

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?*

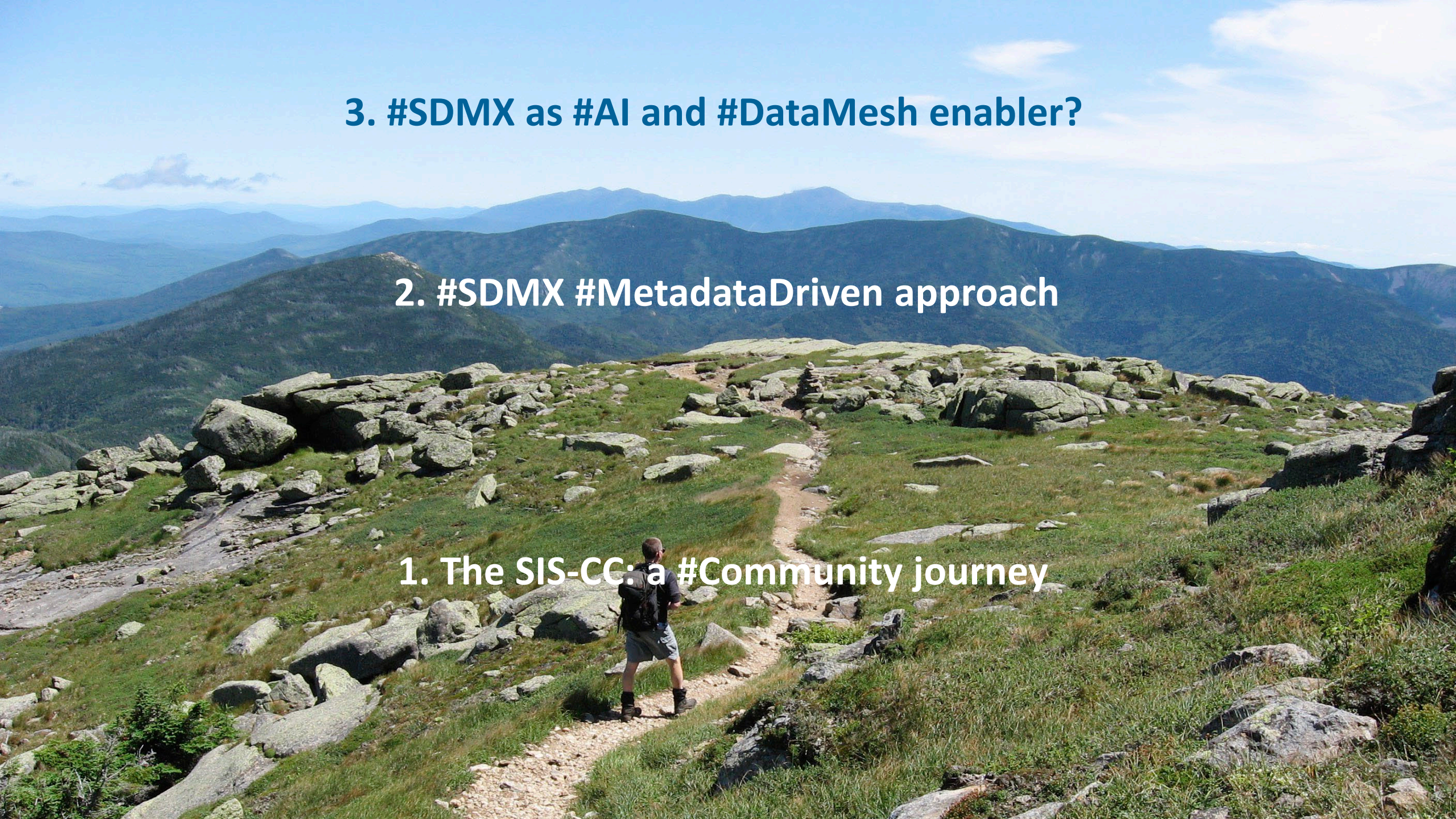


Source: [Why SDMX Matters](https://siscc.org), <https://siscc.org>

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.**

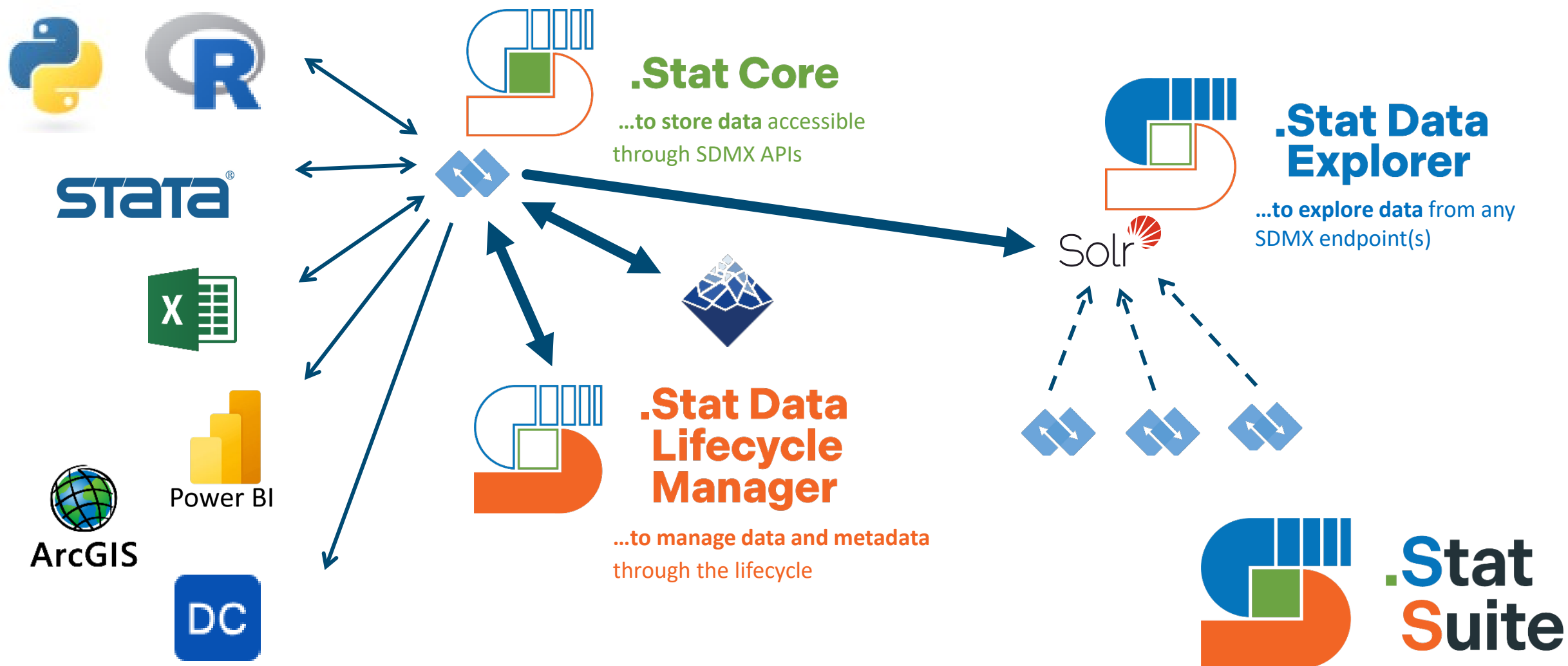


Members

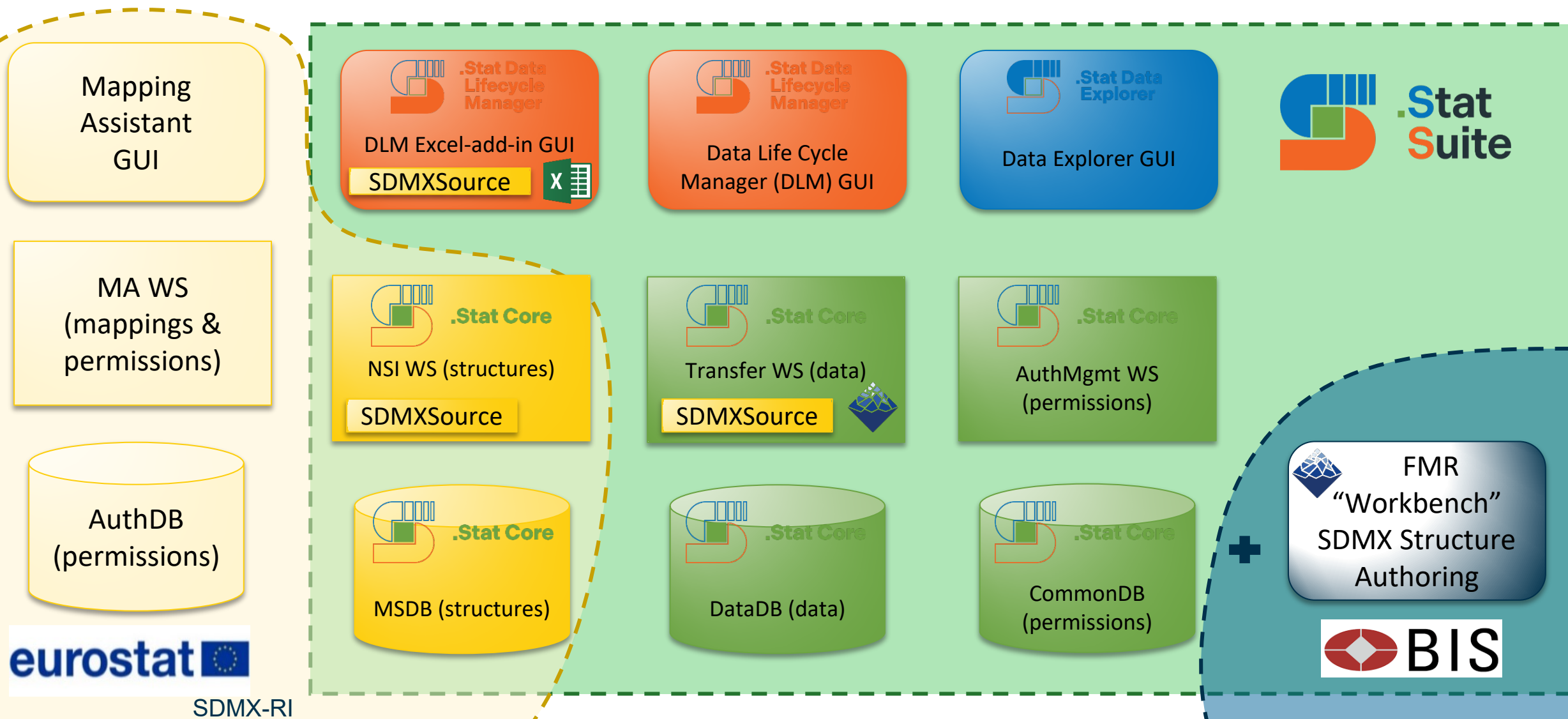


Partners

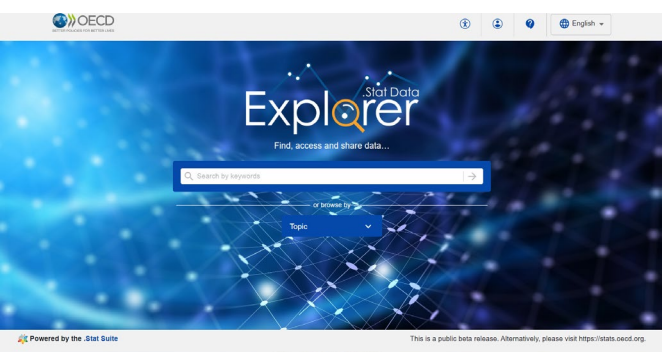
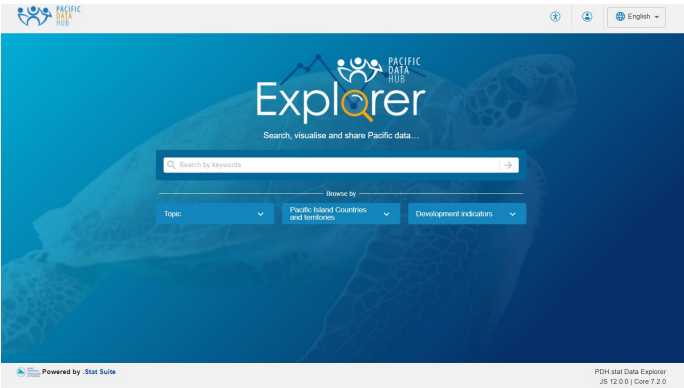
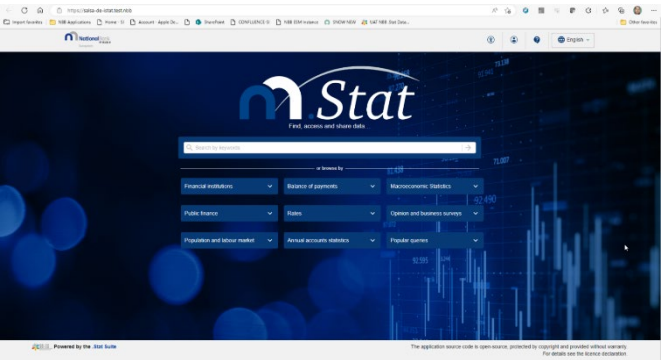
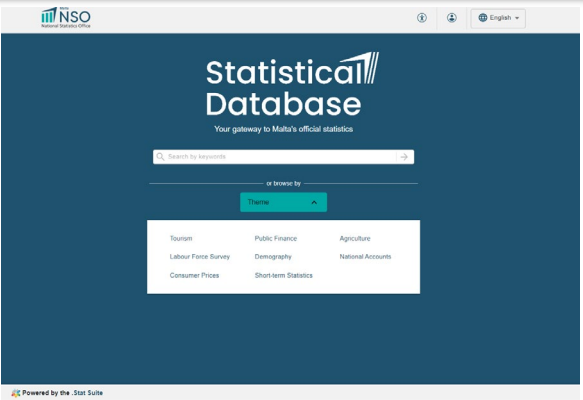
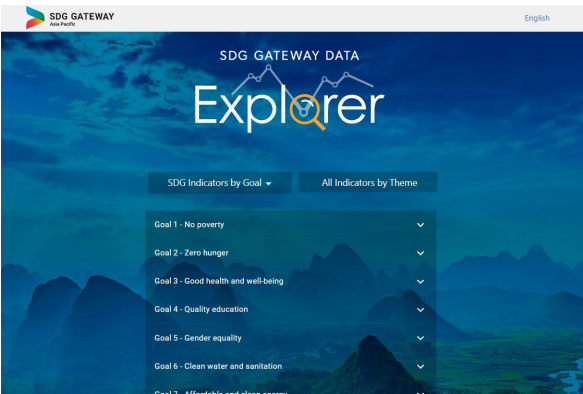
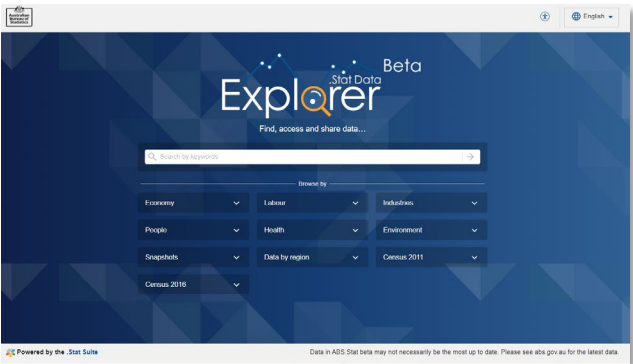
