



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

SDMX in the S-DWH » Layered Architecture

Statistics Portugal
Department of Methodology and Information System
Information Infrastructure Service

Sónia Quaresma



Sep 2015 • Workshop of the CoE • Dublin



Overview

- **Presenting SDMX**
- **GSBPM Model**
- **Data Warehouse Layers**
- **Mapping SDMX uses on GSBPM Model within a Layered Architecture**
- **SDMX tools in the Data Warehouse**

Statistical Data and Metadata Exchange

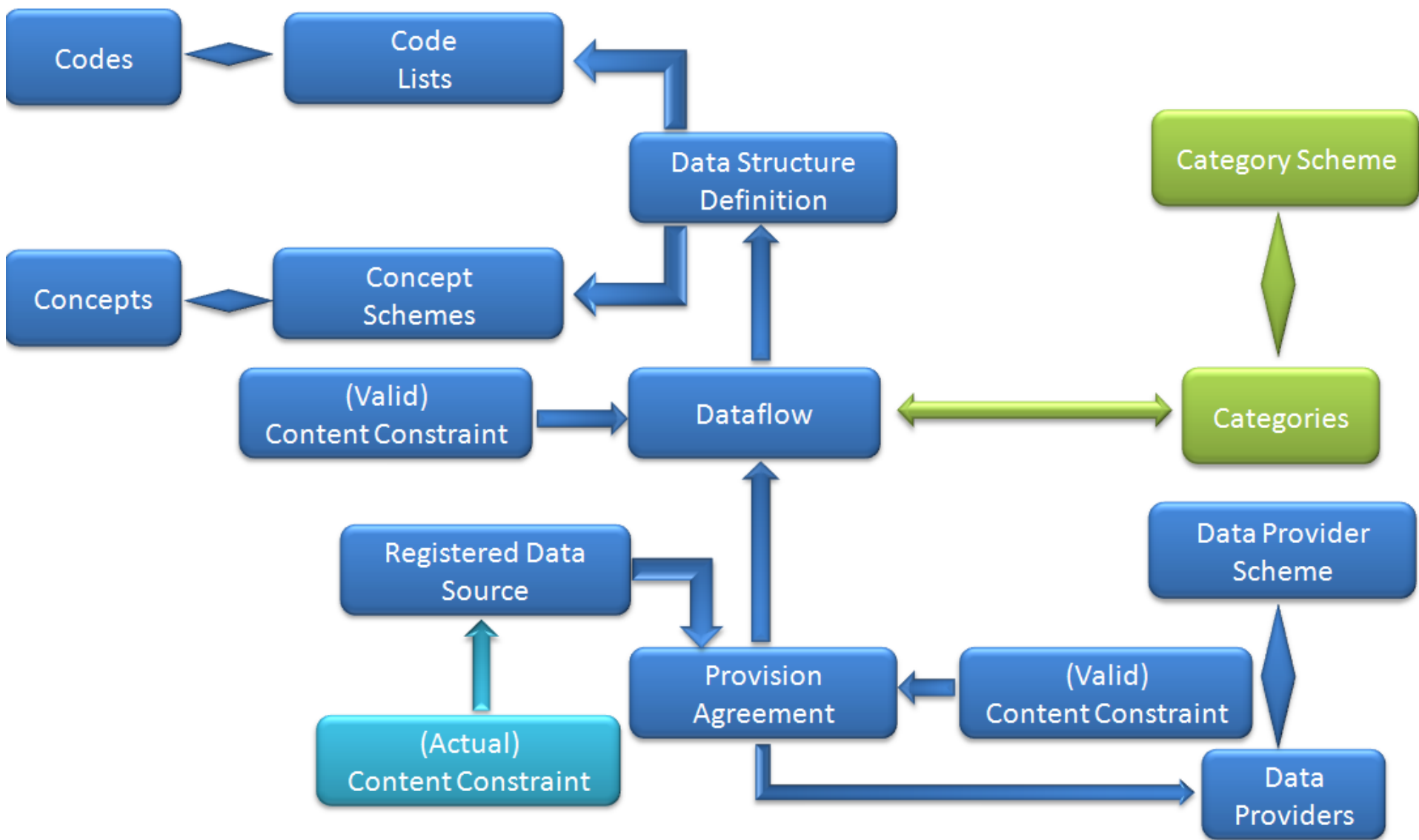
SDMX is an initiative from a number of international organizations, which started in 2001 and aims to set technical standards and statistical guidelines to facilitate the exchange of statistical data and metadata using modern information technology

Statistical Data and Metadata Exchange

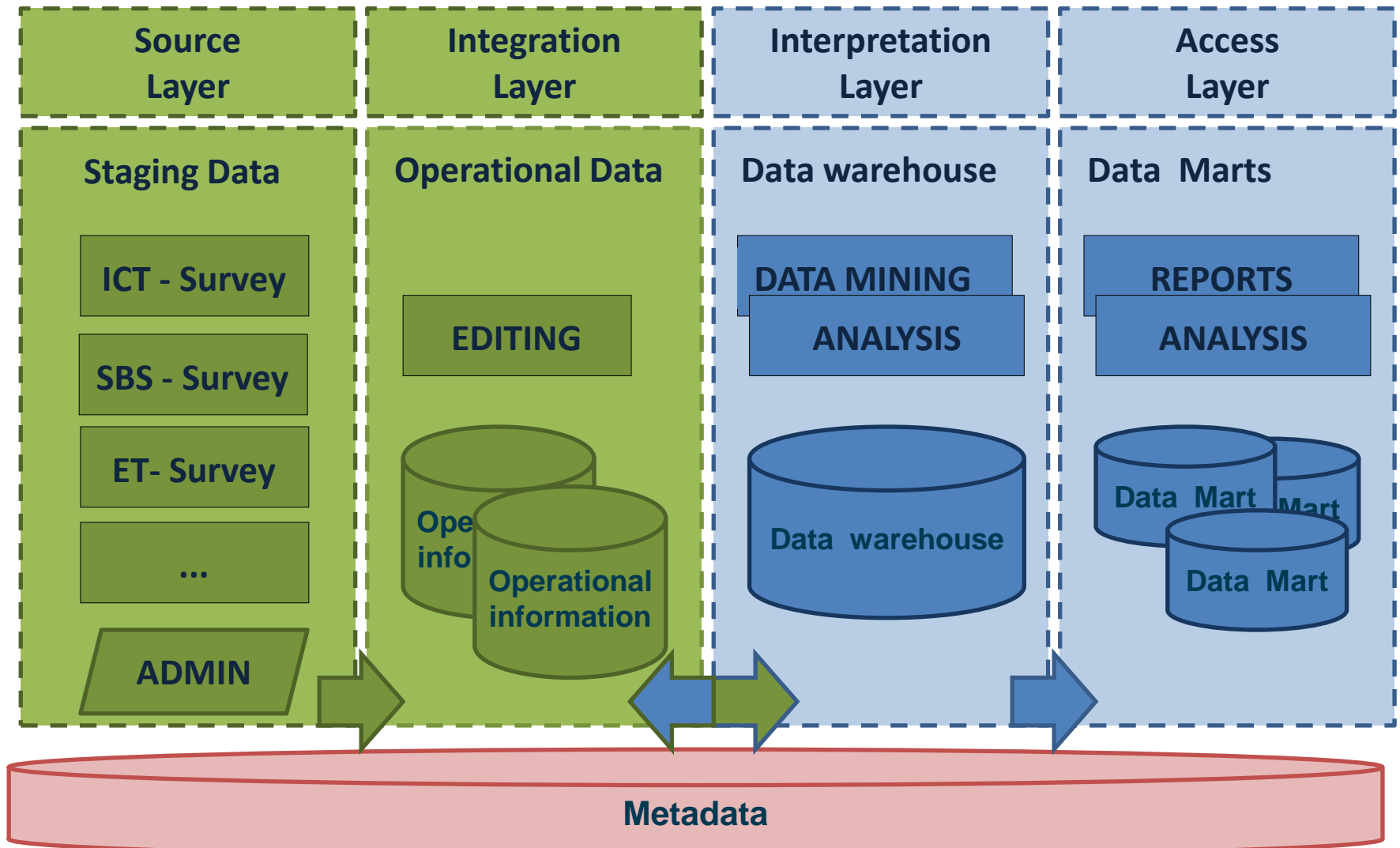
SDMX consists of:

