9th SDMX Global Conference

Empowering Data Communities

SDMX structural metadata governance

Co-chaired by David Barraclough, OECD Smart Data Practices Manager
Luca Gramaglia, Eurostat Team Leader – Data and metadata exchange standards

SDMX Global Conference 2023 | Kingdom of Bahrain

Successful SDMX implementation needs to be grounded in a solid governance foundation

A data mesh architecture requires federated governance

There is an SDMX content-oriented guideline for structural metadata governance

What is Structural Metadata Governance?

It helps decide and describe:

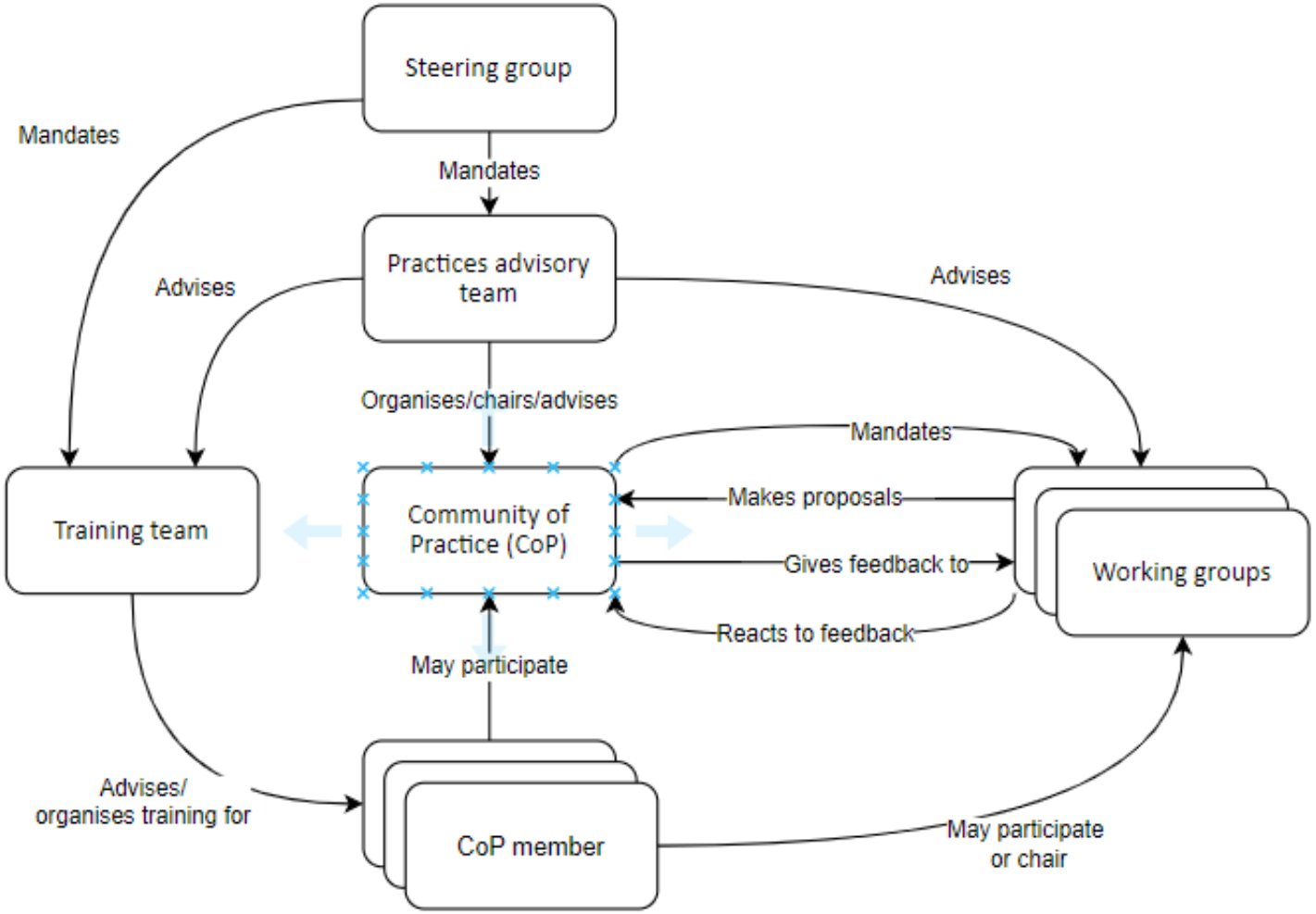
- Clear separation of responsibilities:
  - Who decides the **governance policy** itself?
  - Who **mandates** groups to work on shared structural metadata?
  - Who **reviews** and **approves** changes to structural metadata?
- Clarity on business processes workflows
  - What are the business processes for **maintenance**?
  - How can users **request changes**?
  - What are the **maintenance lifecycle, review periods** and communication lines?

Why is governance needed?

