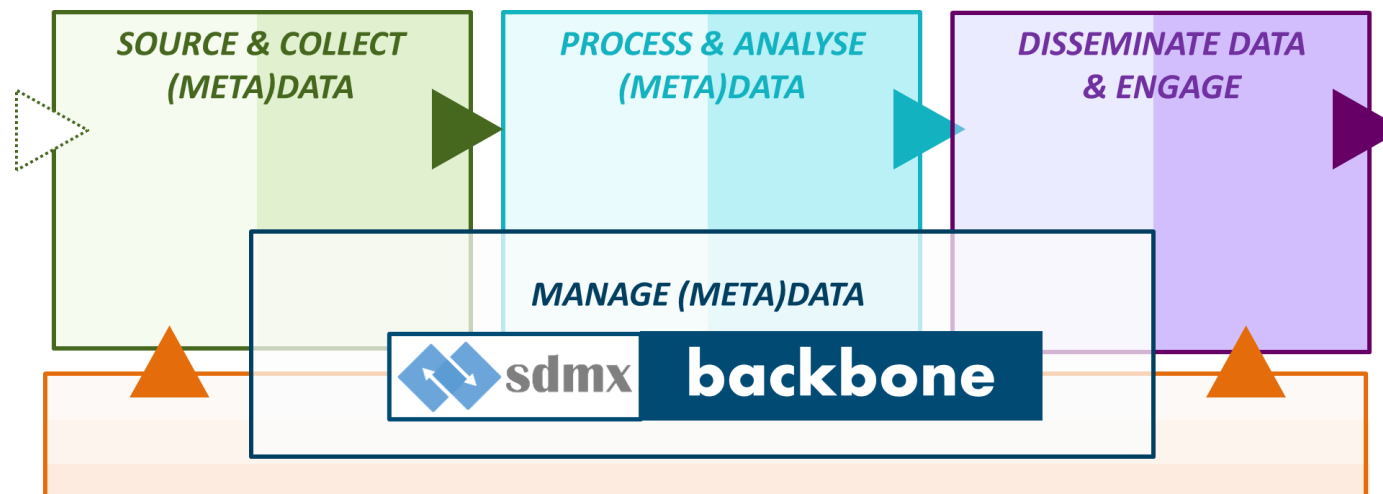
# The #SDMX Backbone

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



# The #SDMX Backbone – Wrap-up

-  **sdmx** data storage becomes a commodity
-  **sdmx** APIs enable universal data interoperability
-  **sdmx** semantics enables “metadata-driven” data user / manager experience



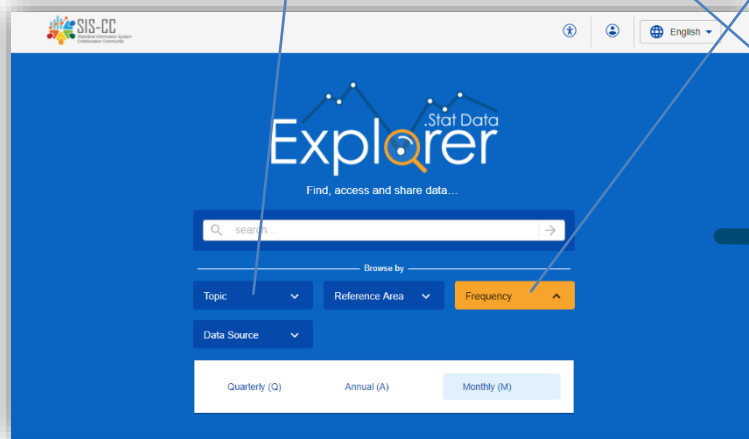


# #SDMX #MetadataDriven Data User Experience

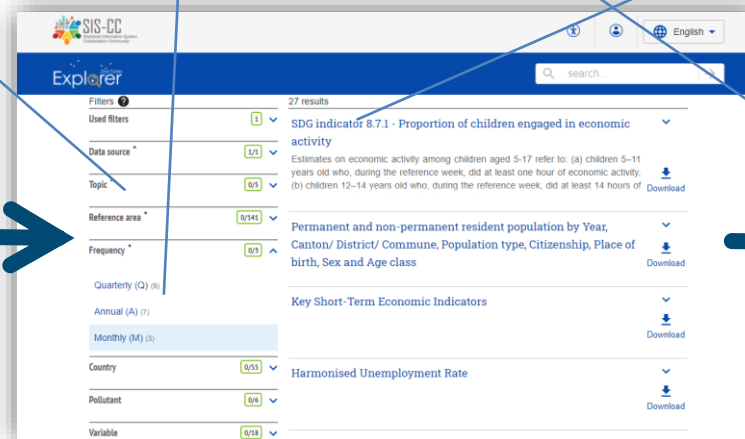
  
category scheme(s)