- Technical standards (Inform. Model),**
- Statistical guidelines, and**
- IT architecture and tools**

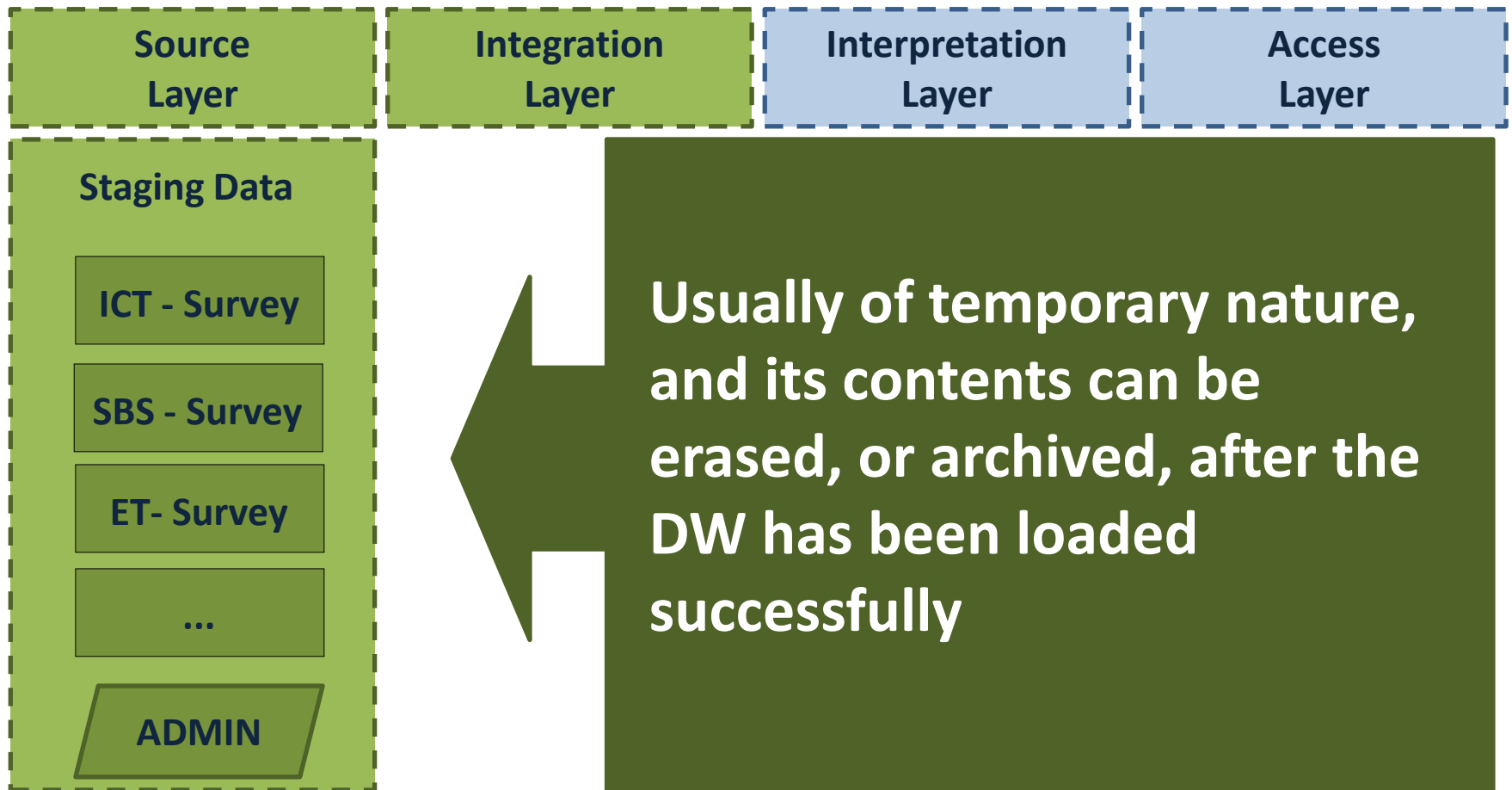
Statistical Data and Metadata Exchange



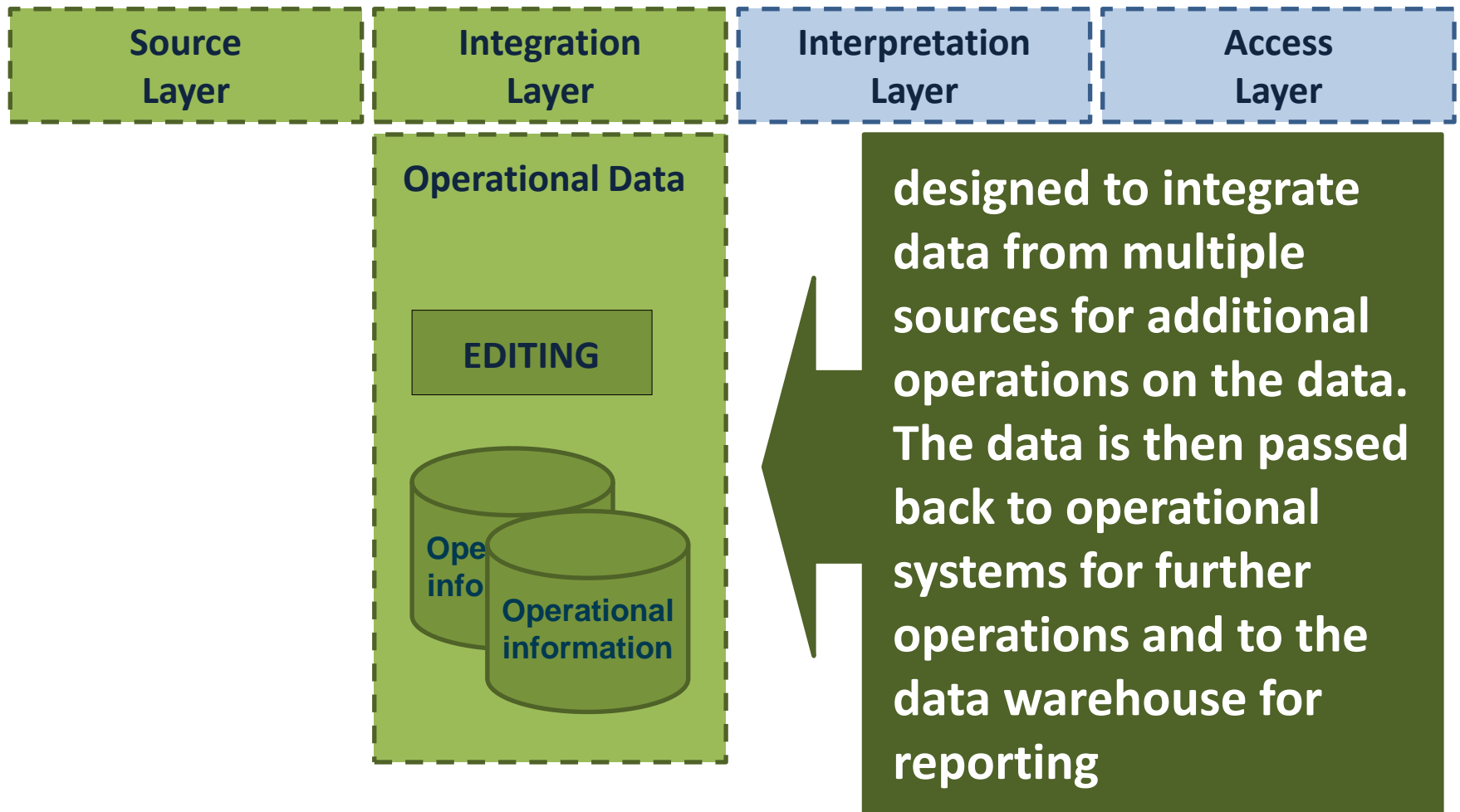
Information Systems Architecture



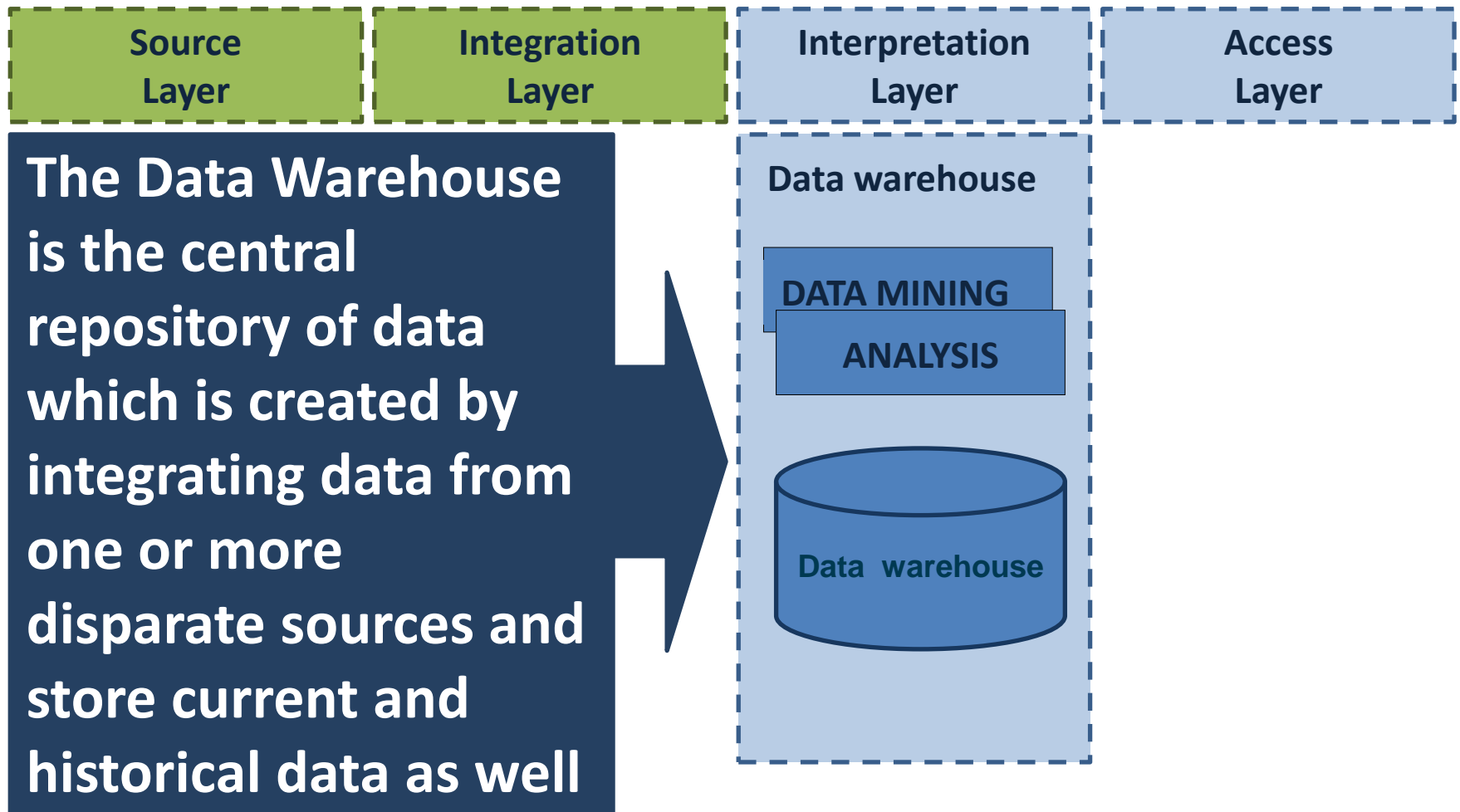
Information Systems Architecture



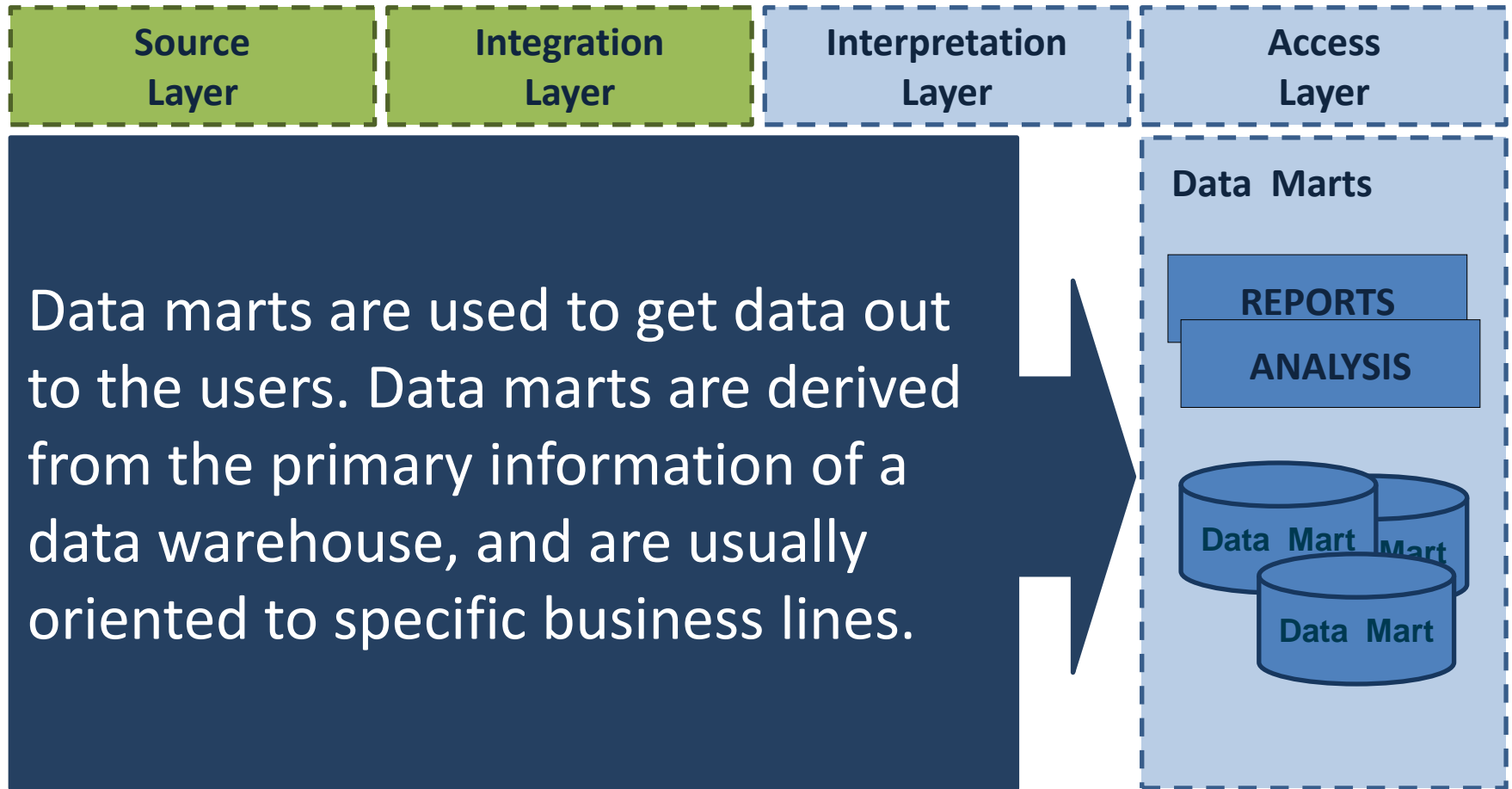