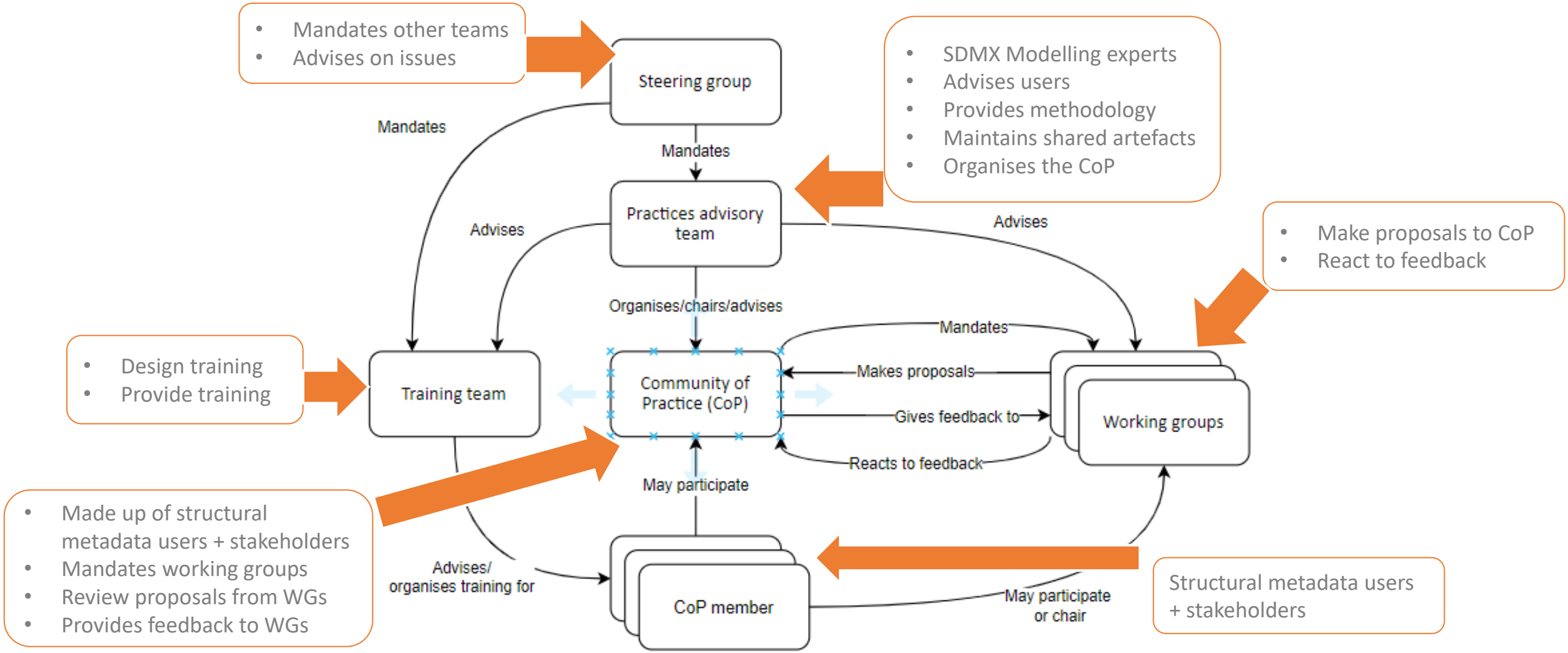
- ✓ Needed to meet **quality**, **coherence** and **harmonisation** requirements
- ✓ Helps stakeholders know the **processes** and their **responsibilities**
- ✓ For data integration **accessibility** and **processing**

A reference framework was developed and available in the SDMX content-oriented guideline: Reference Framework for SDMX Structural Metadata Governance

The reference architecture



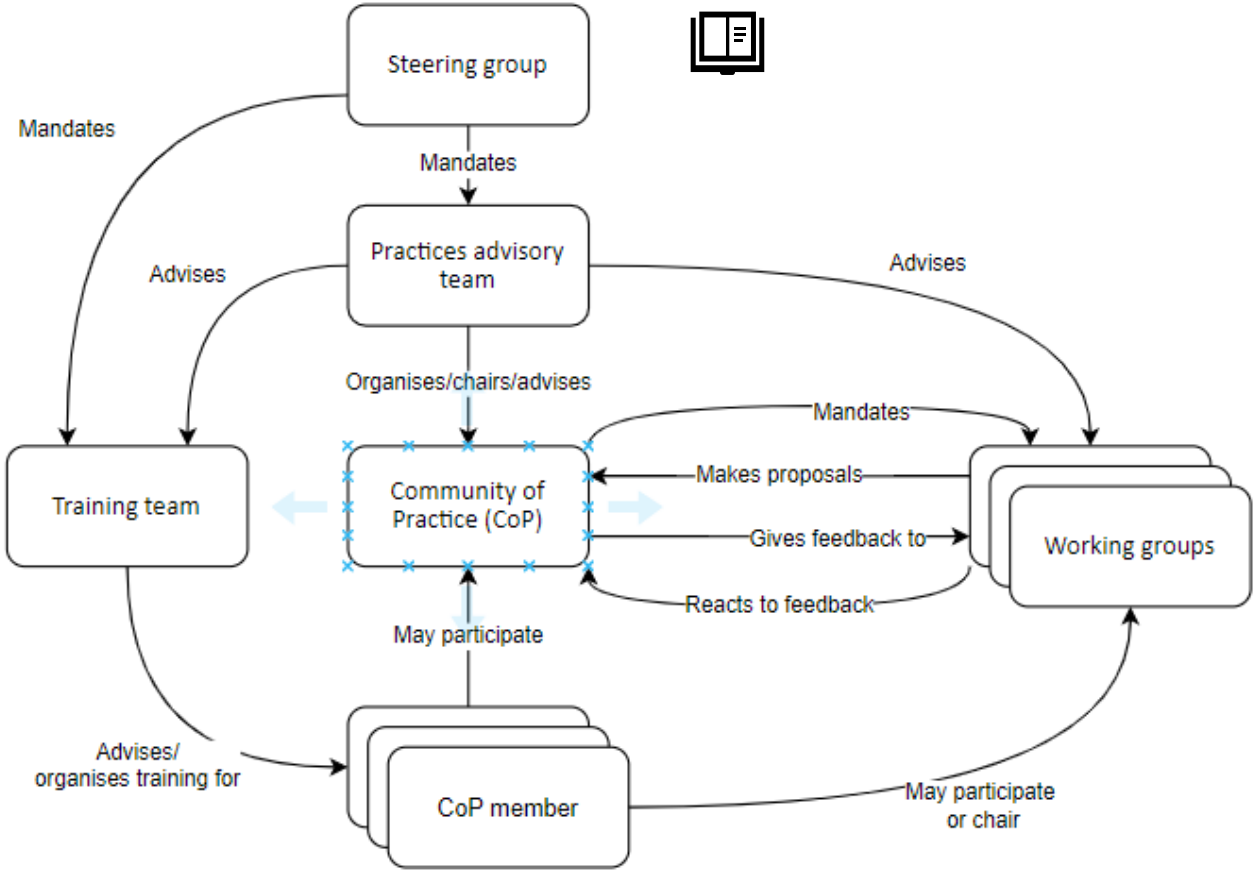
Roles in governance architecture



Variations of the reference framework for implementation

Fully distributed decisions and authority (reference architecture)