The #OpenSource .Stat Suite



Component-based Architecture → Cooperative “Business Model”



Examples of implementations

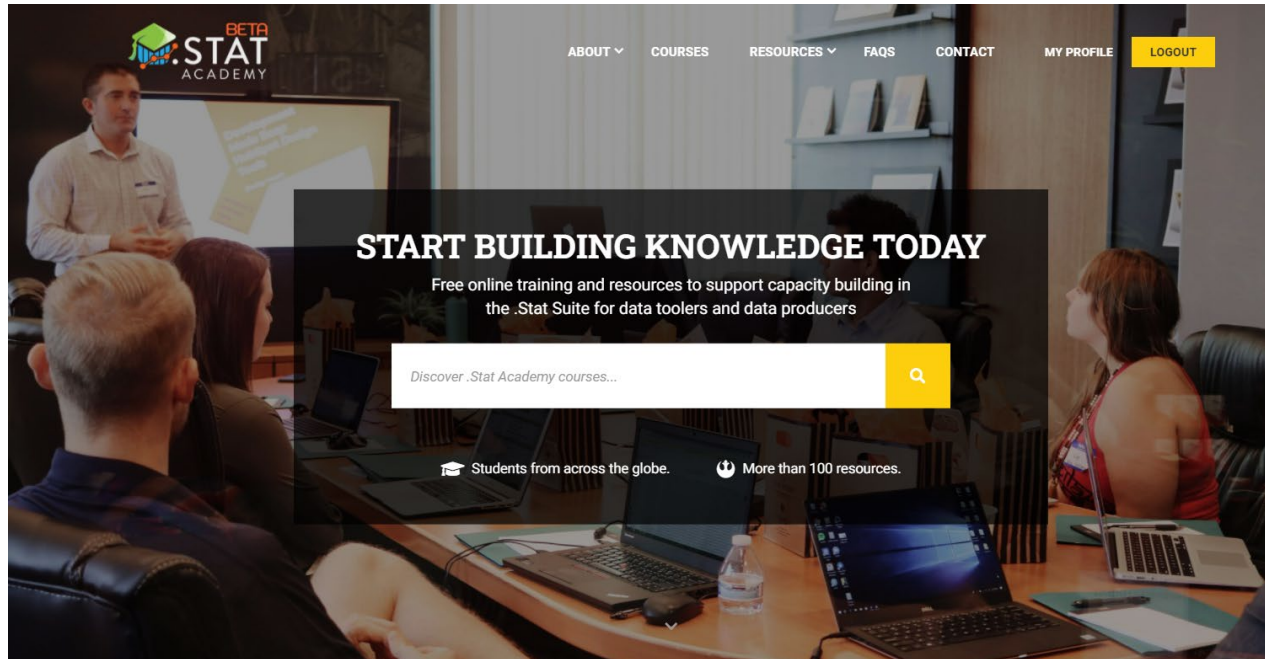




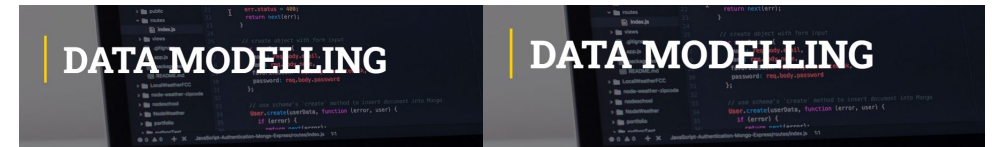
Search, visualise and share Pacific data...

Pacific Data Hub .STAT

The #OpenKnowledge .Stat Academy



<https://academy.siscc.org>

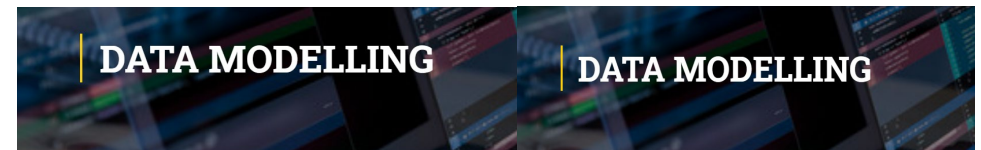


An introduction to SDMX structural model producers

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

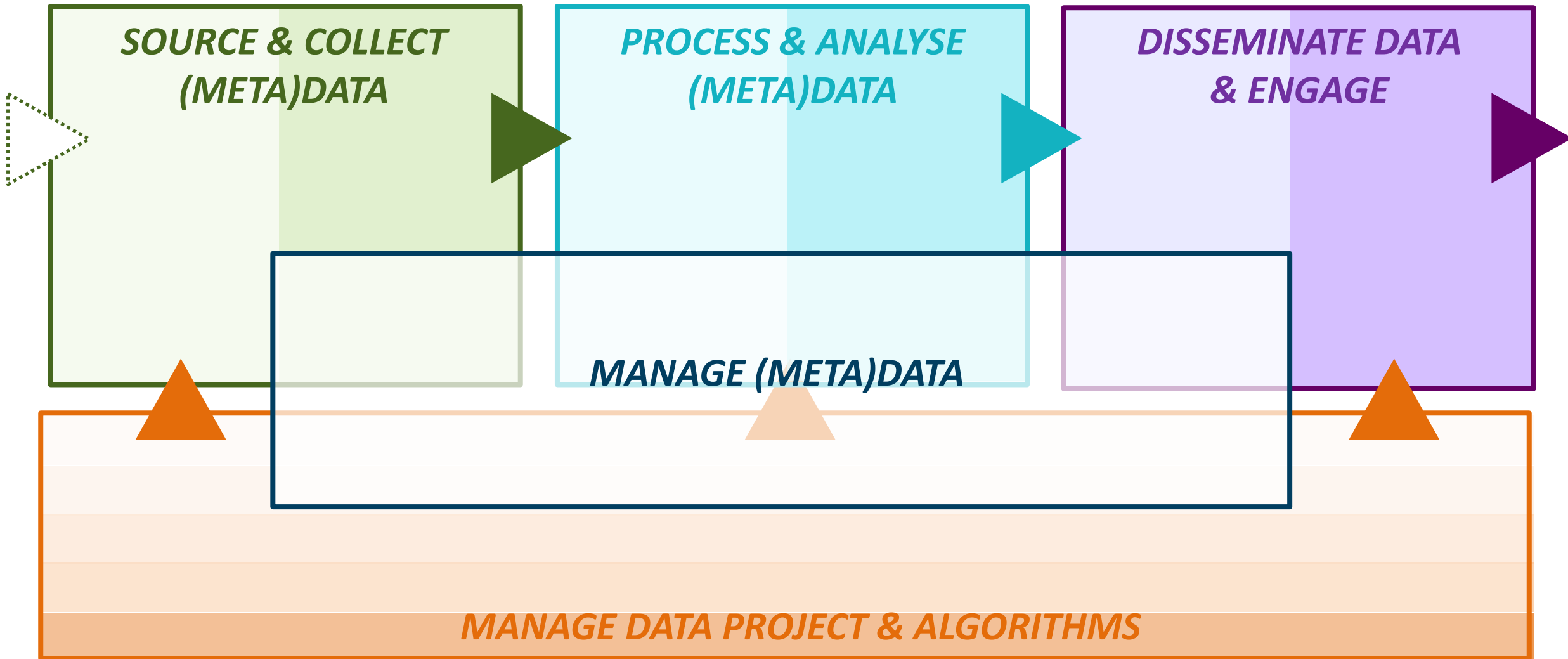
Instructor: Jens Dossé
Categories: Data modelling, SDMX
Review: ★★★★★

