Data sharing using a global data registry
On a place to discover global structured time series, macro and micro data

3rd IFC Workshop on Data Science in Central Banking

Glenn Tice & Matt Nelson (BIS MED IT)

The requirement for global structured data discoverability and access

The Problem

Public structured data is hard to find

- Data scientists need data
- Large global resource of structured data published by institutions worldwide

But

- No single place to discover what's available and how to get access to it

Existing Landscape

Many excellent data services already exist, but are variously regional, domain specific or not tuned for structured data

- EU open data portal
- US data.gov
- Institutional data portals
- European Single Access Portal (planned)
- Emerging initiatives (e.g. Google Data Commons)

The requirement for structured data discoverability and access

Global

Easily discover what datasets are available

Sufficient information to evaluate the suitability of a dataset

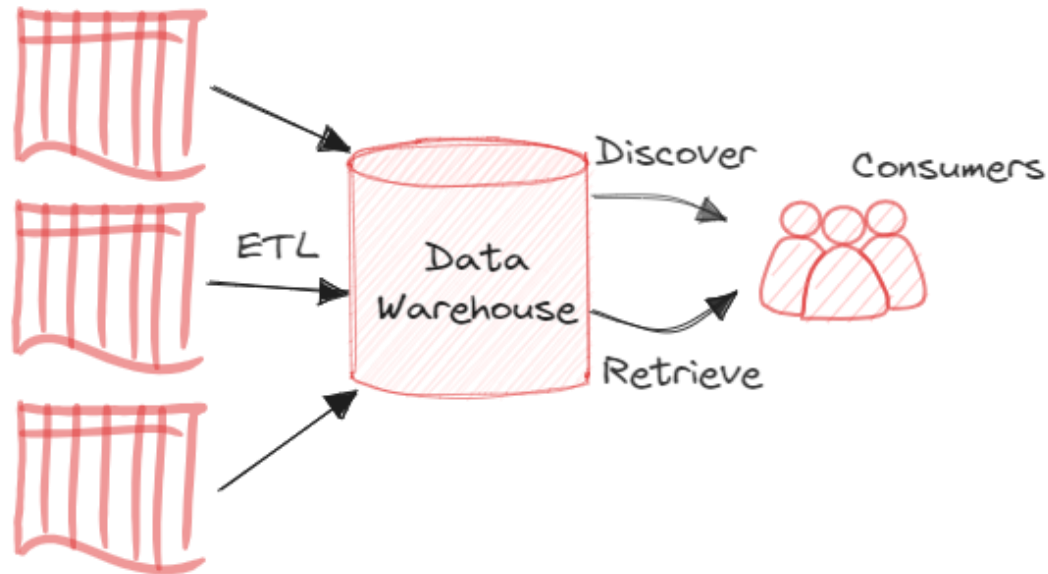Confidently interpret and use datasets

Tuned for structured datasets

Easily retrieve selected datasets
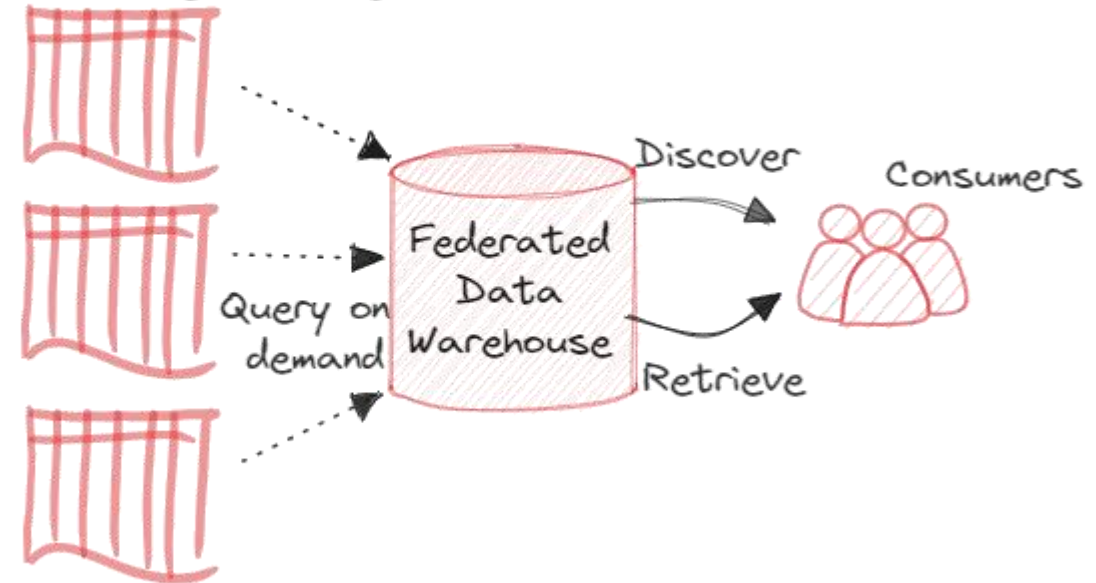
UI for interactive use