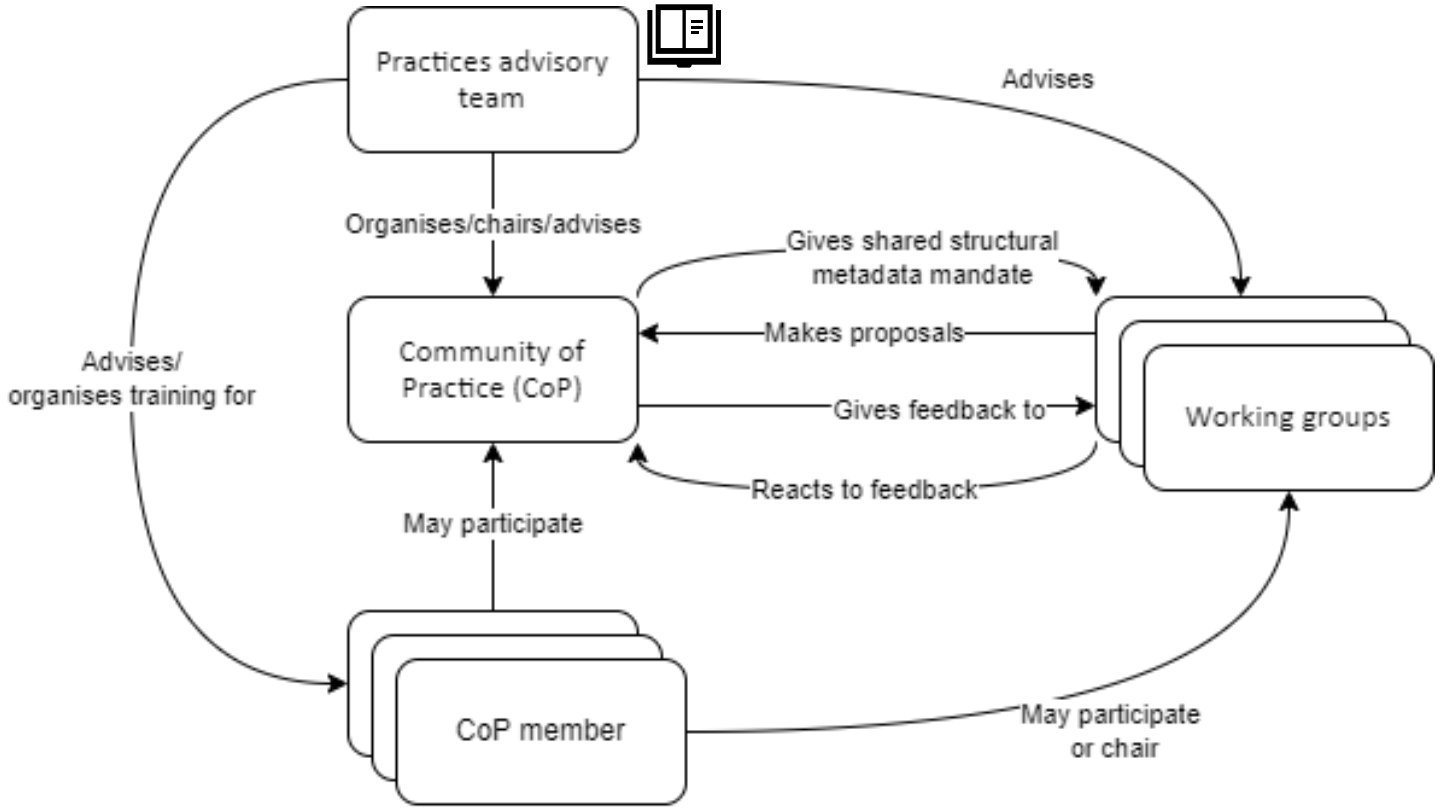
This architecture offers the most quality, reuse, and communication. The trade-off is more resources required and longer time to deliver artefacts and revisions.



Indicator	Score
Metadata quality	●●●●●
Metadata reuse	●●●●●
Agility	●●
Efficiency	●●

Balanced decisions and authority

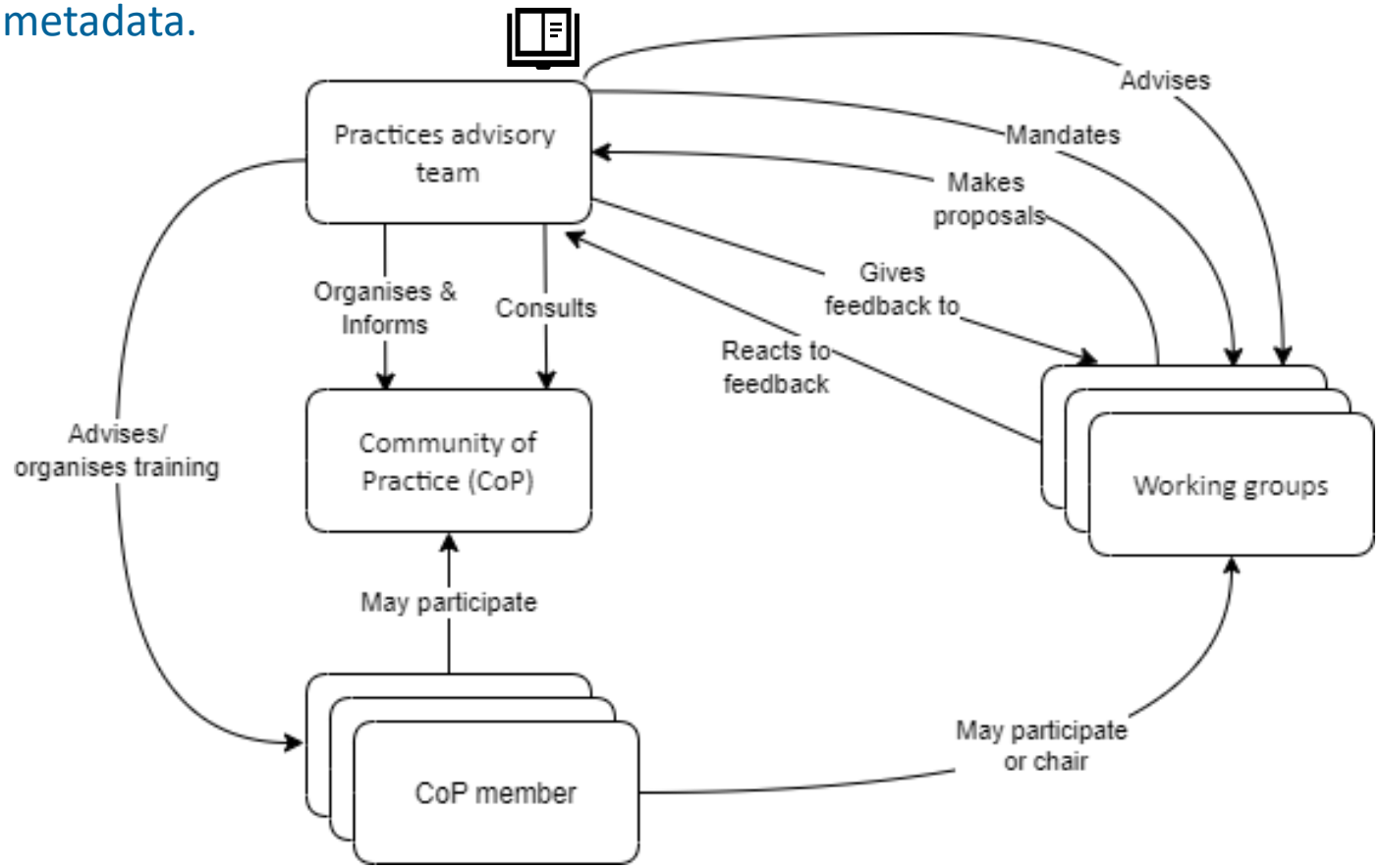
Steering and Training responsibilities are handled by Practices advisory team. A trade-off in quality because there are less checks, and the advisory team must juggle several roles.