Starting point: Build capacity in SDMX data modelling

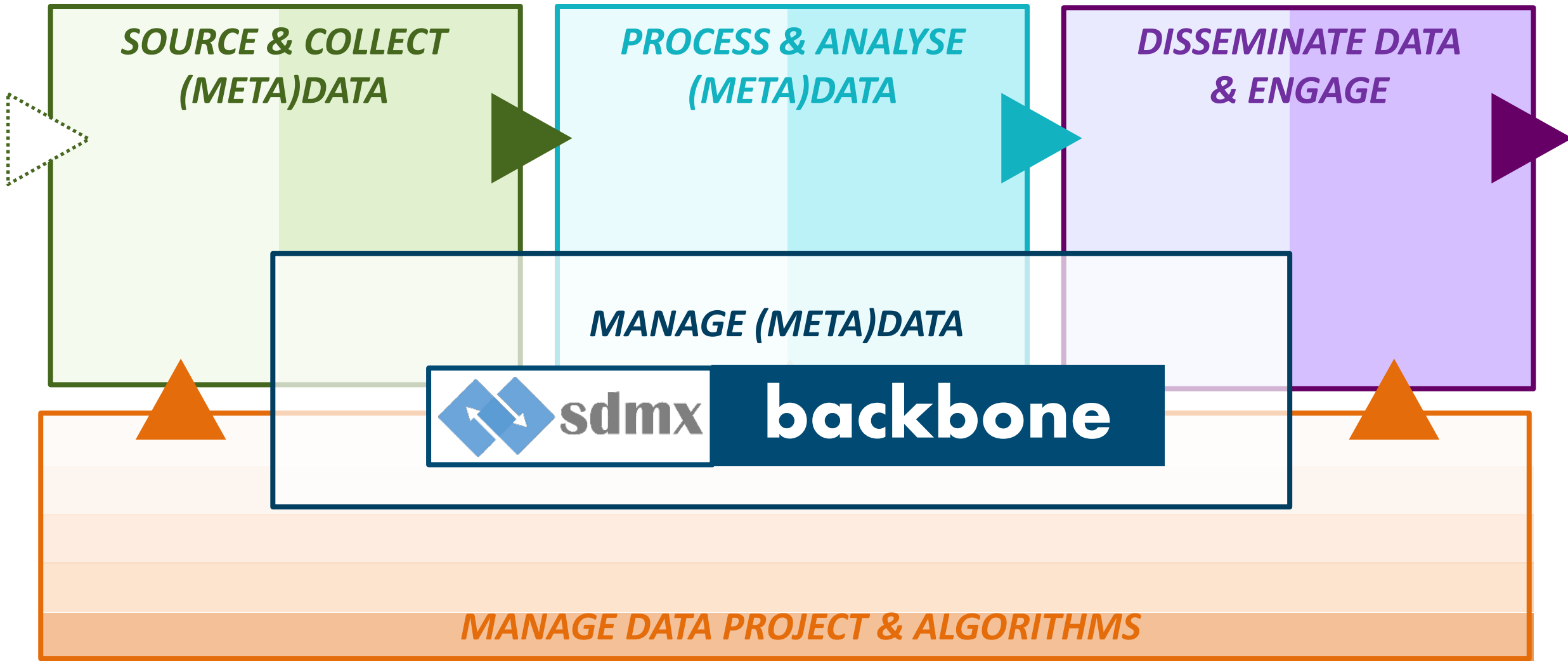
THE #SDMX METADATA-DRIVEN APPROACH



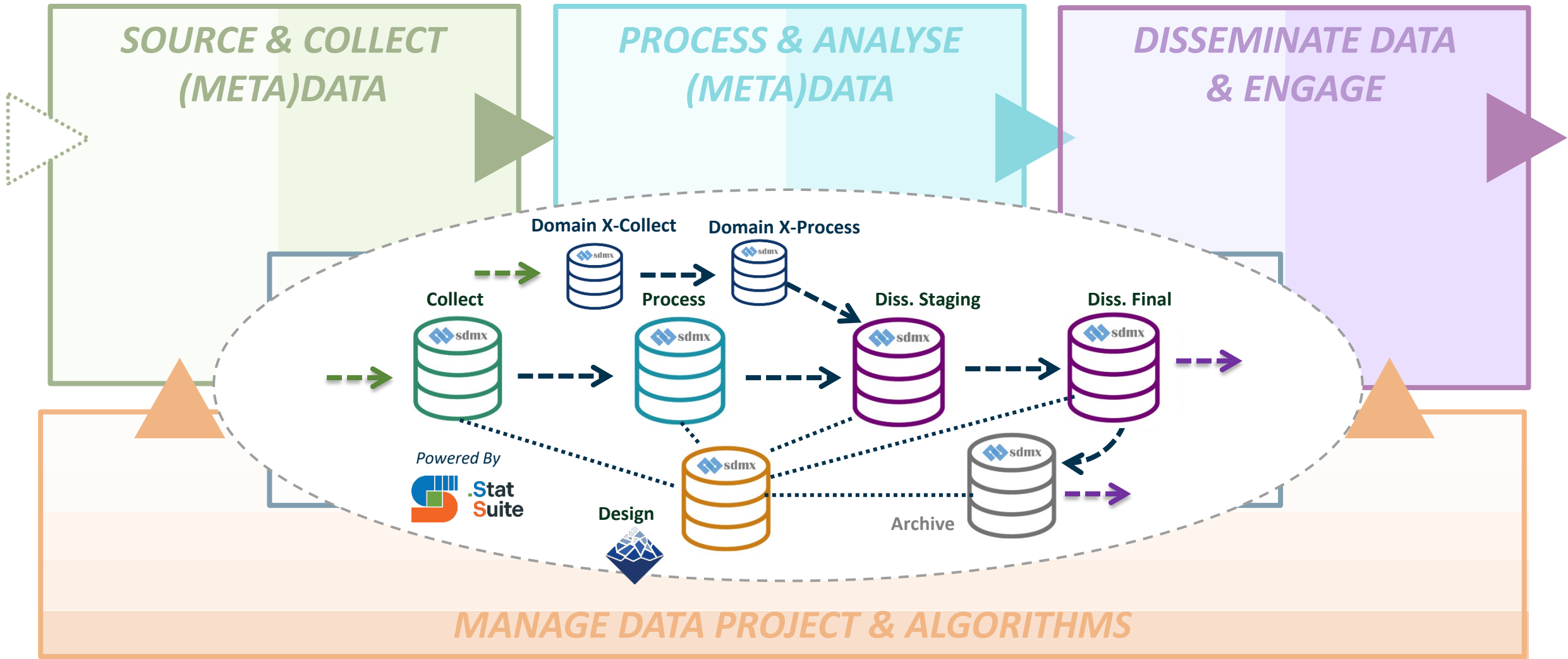
The Data Platform



The #SDMX Backbone

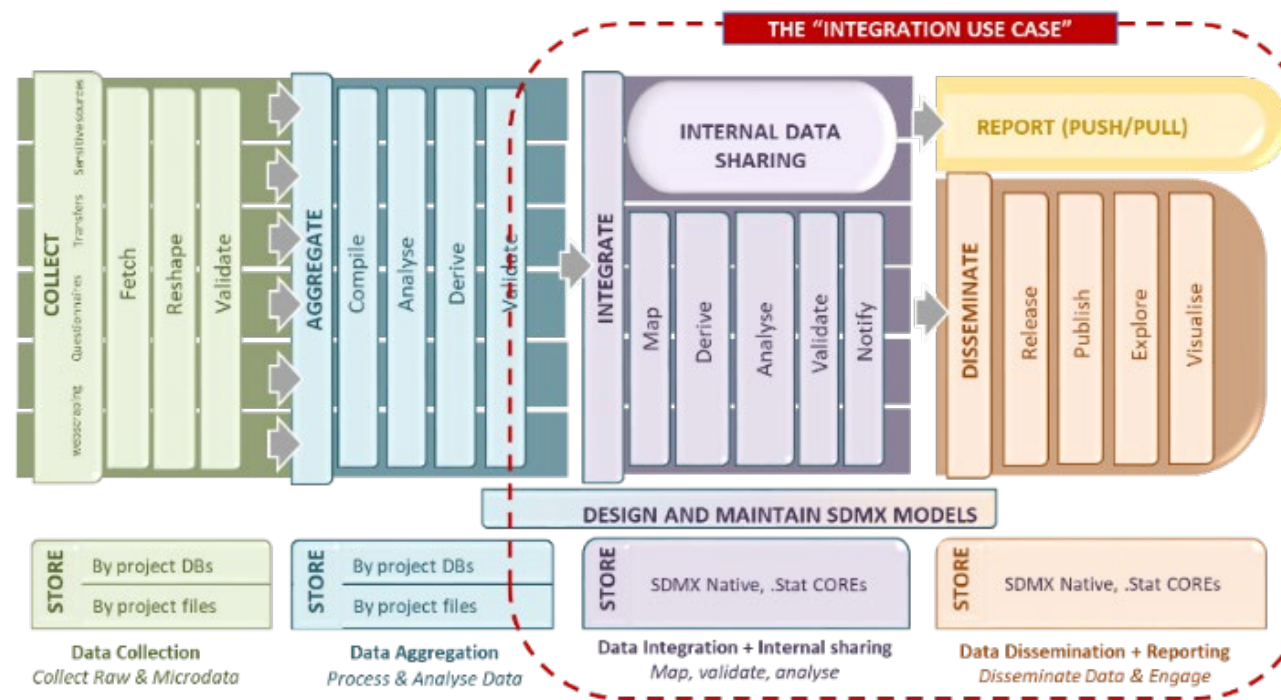


The #SDMX Backbone



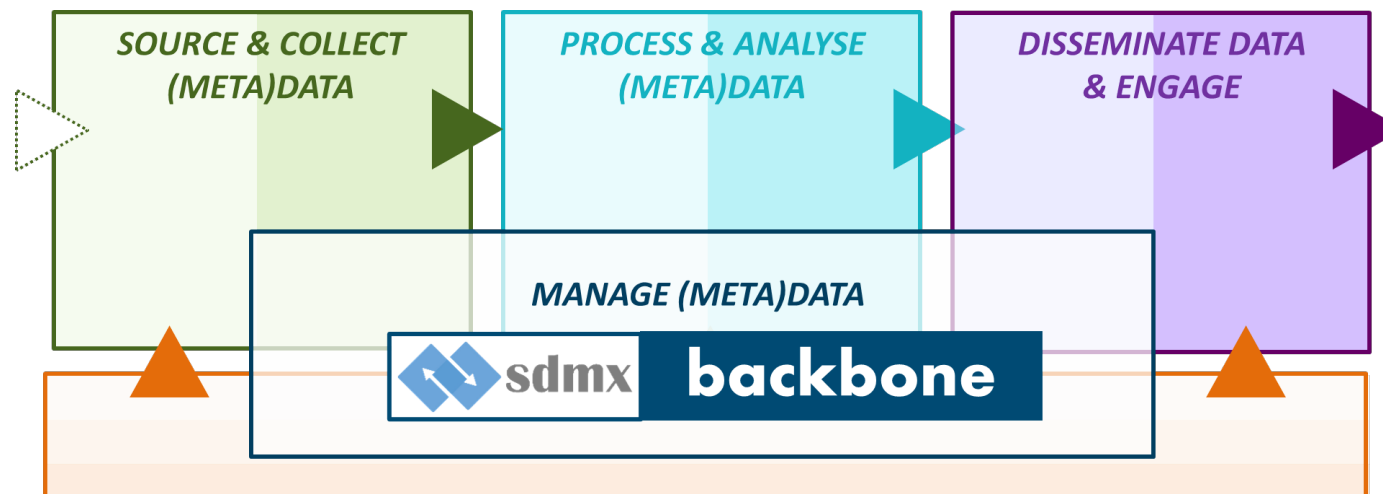
The #SDMX Backbone

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

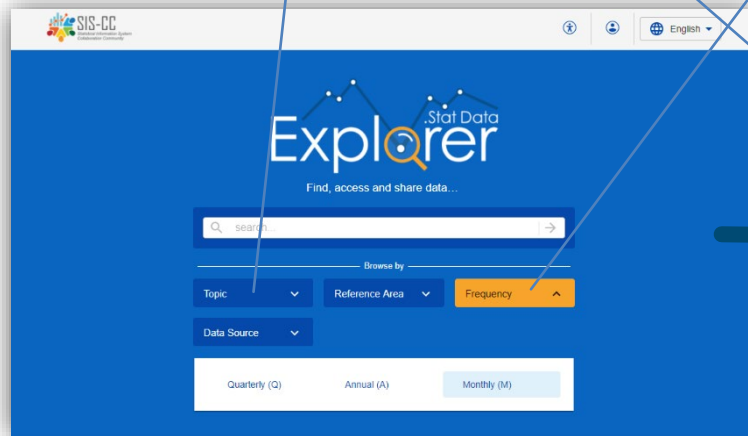


#SDMX #MetadataDriven Data User Experience

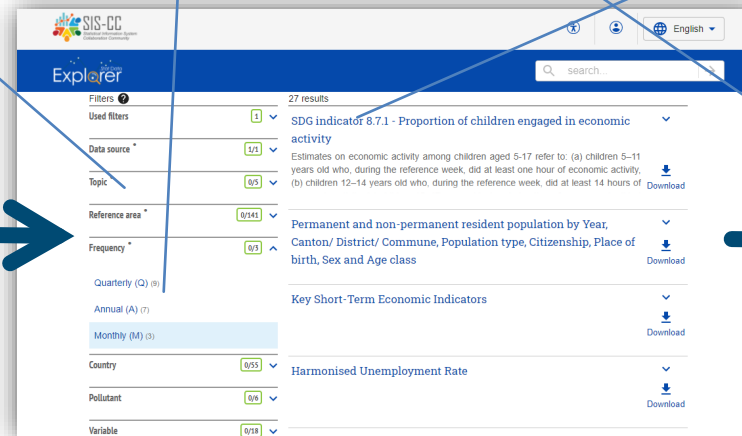
 category scheme(s)