  
dimensions:  
concepts + codelists

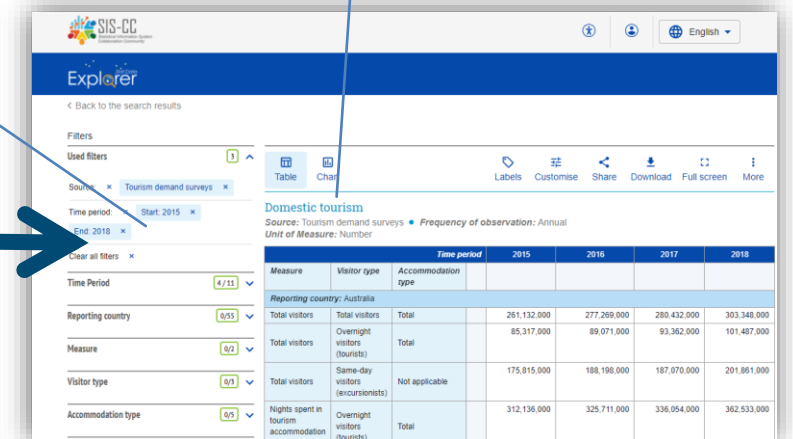
  
dataflow  
+ presentation annotations



**Step 1 Home page**  
Lexical search or Topics

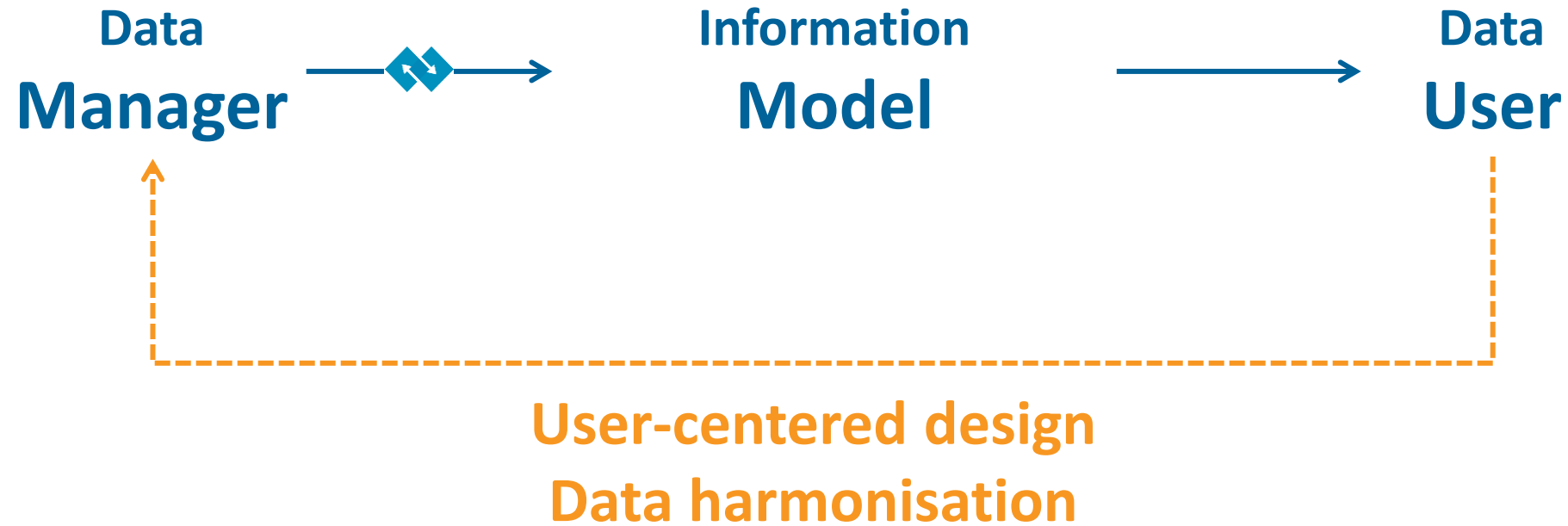


**Step 2 Search results**  
Filtering on dynamic facets



**Step 3 Visualisation page**  
Selecting dimensions / variables

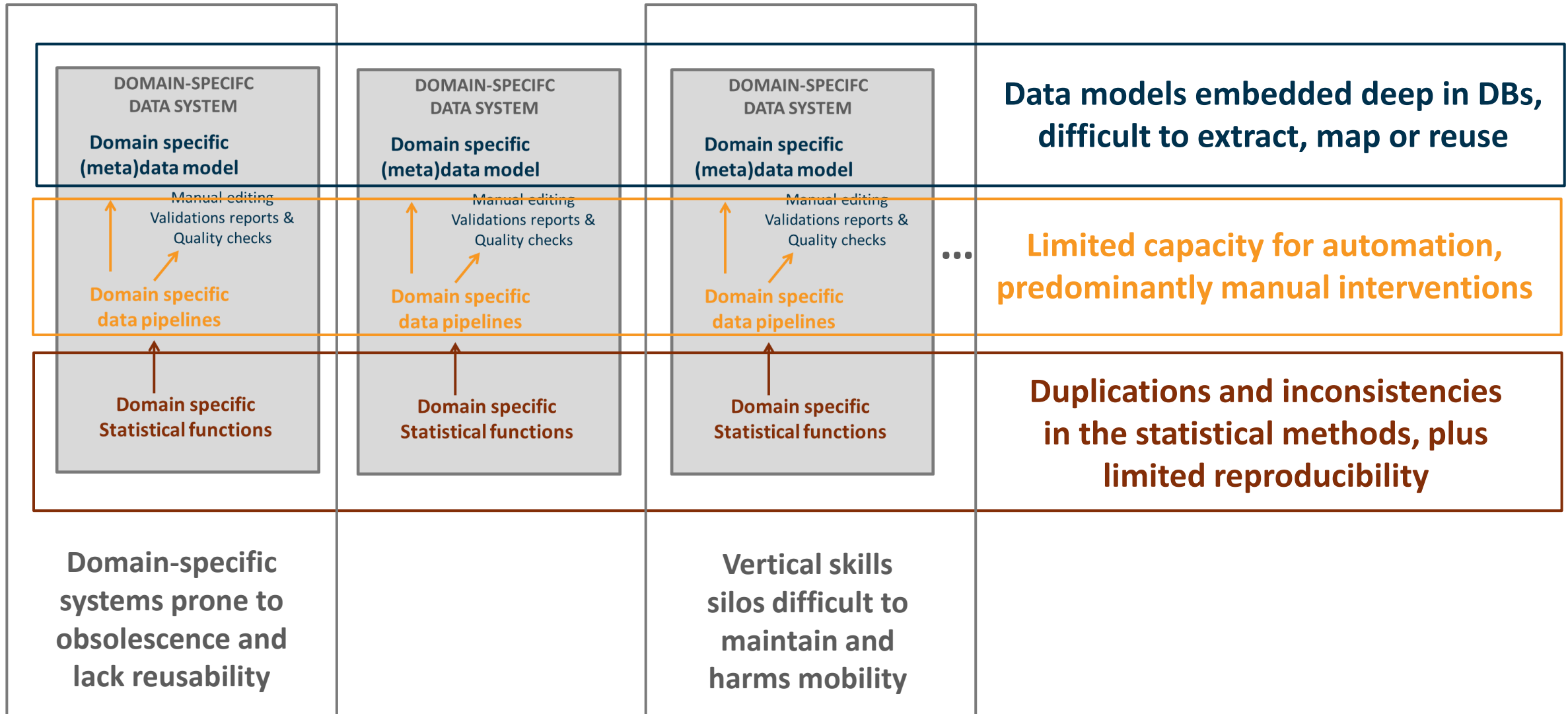
# #SDMX #MetadataDriven Data User Experience



