Fusion Metadata Registry – Data collection using Excel Reporting Templates

On collecting statistical data using SDMX metadata-driven Excel forms

BIS Monetary and Economic Department IT

Topics

- FMR's Excel Reporting Templates principles and use cases
- Where Excel Reporting Templates work well, and when to consider other data collection strategies
- Constraining the universe of data
- Some more detailed features of Excel Reporting Templates

Hands on

- Defining an Excel Template using Fusion Metadata Registry's web user interface
- Generating an Excel form from a Template for a specific data provider
- Validating the completed Excel forms and converting to other SDMX formats

Example Data Structure

- Data Structure used in this presentation is UNICEF Immunisation
- Data Collection Templates will be generated for this Data Structure

Dimension	Example Values
Reference Area	Brazil, Canada, Sweden, …
Indicator	Percentage of surviving infants who received the first dose of…
Vaccine	Hepatitis B first dose, Hepatitis B second dose, Measles first dose, …
Age	Under 1 year, 12-23 Months, 2 years old, …, Total

Example Data Structure

- Labels have corresponding Identifiers
- Humans like labels, machines like IDs

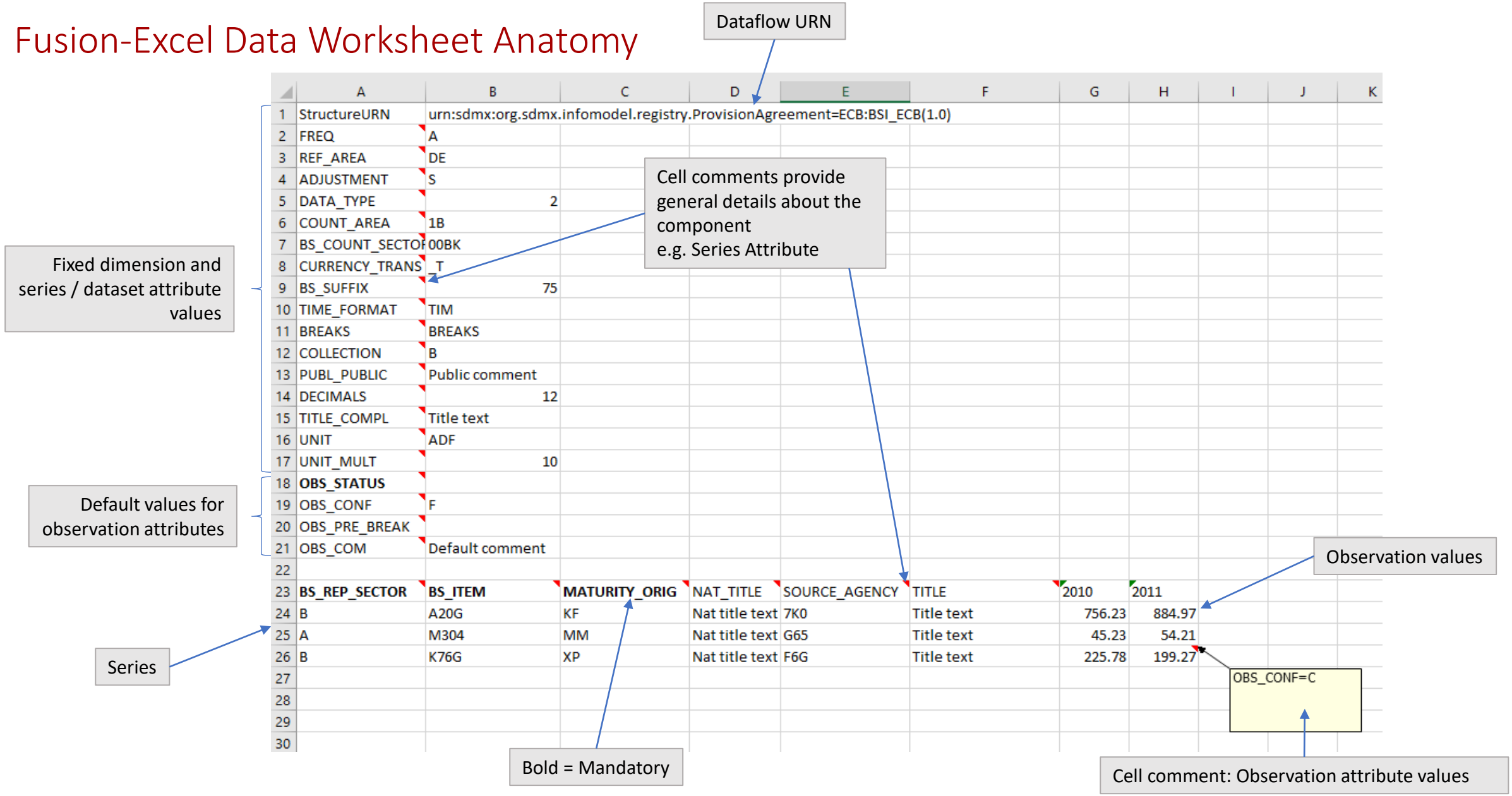
Dimension	Example Values
REF_AREA	BRA, SWE, CAN, …
INDICATOR	IM_DTP1, IM_DTP3
VACCINE	HEPB0, HEPB1, MEA1, …
AGE	Y0, M12T23,Y02,_T

Example Dataset (not a reporting template)

- We can generate a template which consist of column headers
- The data reporter fills in the form