API for automation

Options – single access point

Data consolidated into a central warehouse
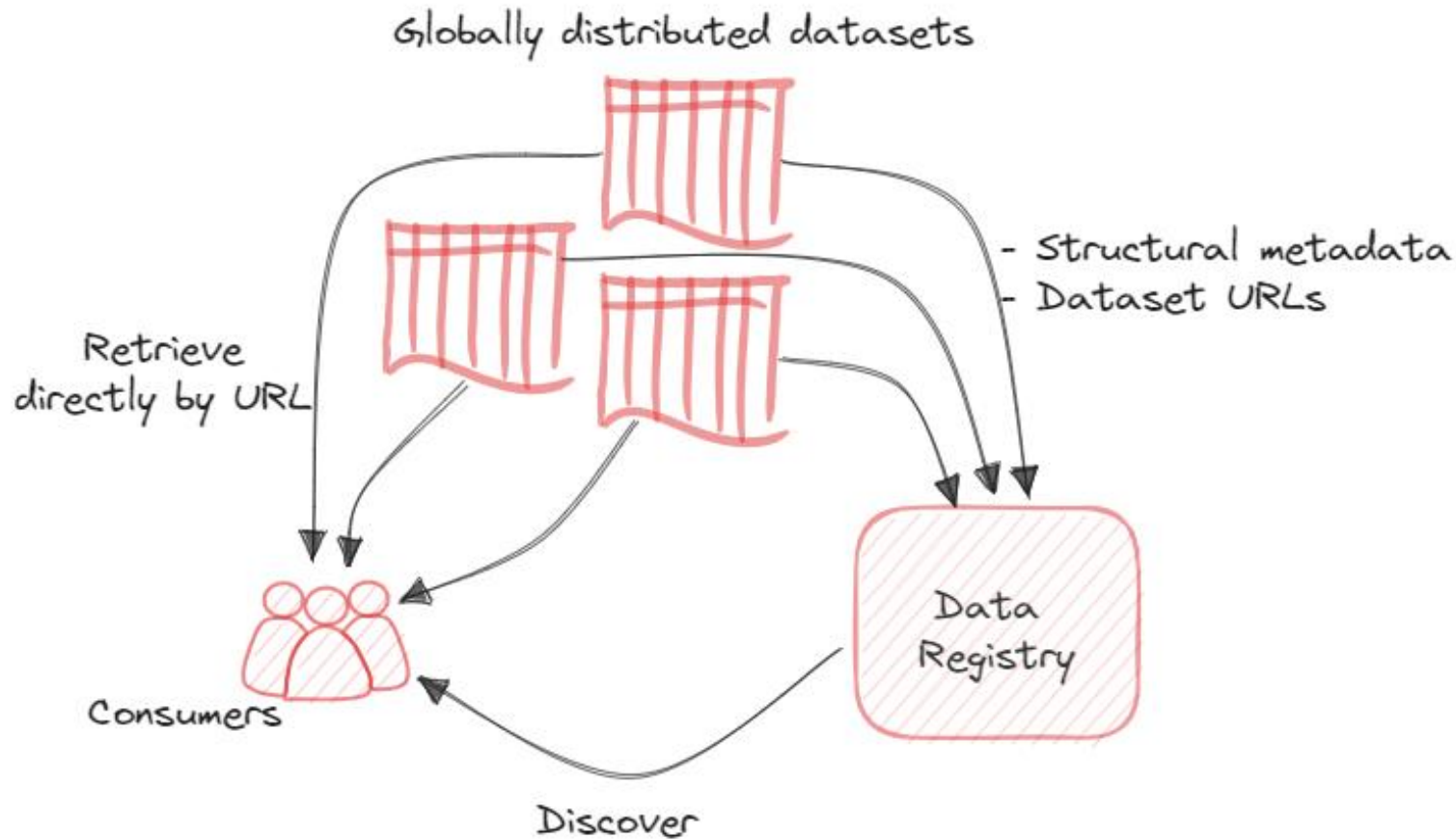
Federated data warehouse

A single access point is attractive but potentially complex and costly to implement

A 'data registry' provides a simpler alternative approach

Globally distributed datasets

- Structural metadata
- Dataset URLs

Retrieve directly by URL

Consumers

Data Registry

Discover

- Catalogue of <u>datasets</u> and where to find them
- Each dataset has:
  - Detailed structural model
  - Data URL
- Decentralised – follows Data Mesh principles
- Technically simpler
- Legally simpler - risk of third party IP or personal data privacy claims reduced