# Establish a systematic User Research



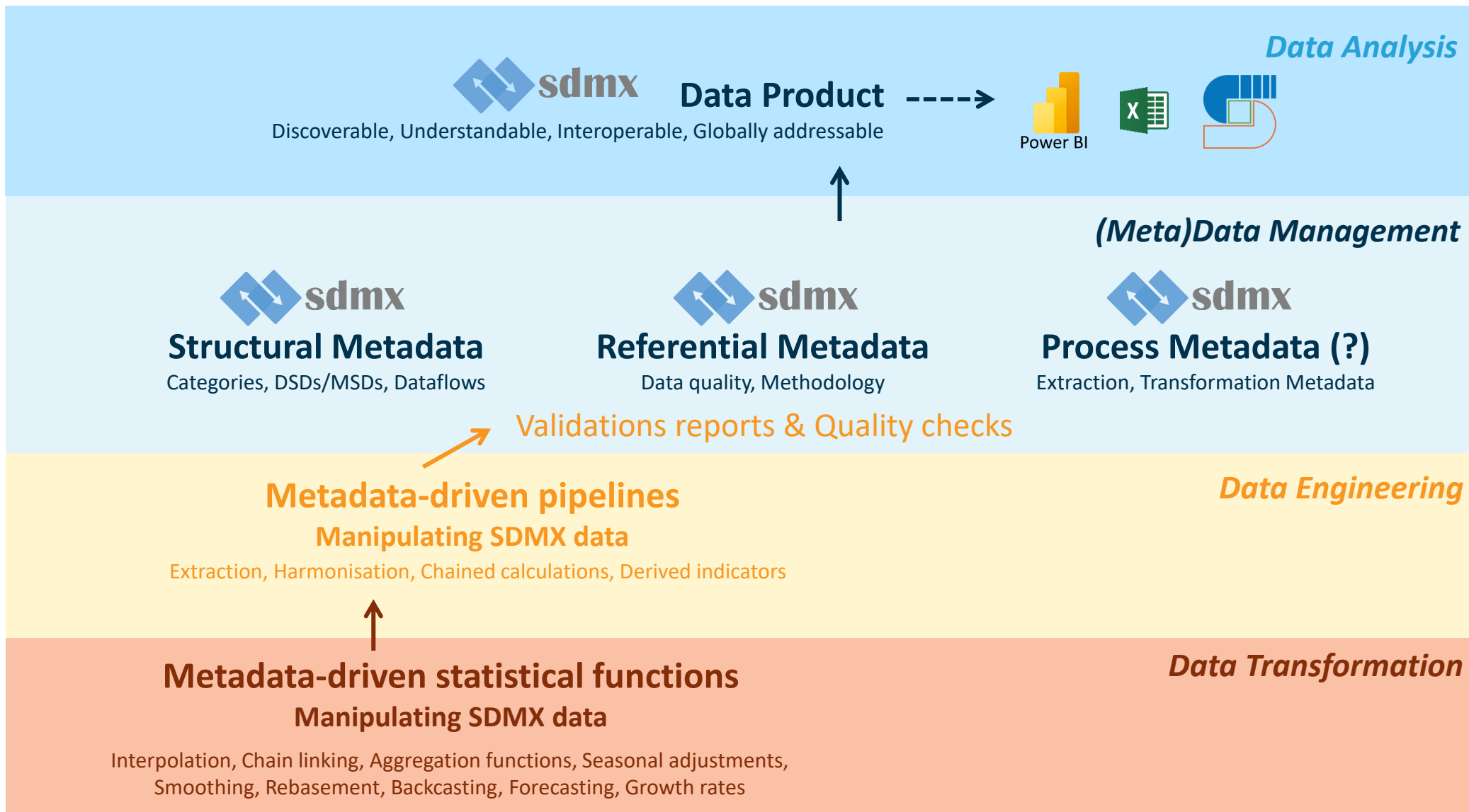
# The Data Platform is fragmented



# #SDMX #MetadataDriven Data Manager Experience

MANAGE DATA  
& METADATA

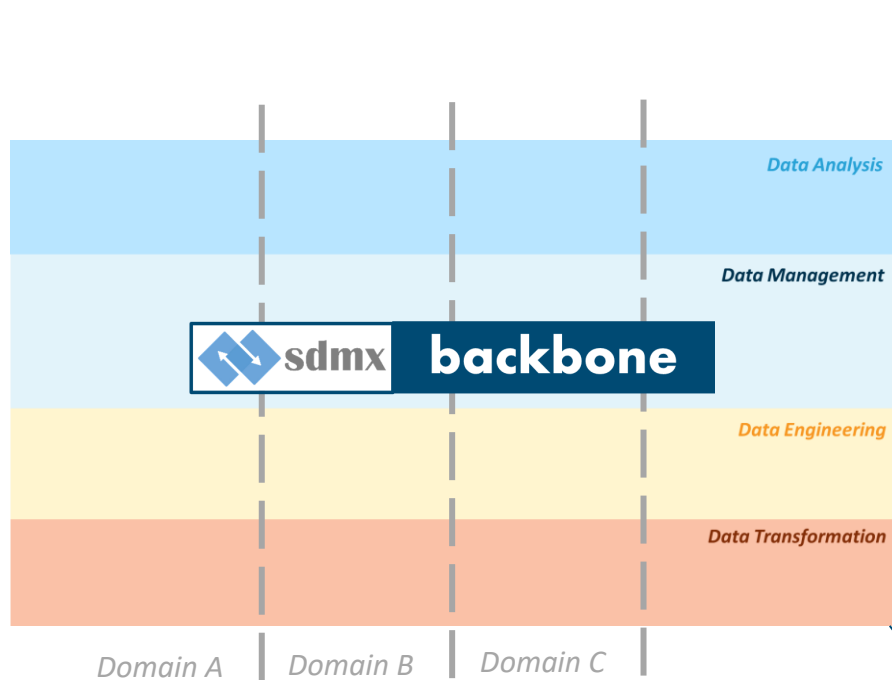
MANAGE DATA PROJECT  
& ALGORITHMS



(?) 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 multiple benefits:*



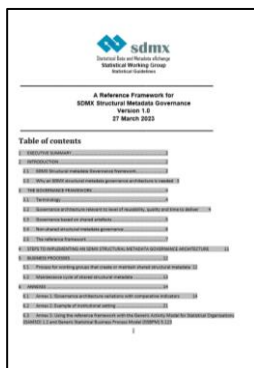
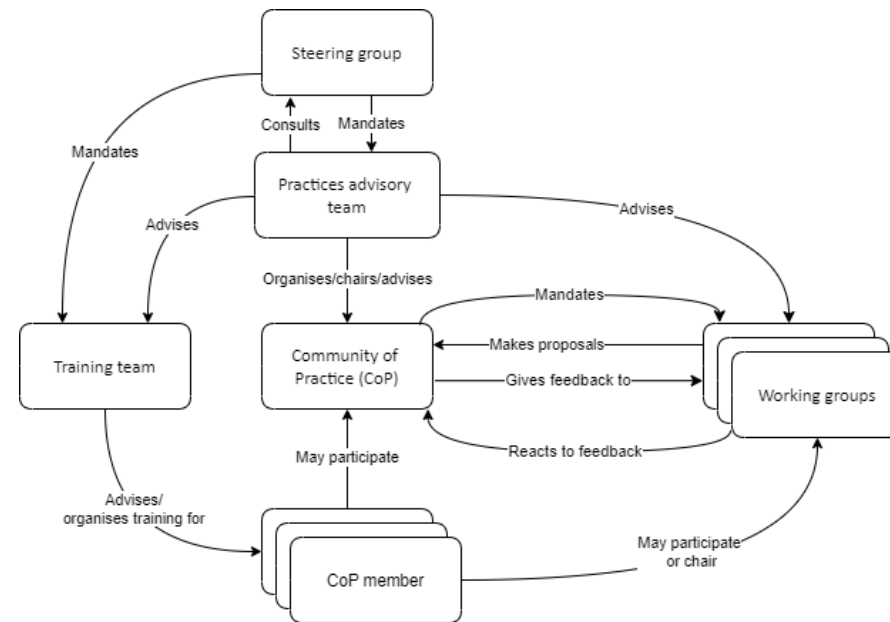
- Decoupled roles and specializations
- Data mgt by non coding-experts
- Facilitation of staff mobility
- Possibility of data tasks mutualization
- Facilitation of platform migration
- Domain specific coding language
- Domain specific interfaces
- Capitalize on shared methods / reuse
- Capitalize on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



## Structure your Data governance

# #SDMX to structure your Data Governance

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





## **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**

**Drive**  
**Data User experience**  
**Data Manager experience**  
**With SDMX metadata**

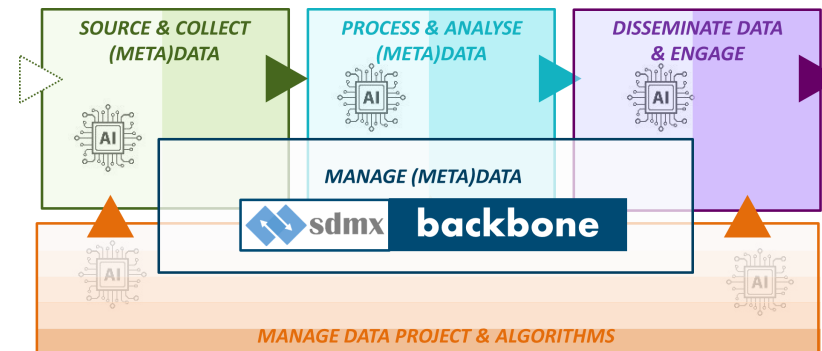
**#SDMX #AI**  
**SDMX AS AN AI ENABLER?**