REF_AREA	INDICATOR	VACCINE	AGE	2002	2003

Fusion-Excel Data Worksheet Anatomy



	A	B	C	D	E	F	G	H	I	J	K
1	StructureURN	urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ECB:BSI_ECB(1.0)									
2	FREQ	A									
3	REF_AREA	DE									
4	ADJUSTMENT	S									
5	DATA_TYPE	2									
6	COUNT_AREA	1B									
7	BS_COUNT_SECTOR	00BK									
8	CURRENCY_TRANS	_T									
9	BS_SUFFIX	75									
10	TIME_FORMAT	TIM									
11	BREAKS	BREAKS									
12	COLLECTION	B									
13	PUBL_PUBLIC	Public comment									
14	DECIMALS	12									
15	TITLE_COMPL	Title text									
16	UNIT	ADF									
17	UNIT_MULT	10									
18	**OBS_STATUS**										
19	OBS_CONF	F									
20	OBS_PRE_BREAK										
21	OBS_COM	Default comment									
22											
23	**BS_REP_SECTOR**	**BS_ITEM**	**MATURITY_ORIG**	NAT_TITLE	SOURCE_AGENCY	TITLE	2010	2011			
24	B	A20G	KF	Nat title text	7K0	Title text	756.23	884.97			
25	A	M304	MM	Nat title text	G65	Title text	45.23	54.21			
26	B	K76G	XP	Nat title text	F6G	Title text	225.78	199.27			
27											
28											
29											
30											

Fusion-Excel data format vs Excel Reporting Templates

Dataflow	BIS:CIBL_IB(1.0)		
L_REP_CTY	Reporting Country	AR	Argentina
CBS_BASIS	BBS reporting basis	F	Immediate Counterparty Basis
REM_MATURITY	Remaining Maturity	A	All Maturities
CURR_TYPE_BOOK	Currency Type of Booking Location	TO1	All currencies
L_MEASURE	Measure	S	Amounts outstanding / Stocks

Confidentiality Status	
F	Freely Available
R	Restricted, for use by BIS and sharing central banks
C	Confidential statisitical information (for BIS only)

	Balance sheet position	Total Claims						International Claims	
	CBS Bank Type	Domestic Banks			All excluding 4C banks			Domestic Banks	
	Counterparty Sector	All Sectors	Total	Other Sector	All Sectors	Total	Other Sector	All Sectors	Total
L_CP_SECTOR	Types of Intrument / Counterpart Country	M:AR::F::A:A	M:AR::F::A:A	M:AR::F::A:A	M:AR::F::A:A	M:AR::F::A:A	M:AR::F::A:A	M:AR::F::A:A	M:AR::F::A:A
	Bahamas								
	Aerbaijan								
C	All Instruments / All Countries								
	All Countries excluding residents								
	Australis								
	Austria								
	Bahamas								
	Aerbaijan								
	Derivatives / All Countries								
	All Countries excluding residents								
	Australis								
	Austria								
	Bahamas								
	Aerbaijan								

Data Table

Pre-Break Value

	Balance Sheet Position	Total Claims						International Claims	
	CBS Bank Type	Domestic Banks			All excluding 4C banks			Domestic Banks	
	Counterpart Sector	All Sectors	Total	Other Sector	All Sectors	Total	Other Sector	All Sectors	Total
L_CP_SECTOR	Types of Intrument / Counterpart Country								
A	All Instruments / All Countries								
	All Countries excluding residents								
	Australis								
	Austria								
	Bahamas								
	Aerbaijan								
	Derivatives / All Countries								
	All Countries excluding residents								
	Australis								

Pre-Break Value Table

Reporting Template

Excel reporting templates use cross tabulations

FusionXL uses a columnar structure with one row per series

FusionXL data set worksheet

	B	C	D	E	F	G	H	I	J	K
	urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ECB:BSI_ECB(1.0)									
	A									
	DE									
	S									
	2									
	1B									
	00BK									
	_T									
	75									
	TIM									
	BREAKS									
	B									
	Public comment									
	12									
	Title text									
	ADF									
17	UNIT_MULT	10								
18	**OBS_STATUS**									
19	OBS_CONF	F								
20	OBS_PRE_BREAK									
21	OBS_COM	Default comment								
22										
23	**BS_REP_SECTOR**	**BS_ITEM**	**MATURITY_ORIG**	NAT_TITLE	SOURCE_AGENCY	TITLE	2010	2011		
24	B	A20G	KF	Nat title text	7K0	Title text	756.23	884.97		
25	A	M304	MM	Nat title text	G65	Title text	45.23	54.21		
26	B	K76G	XP	Nat title text	F6G	Title text	225.78	199.27		
27										
28										
29										
30										

OBS_CONF=C

Example Reporting Template – Overview

- Cross tabulation
- Key Differences
  - The layout of the form (Dimensions in the Header/Rows) defined by the data collector
  - The reporting universe (white cells) can be controlled by the data collector
  - The user fills in the white cells (**observation values**)

Vaccine		Bacille calmette-guerin (tuberculosis)	Diphtheria, tetanus, and pertussis first dose
Geographic area	**Current age**	::BCG:	::DTP1:
Afghanistan	12 to 23 months old	11	12
	Under 1 year old	13	14
	2 years old	15	16
	3 years old		
	4 years old		
	5 years old		
	6 years old		
	7 years old		
	8 years old		
	9 years old		