Information Systems Architecture



Information Systems Architecture



Information Systems Architecture



GSBPM Model

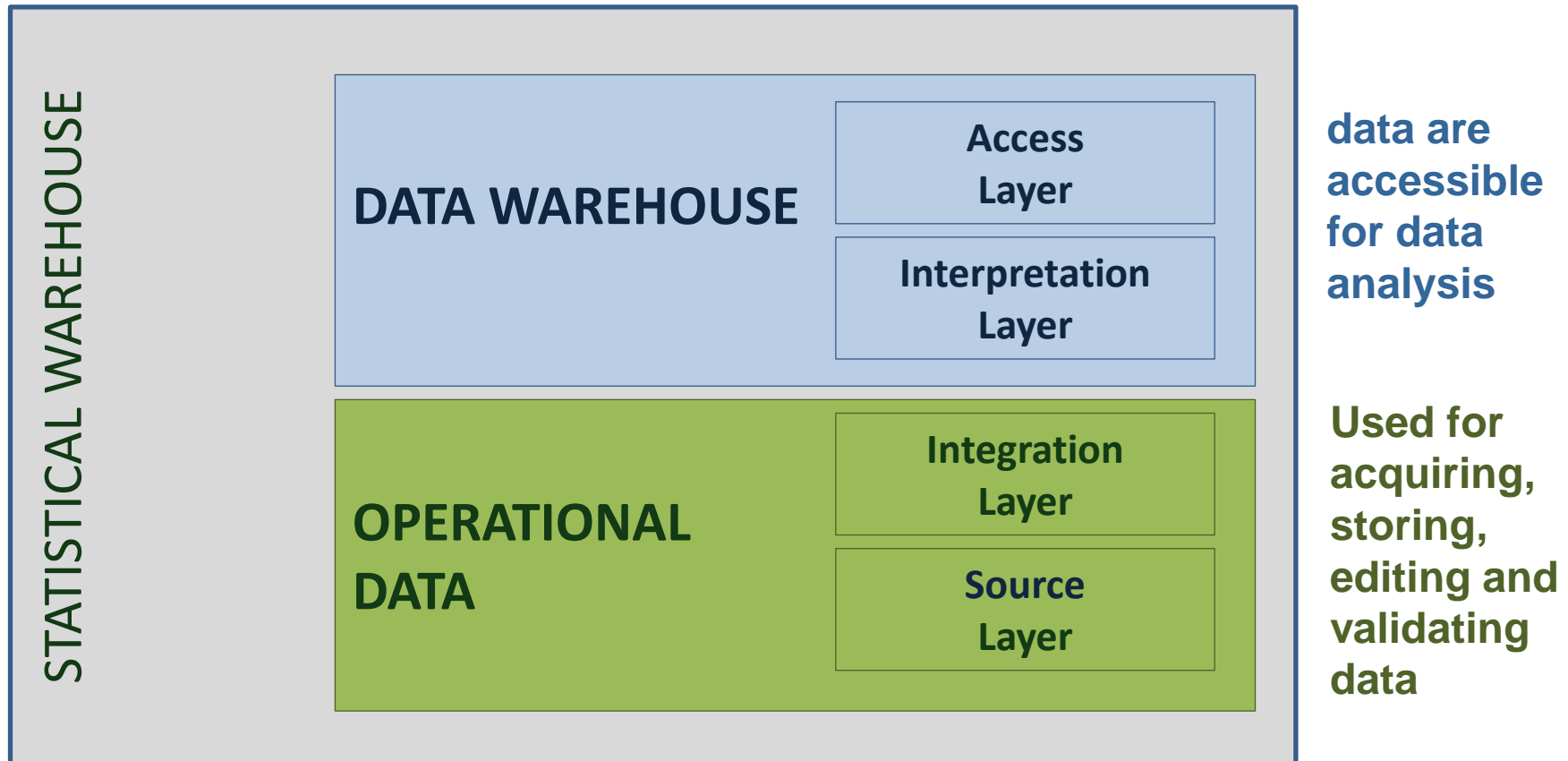
- **The Generic Statistical Business Process Model defines and describes statistical processes.**
- **It is a matrix and a strict order between it's sub processes does not exist.**

