



# Fusion Metadata Registry – validating SDMX data

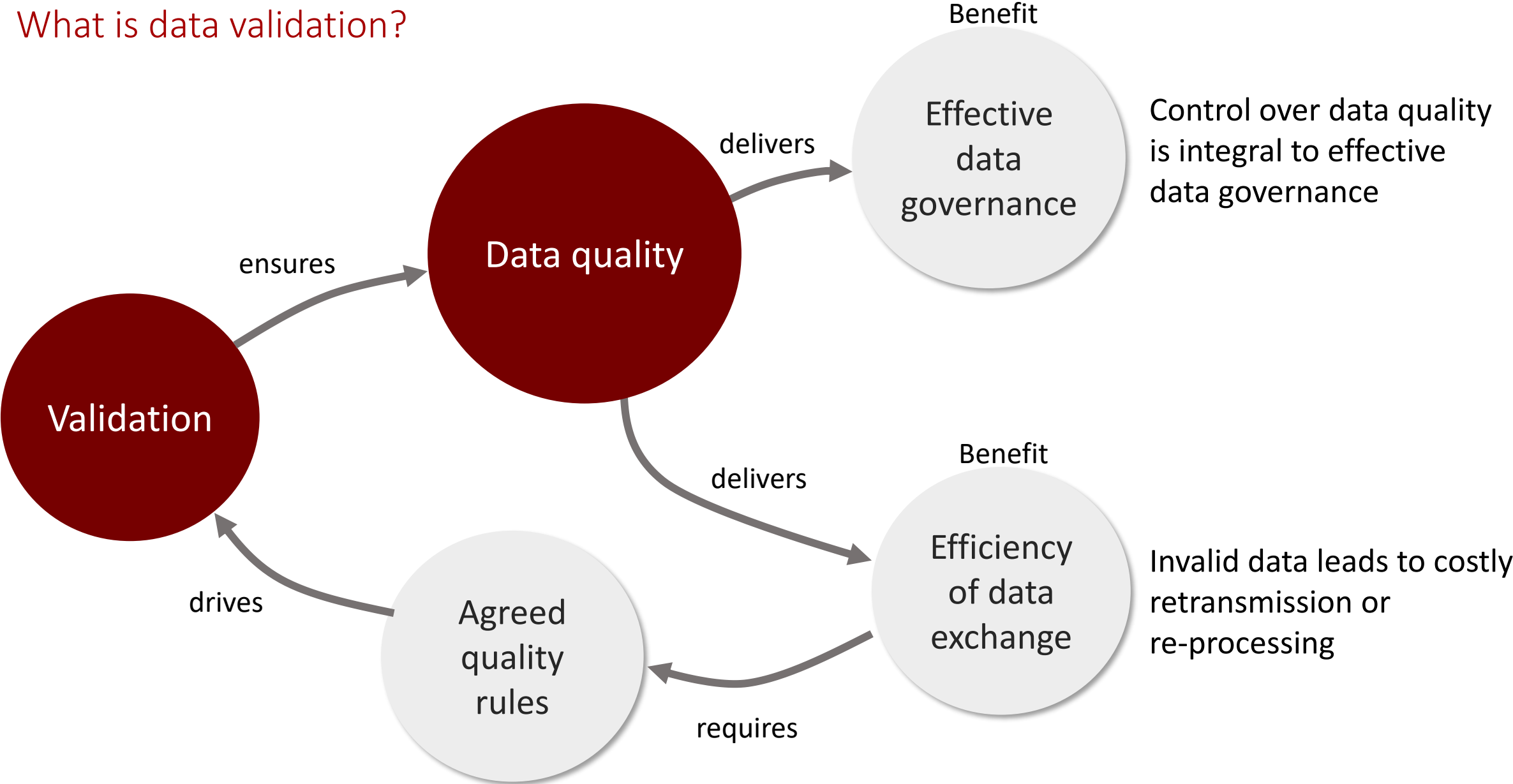
An introduction to SDMX data validation using Fusion Metadata Registry