FusionXL data authoring vs Excel Reporting Templates

Excel Reporting Templates

- Fixed cross tabulation forms
  reporter fills in the blanks
  i.e. enter just the observation values
- Templates are designed by the data collector from which the Excel forms are generated
- Validation: Server side validation plus optional checksums as Excel formulas
- Cell colour has meaning
- Better suited to smaller universes of data to avoid the tables becoming too large – although an Excel workbook can contain multiple worksheets / tables

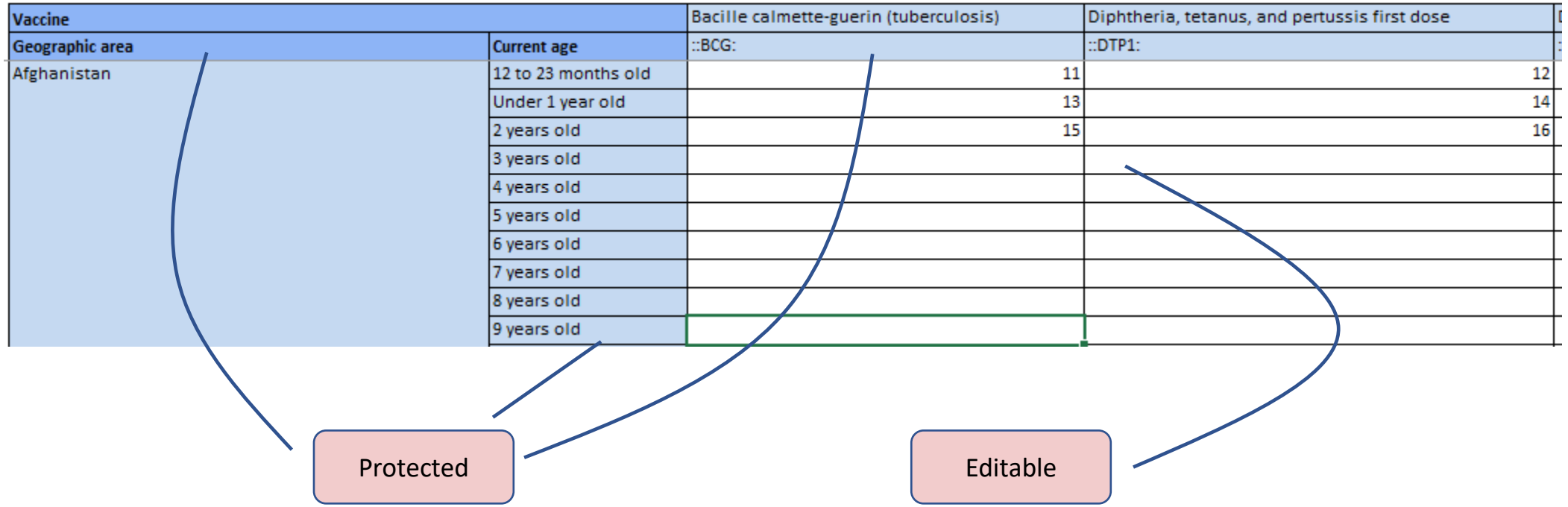
FusionXL data authoring

- Free-form columnar layout
  reporter adds a new row for each series
  i.e. must enter both the series metadata and observations
- Reporter creates a dataset worksheet using the FusionXL Create Dataset tool
- Validation: server-side validation only
- Suitable for larger universes of data, i.e. where there are a large number of possible dimension values

Important to Note

- No VBA, 100% Excel, no plugins required
- Metadata required to read the workbook is embedded in the workbook
- Worksheet is protected to prevent tampering

Vaccine		Bacille calmette-guerin (tuberculosis)	Diphtheria, tetanus, and pertussis first dose	
Geographic area	Current age	::BCG:	::DTP1:	:
Afghanistan	12 to 23 months old	11	12	
	Under 1 year old	13	14	
	2 years old	15	16	
	3 years old			
	4 years old			
	5 years old			
	6 years old			
	7 years old			
	8 years old			
	9 years old			



Reporting Template – Metadata Driven Solution

- Makes use of SDMX
  - Data Structure
  - Dataflow
  - Concepts / Codelists
  - Content Constraints
  - Hierarchical Codelists
  - Validation Rules