 dimensions:
concepts + codelists

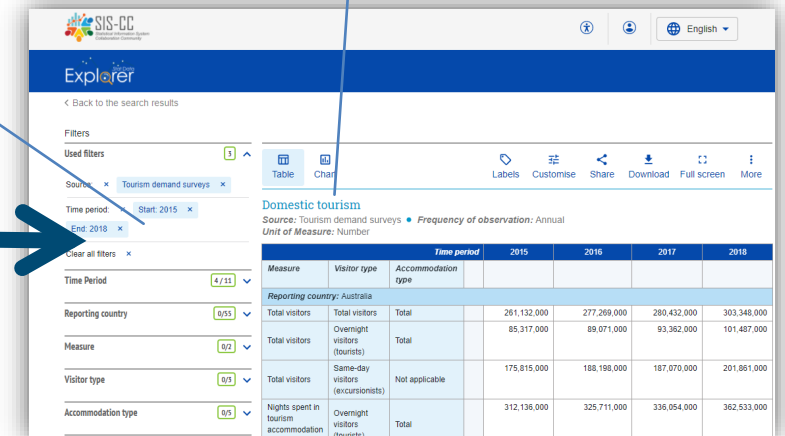
 dataflow
+ presentation annotations



Step 1 Home page
Lexical search or Topics



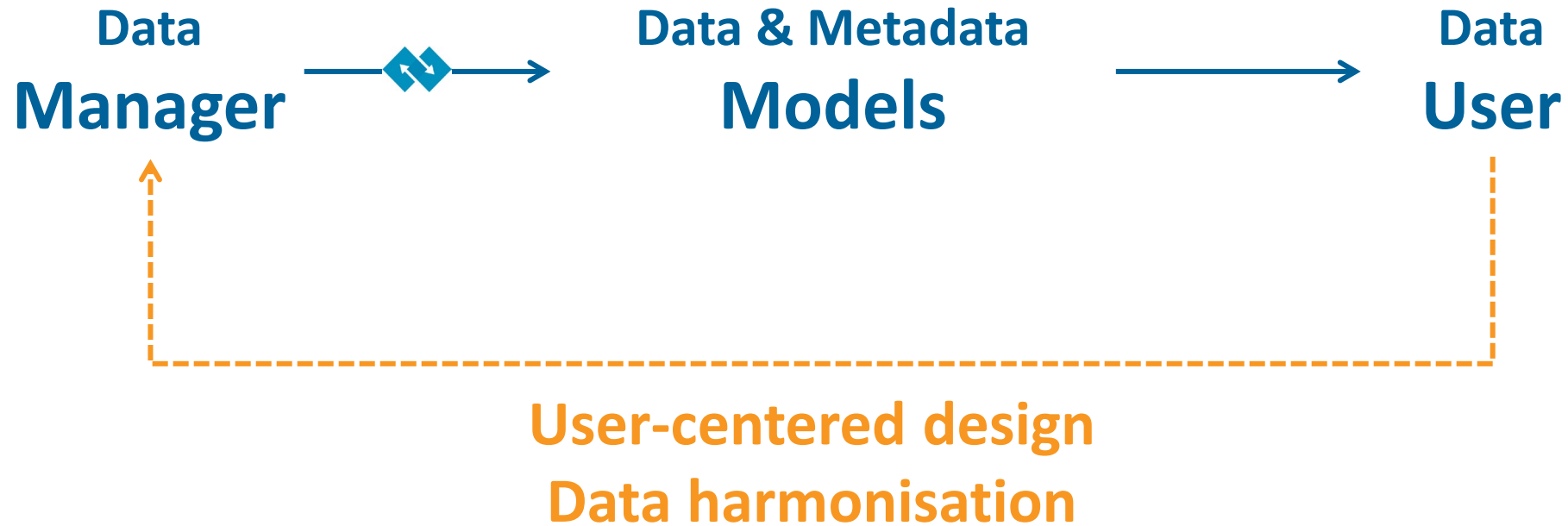
Step 2 Search results
Filtering on dynamic facets



		Time period			
		2015	2016	2017	2018
Reporting country: Australia					
Total visitors	Total visitors	261,132,000	277,269,000	280,432,000	303,348,000
Total visitors	Overnight visitors (tourists)	85,317,000	89,071,000	93,362,000	101,487,000
Total visitors	Same-day visitors (excursionists)	175,815,000	188,198,000	187,070,000	201,861,000
Nights spent in tourism accommodation establishments	Overnight visitors (tourists)	312,136,000	325,711,000	336,054,000	362,533,000

Step 3 Visualisation page
Selecting dimensions / variables

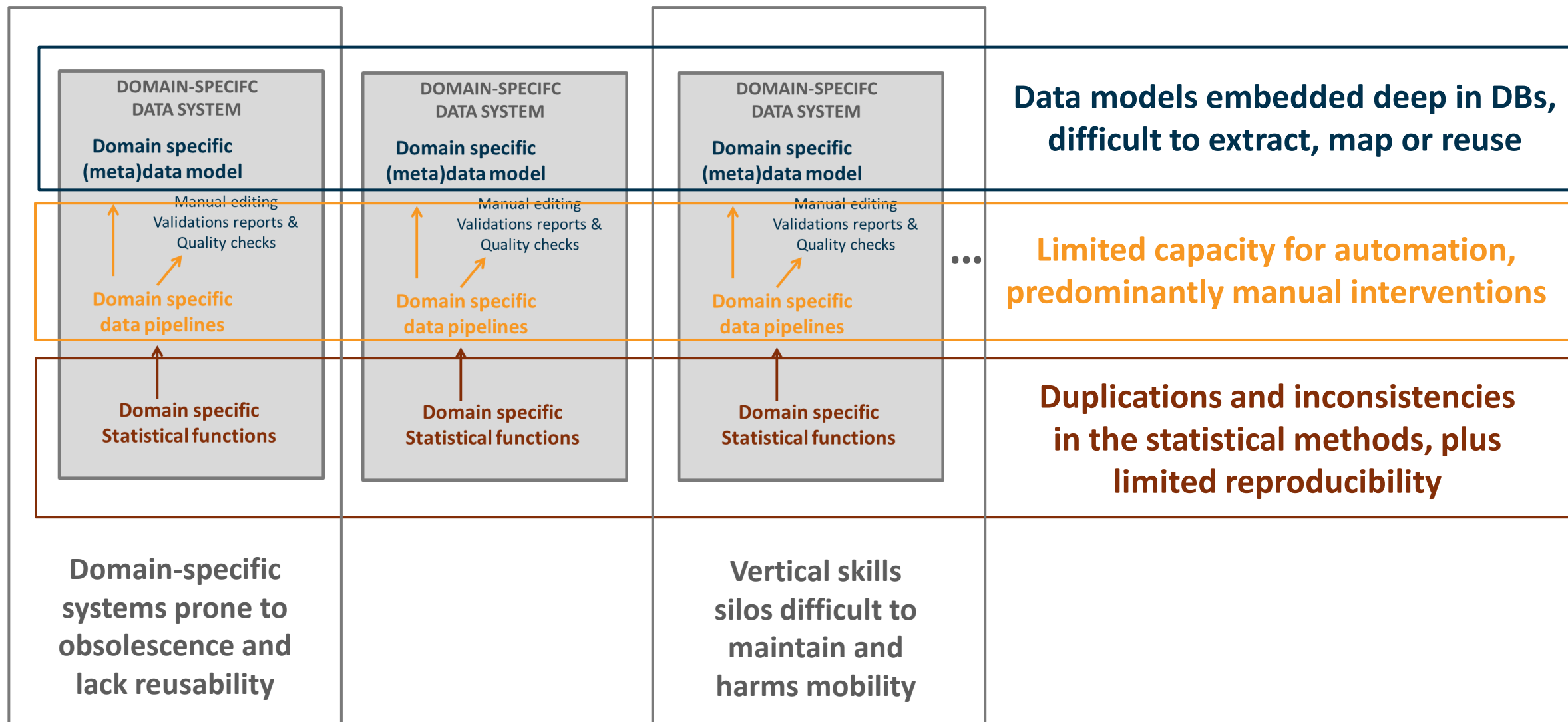
#SDMX #MetadataDriven Data User Experience