Concrete solutions

Data modelling

SDMX 3.0

- Data modelling
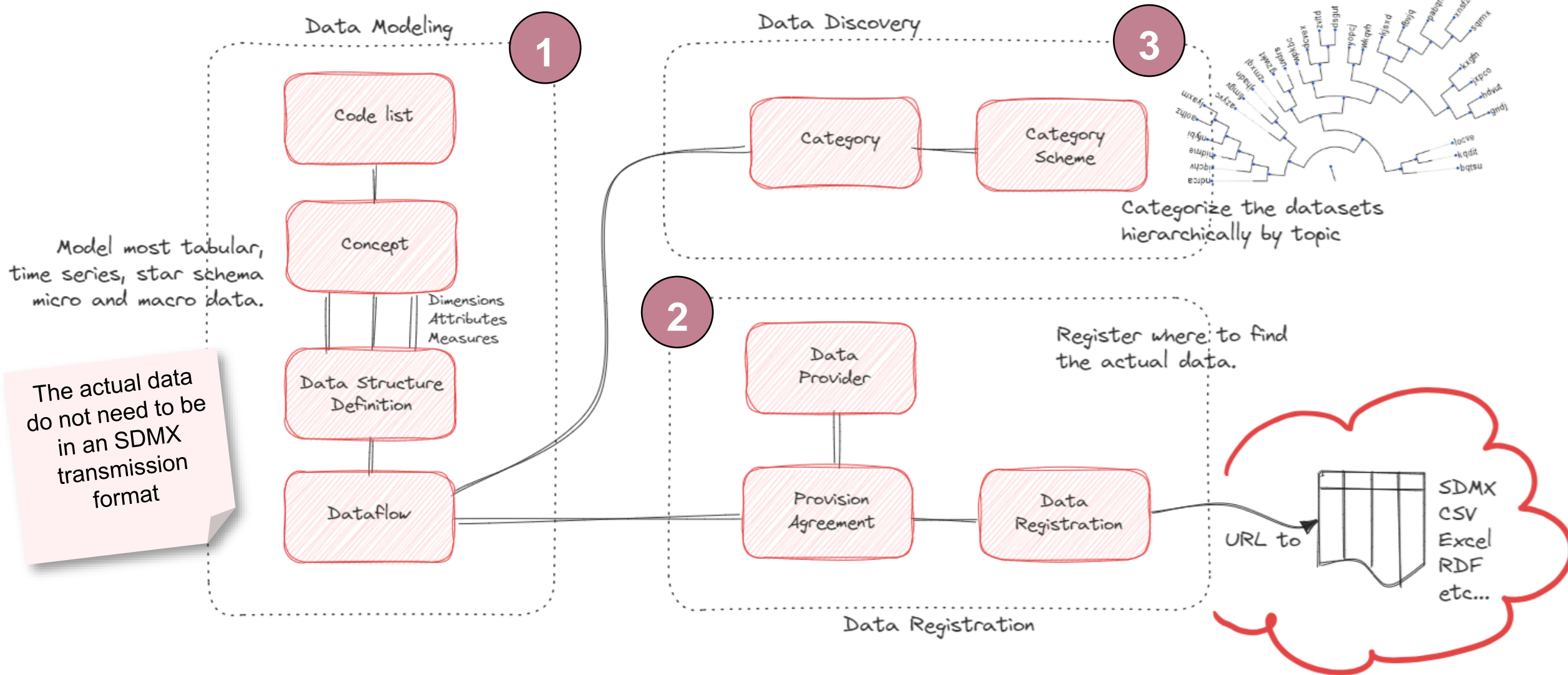- Data discovery
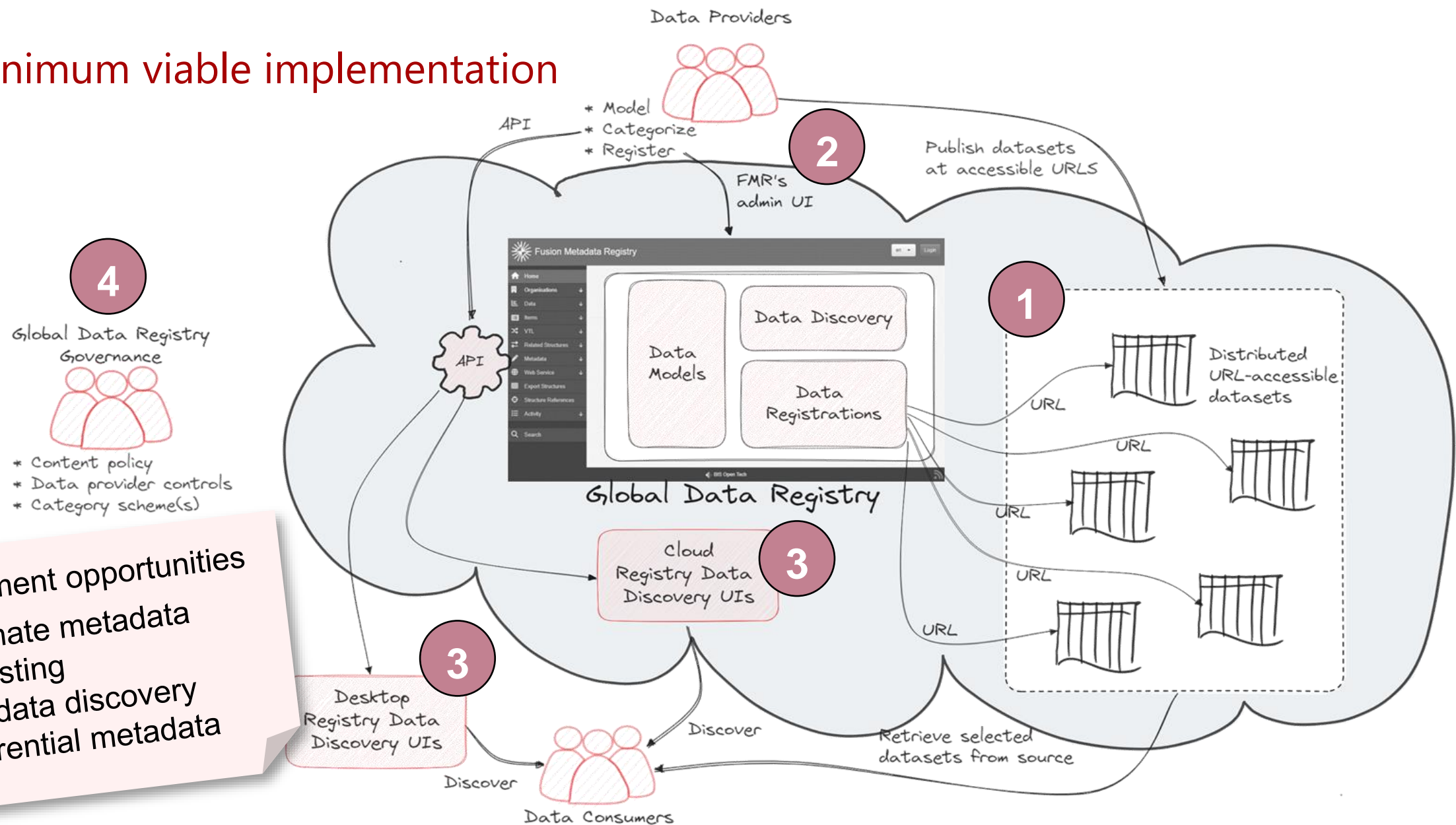- Data registrations – URLs link to the data