GSBPM Model

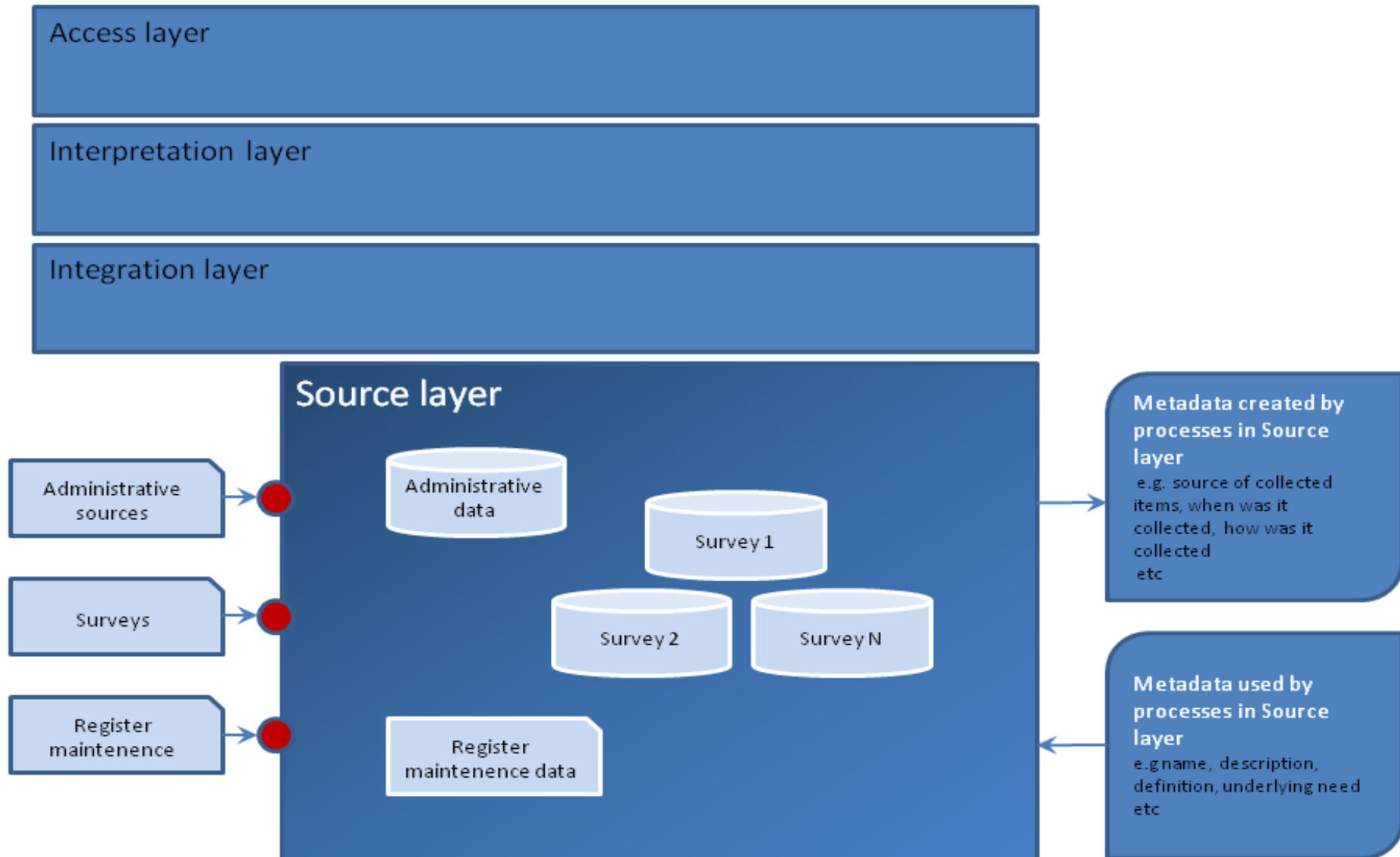
Quality Management / Metadata Management

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Layered architecture



Source Layer



SDMX and Collection Phase (Step4)

SDMX is more appropriate for Macro Data and as such does not relate directly to the source layer where Micro Data is primarily concerned.