Establish a systematic User Research



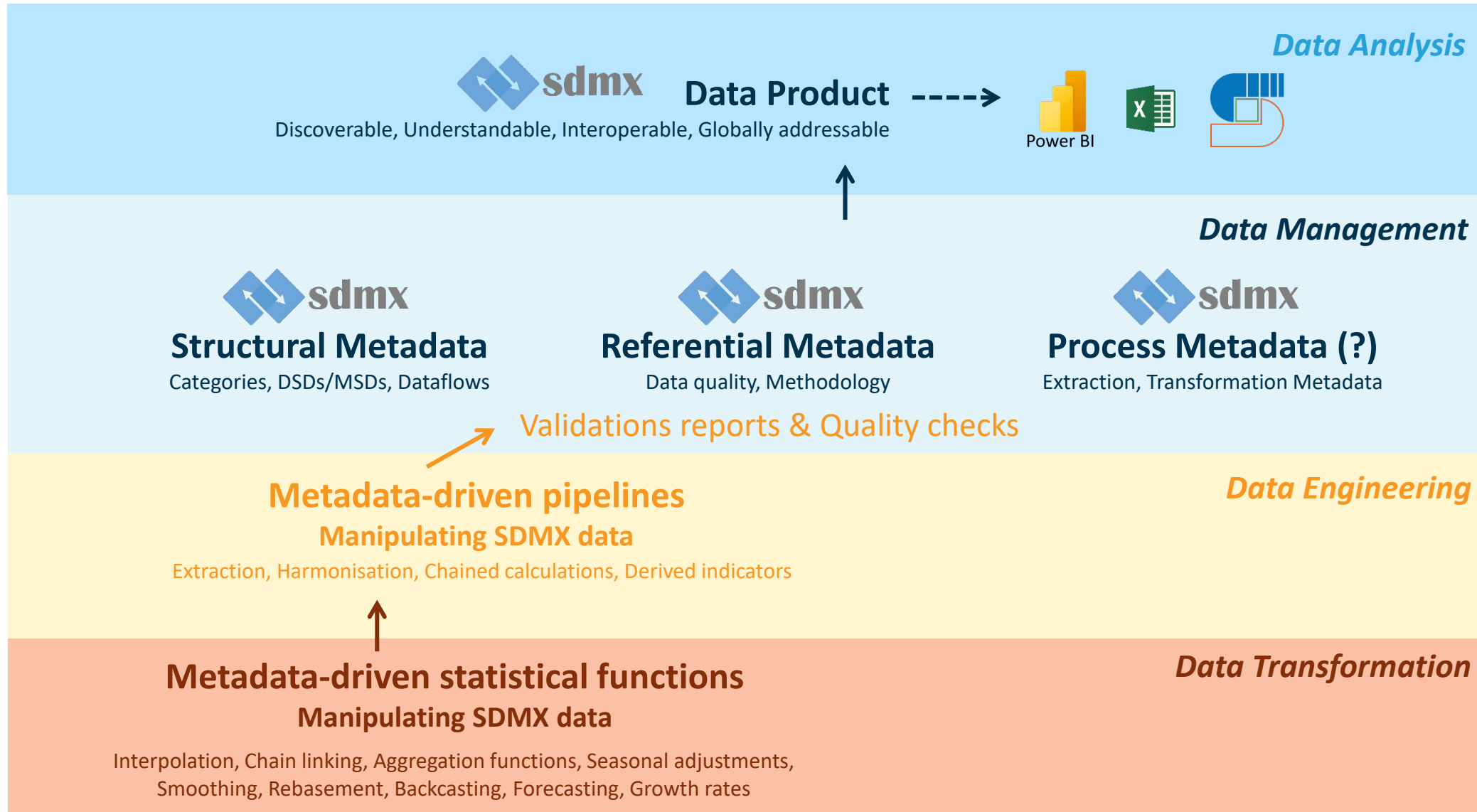
The Data Platform is fragmented



#SDMX #MetadataDriven Data Manager Experience

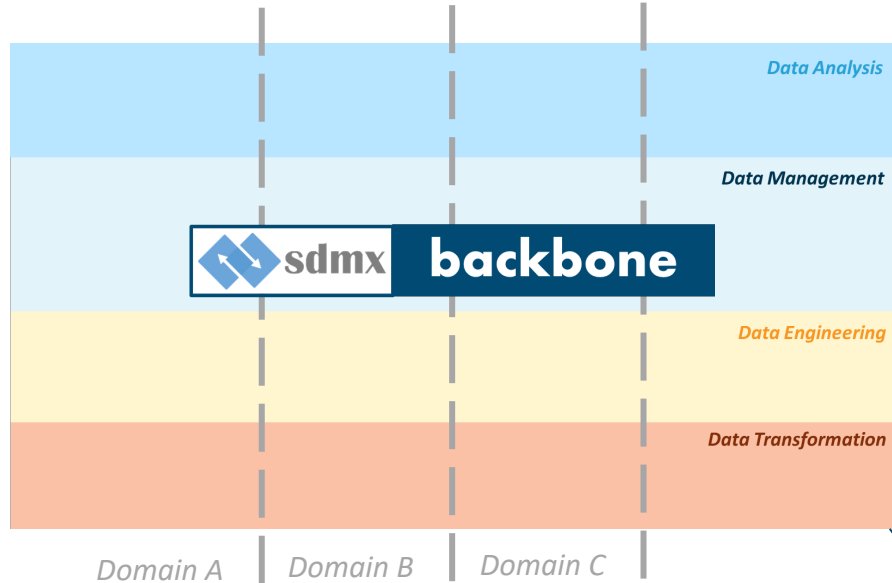
MANAGE DATA
& METADATA

MANAGE DATA PROJECT
& ALGORITHMS



#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




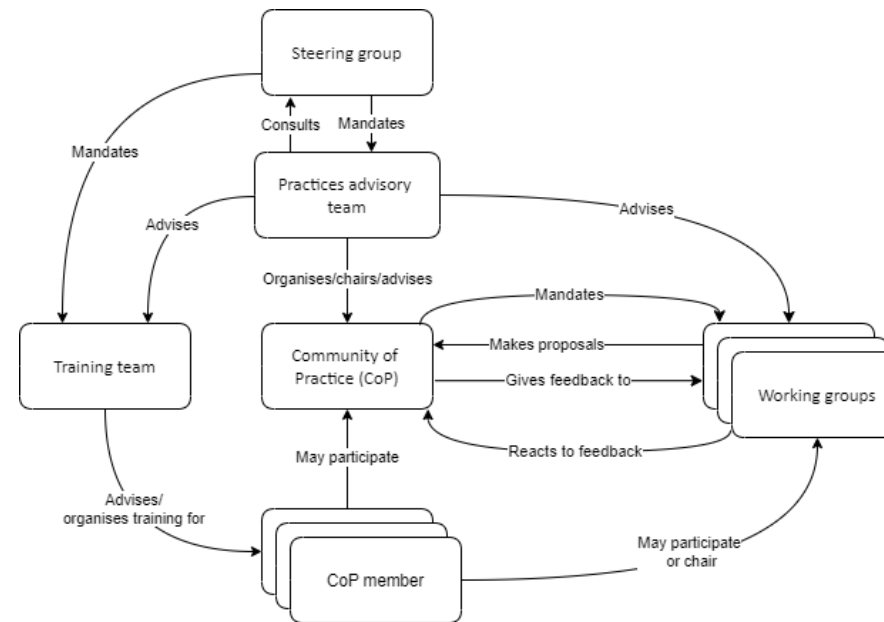
- Decoupled roles and specializations
- Data mgt by non coding-experts
- Facilitation of staff mobility
- Possibility of data tasks mutualization
- Facilitation of platform migration
- Domain specific coding language
- Domain specific interfaces
- Capitalize on shared methods / reuse
- Capitalize on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



Statistical Data and Metadata Exchange
Statistical Information System
Collaboration Community

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

Table of contents

- 1. EXECUTIVE SUMMARY
- 2. INTRODUCTION
- 3. SDMX STRUCTURAL METADATA GOVERNANCE FRAMEWORK
- 4. WHY WE NEED AN SDMX STRUCTURAL METADATA GOVERNANCE FRAMEWORK
- 5. THE GOVERNANCE FRAMEWORK
- 6. THE GOVERNANCE FRAMEWORK
- 7. THE GOVERNANCE FRAMEWORK
- 8. THE GOVERNANCE FRAMEWORK
- 9. THE GOVERNANCE FRAMEWORK
- 10. THE GOVERNANCE FRAMEWORK
- 11. THE GOVERNANCE FRAMEWORK
- 12. THE GOVERNANCE FRAMEWORK
- 13. THE GOVERNANCE FRAMEWORK
- 14. THE GOVERNANCE FRAMEWORK
- 15. THE GOVERNANCE FRAMEWORK
- 16. THE GOVERNANCE FRAMEWORK
- 17. THE GOVERNANCE FRAMEWORK
- 18. THE GOVERNANCE FRAMEWORK
- 19. THE GOVERNANCE FRAMEWORK
- 20. THE GOVERNANCE FRAMEWORK
- 21. THE GOVERNANCE FRAMEWORK
- 22. THE GOVERNANCE FRAMEWORK
- 23. THE GOVERNANCE FRAMEWORK
- 24. THE GOVERNANCE FRAMEWORK
- 25. THE GOVERNANCE FRAMEWORK
- 26. THE GOVERNANCE FRAMEWORK
- 27. THE GOVERNANCE FRAMEWORK
- 28. THE GOVERNANCE FRAMEWORK
- 29. THE GOVERNANCE FRAMEWORK
- 30. THE GOVERNANCE FRAMEWORK
- 31. THE GOVERNANCE FRAMEWORK
- 32. THE GOVERNANCE FRAMEWORK
- 33. THE GOVERNANCE FRAMEWORK
- 34. THE GOVERNANCE FRAMEWORK
- 35. THE GOVERNANCE FRAMEWORK
- 36. THE GOVERNANCE FRAMEWORK
- 37. THE GOVERNANCE FRAMEWORK
- 38. THE GOVERNANCE FRAMEWORK
- 39. THE GOVERNANCE FRAMEWORK
- 40. THE GOVERNANCE FRAMEWORK
- 41. THE GOVERNANCE FRAMEWORK
- 42. THE GOVERNANCE FRAMEWORK
- 43. THE GOVERNANCE FRAMEWORK
- 44. THE GOVERNANCE FRAMEWORK
- 45. THE GOVERNANCE FRAMEWORK
- 46. THE GOVERNANCE FRAMEWORK
- 47. THE GOVERNANCE FRAMEWORK
- 48. THE GOVERNANCE FRAMEWORK
- 49. THE GOVERNANCE FRAMEWORK
- 50. THE GOVERNANCE FRAMEWORK
- 51. THE GOVERNANCE FRAMEWORK
- 52. THE GOVERNANCE FRAMEWORK
- 53. THE GOVERNANCE FRAMEWORK
- 54. THE GOVERNANCE FRAMEWORK
- 55. THE GOVERNANCE FRAMEWORK
- 56. THE GOVERNANCE FRAMEWORK
- 57. THE GOVERNANCE FRAMEWORK
- 58. THE GOVERNANCE FRAMEWORK
- 59. THE GOVERNANCE FRAMEWORK
- 60. THE GOVERNANCE FRAMEWORK
- 61. THE GOVERNANCE FRAMEWORK
- 62. THE GOVERNANCE FRAMEWORK
- 63. THE GOVERNANCE FRAMEWORK
- 64. THE GOVERNANCE FRAMEWORK
- 65. THE GOVERNANCE FRAMEWORK
- 66. THE GOVERNANCE FRAMEWORK
- 67. THE GOVERNANCE FRAMEWORK
- 68. THE GOVERNANCE FRAMEWORK
- 69. THE GOVERNANCE FRAMEWORK
- 70. THE GOVERNANCE FRAMEWORK
- 71. THE GOVERNANCE FRAMEWORK
- 72. THE GOVERNANCE FRAMEWORK
- 73. THE GOVERNANCE FRAMEWORK
- 74. THE GOVERNANCE FRAMEWORK
- 75. THE GOVERNANCE FRAMEWORK
- 76. THE GOVERNANCE FRAMEWORK
- 77. THE GOVERNANCE FRAMEWORK
- 78. THE GOVERNANCE FRAMEWORK
- 79. THE GOVERNANCE FRAMEWORK
- 80. THE GOVERNANCE FRAMEWORK
- 81. THE GOVERNANCE FRAMEWORK
- 82. THE GOVERNANCE FRAMEWORK
- 83. THE GOVERNANCE FRAMEWORK
- 84. THE GOVERNANCE FRAMEWORK
- 85. THE GOVERNANCE FRAMEWORK
- 86. THE GOVERNANCE FRAMEWORK
- 87. THE GOVERNANCE FRAMEWORK
- 88. THE GOVERNANCE FRAMEWORK
- 89. THE GOVERNANCE FRAMEWORK
- 90. THE GOVERNANCE FRAMEWORK
- 91. THE GOVERNANCE FRAMEWORK
- 92. THE GOVERNANCE FRAMEWORK
- 93. THE GOVERNANCE FRAMEWORK
- 94. THE GOVERNANCE FRAMEWORK
- 95. THE GOVERNANCE FRAMEWORK
- 96. THE GOVERNANCE FRAMEWORK
- 97. THE GOVERNANCE FRAMEWORK
- 98. THE GOVERNANCE FRAMEWORK
- 99. THE GOVERNANCE FRAMEWORK
- 100. THE GOVERNANCE FRAMEWORK

#SDMX #MetadataDriven approach – Wrap-up

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?