Nowcasting  
Federated Learning



Outlier detection  
Clustering  
Data Curation

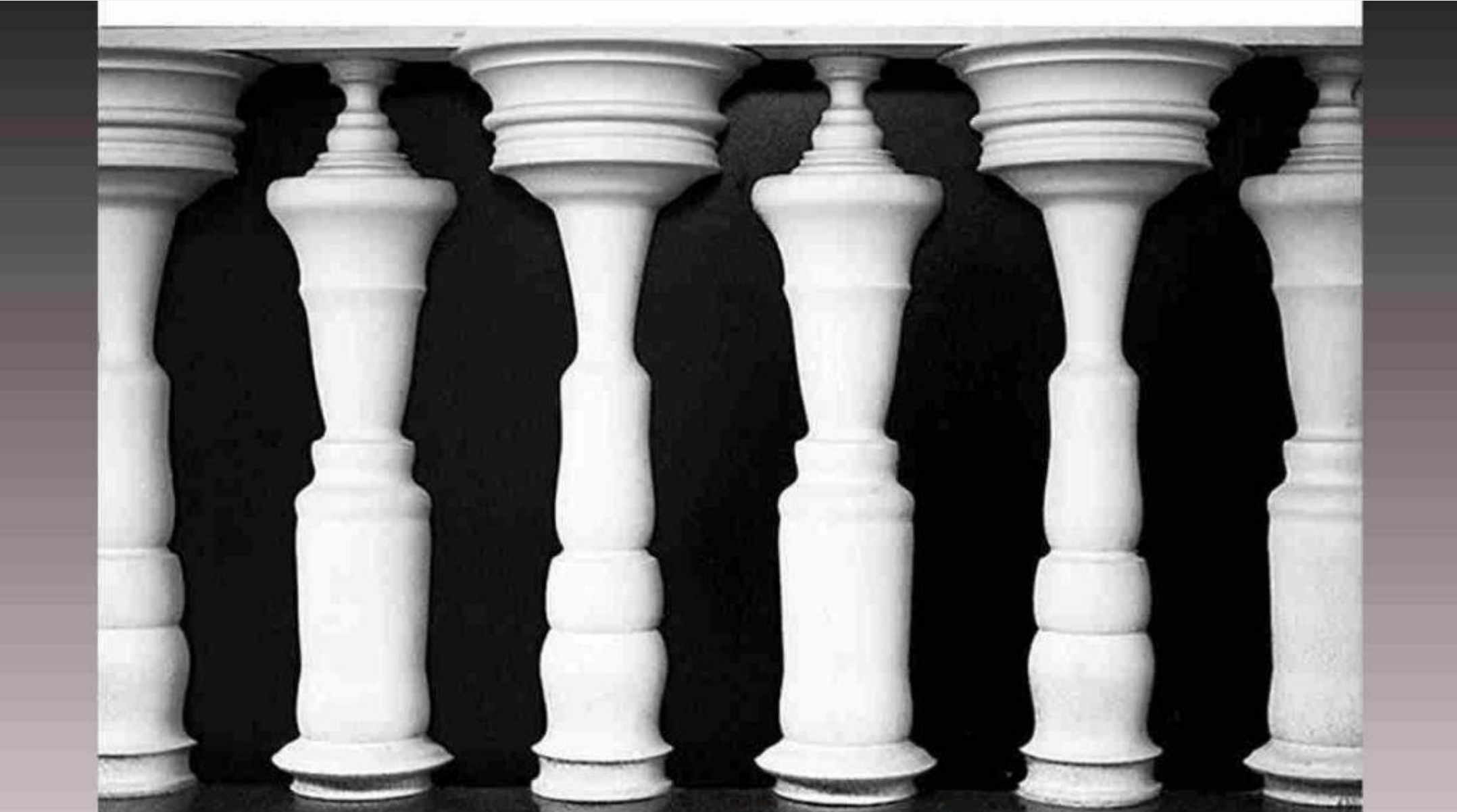


Natural Language  
Conversational  
Data Access



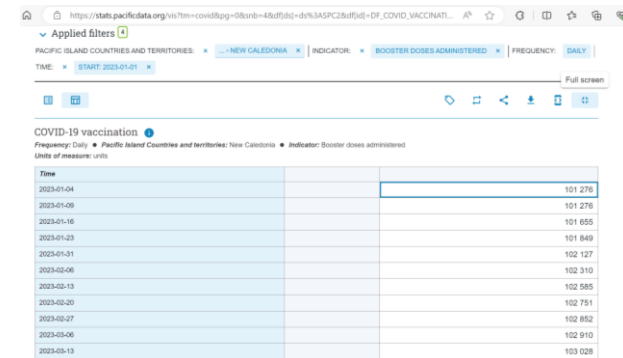
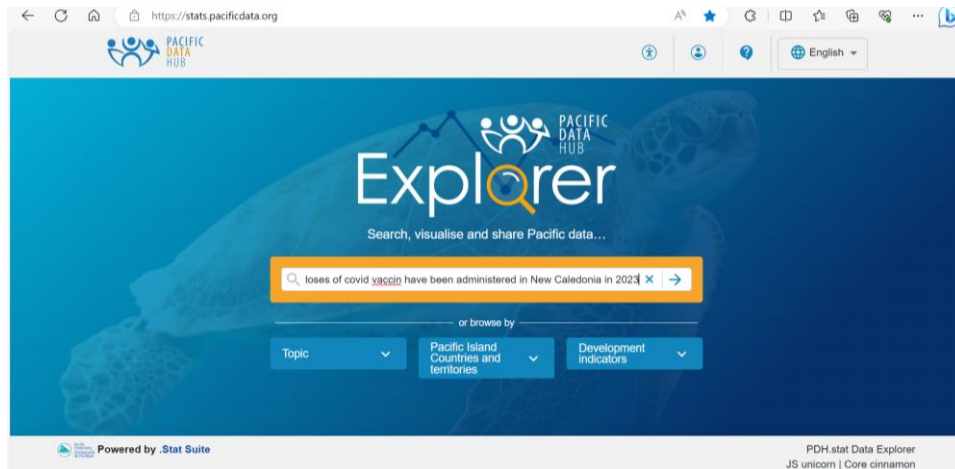
Pair programming  
Pair modelling

# The cognitive gap



# Develop a natural language capable Data Explorer (I)

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



The screenshot shows the Pacific Data Explorer interface with the search query "loses of covid vaccin have been administered in New Caledonia in 2023" applied. The interface displays a data table