Glenn Tice

## Topics

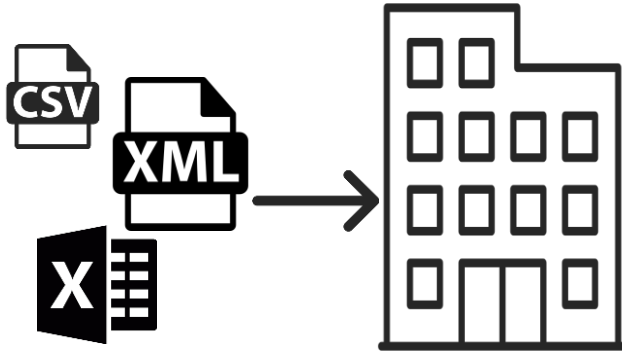
- FMR data validation use cases
- Validating data interactively using FMR's web user interface
- The validation rules available in FMR
- Using Concept representations and Constraints to define the universe of valid data
- Checking balance equalities using Validation Schemes

What is data validation?



## FMR data validation use cases

### Data Reporting



Data producers use FMR to validate reporting data using the collectors' rules prior to transmission

### Data Collection



Data collectors validate data received from data reporters

### Statistics Production



Set and enforce dataset quality rules

# Using FMR's web UI to validate a data set

Home

Organisations

Data

Data Definitions

Data Reporting

Data Dissemination

Convert Data

Items

VTL

Data Set Details

Data Source

Load From File

Data Format

Auto Detect

Structure Details

Data Structure

Auto Detect

Data

Data File

Choose File

Load Data

Home

Organisations

Data

Data Definitions

Dataflows

Data Reporting

Provision Agreements

Reporting Constraints

Reporting Templates

Validation Schemes

Data Dissemination

Publication Tables

Convert Data

Convert Data

Items

VTL

Related Structures

Metadata

Web Service

Bulk Actions

Structure References

Activity

Server Integrity

Dataset Details

Filename

mapping\_data\_invalid.csv

File Format

SDMX-CSV

Dataset

1

Data Structure

METATECH.BANKING\_STATISTICS(1.0) - Banking Statistics

Data Flow

METATECH.BANKING\_STATISTICS(1.0) - Macro economic banking statistics [change](#)

Provision Agreement

Select Provision Agreement (optional)

Data Provider

-

Number of Series

5

Number of Observations

96

Number of Groups

0

Action

Load Data

Re-Verify Data

Convert Data

Semantically Compliant

✓

Duplicate Observations

✓

Mandatory Attributes

✓

Obs Status

✓

Time Period Format

✗ 1 Error

Valid Calculations

✓

Valid Constraint

✓

Valid Representation

✗ 1 Error

Valid Structure

✓

Download Data

Map Data

BIS:BIS\_MACRO(1.0)