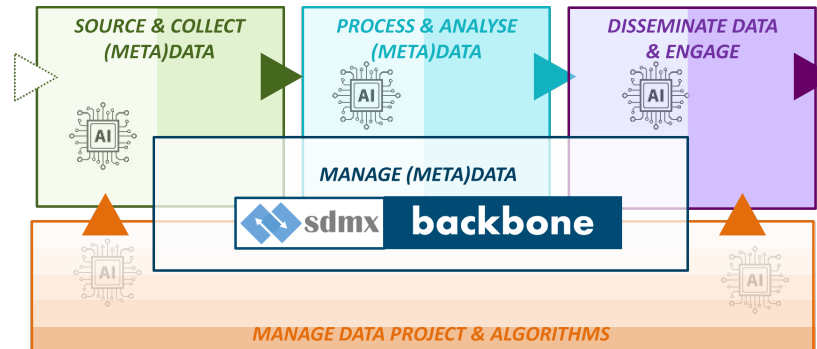
Opportunities #SDMX #AI



**Nowcasting
Federated Learning**



**Outlier detection
Clustering**

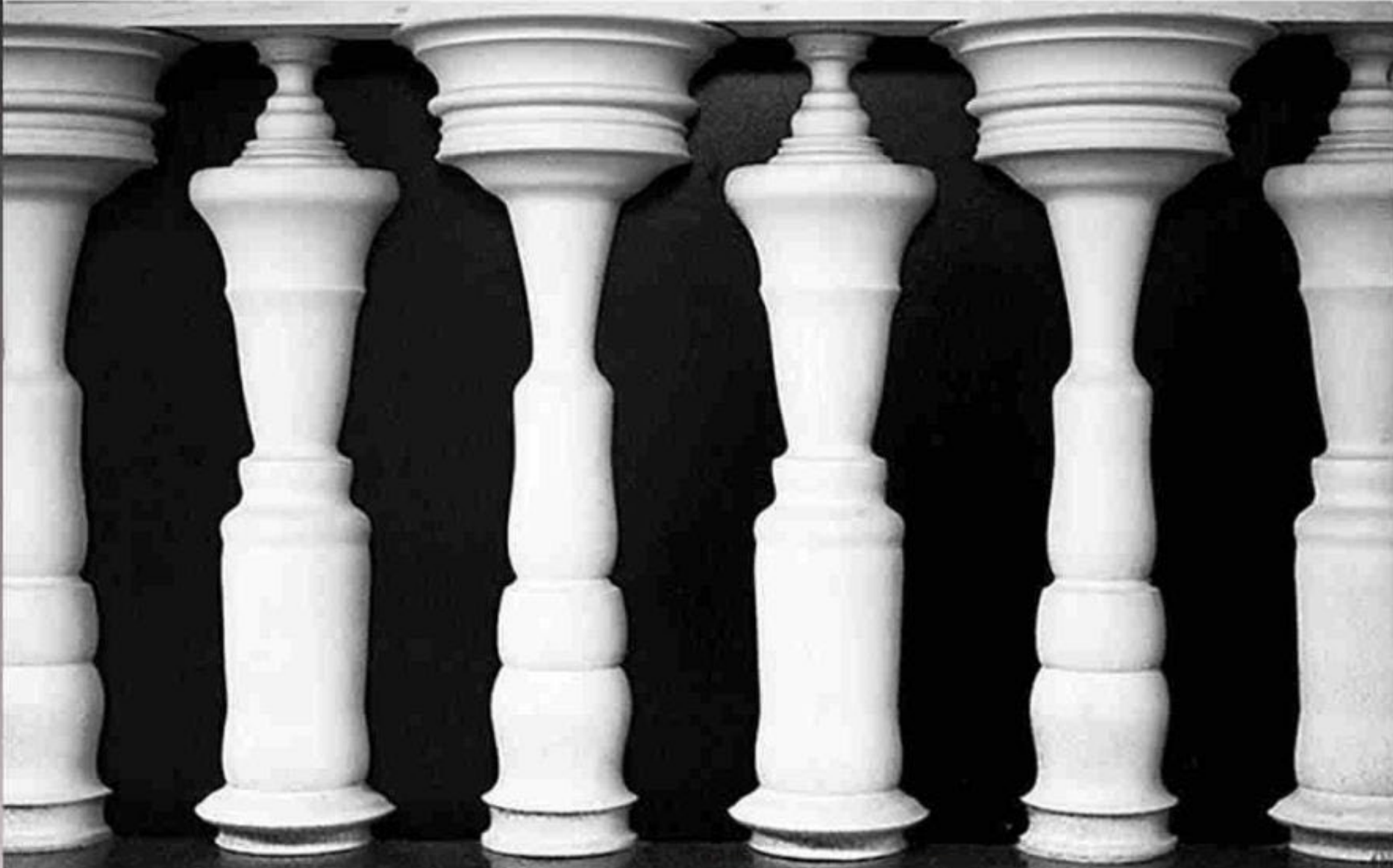


**Natural Language
Conversational
Data Access**



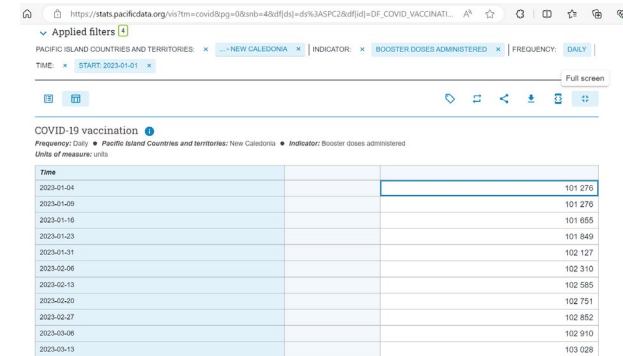
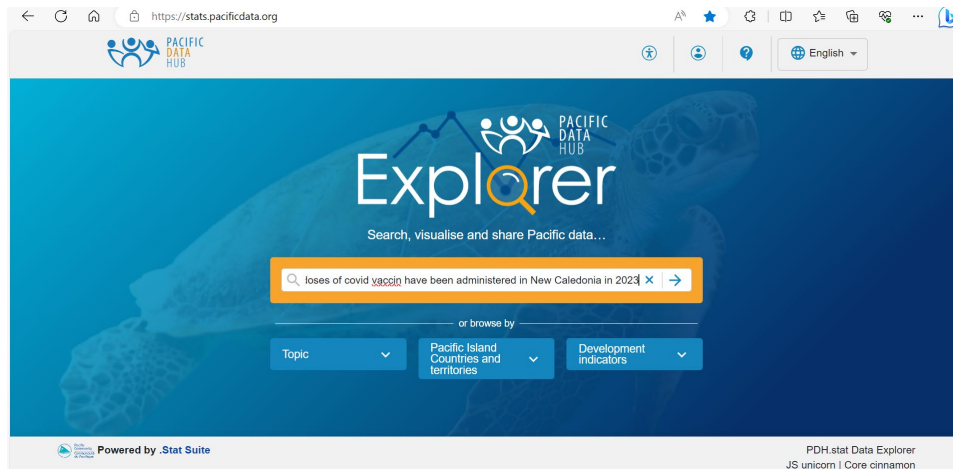
**Code generation
Data Modelling**

The cognitive gap



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

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

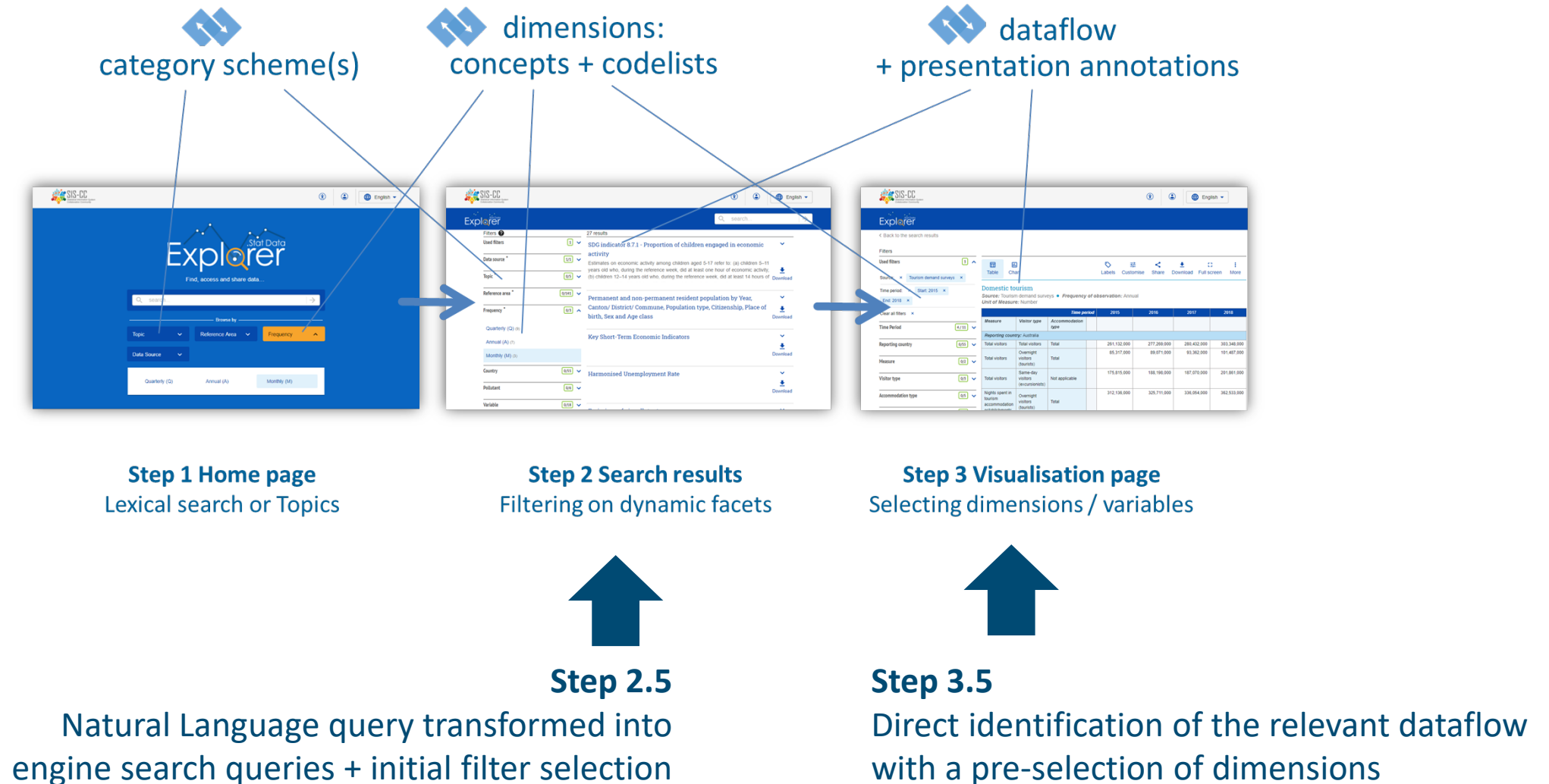


Time	Value	Units of measure
2023-01-04	101 276	
2023-01-09	101 276	
2023-01-16	101 665	
2023-01-23	101 949	
2023-01-31	102 127	
2023-02-06	102 310	
2023-02-13	102 585	
2023-02-20	102 751	
2023-02-27	102 852	
2023-03-06	102 910	
2023-03-13	103 028	

Step 2.5 or 3.5

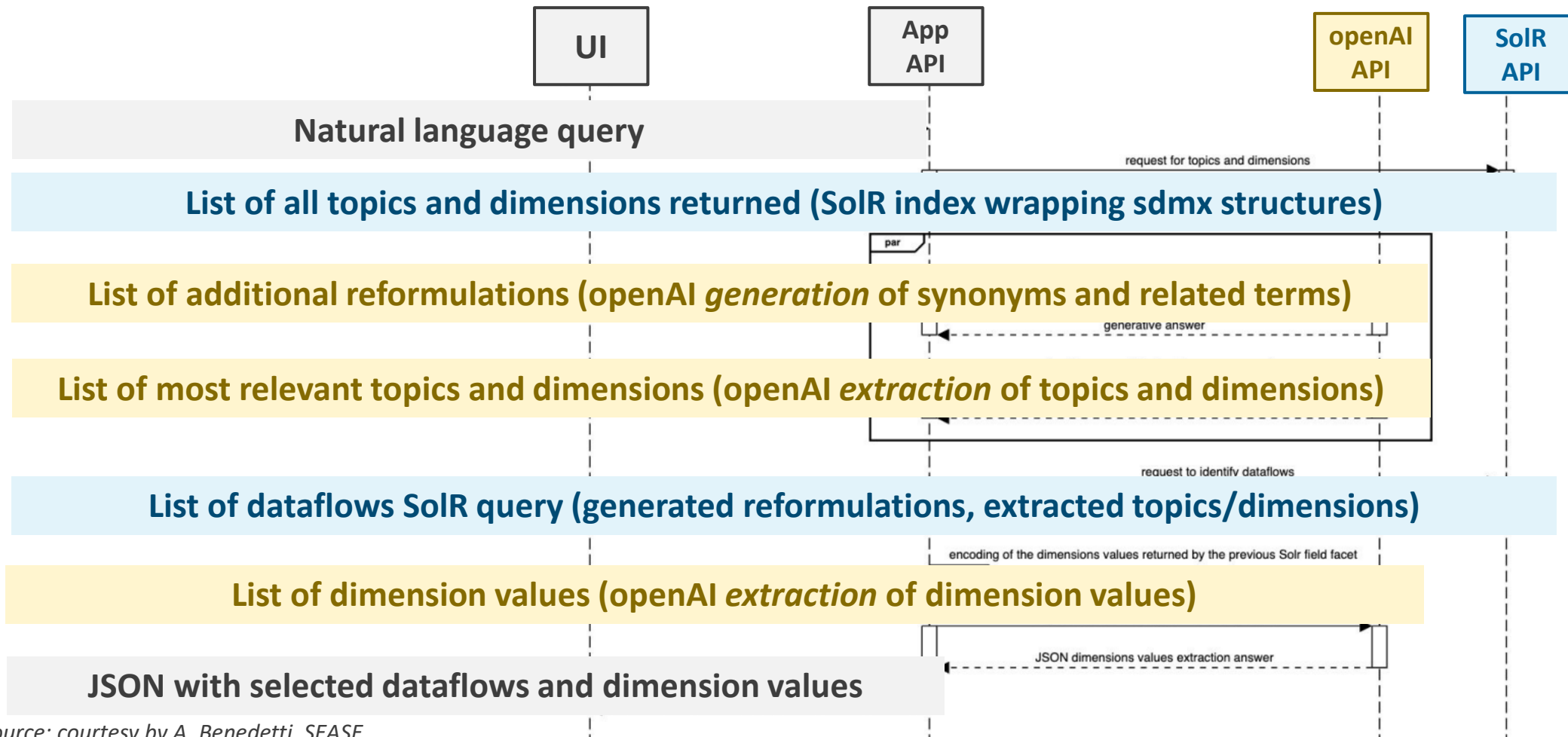
Natural Language query transformed into search queries
+ Pre-selection of dimensions on target dataflow(s)