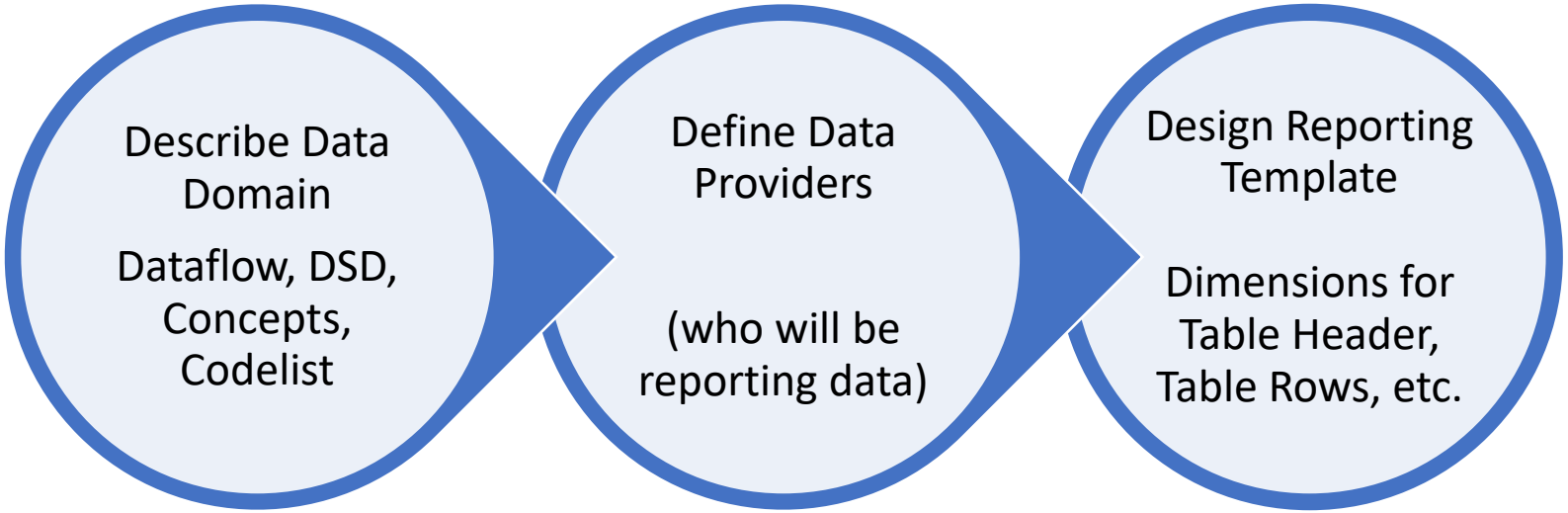
Any changes to SDMX metadata are reflected in generated Reporting Template

Multilingual Labels supported if the metadata has this

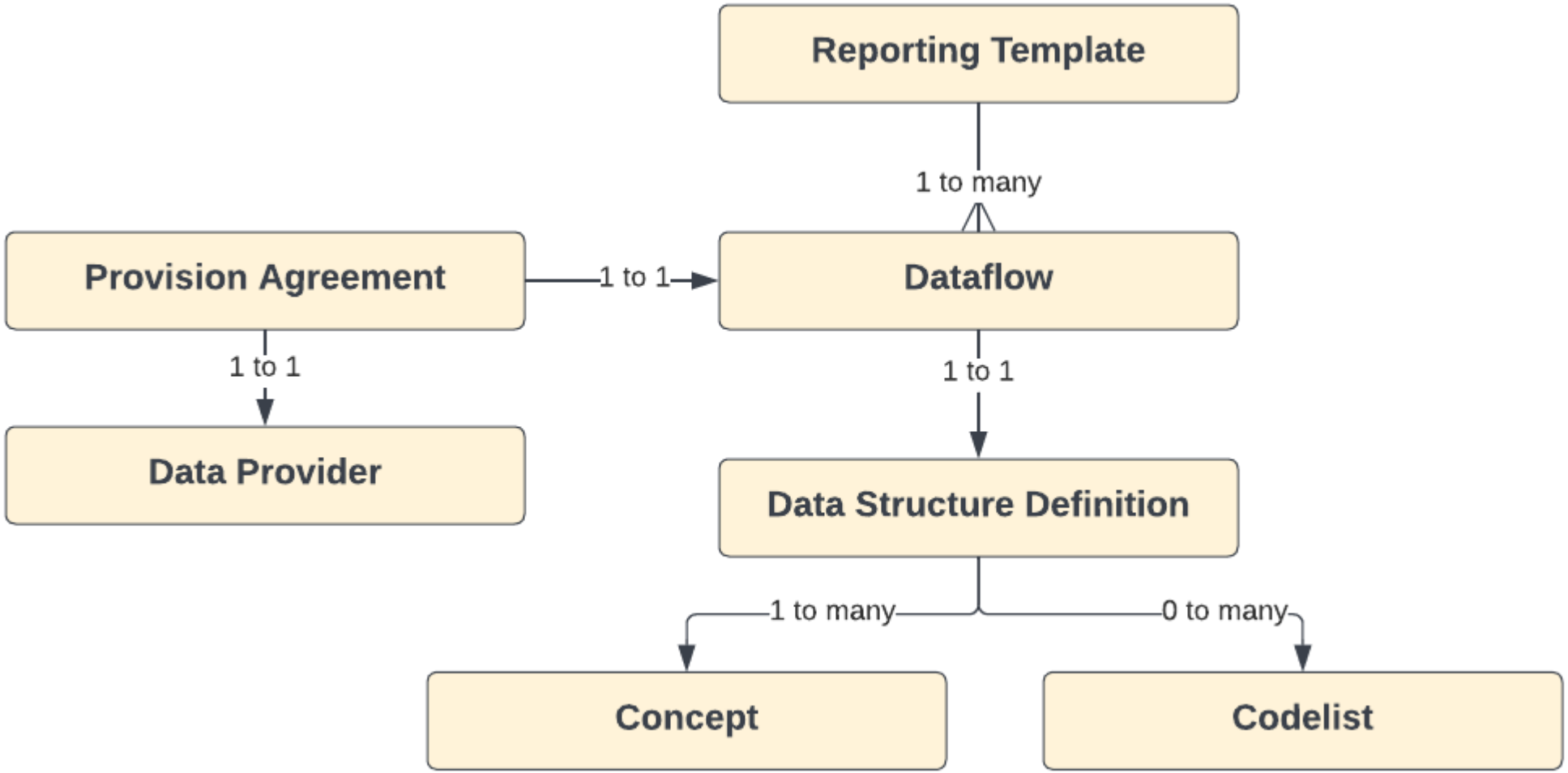
Same rules used to validate SDMX Dataset are used to generate the reporting universe in an Excel Reporting Template