Indicator	Score
Metadata quality	●●●●
Metadata reuse	●●●●●
Agility	●●●
Efficiency	●●●

Centralised ownership and maintenance with community

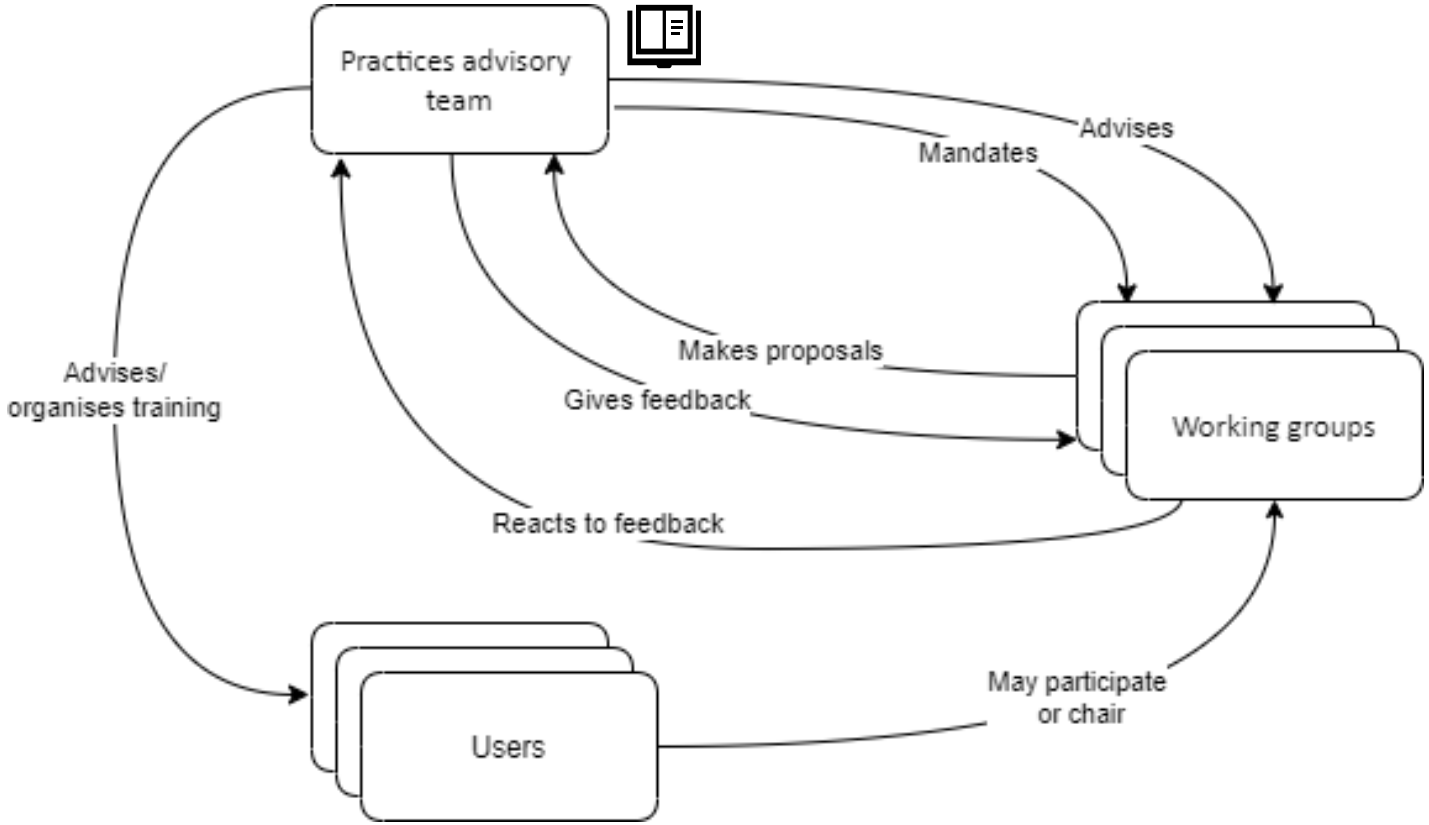
Practices advisory team is the owner rather than CoP. This architecture reduces resources and time for managing the CoP. Practices advisory team must have the expertise, capacity and be mandated to make decisions on the shared metadata.



Indicator	Score
Metadata quality	●●●
Metadata reuse	●●●
Agility	●●●●
Efficiency	●●●

Centralised ownership and maintenance without community

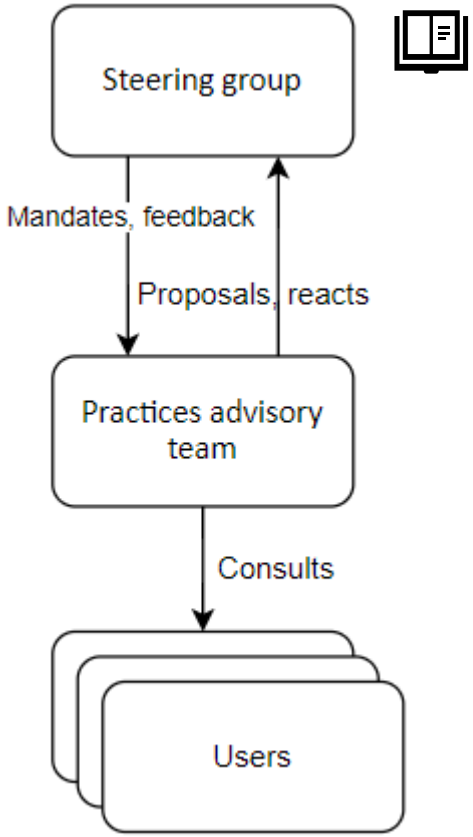
CoP is removed. Working groups make proposals directly to the Practices advisory team who make the final decisions.