Data Format

SDMX

Sub-Format

v2.1 Structure Specific

Compression

v2.1 Structure Specific ✓

v2.1 Generic

v2.0 Compact

v2.0 Generic

v1.0 CSV

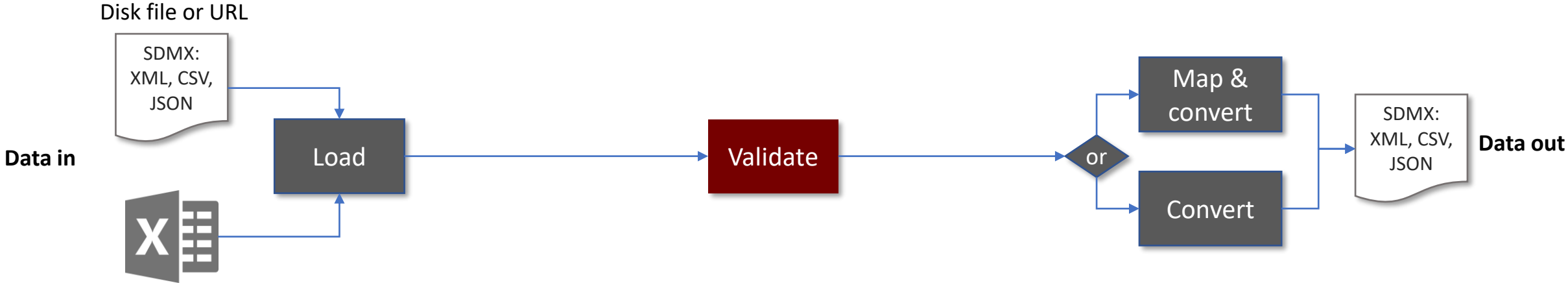
v2.0 CSV

JSON

Sender Id

Cancel

Download





## FMR's nine validation rules – a quick overview

Rules		Test applied
② Semantically Compliant	✓	The XML, JSON, CSV or Excel is well formed
② Duplicate Observations	✓	Uniqueness - there is only one observation value reported for each time period
② Mandatory Attributes	✓	All mandatory attributes are reported
② Obs Status	✓	<a href="#">OBS STATUS</a> is consistent with the observation value
② Time Period Format	✓	E.g. FREQ=M means the TIME_PERIOD format must be YYYY-MM
② Valid Calculations	observation consistency	Balance equalities defined using Validation Schemes
② Valid Constraint	valid universe of data (ID)	The data is within the universe defined by <b>Data Constraints</b>
② Valid Representation		Each component complies with the <b>representation</b> defined in the DSD
② Valid Structure	✓	The dimensions and attributes are consistent with the DSD


Defining the valid universe of data ( $\mathbb{D}$ ) – DSD component representations

An example DSD – BIS Macro-economic series

	Infinite universe of data Everything is valid!	The dataset's universe is more specific
Dimensions	Representation 	Representation 
Frequency	String	Codelist: CL_FREQ
BIS economic phenomenon	String	Codelist: CL_BIS_TOPIC
Reference area	String	Codelist: CL_REF_AREA
BIS suffix	String	Codelist: CL_BIS_SUFFIX
Measures		
Observation value	String	Number: $\mathbb{R} > 0$
Attributes		
Observation comment	String	String: $1 \leq \text{length} \leq 1050$



# Setting DSD component representations in practice using FMR

 Fusion Metadata Registry

en

Home

Organisations

Data

Data Definitions

Data Structures

Dataflows

Data Reporting

Provision Agreements

Reporting Constraints

Reporting Templates

Validation Schemes

Data Dissemination

Publication Tables

Convert Data

Convert Data

Items

VTL

Related Structures

Metadata

Web Services

Bulk Actions

Structure References

Activity

Change allowable content for: 'REF\_AREA'

Enumerated

Agency	Id	Name	Version
BIS	CL_BIS_GL_REF_AREA	Reference Area Code for BIS General Economics and Block L	1.0
BIS	CL_BIS_SUFFIX	Suffix	1.0
BIS	CL_BIS_TOPIC	BIS Topic code list	1.0
BIS	CL_BIS_UNIT	BIS_Unit	1.0
BIS	CL_COLLECTION	Collection	1.0
BIS	CL_CONF_STATUS	Observation confidentiality code list	1.0
BIS	CL_DECIMALS	Decimals codelist (BIS, ECB)	1.0
BIS	CL_FREQ	Code list for Frequency (FREQ)	1.0

Showing 2 to 10 of 122 entries 1 row selected

Search:

A coded representation may be limited further by specifying a further restriction:

Data Format: String (UTF-8)

Min Length2

Max Length2

RegEx Pattern

Cancel

Save

Choose between coded (enumerated) or un-coded (described)