Data in a Reporting Template is convertible into an SDMX Dataset

Reporting Template – Template Design (simple) Workflow



Reporting Template – Template Design (simple) model



Reporting Template – Universe of Data

Dimension	Unique Values
Reference Area	343
Indicator	16
Vaccine	32
Age	16

Grid Size = 343 x 16 x 32 x 16
= 2,809,856 Cells **!Limit Exceeded**

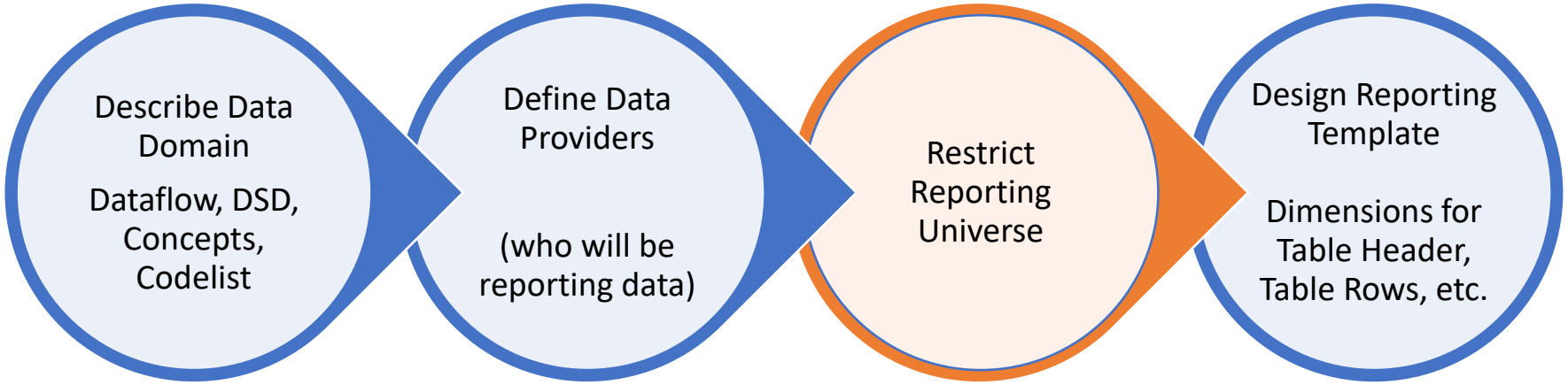
The universe of data may be a limitation for some data collections, and becomes more prevalent for Data Structures with a large number of Dimensions / large number of possible values per Dimension

Reporting Template – Constraining Universe of Data - Constraints

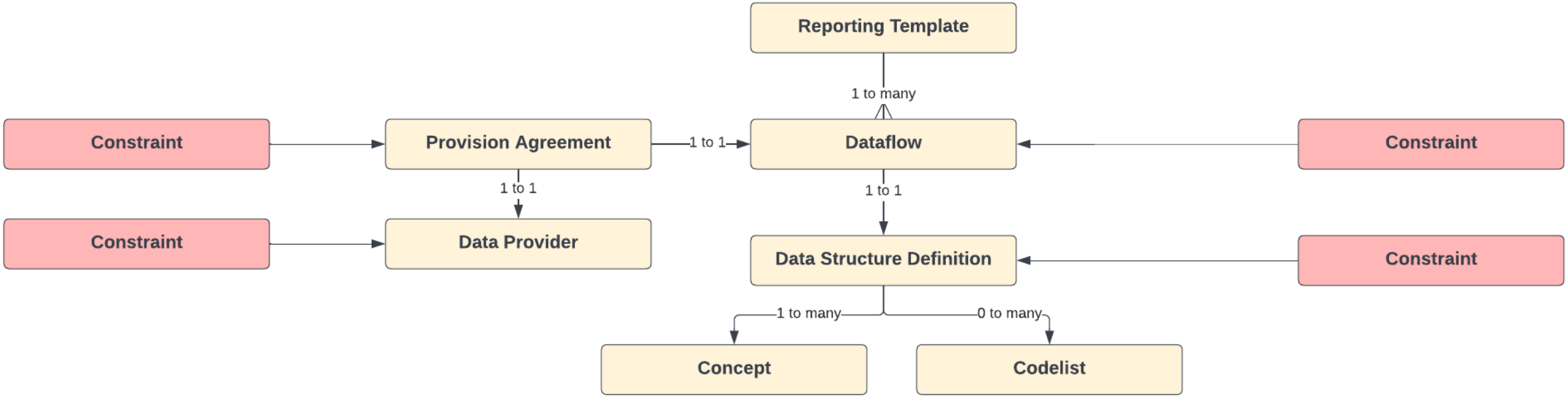
Dimension	Unique Values
Reference Area	~~343~~ 1 (when Canada reports)
Indicator	16
Vaccine	32
Age	16

Brazil can only report data for Reference Area ‘CAN’
Grid Size = ~~343~~ 1 x 16 x 32 x 16
= 8,192 Cells