for "COVID-19 vaccination" with the following columns: "Time", "Pacific Island Countries and territories: New Caledonia", and "Indicator: Booster doses administered". The table shows the number of booster doses administered in New Caledonia for various dates in 2023.

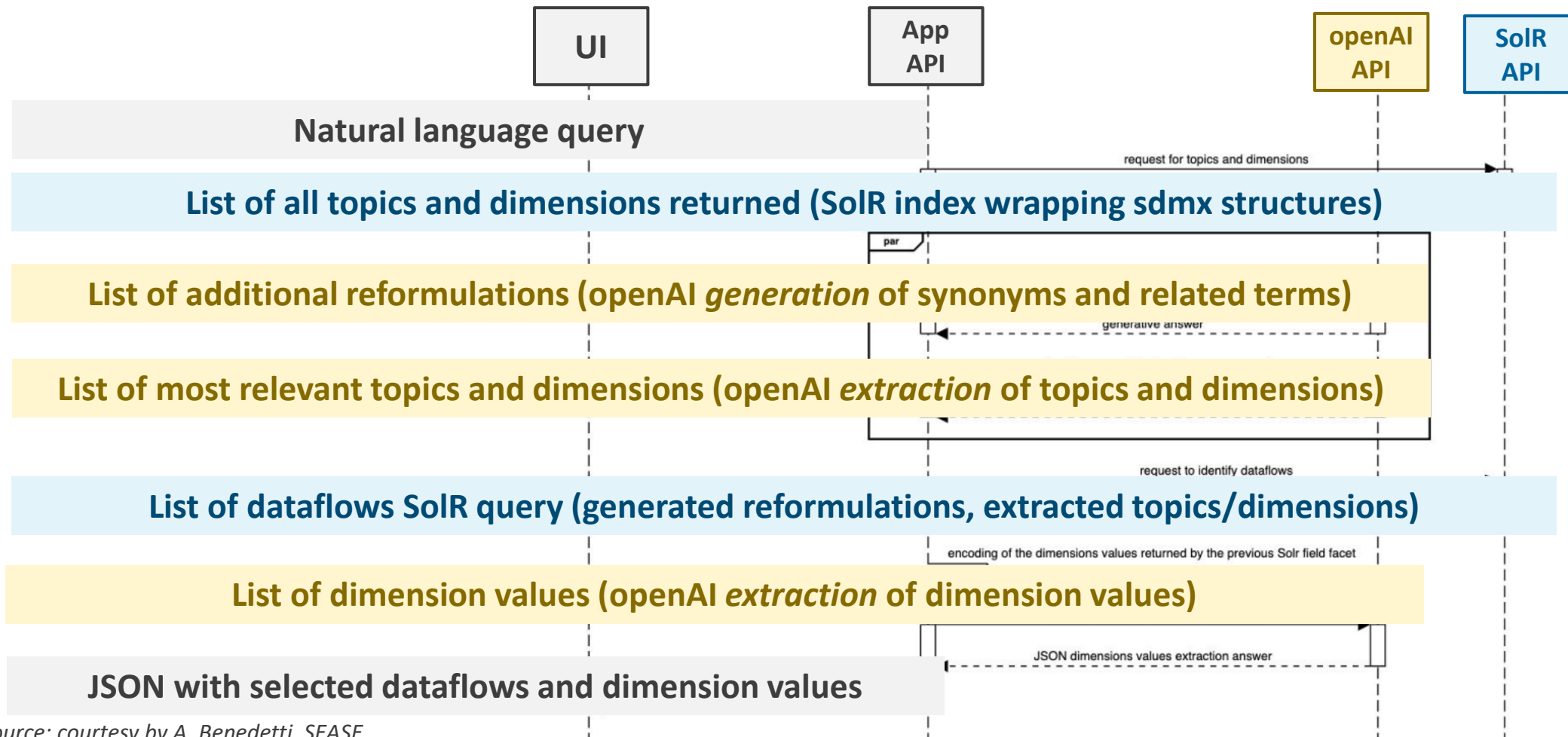
Time	Pacific Island Countries and territories: New Caledonia	Indicator: Booster doses administered
2023-01-04		101 276
2023-01-09		101 276
2023-01-16		101 655
2023-01-23		101 849
2023-01-31		102 127
2023-02-06		102 310
2023-02-13		102 595
2023-02-20		102 751
2023-02-27		102 852
2023-03-06		102 910
2023-03-13		103 028