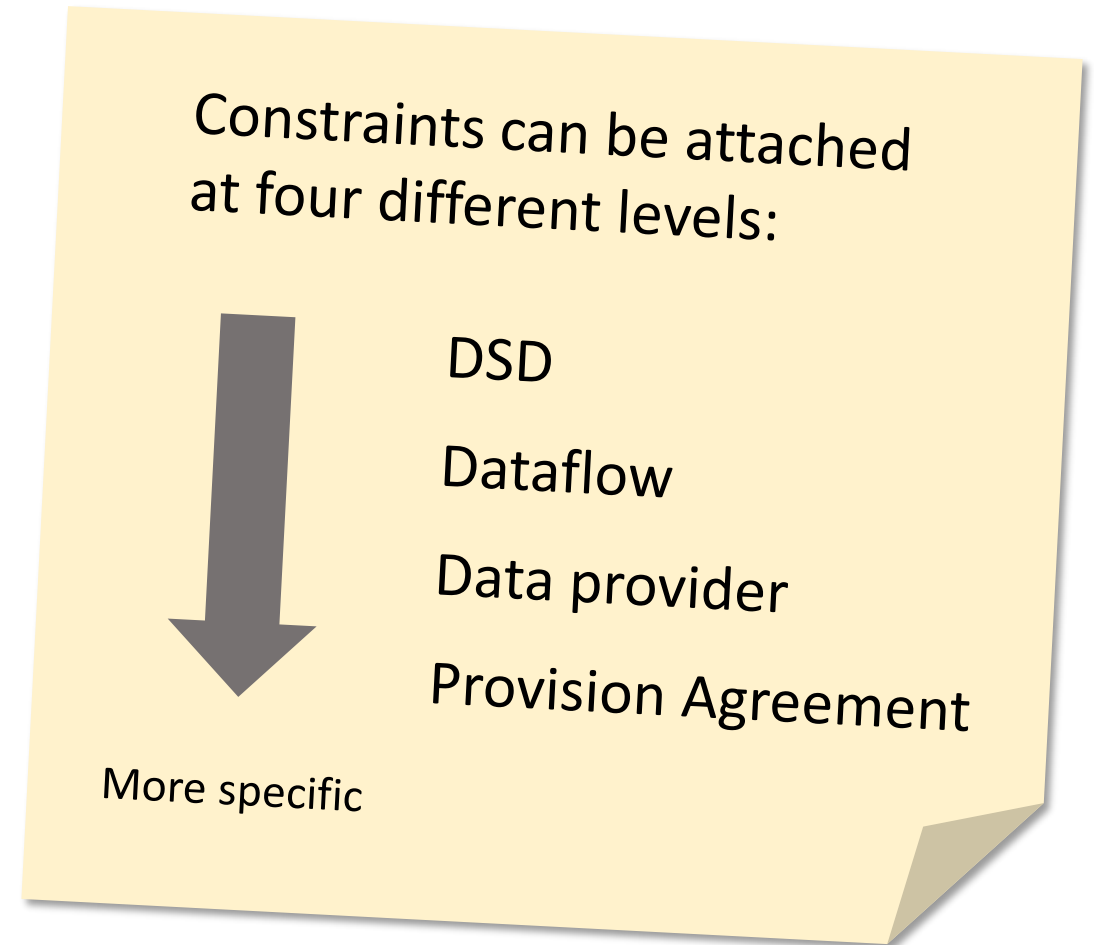
For coded components, pick the Codelist to use from those available in the registry

For coded representation, SDMX also allows further restriction rules to be applied, e.g. only codes that start with a specific string



## Further refining the valid universe of data ( $\mathbb{D}$ ) – Data Constraints

- A dataset's universe of valid data can still be large even with carefully designed representations
- **SDMX Data Constraints** allow further restrictions on the valid universe
- Constraints come in two flavours:
  - Cube Region
  - Series



## Further refining the valid universe of data ( $\mathbb{D}$ ) – Data Constraints

Two common use cases for Data Constraints

### 1. Restrict the domain for a specific Dataflow

- Problem: generic DSD that can be used for different datasets  
e.g. World Bank World Development Indicators DSD [WB:WDI\(1.0\)](#)
- Solution: Add constraint to the Dataflow to make the domain specific

### 2. Restrict what specific data providers can report

- Problem: Each data provider must only report certain values
- Solution: Add a constraint to each Provision Agreement

# Defining Data Constraints in practice using FMR

Set the valid values for each component:

- Included values
- Excluded values

Select one of the four attachment levels

Choose which components to constrain

## Reporting Constraint Wizard

OECD:CC\_SMEE\_OUTLOOK(1.0)

1. Details

2. Constrained Structure

3. Included Values

4. Excluded Values

Structure Type

Dataflow

Constrained Structure(s)

OECD:DF\_SMEE\_OUTLOOK(1.0)

Add New Remove Selected

Select Components to Include in Constraint

Selected	Component Id	Component Name	Codelist	Type
<input checked="" type="checkbox"/>	COU	Country	CL_SMEE_OUTLOOK__COU[1.0]	Dimension
<input checked="" type="checkbox"/>	INDICATOR	Indicator	CL_SMEE_OUTLOOK__INDICATOR[1.0]	Dimension
<input checked="" type="checkbox"/>	YEAR	Year	CL_SMEE_OUTLOOK__YEAR[1.0]	Dimension
<input type="checkbox"/>	OBS_STATUS	Observation Status	CL_SMEE_OUTLOOK__OBS_STATUS[1.0]	Attribute
<input type="checkbox"/>	UNIT_MEASURE	Unit of Measures	CL_SMEE_OUTLOOK__UNIT_MEASURE[1.0]	Attribute
<input type="checkbox"/>	UNIT_MULT	Multiplier	CL_SMEE_OUTLOOK__UNIT_MULT[1.0]	Attribute