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



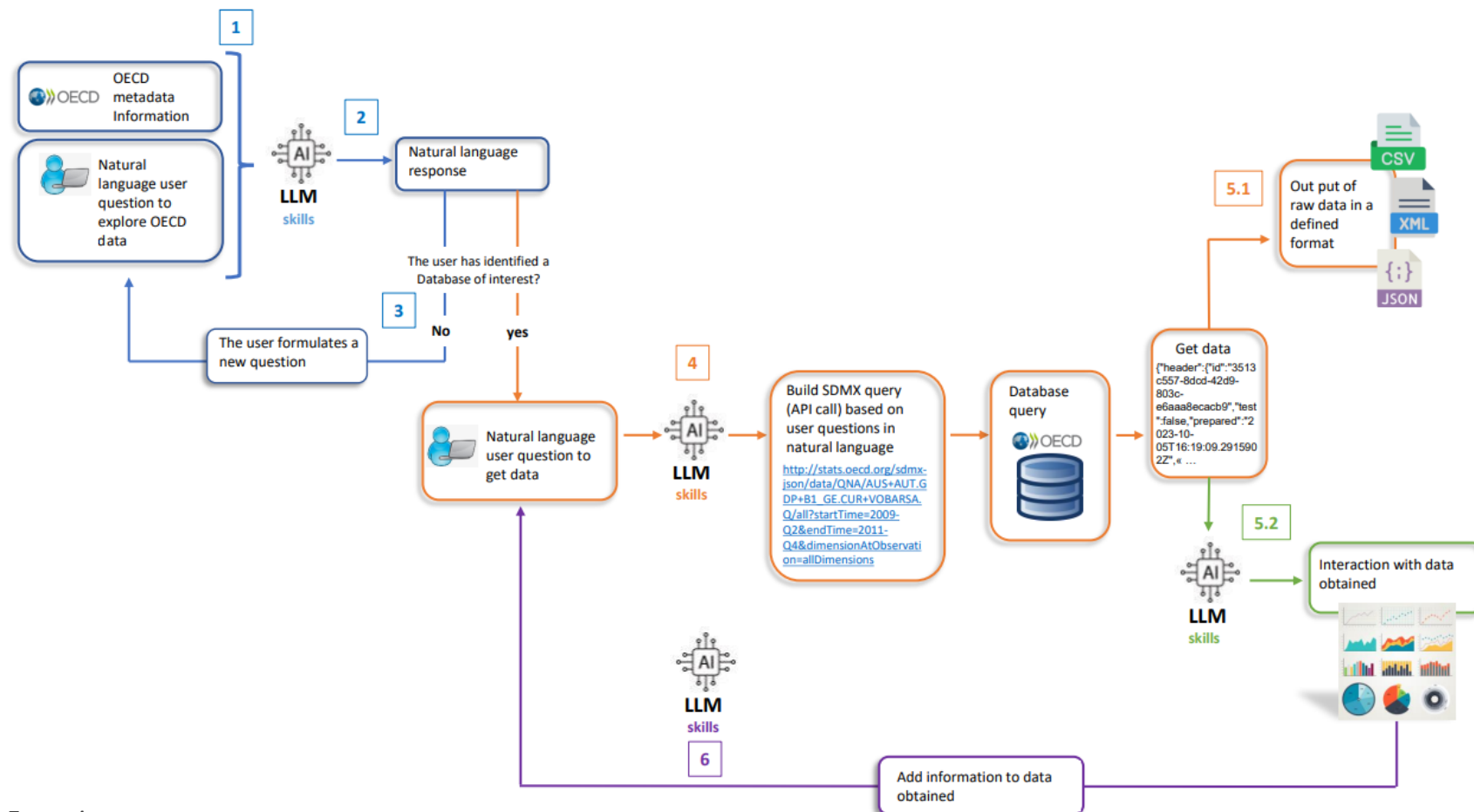
Augmenting the Data Explorer with #SDMX #AI (I)

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

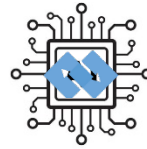
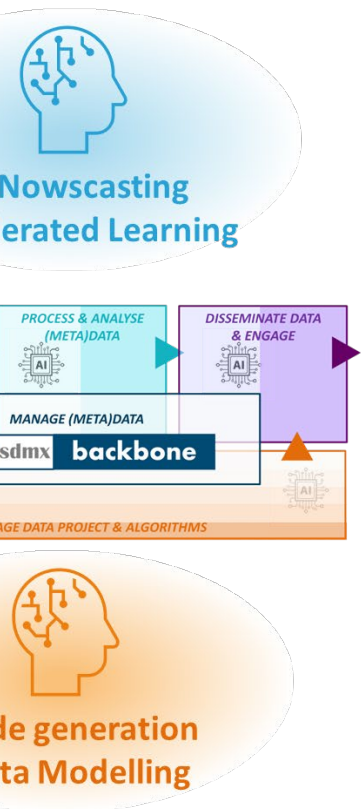


Replacing the Data Explorer with #SDMX #AI (II)

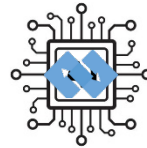
The StatsBot complements / replaces the Data Explorer with a conversational data assistant



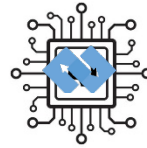
Creating #SDMX #AI open and reusable assets



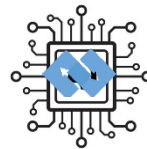
Natural language capable Data Explorer (I)



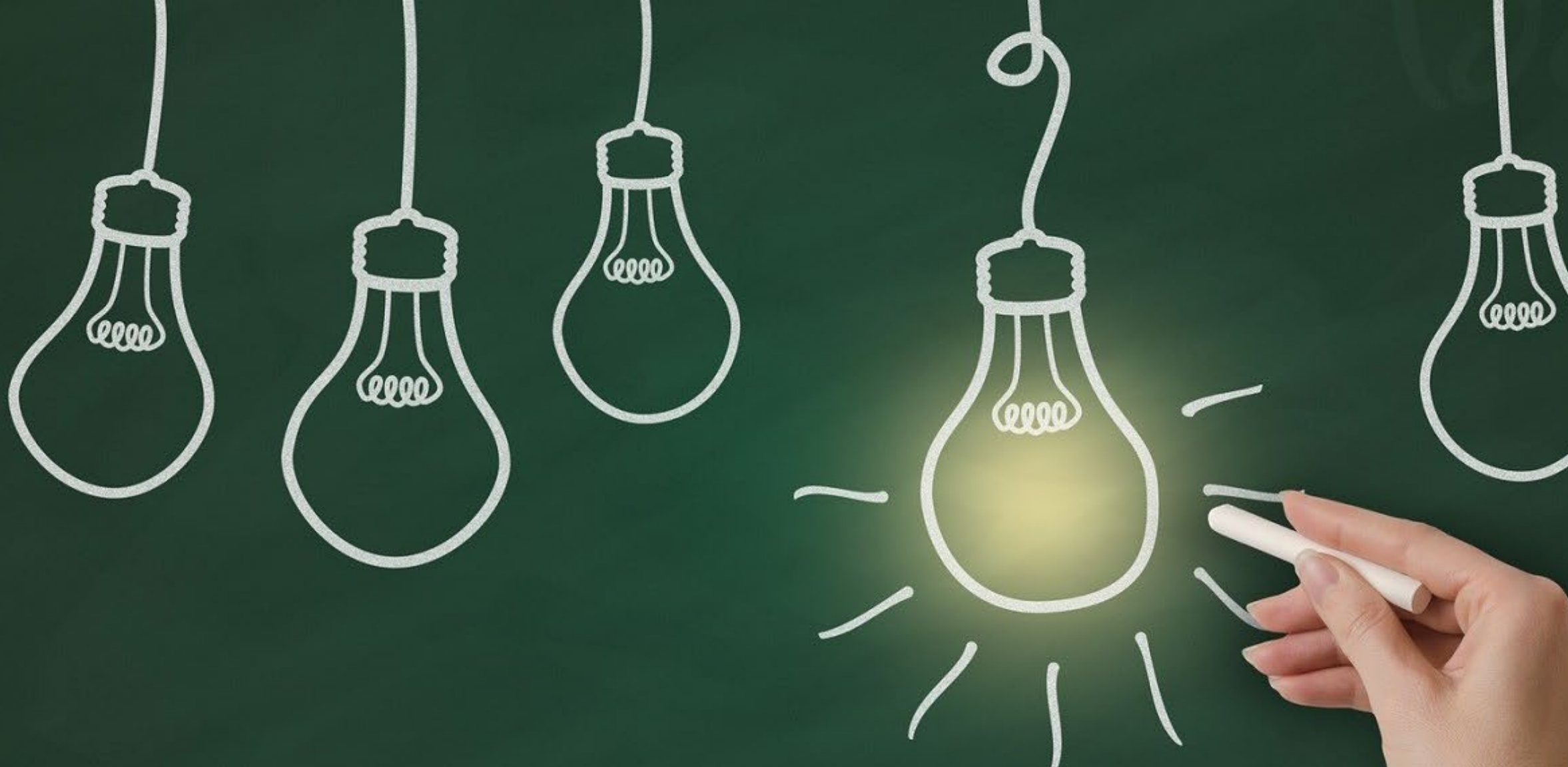
Universal StatsBot to access any SDMX source through a conversation (II)



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

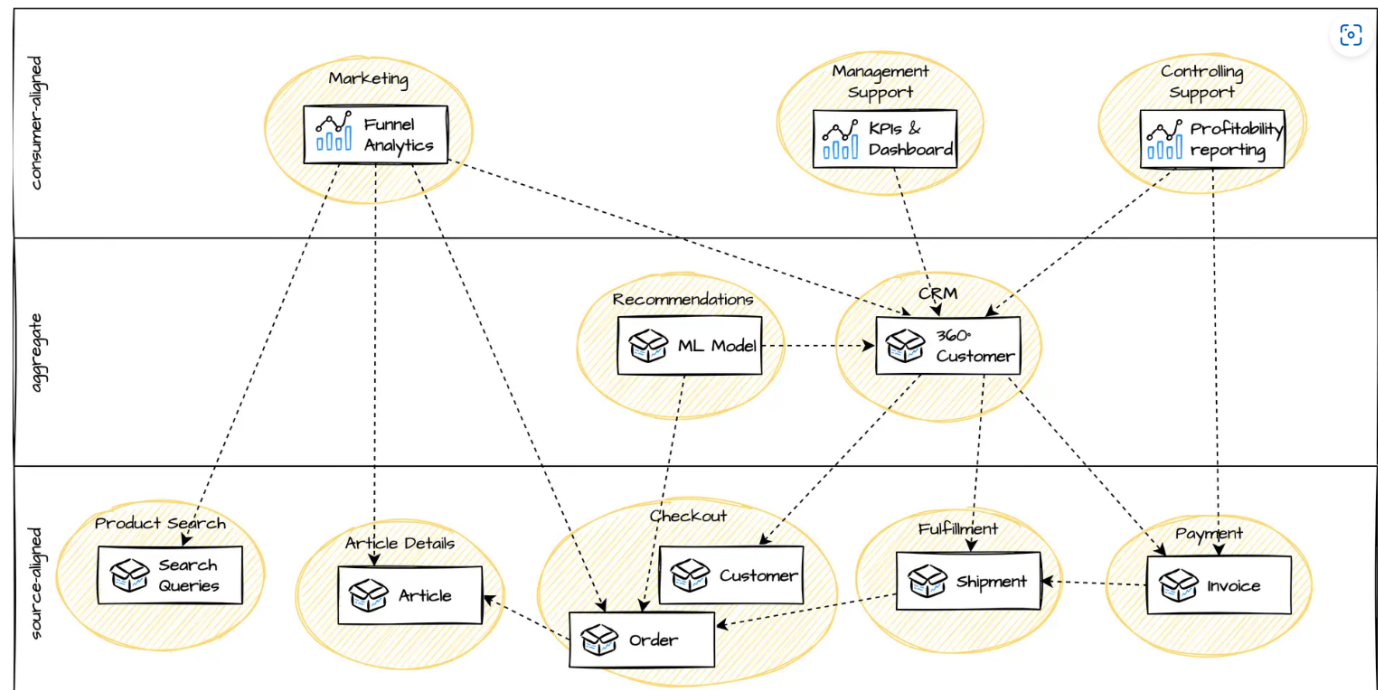


Zhamak Dehghani

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.