SDMX and Collection Phase (Step4)

There are some exercises in which SDMX is being used with MicroData – data exchange exercise on Business Registers.

SDMX and Collection Phase (Step4)



**ISTAT metadata experts are modelling some
microdata sets with SDMX**

SDMX in the Source Layer

The structure of a dataset in SDMX is described using a Data Structure Definition (DSD), in which the metadata elements are:

- 1. dimensions, which form the identifiers for the statistical data**
- 2. attributes, which provide additional descriptive information about the data.**

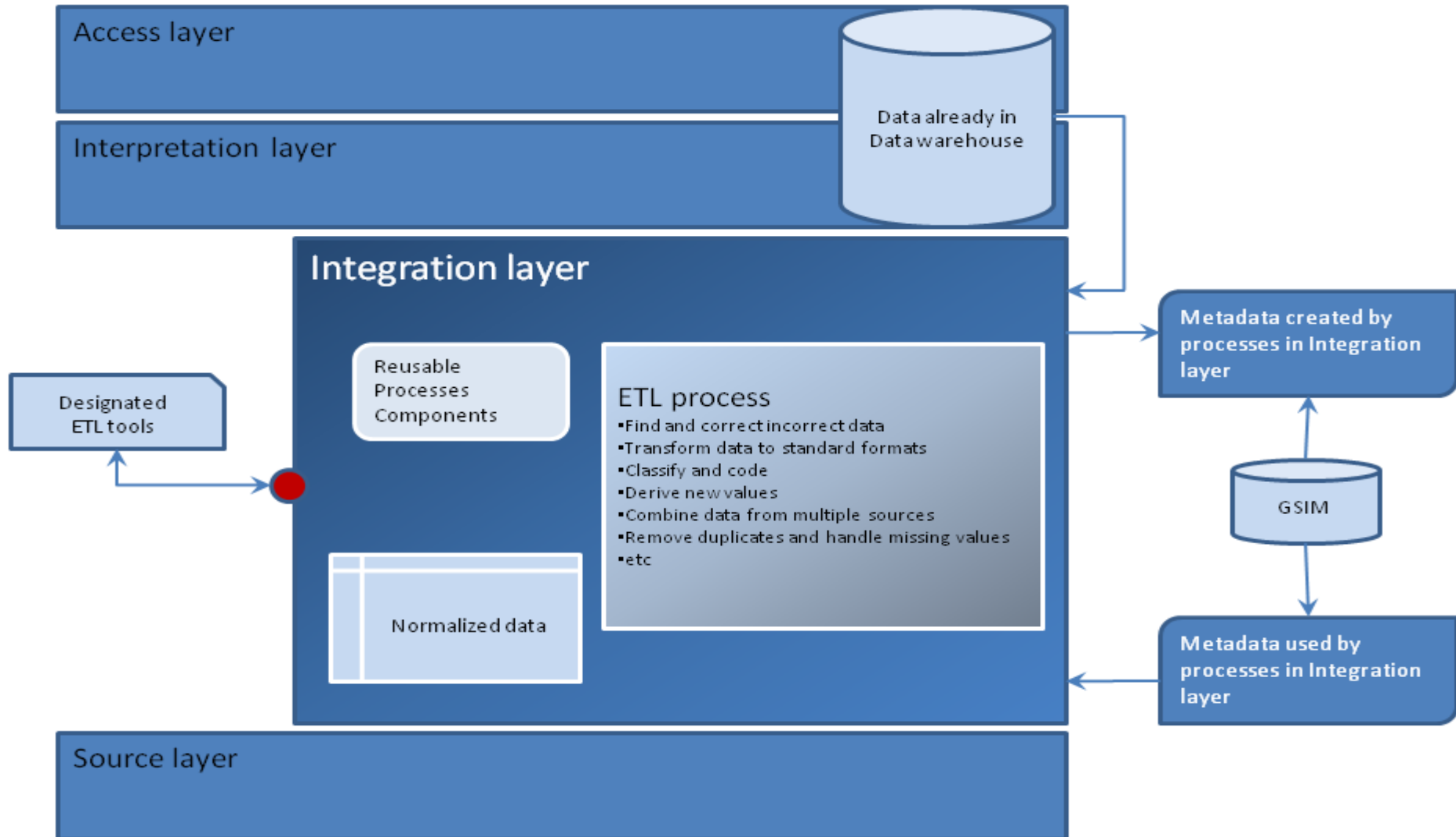
SDMX in the Source Layer

When several actors are involved it is easier to achieve coherence and guarantee integrity if all abide to the same DSD.

SDMX Potentialities

**SDMX as a Model for the Structures of the
Metadata Repository or/and the Statistical
DataWarehouse**

Integration Layer



SDMX and Process Phase (Step5)

Aggregates can be captured in a standard

SDMX format

5.7 Calculate Aggregates



No direct use of SDMX but derived variables and recodes must match the requirements of the standard DSD to ensure comparison

5.8 Finalise Data Files



Use of SDMX-ML DSD and data formats to format aggregates. There are several “flavours” of SDMX to create SDMX-ML data sets.

5.8 Finalise Data Files



“Flavours” of SDMX:

- SDMX-EDI (also known as GESMES/TS)**
- SDMX-ML (the XML version)**

5.8 Finalise Data Files

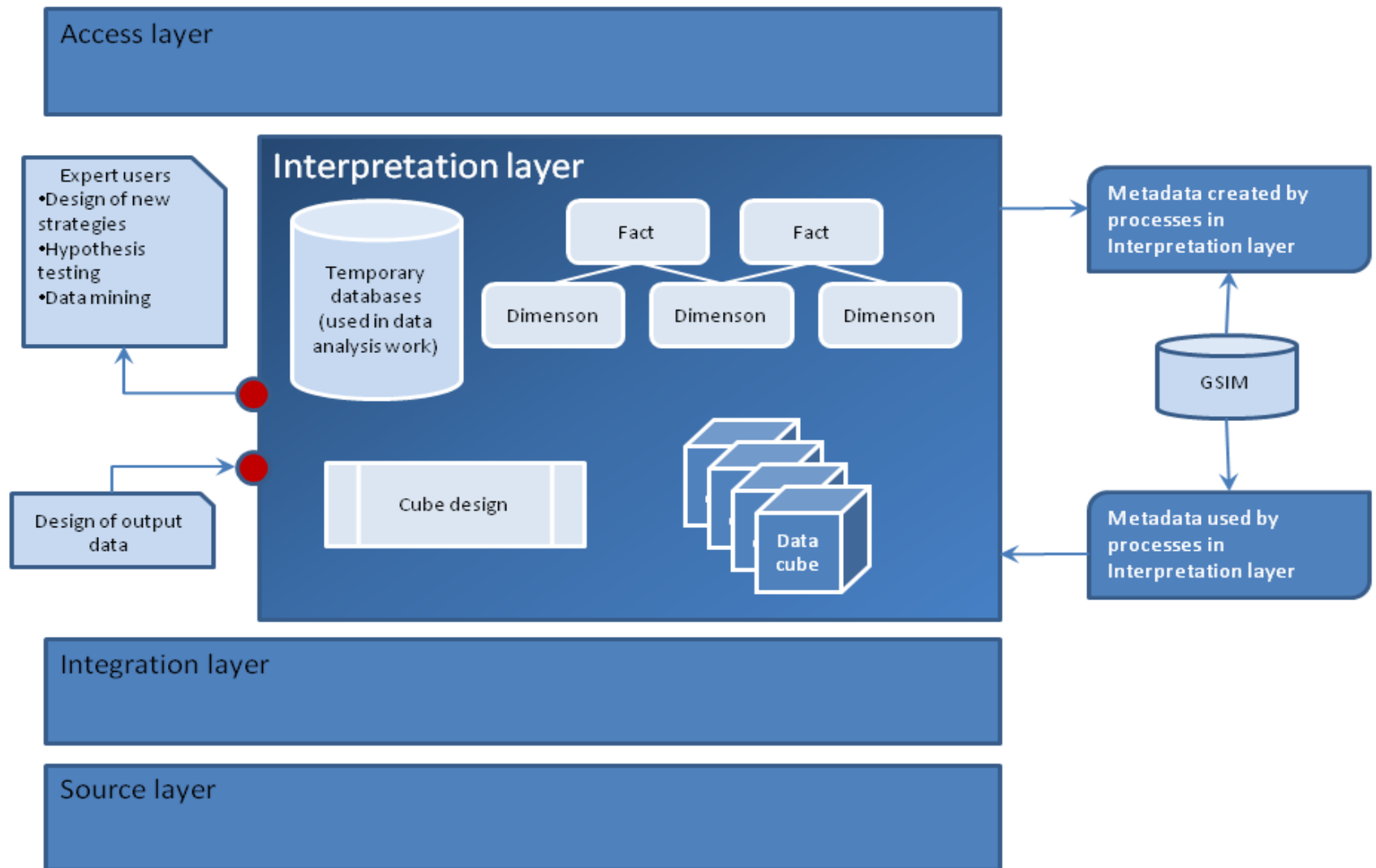
The SDMX Technical Working Group is extending the formats of SDMX:

- JSON – there's already a published proposal**
- CSV – could be a very important format for microdata exercises**

SDMX Potentialities

**SDMX usage for EXTRACTION,
TRANSFORMATION and LOAD of DATA**

Interpretation Layer

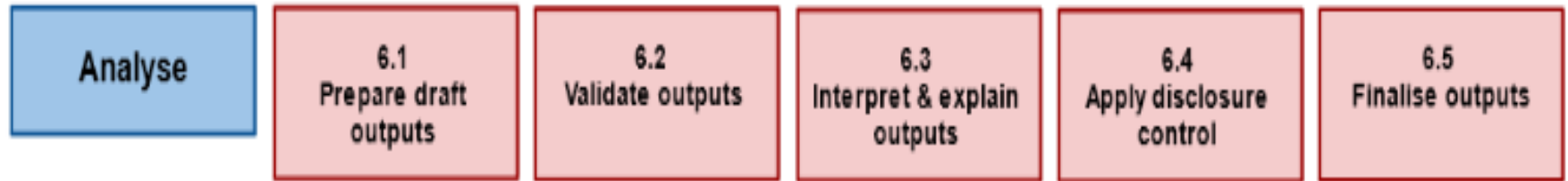


SDMX and Analyze Phase (Step6)

**SDMX can provide some useful functions
for the Analysis of Aggregates**



6.1 Prepare Draft Outputs

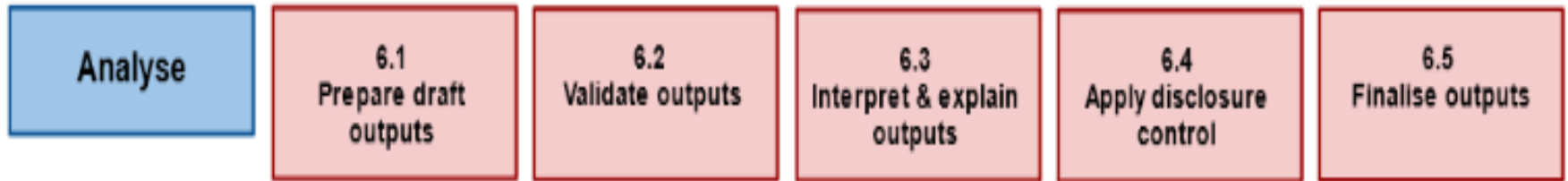


SDMX can help to visualize and process data, and can be used as a source format for outputs

6.1 Prepare Draft Outputs

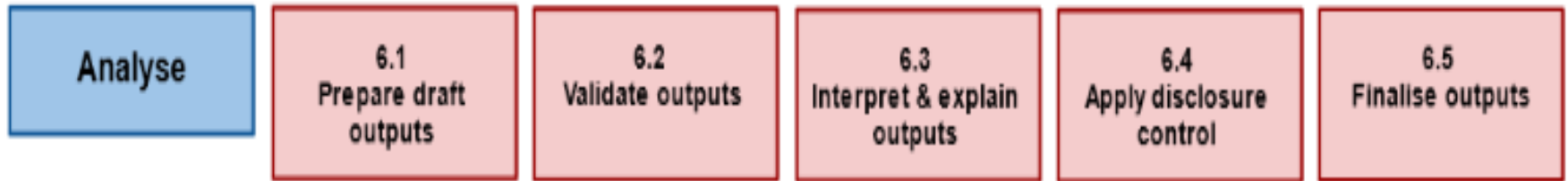
**Relies on technologies which easily
transform XML into other output formats**

6.2 Validate Outputs



SDMX-ML provides validation of all rules in the DSD (correct codes, complete and valid descriptions and keys, etc.)

6.2 Validate Outputs



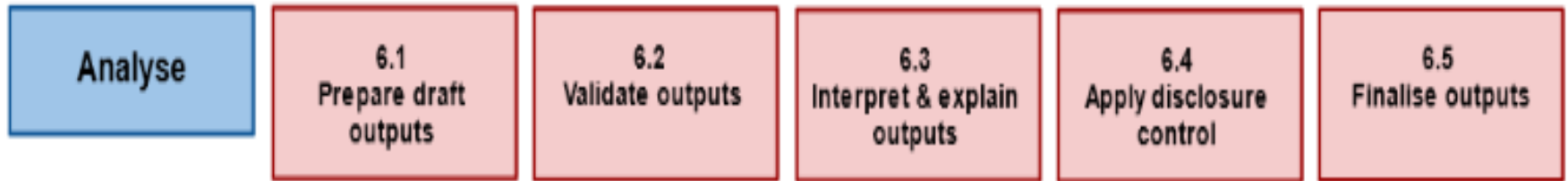
Some validation can be performed by XML schema (e.g. use of valid codes and dimension Ids)

6.3 Interpret and Explain Outputs



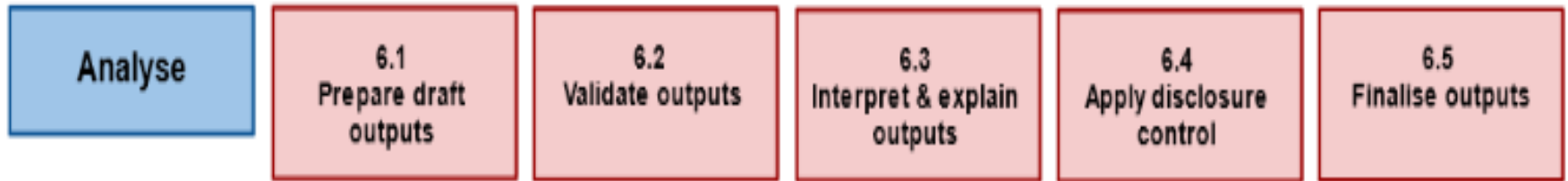
**SDMX visualizations may help to easily
view data**

6.3 Interpret and Explain Outputs



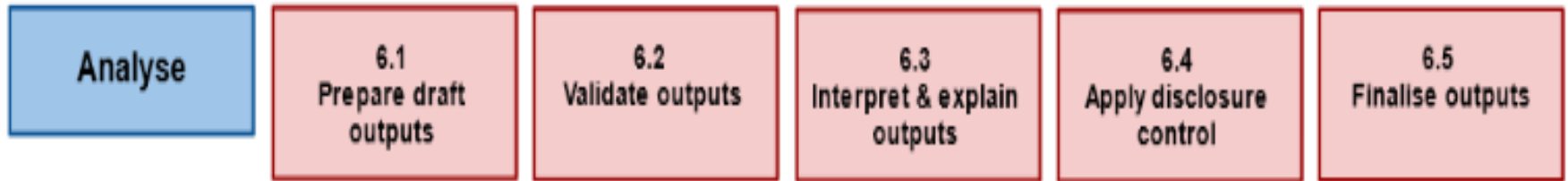
**SDMX visualizations may help to easily
view data and generate views for output
products**

6.4 Apply Disclosure Control



It is not a primary application for SDMX but visualizations can help to verify disclosure processing.