Software

FMR 11

- SDMX 3.0 structural metadata registry
- Free and open source
- Cloud native
- BIS owned and maintained

Using SDMX 3.0 for the global data registry use case

Data Modeling

1

Code list

Concept

Model most tabular,
time series, star schema
micro and macro data.

Dimensions
Attributes
Measures

Data Structure
Definition

The actual data
do not need to be
in an SDMX
transmission
format

Dataflow

2

Data Discovery

3

Category

Category
Scheme

Categorize the datasets
hierarchically by topic

Register where to find
the actual data.

Data
Provider

Provision
Agreement

Data
Registration

URL to

SDMX
CSV
Excel
RDF
etc...

Data Registration

A minimum viable implementation

Conclusion

Plenty of public structured datasets published by institutions globally, but **hard to find and retrieve.**

The **global <u>data registry</u>** approach provides a centralised place to discover and retrieve datasets without the complexity of a single access point.

A practical implementation could be based on **SDMX 3.0** and existing software tools like **Fusion Metadata Registry.**

Is there **demand** from data consumers?
More consultation needed.

Thank you

Glenn Tice   glennphilip.tice@bis.org
Matt Nelson   matt.nelson@bis.org

BIS MED IT

References

About Fusion Metadata Registry (FMR)   Link

FMR quick start using Docker   Link

European open data portal data.europa.eu   Link

Banca d'Italia public Aggregated Data – Statistical Database   Link

US open data portal data.gov   Link

Google Data Commons   Link

European Single Access Point (ESAP) regulation   PDF Link

SDMX data registrations API query example (returns XML)   Link