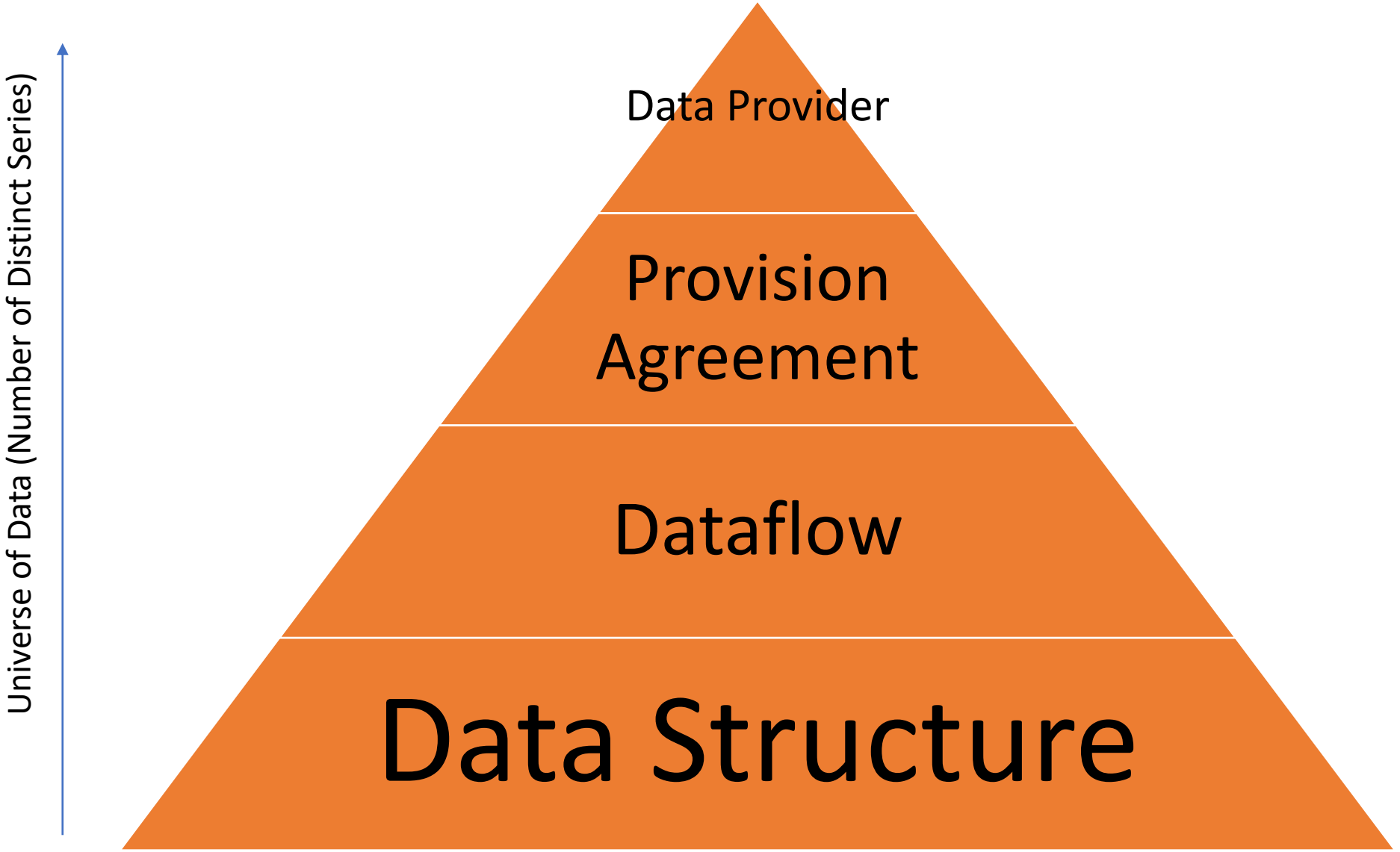
Reporting Template – Template Design updated Workflow