6.5 Finalize Outputs



SDMX visualizations may provide views of data for final outputs, which may be generated on-demand for dissemination on Website for example.

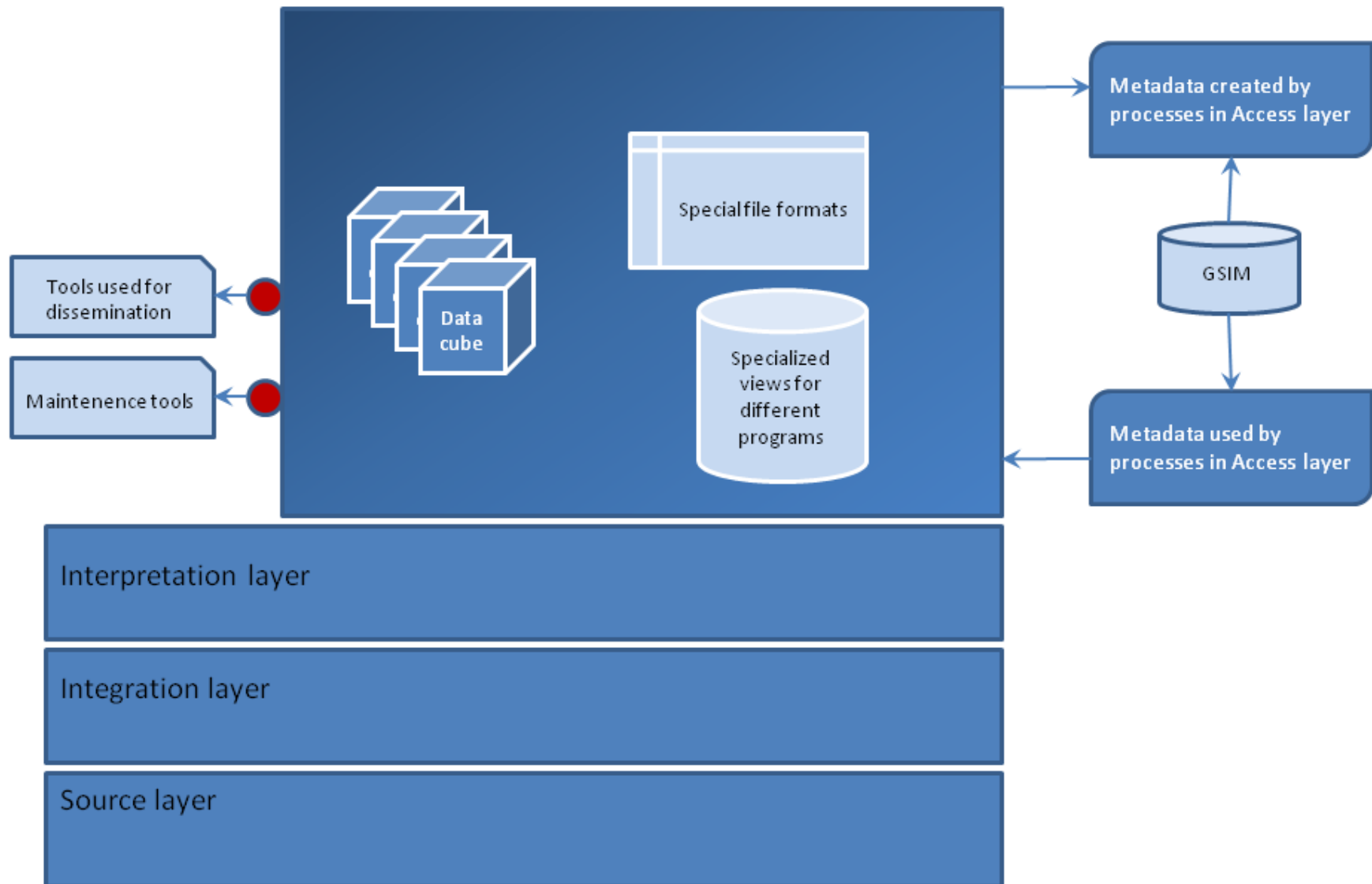
SDMX Potentialities

Reporting with SDMX:

- Push reporting format for data and metadata**

- Pull reporting format for data and metadata**

Access Layer



SDMX and Disseminate Phase (Step7)

SDMX most immediate usage in S-DWH is in the access layer which is intended for the final presentation, dissemination and delivery of information to end users.

7.1 Update Output Systems



**SDMX provides useful format for loading
into output systems.**

7.1 Update Output Systems



SDMX can be used as a format for the exchange of data between systems, whether these systems are internal to an organization, or external.

7.1 Update Output Systems



Most tools and databases provide good support for XML formats such as SDMX-ML, so SDMX-ML can be used as input to systems for creating HTML, PDF, Excel, and other output formats.

7.1 Update Output Systems



SDMX Registry can make the reporting of data more automated by using the data registration mechanism supported by a registry.

Once new data has been registered, the data user can simply query the service for the new data.

This helps to ease the burden of data reporting.

7.2 Produce Dissemination Products



SDMX visualizations may provide views of data for final outputs.

Outputs may be generated on-demand for dissemination on Websites, etc.

7.3 Manage Release of Dissemination Products



SDMX serves as a format for reporting and dissemination to some users/data collectors.

SDMX serves also as basis for generating other outputs; static or on-demand.

7.4 Promote Dissemination Products



The use of SDMX Registry Services provides a high level of visibility for data.

7.4 Promote Dissemination Products



The use of SDMX Registry Services provides a high level of visibility for data.

It depends on the availability of a domain registry for this purpose – requires the new data to be registered.

SDMX Potentialities

Discovery and Visualization:

- To drive website presentation of data and metadata**
- As a queryable data source**
- For standardized file downloads**

SDMX tools for the Source Layer

Some SDMX tools for the Registry/Metadata Repository:

- **Eurostat SDMX-RI**
- **Eurostat SDMX Registry**
- **Metadata Technology Fusion Registry**
- **ISTAT SDMX-RI Mapping Store Extension**
- **ISTAT SDMX-RI Web Service Extension**

SDMX tools for the Integration Layer

Some relevant SDMX tools for modelling and building the Structures of the S-DWH:

- **Eurostat DSW**
- **Metadata Technology Fusion Matrix**
- **Metadata Technology Fusion Weaver, Transformer, Cloud**
- **ISTAT Loader Builder**

SDMX tools for the Interpretation Layer

Some relevant SDMX tools for reporting:

- **Eurostat SDMX-RI**
- **Eurostat SDMX Converter**
- **ECB SDMX Java suite**
- **PandaSDMX**