Indicator	Score
Metadata quality	●●
Metadata reuse	●●
Agility	●●●●
Efficiency	●●●●

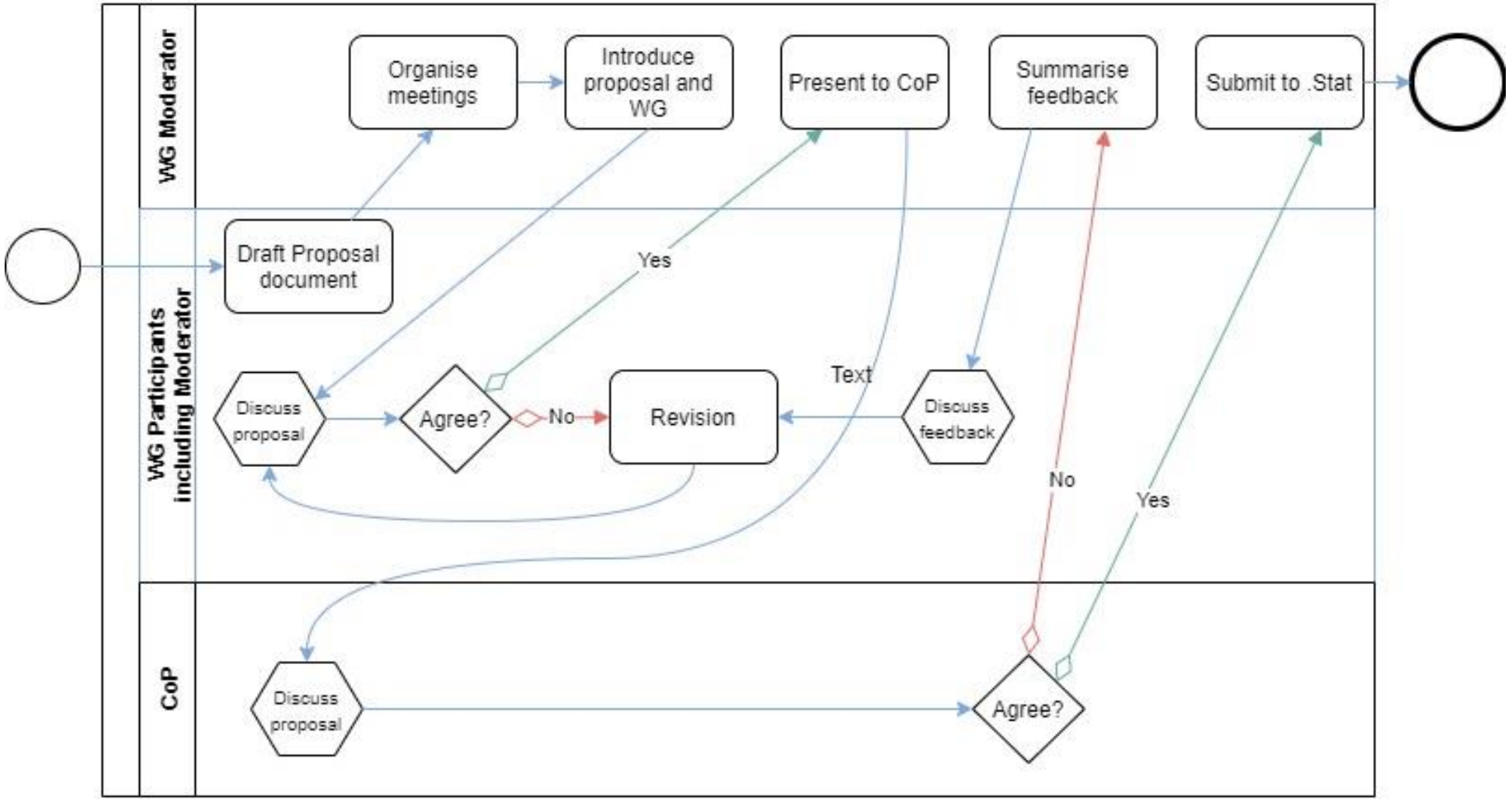
Sole authority and decision making

Practices advisory team is sole author and owner. Risk of reduced quality as no formal consultation on changes. Reuse is still possible because users are notified of shared metadata.



Indicator	Score
Metadata quality	●
Metadata reuse	●●
Agility	●●●●●
Efficiency	●●●●