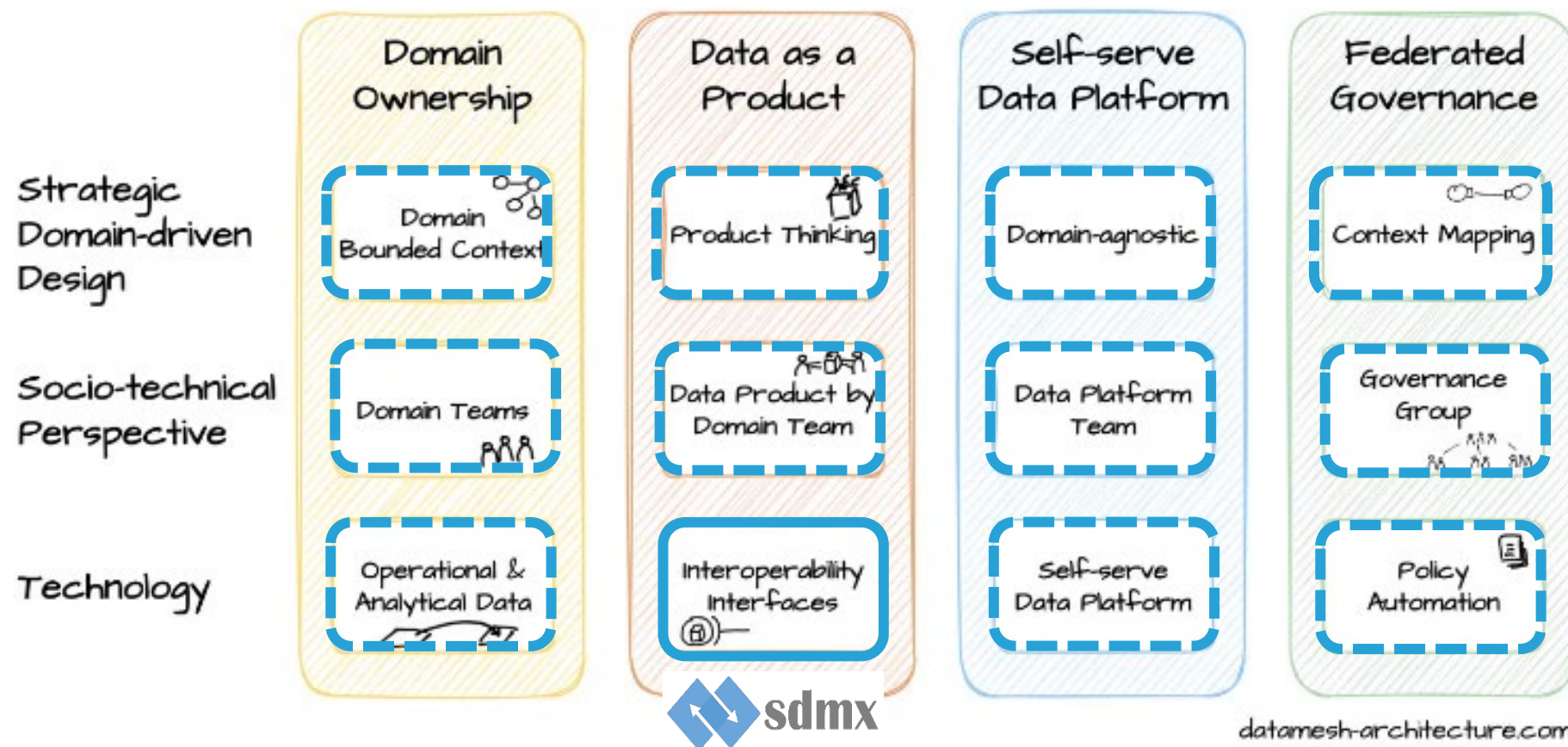


#DataMesh requires common data semantics

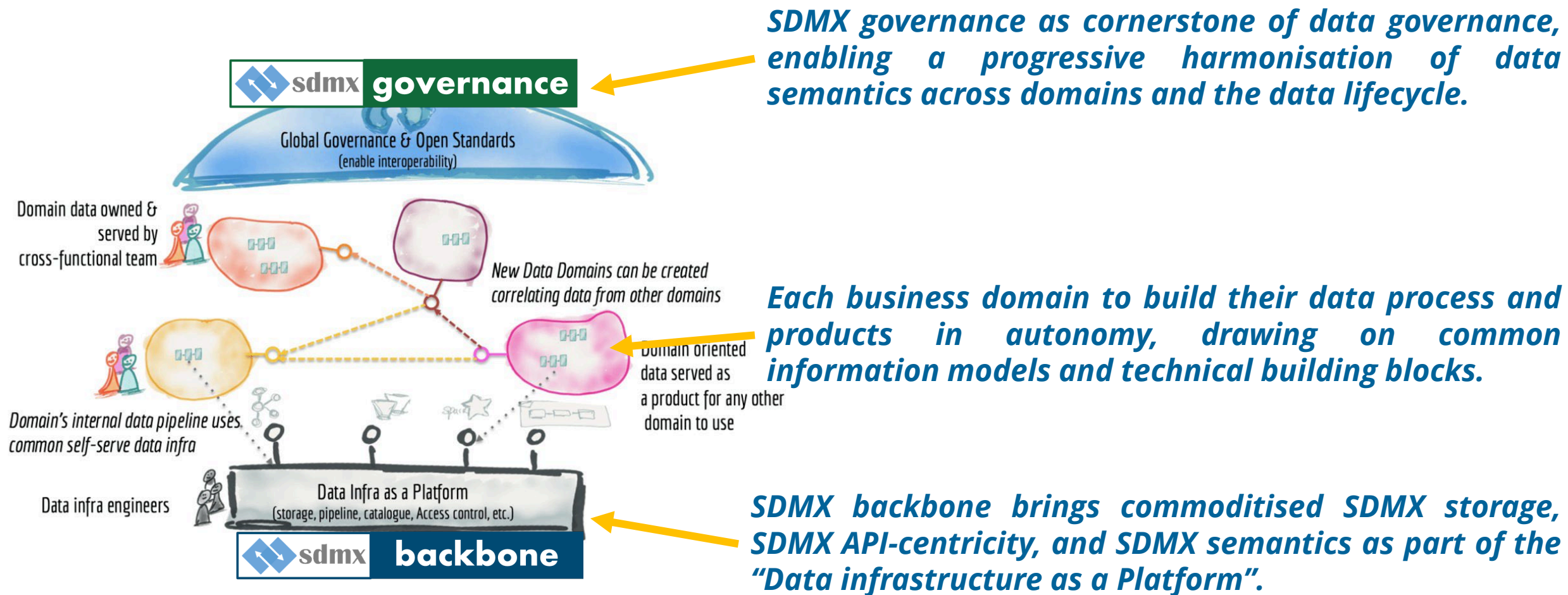
*“Data mesh attempts to **strike a balance between team autonomy and interteam interoperability and collaboration**. It gives domain teams autonomy to take control of their local decision making, for example, **choosing the best data model for their data products**, while it uses computational governance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

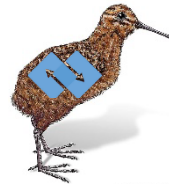
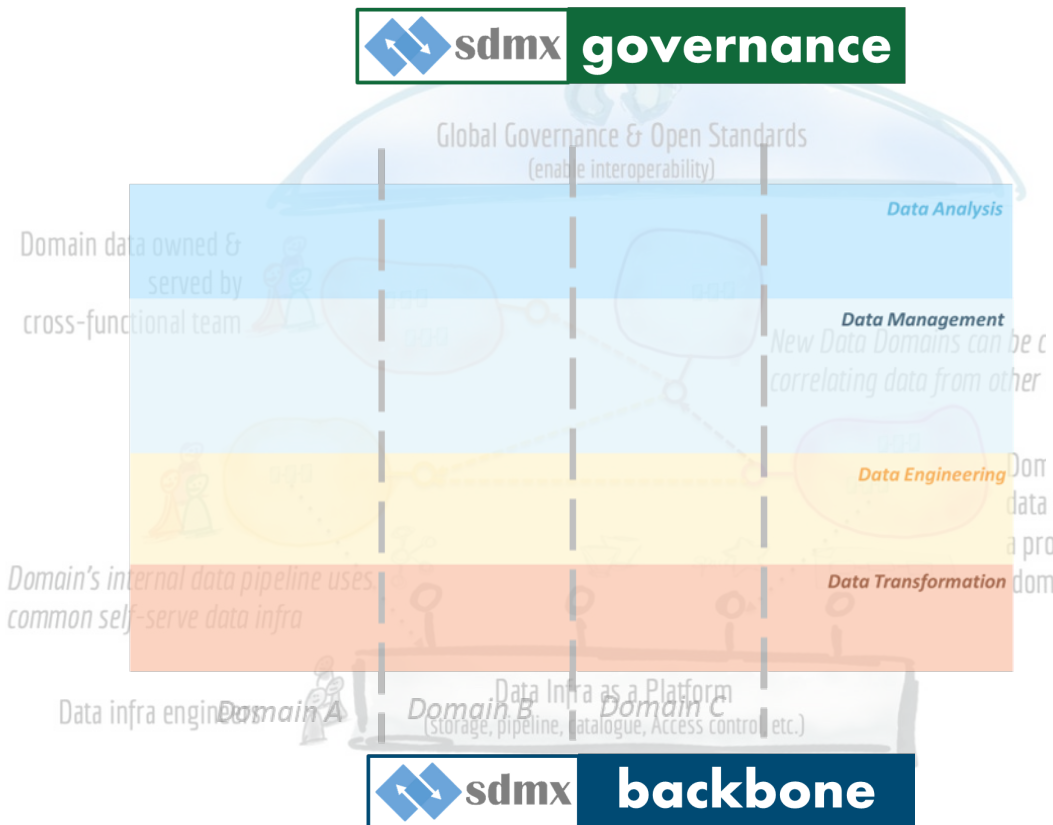
#SDMX as a #DataMesh enabler?



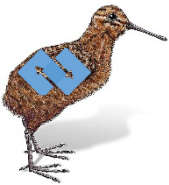
#SDMX as a #DataMesh enabler?



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

Reference resources

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



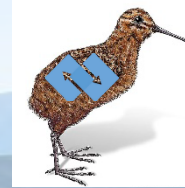
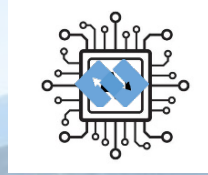
- .Stat Suite – [High level product overview and flight planner](#)
- .Stat Suite – [Kanban Board](#)
- .Stat Suite – [Changelog](#)
- .Stat Suite – [Extensive documentation](#)
- .Stat Suite – [Code repository](#)

.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