SDMX tools for the Access Layer

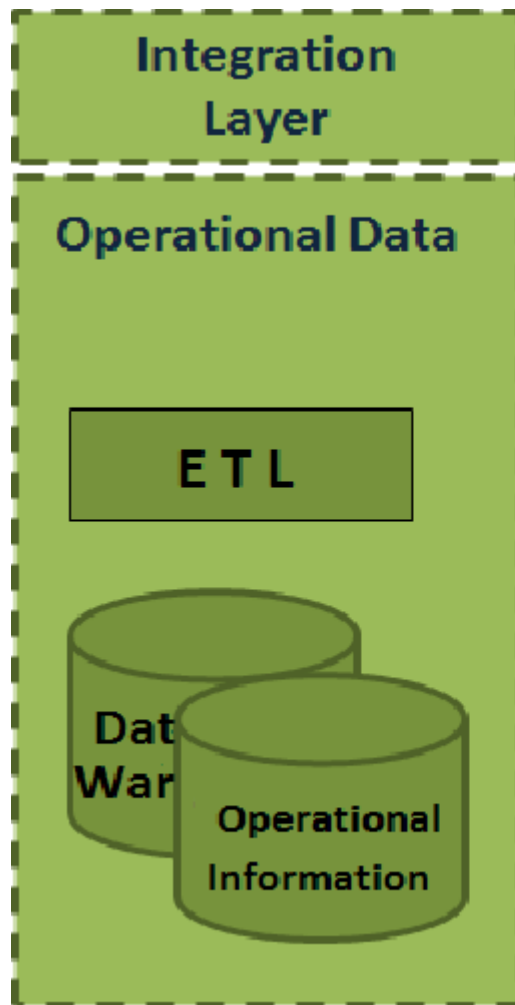
Some SDMX tools for discovery and visualization as well as Machine-Actionable Dissemination:

- **Eurostat SDMX-RI**
- **Metadata Technologies Fusion Matrix**
- **ISTAT Web Browser**
- **ISTAT SDMX-RI Web Service Extension**
- **Flex CB Visualization**

SDMX tools

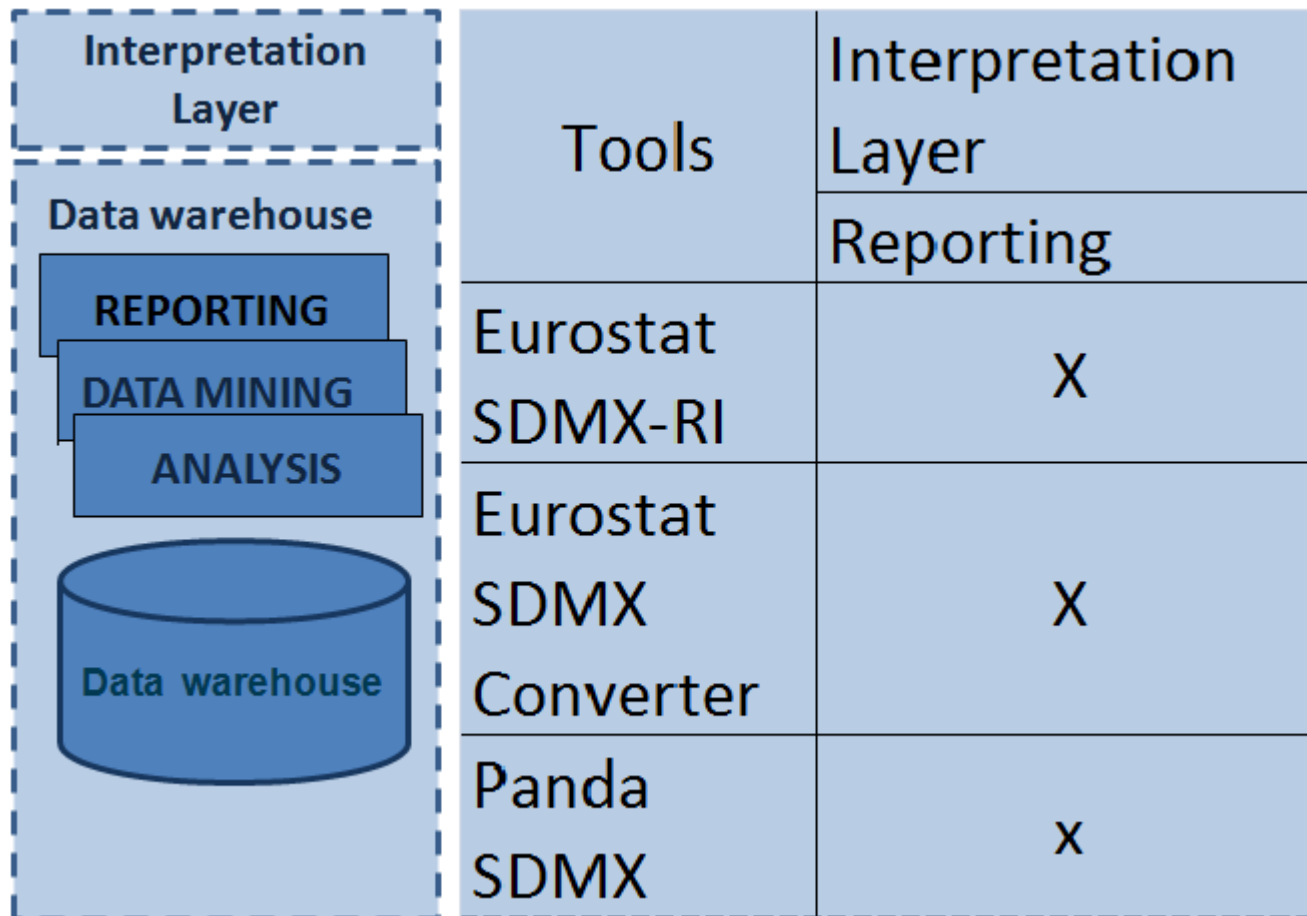
Source Layer	Tools	Source Layer
		Registry/Metadata repository
Staging Data	Eurostat SDMX-RI	X
	Eurostat SDMX Registry	X
	MT Fusion Registry	X
	ISTAT SDMX-RI Mapping Store extension	X
	ISTAT SDMX-RI Web Service extension	X
SURVEYS		
REGISTER		
...		
ADMIN		

SDMX tools

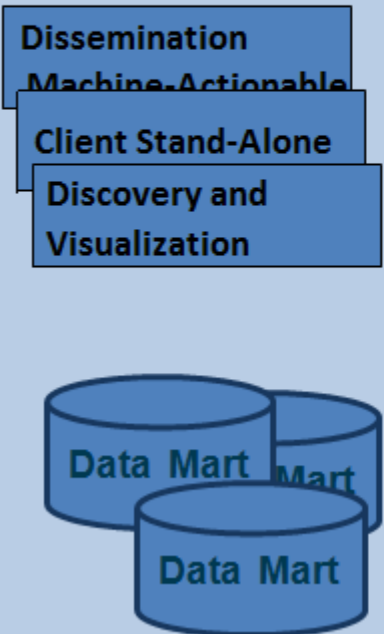


Tools	Integration Layer	
	Modelling SDWH	Building SDWH
Eurostat DSW	X	
MT Fusion Matrix	X	X
MT Fusion Weaver, Transformer, Cloud	X	
ISTAT Loader Builder	X	X

SDMX tools



SDMX tools

Access Layer	Tools	Access Layer		
		Dissemination machine-actionable	Discovery and visualization	Client stand-alone tool
	Eurostat SDMX-RI	X	X	X
	MT Fusion Matrix		X	
	MT Fusion XL			X
	Istat Web Browser		X	
	Istat SDMX-RI Web Service extension	X		
	Istat Excel Plug-in			X
	Flex CB Visualization		x	

SDMX tools

Layers / Tools		Eurostat SDMX-RI	Eurostat SDMX Converter	Eurostat SDMX Registry	Eurostat DSW	MT Fusion Matrix	MT Fusion Registry	MT Fusion Weaver, Transformer, Cloud	MT Fusion XL	ISTAT Loader Builder	ISTAT Web Browser	ISTAT SDMX-RI Mapping Store extension	ISTAT SDMX-RI Web Service extention	ISTAT Excel Plug-in	Flex CB Visualization	Panda SDMX
Access Layer	Machine-Actionable Dissemination	X											X			
	Discovery and Visualization	X				X					X				X	
	Client stand-alone	X							X					X		
Interpretation Layer	Reporting	X	X													X
Integration Layer	Modelling SDWH				X	X		X		X						
	Building SDWH					X				X						
Source Layer	Registry/Metadata repository	X		X			X					X	X			

Thank you for
your attention

Questions ?