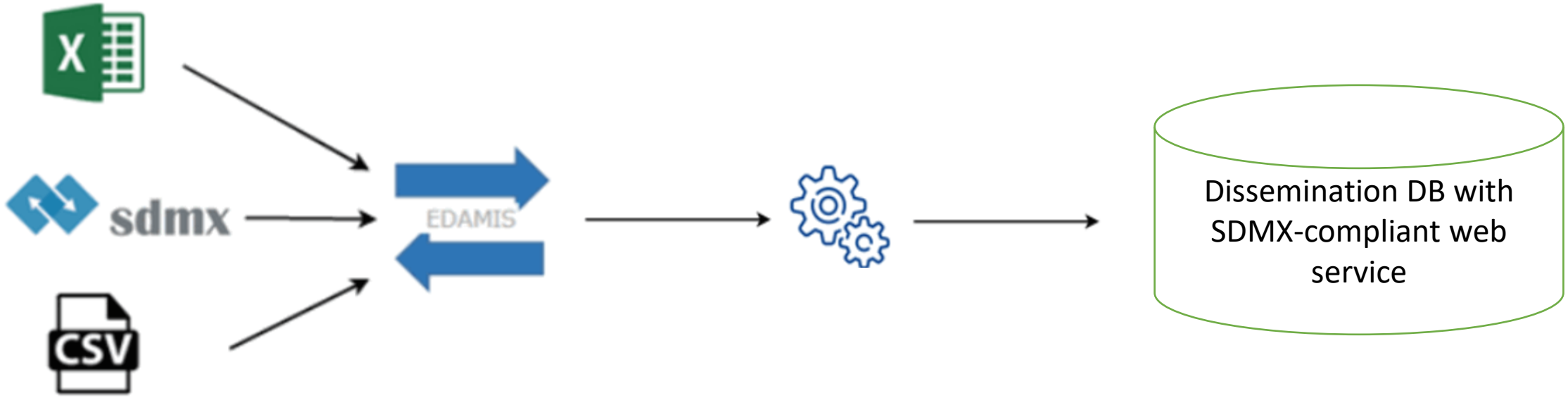
Business process for working groups



Governance Architecture in practice

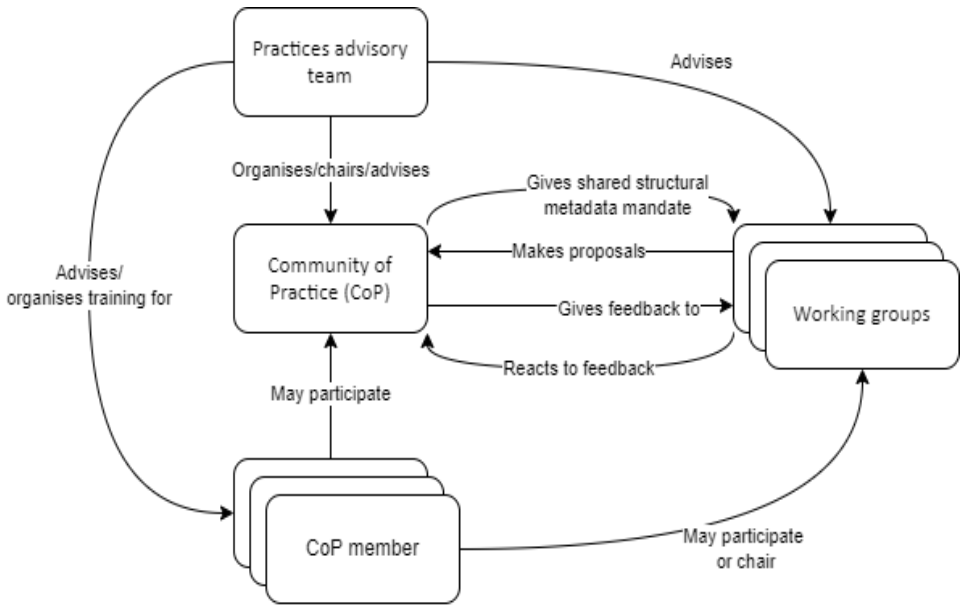
The Eurostat experience

SDMX in EUROSTAT



Eurostat uses SDMX for both data collection and data dissemination – and the different environments present different challenges

Eurostat – Data DISSEMINATION

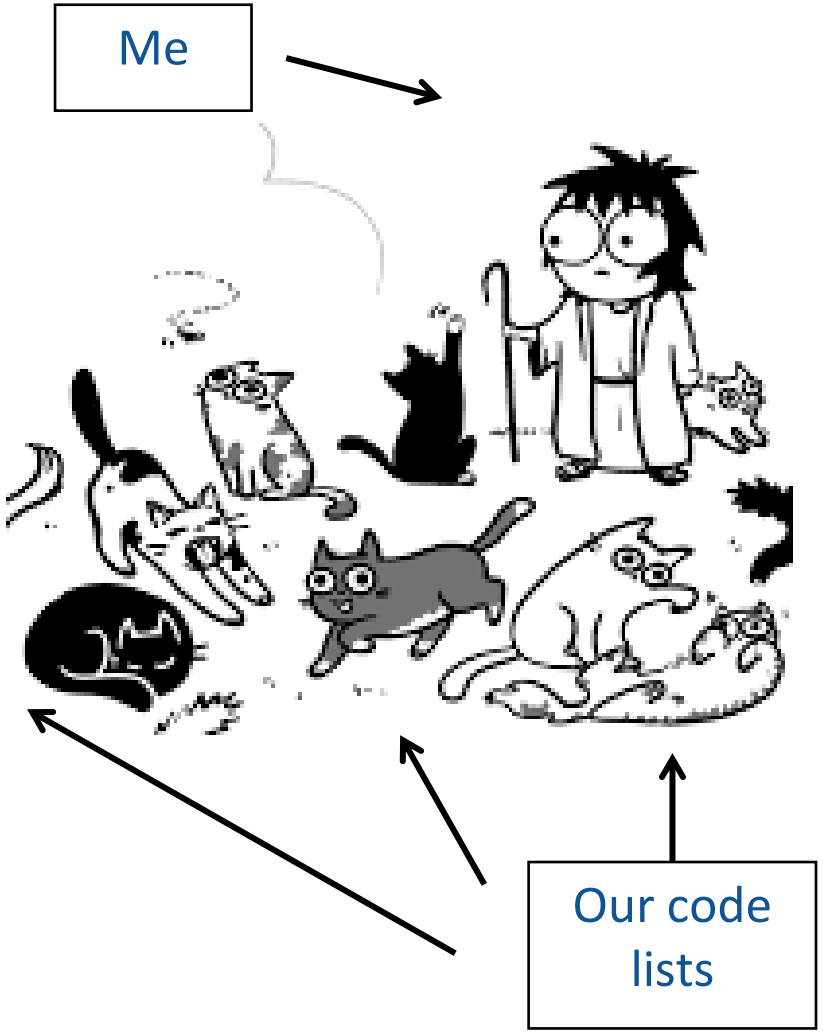


Eurostat dissemination is characterised by:

- A high control over the data sources (data comes from Eurostat's different production systems)
- A high emphasis on harmonisation of structural metadata, in order to present comparable data to our users.

A balanced governance model, with a high degree of collaboration between different groups within Eurostat, is the best model to support the dissemination goals in our environment and has been gradually adopted since ~15 years.