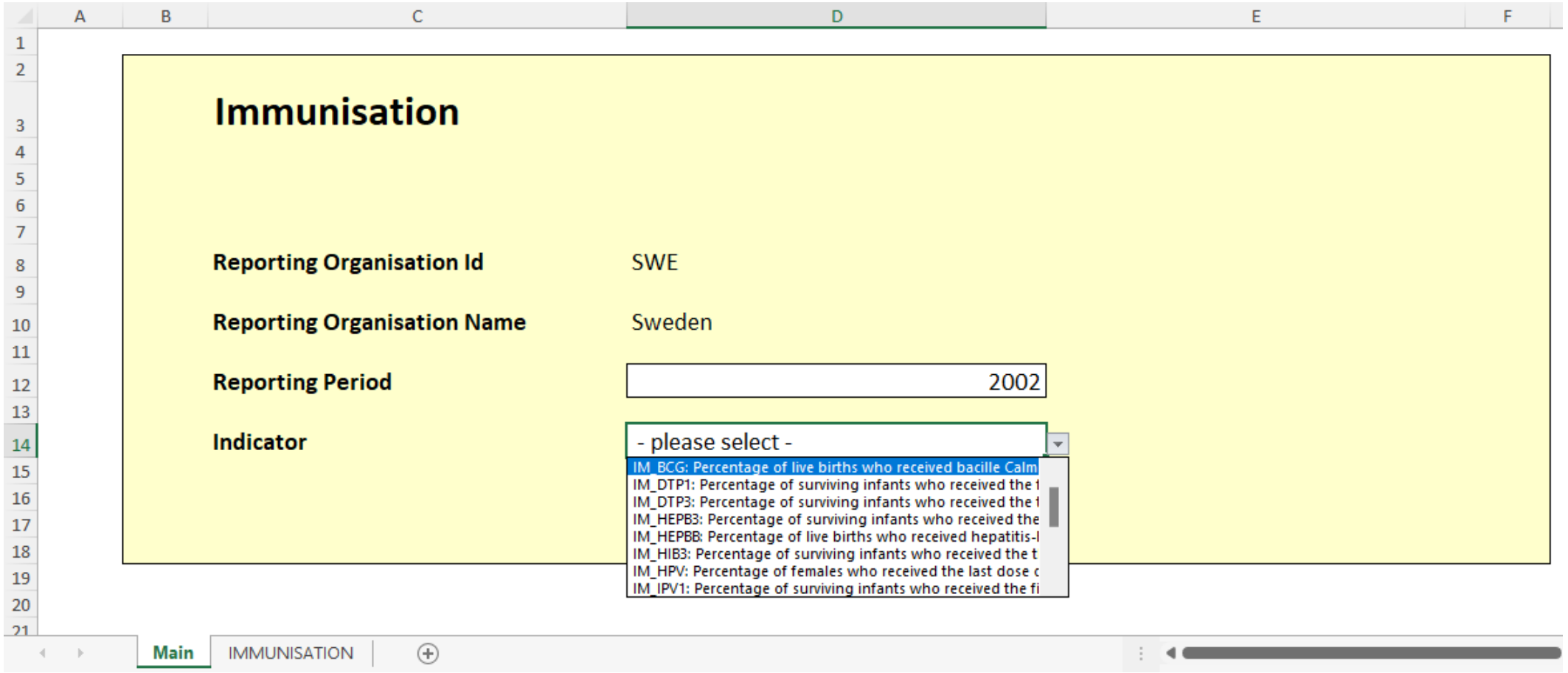
Reporting Template – Template Design updated Model



Reporting Template – A way to thing about Constraints



Reporting Template – Constraining Universe of Data – Variable Dimension



Grid Size = ~~343~~ 1 x 1 ~~16~~ x 32 x 16
= 512 Cells

Reporting Template – Constraining Universe of Data – Multiple Dataflows

L_REP_CTY	Reporting country	KR	Korea	
CBS_BANK_TYPE	CBS bank type	4M	All banks (=4B +4C + 4D +4E)	
COLLECTION	Collection Indicator	E	End of period	

Balance sheet position	Total claims			
Remaining maturity	Total (all maturities)			
Counterparty sector	All sectors			
Type of instruments	All instruments			
Currency type of booking location	All currencies			
CBS reporting basis	Immediate counterparty basis	Outward risk transfers	Inward risk transfers	Net risk transfers (Inward-Outward)
L_CP_COUNTRY|Counterparty country	Q:S:KR:4M:F:C:A:A:TO1:A:	Q:S:KR:4M:O:C:A:A:TO1:A:	Q:S:KR:4M:P:C:A:A:TO1:A:	Q:S:KR:4M:Q:C:A:A:TO1:A:
5J|All countries (total)				
AF|Afghanistan				
AL|Albania				
DZ|Algeria				
AD|Andorra				
AO|Angola				
AI|Anguilla				
AG|Antigua and Barbuda				
AR|Argentina				
AM|Armenia				
AW|Aruba				
AU|Australia				
AT|Austria				
AZ|Azerbaijan				
BS|Bahamas				
BH|Bahrain				

Main | INSTRUCTIONS | CBSI_RT_ALL_BANKS | CBSI_RT_DOMESTIC | CBSI_CONS_INSIDE | CB ...

Worksheet for each Dataflow

Features – Attribute Support

- Attributes can be
  - Excluded from the table
  - Given a fixed value or a conditional value (if obs=NaN, obs status=M)
  - Reported in the same table as the observation cells
  - Reported in a separate table/dedicated worksheet

Vaccine		All vaccinations	Diphtheria, tetanus, and pertussis	
			Tetanus protection at birth	Diphtheria, tetanus, and pertussis first dose
Current age	Observation	SWE::FULL-VAC:	SWE::PAB:	SWE::DTP1:
Total	Obs Value			
	Observation confidentaility			
	Observation Status			
12 to 23 months old	Obs Value			
	Observation confidentaility			
	Observation Status			
Under 1 year old	Obs Value			
	Observation confidentaility			
	Observation Status			
2 years old	Obs Value			
	Observation confidentaility			
	Observation Status			

Observation Value

Observation Confidentiality

Observation Status

Features – Attribute Support

Colour used, as attributes are reported – table colour updated automatically

Observation confidentaility	
F	Free
N	Not for publication, restricted for internal use only
C	Confidential statistical information

Current age		12 to 23 months old	Under 1 year old	2 years old	3 years old	4 years old	5 years old	6 years old	7 years old	8 years old	9 years old
Vaccine		SWE:::M12T23	SWE:::Y0	SWE:::Y02	SWE:::Y03	SWE:::Y04	SWE:::Y05	SWE:::Y06	SWE:::Y07	SWE:::Y08	SWE:::Y09
Bacille calmette-guerin (tuberculosis)											
Diphtheria, tetanus, and pertussis first dose											
Diphtheria, tetanus, and pertussis second dose											
Diphtheria, tetanus, and pertussis third dose											
Hepatitis B birth dose											
Hepatitis B first dose											
Hepatitis B second dose											
Hepatitis B third dose											
Haemophilus influenzae type b first dose											
Haemophilus influenzae type b second dose											
Haemophilus influenzae type b third dose											
Polio birth dose											
Polio first dose											
polio second dose											
Polio third dose											
Pneumococcal conjugate first dose											
Pneumococcal conjugate second dose											
Pneumococcal conjugate third dose											
Rotavirus first dose											
Rotavirus second dose											
Rotavirus third dose											
Rotavirus last dose											
Tetanus protection at birth											
All vaccinations											
No vaccinations											