## Step 2.5 or 3.5

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

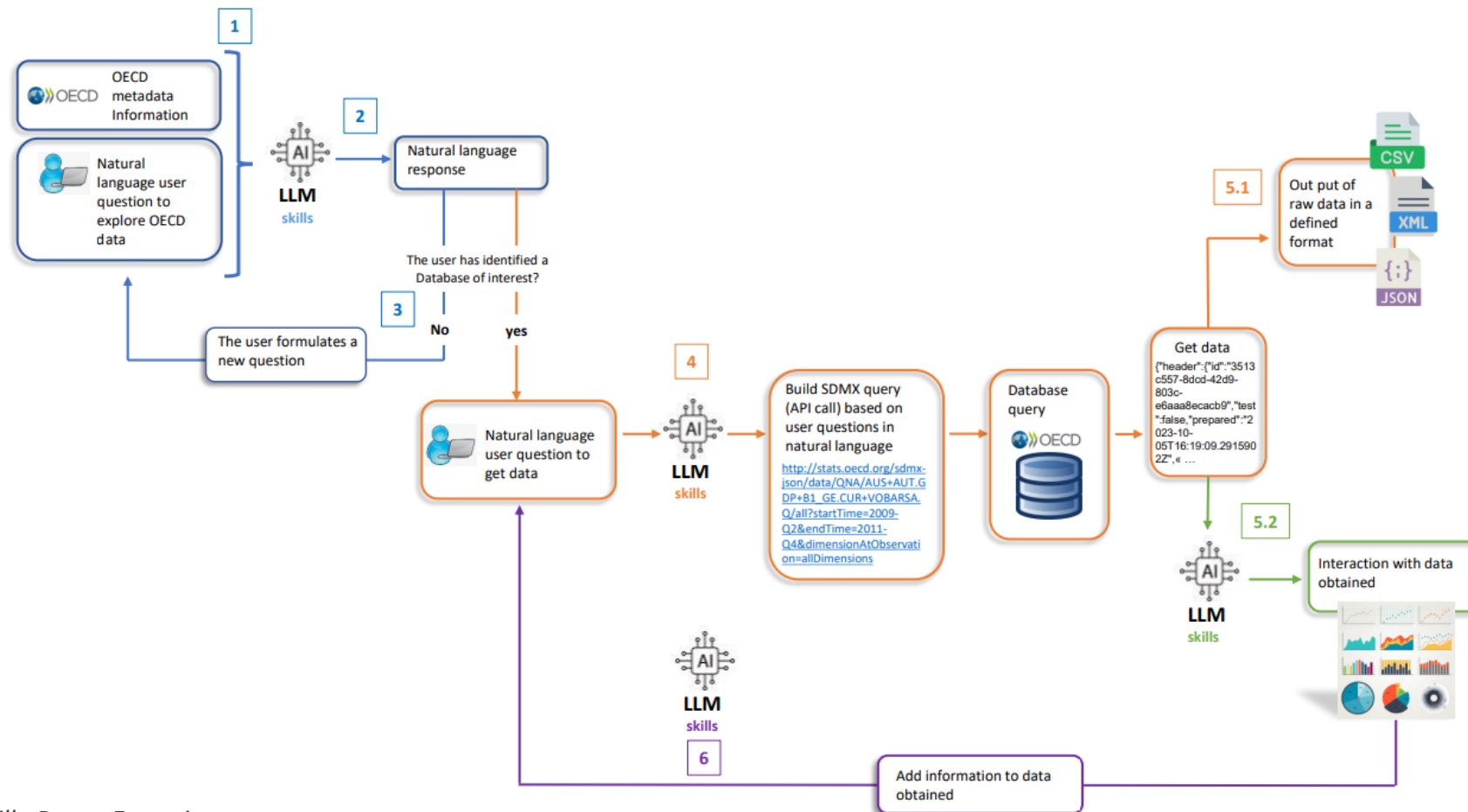
# Develop a natural language capable Data Explorer (I)

*Proof of Concept ongoing, mixing Generative AI & Solr techniques.*

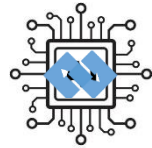
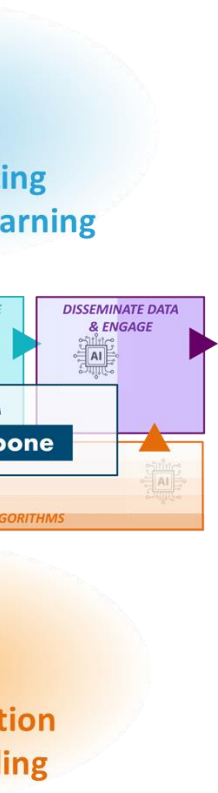


# The universal StatsBot to access any SDMX source (II)

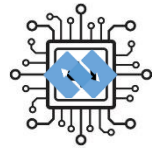
*The StatsBot is within reach... we could develop it as an open and reusable asset.*



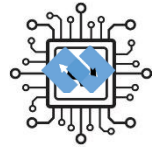
# #SDMX #AI Co-investment and co-innovation opportunities



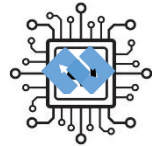
***Develop a Natural language capable Data Explorer (I)***