Eurostat – Data COLLECTION

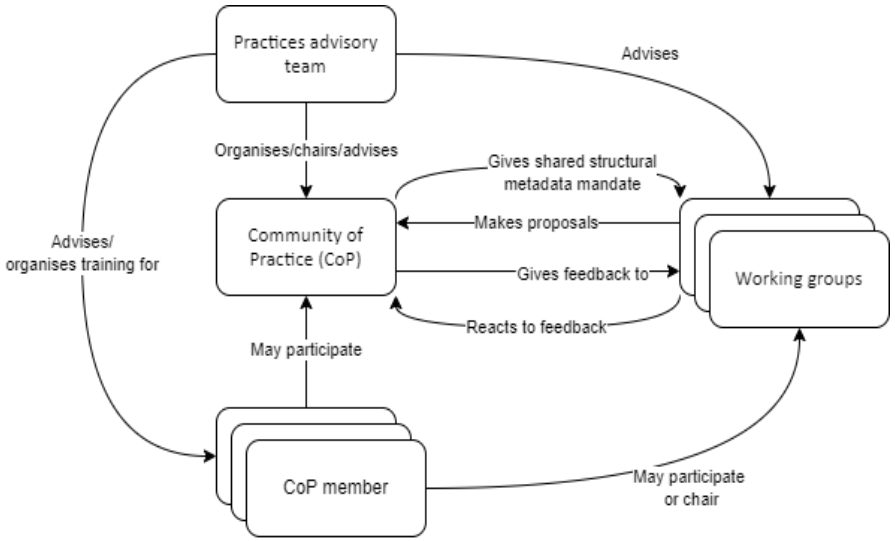


Data collection is characterised by a much more dynamic environment:

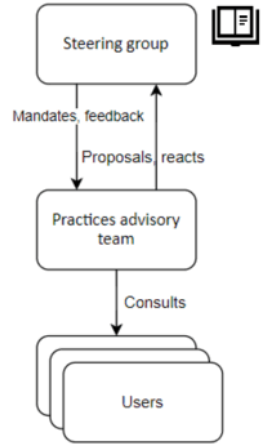
- High diversity of providers and legal basis.
- More frequent changes in data needs, experimental data collections.

Eurostat – Data COLLECTION

As a result, different governance models in different domains



More centralisation and harmonisation for established domains with high level of buy-in



More agile approach with less emphasis on harmonisation in developing domains

Key message

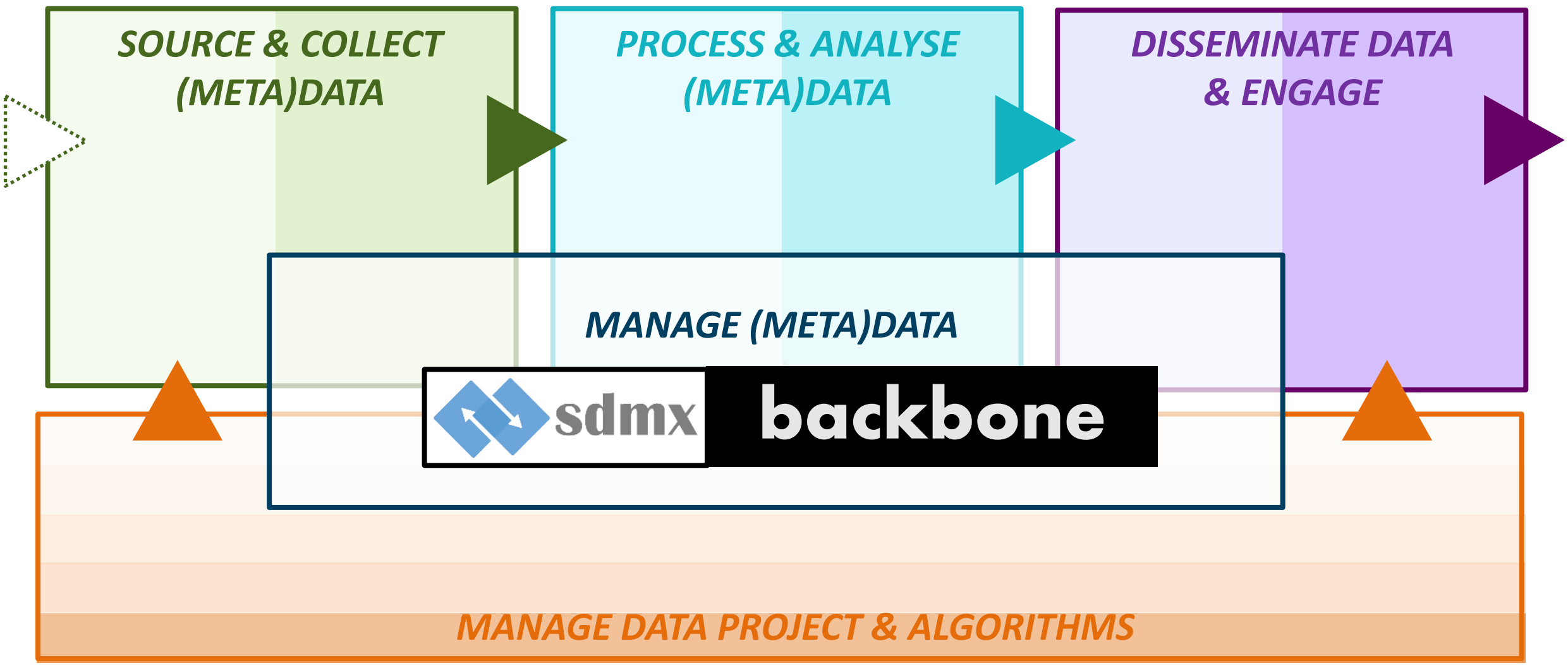
- **The “right” governance model depends on your goals** and on the conditions under which data exchange and data dissemination occurs

- The reference framework provides guidance for the “right” model depending on your goals. The **SDMX standard also provides a common language to underpin the governance.**