Showing 1 to 6 of 6 entries 3 rows selected

Search:

Cube Region  
constraint  
example

Checking data set ‘balance equalities’ using FMR Validation Schemes

In some datasets, reported observations must be in balance

Dimension      REF\_AREA  
Balance rule    EUR = DE + FR + ES + IT

Reported values	REF_AREA	2019	2020	2021
	DE	5	6	4
	FR	3	4	5
	ES	7	5	5
	IT	2	7	2
Balance equality	EUR	17	20	16
	EUR = DE + FR + ES + IT		20≠22	

# Defining balance equalities in practice using FMR Validation Schemes

## Validation Scheme Wizard

BIS:VS1(1.0)

1. Details

2. Attachment

3. CSV Import

4. Expression builder

This step is optional. Use this step to import custom validation expressions. The CSV may include quotes for each value, example "AN\_ID","A Name","A description". Any CSV rules will be added in addition to any rules which may exist for this Validation Scheme.

Each validation expression can contain 'meta' information such as an **Id**, **Name**, and **Dimension Id**.

The rules' expression is then broken down into: **Dimension Id** on which the rule is operating, e.g REF\_AREA; **Output**, either numerical, or Code Id in square brackets e.g [EUROPE]; **Equality Operator**, valid values are =, <=, <, >=, >; and **Expression**, where each Code Id in the expression must be placed in square brackets, for example [UK]+[FR]+[DE]. Valid operators are +, -, /, \* and brackets '()' are supported.

Rule Details	Column Index	Rule Expression	Column Index
Rule Id	<input type="text" value="1"/>	Dimension Id	<input type="text" value="3"/>
Rule Name	<input type="text" value="2"/>	Output	<input type="text" value="4"/>
Rule Description	<input type="text"/>	Equality Operator	<input type="text" value="5"/>
		Expression	<input type="text" value="6"/>

REF\_AREA\_BALANCE\_Europe balance check REF\_AREA[EUR]=,"[DE]+[FR]+[ES]+[IT]"

Equalities are defined for specific dimensions in a DSD - REF\_AREA in this example

An expression defines the balance calculation – standard arithmetic operators are allowed (+ - / \*)

Balances of the following form are also allowed:

$$0 = [\text{EUR}] - ([\text{DE}] + [\text{FR}] + [\text{ES}] + [\text{IT}])$$

## Topics

We saw:

- FMR data validation use cases
- Validating data interactively using FMR's web user interface
- The validation rules available in FMR
- Using Concept representations and Constraints to define the universe of valid data
- Checking balance equalities using Validation Schemes

# Questions

BIS MED IT  
Glenn Tice  
[glennphilip.tice@bis.org](mailto:glennphilip.tice@bis.org)



## FMR data validation follow-on topics

- Series Constraints
- + and % operators for efficiently defining Series Constraints
- Code validity periods – Constraints, and at the code level
- Balance equalities using Hierarchical Codelists (Hierarchies in SDMX 3.0)
- Automating data validation using FMR's REST API
- Using FMR SDMX Codelists for validation in R

## References

FMR Docker image	<a href="https://hub.docker.com/r/metadatatechnology/fmr-mysql">https://hub.docker.com/r/metadatatechnology/fmr-mysql</a>
Download FMR	<a href="https://fusionsoftware.s3.us-west-2.amazonaws.com/fmr">https://fusionsoftware.s3.us-west-2.amazonaws.com/fmr</a>
FMR product page	<a href="https://metadatatechnology.com/software/FMR.php">https://metadatatechnology.com/software/FMR.php</a>
FMR quick start guide	<a href="https://fmrwiki.sdmxcloud.org/Quick_start_guide_-_Windows,_Linux_or_Mac">https://fmrwiki.sdmxcloud.org/Quick_start_guide_-_Windows,_Linux_or_Mac</a>
FMR Wiki – general reference	<a href="https://fmrwiki.sdmxcloud.org/Main_Page">https://fmrwiki.sdmxcloud.org/Main_Page</a>
Data Validation Cookbook	<a href="https://data-cleaning.github.io/validate/">https://data-cleaning.github.io/validate/</a>