***Develop the universal StatsBot to access any SDMX source (II)***



***Develop a “prompt engineering in an SDMX context” practice***



***Develop the specialized, “SDMX-literate” open source LLM***



# #SDMX #DATAMESH SDMX AS A DATA MESH ENABLER?

O'REILLY®

## Data Mesh

Delivering Data-Driven Value at Scale

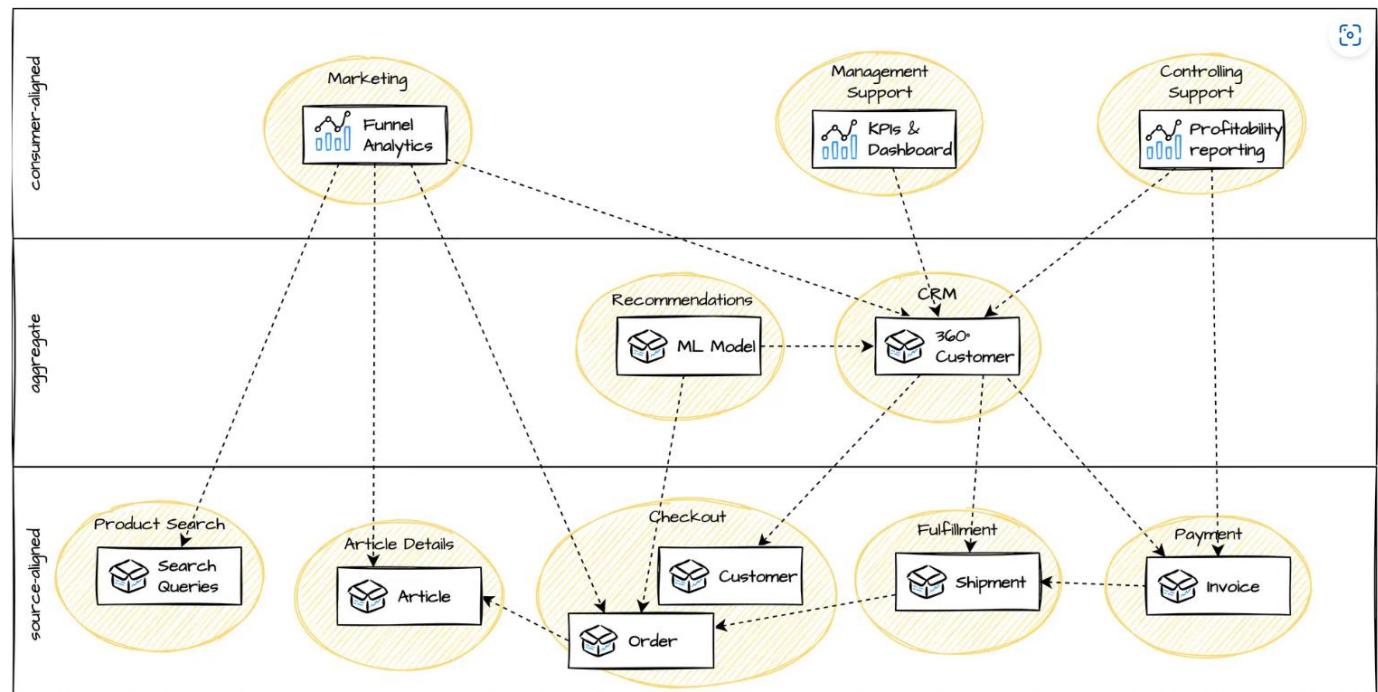


Zhamak Dehghani

# 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.**

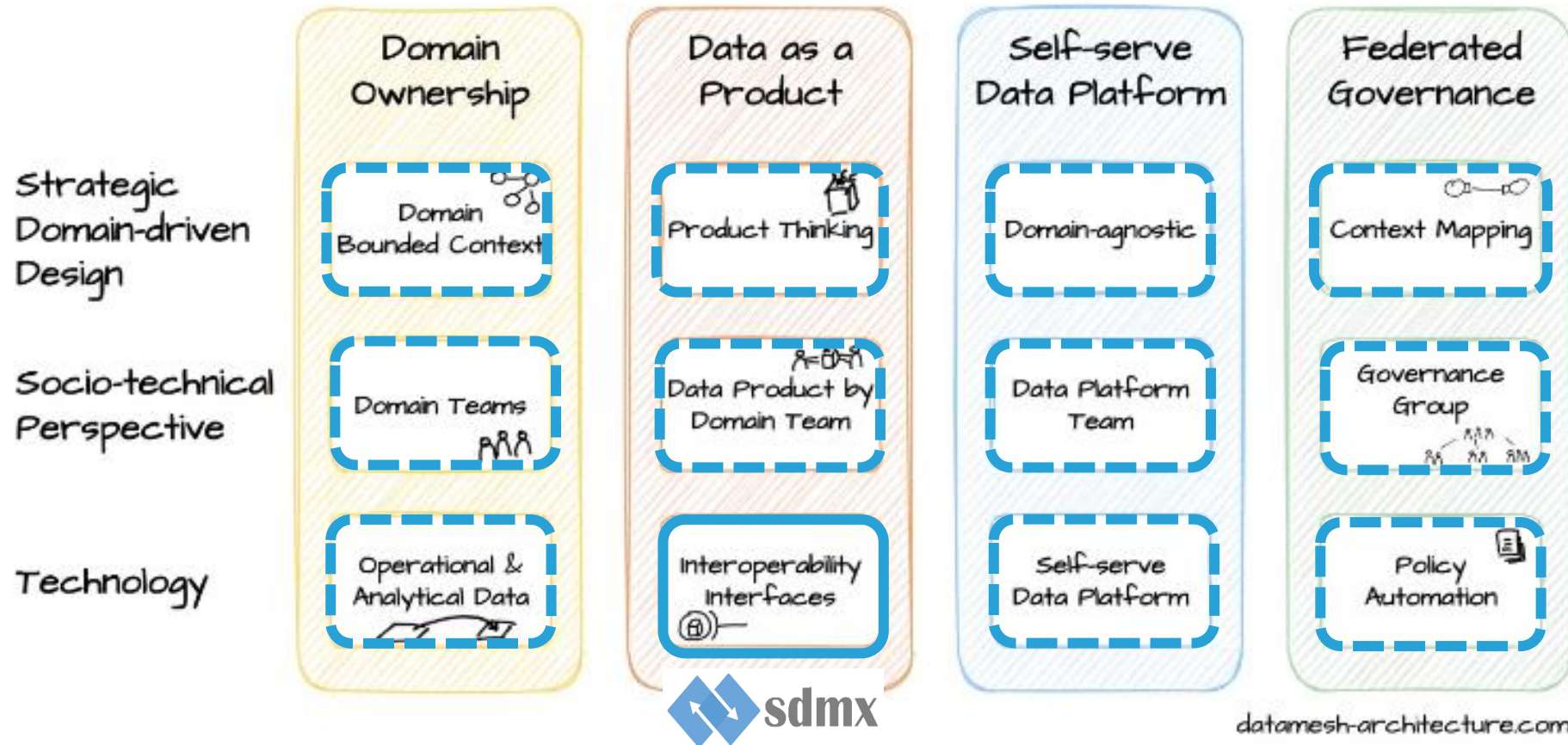


## #DataMesh requires common data semantics

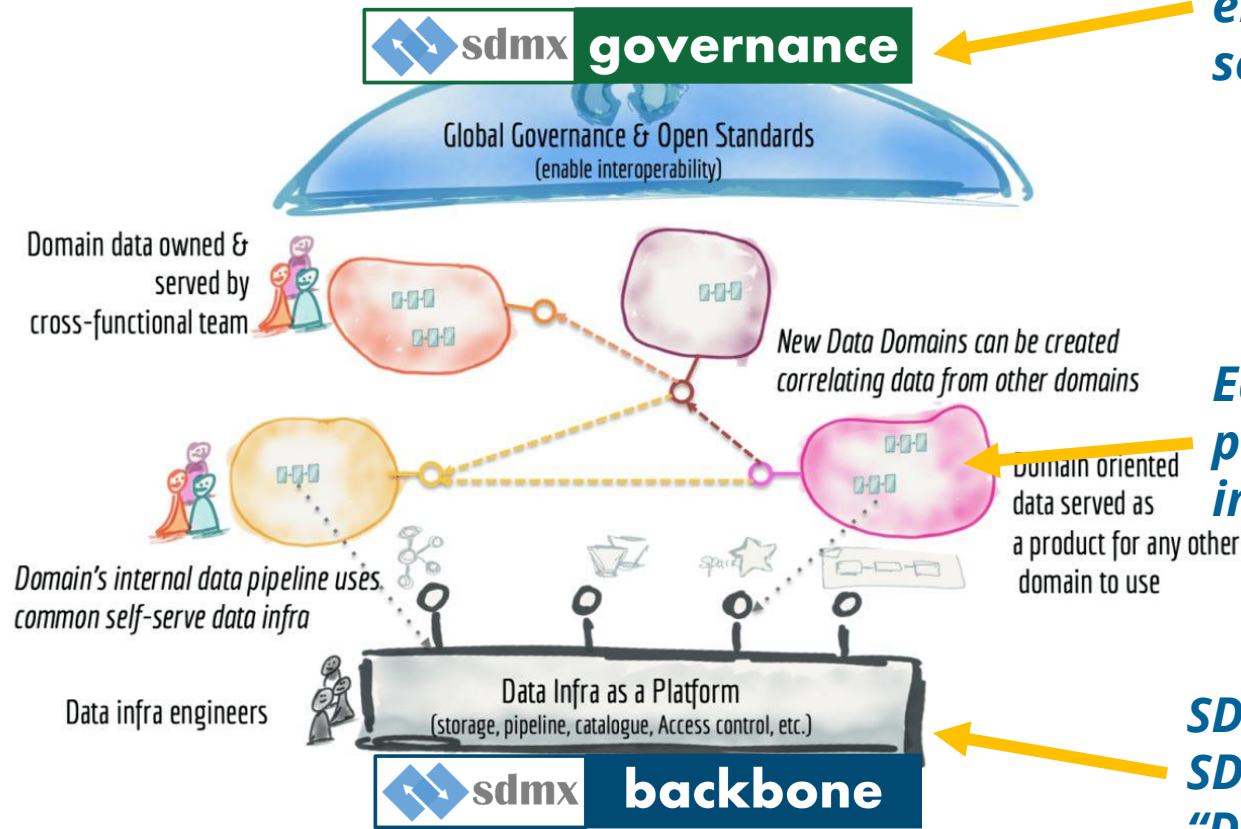
*“Data mesh attempts to strike a balance between team autonomy and interteam interoperability and collaboration. **It gives domain teams autonomy to take control of their local decision making, for example, choosing the best data model for their data products, while it uses computational governance policies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



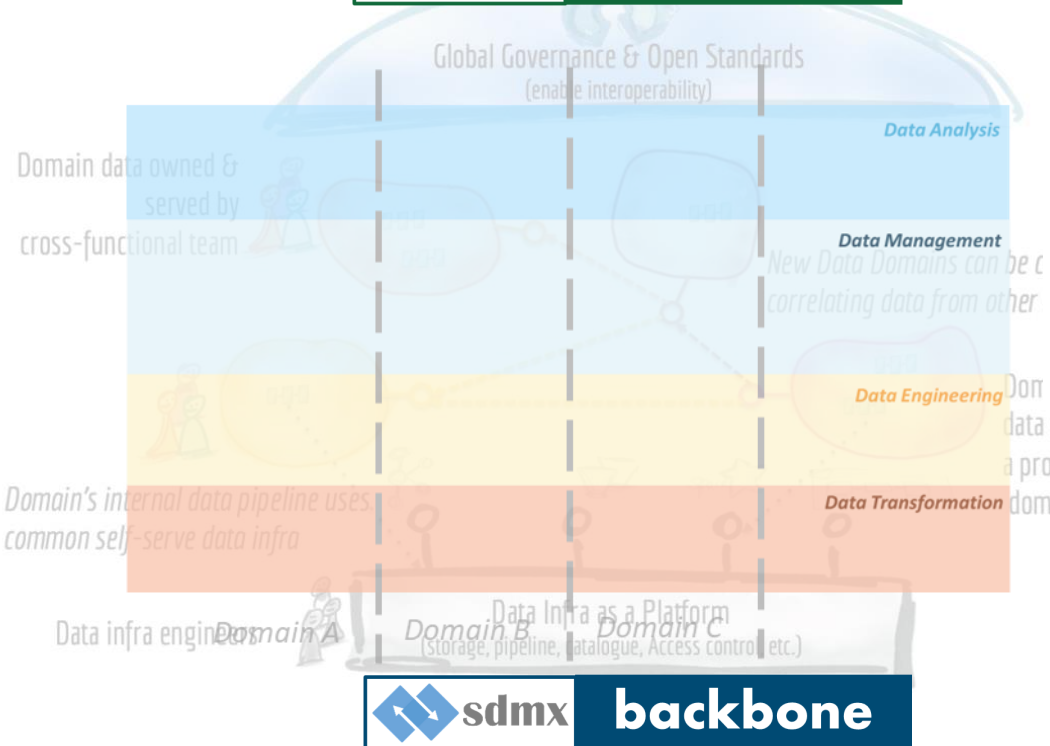
*SDMX governance as cornerstone of data governance, enabling a progressive harmonisation of data semantics across domains and the data lifecycle.*

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

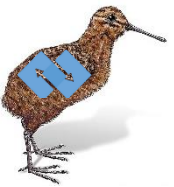
*SDMX backbone brings commoditised SDMX storage, SDMX API-centricity, and SDMX semantics as part of the "Data infrastructure as a Platform".*

# #SDMX #AI Co-investment and co-innovation opportunities

## sdmx governance

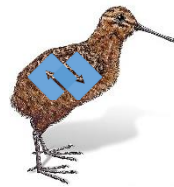


**Develop the architecture vision & business case for #DataMesh in Official Statistics**



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

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



**Develop implementation guidelines for the #SDMX #DataMesh & Experiment**

## 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. The #SDMX #MetadataDriven approach

 **sdmx governance**

 **sdmx backbone**



## 1. Creating value with #SDMX – A #Community journey

**co-investment**

**Co-innovation**