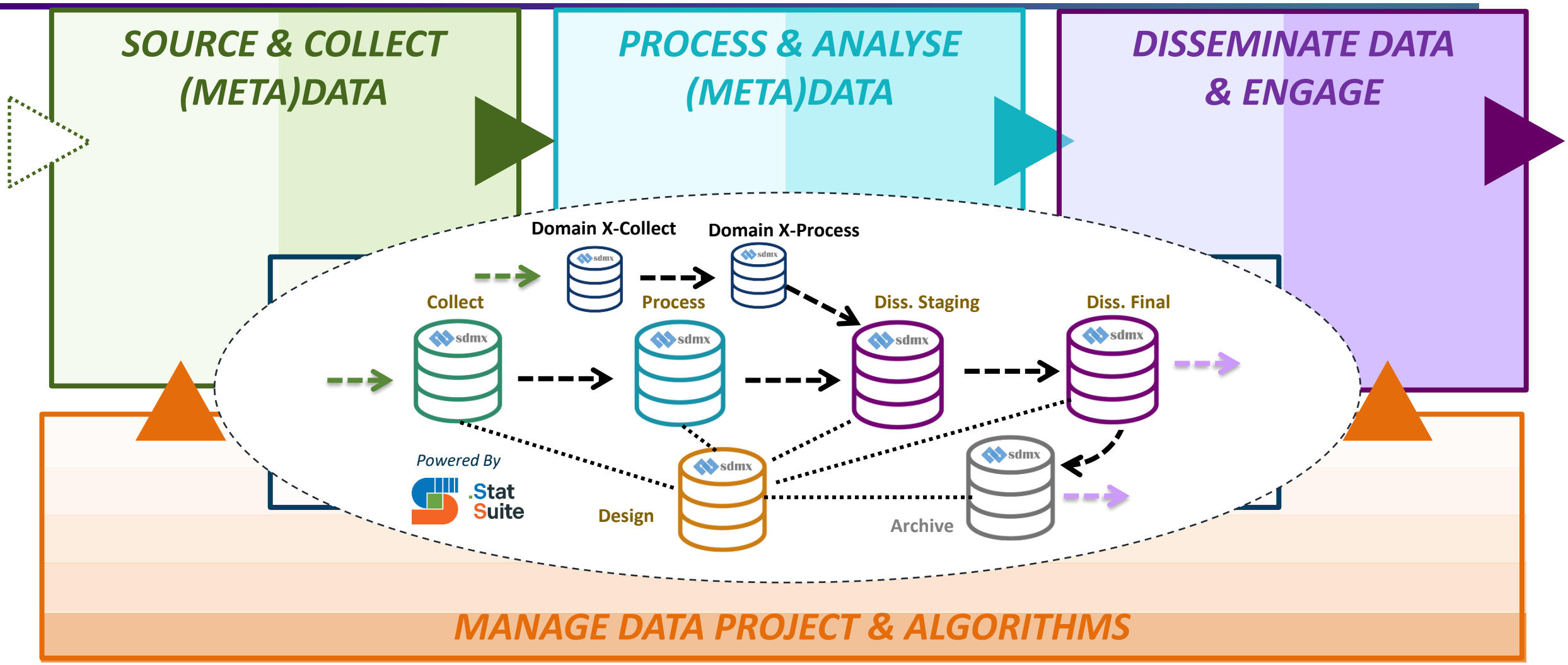
- **Adopt the best governance model for your circumstances –** and even within the same organisation, different models may be needed for different cases

Integration with Data Mesh

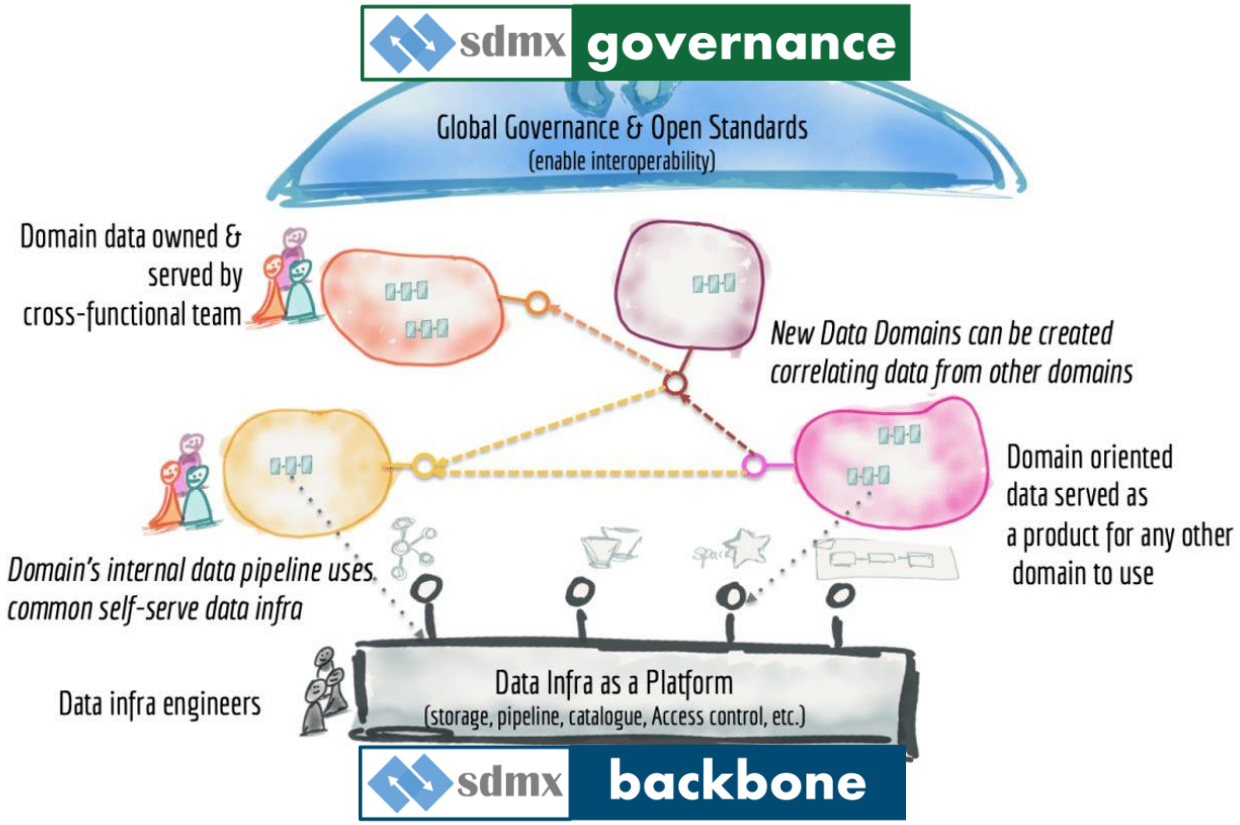
The #SDMX Backbone



The #SDMX Backbone



#SDMX as a #DataMesh enabler?



Source: Data Mesh, O'Reilly, Z. Dehghani

DATA MESH…

…strike a balance between team autonomy and interteam interoperability and collaboration.

…choosing the best data model for their data products…

… standardising the data modelling language that all domains utilize…

domain-agnostic data platform in place for teams

9th SDMX Global Conference

Empowering Data Communities

Thank you

David Barraclough, OECD Smart Data Practices Manager

SDMX Global Conference 2023 | Kingdom of Bahrain