Observation Attributes

Current age		12 to 23 months old	Under 1 year old	2 years old	3 years old	4 years old	5 years old	6 years old	7 years old	8 years old	9 years old
Vaccine	Obs Attribute	SWE:::M12T23	SWE:::Y0	SWE:::Y02	SWE:::Y03	SWE:::Y04	SWE:::Y05	SWE:::Y06	SWE:::Y07	SWE:::Y08	SWE:::Y09
Bacille calmette-guerin (tuberculosis)	Observation confidentaility	C	N								
Diphtheria, tetanus, and pertussis first dose	Observation confidentaility	C	N								
Diphtheria, tetanus, and pertussis second dose	Observation confidentaility	C									

Features – Hierarchy Support on Header

SDMX Hierarchical Codelist applied to Header VACCINE Dimension

Haemophilus influenzae type b			Polio			
Haemophilus influenzae type b first dose	Haemophilus influenzae type b second dose	Haemophilus influenzae type b third dose	Polio birth dose	Polio first dose	polio second dose	Polio third dose
SWE::HIB1:	SWE::HIB2:	SWE::HIB3:	SWE::POL0:	SWE::POL1:	SWE::POL2:	SWE::POL3:

Features – Hierarchy Support on Rows

Hierarchical Codelist in this example includes codes sourced from 2 Codelists to build the hierarchy

Current age	12 to 23 months old	Under 1 year old	2 years old
Vaccine	SWE:::M12T23	SWE:::Y0	SWE:::Y02
All vaccinations			
Diphtheria, tetanus, and pertussis			
Tetanus protection at birth			
Diphtheria, tetanus, and pertussis first dose			
Diphtheria, tetanus, and pertussis second dose			
Diphtheria, tetanus, and pertussis third dose			
Hepatitis B			
Hepatitis B birth dose			
Hepatitis B first dose			
Hepatitis B second dose			
Hepatitis B third dose			
Haemophilus influenzae type b			
Haemophilus influenzae type b first dose			
Haemophilus influenzae type b second dose			
Haemophilus influenzae type b third dose			
Polio			
Polio birth dose			
Polio first dose			
polio second dose			
Polio third dose			

Grouping Codes are not part of the data collection Codelist

Cells which should not have data reported against them are disabled

Features – Client side validation – Check Table

Vaccine	All vaccinations	Diphtheria, tetanus, and pertussis		
		Tetanus protection at birth	Diphtheria, tetanus, and pertussis first dose	Diphtheria, tetanus, and pertussis second dose
Current age	SWE::FULL-VAC:	SWE::PAB:	SWE::DTP1:	SWE::DTP2:
Total	5		10	20
12 to 23 months old	2			
Under 1 year old	2			
2 years old				
3 years old				
4 years old				
5 years old				
6 years old				
7 years old				
8 years old				
9 years old				
10 years old				
11 years old				
13 years old				
14 years old				
12 years old				

Those receiving second dose cannot exceeded those that received first dose

Check Table

Vaccine	All vaccinations	Diphtheria, tetanus, and pertussis		
		Tetanus protection at birth	Diphtheria, tetanus, and pertussis first dose	Diphtheria, tetanus, and pertussis second dose
Current age	SWE::FULL-VAC:	SWE::PAB:	SWE::DTP1:	SWE::DTP2:
Total	1		0	DTP_D2_LTE_D1
12 to 23 months old				0
Under 1 year old				0
2 years old				0

Features – Client side validation – Summary of Rules

Relationship Validation Rules Summary

These checks use formula which allow a range of possible reported values. Failures will display error codes in the Check Table of the respective worksheet. This summary shows the number of errors detected for each rule, per worksheet that has relationship rules.

Worksheet	Rule Id	Dimension	Formula	Error Count
IMMUNISATION	DTP_D2_LTE_D1	VACCINE	DTP2<=DTP1	1
IMMUNISATION	DTP_D3_LTE_D2	VACCINE	DTP3<=DTP2	0
IMMUNISATION	HEPB_D2_LTE_D1	VACCINE	HEPB2<=HEPB1	0
IMMUNISATION	HEPB_D3_LTE_D2	VACCINE	HEPB3<=HEPB2	0
IMMUNISATION	HIB_D2_LTE_D1	VACCINE	HIB2<=HIB1	0
IMMUNISATION	HIB_D3_LTE_D2	VACCINE	HIB3<=HIB2	0

Equlaity Check Summary

Equality check failures result in the difference between the reported value and the expected value being displayed in the Check Table. Zero values indicate there is no difference between the reported value and the expected value. This summary shows the maximum difference, for each worksheet that define equality checks.

Worksheet	Max Difference
IMMUNISATION	1

Features – Client side validation – Summary of Rules

Relationship Validation Rules Summary

These checks use formula which allow a range of possible reported values. Failures will display error codes in the Check Table of the respective worksheet. This summary shows the number of errors detected for each rule, per worksheet that has relationship rules.

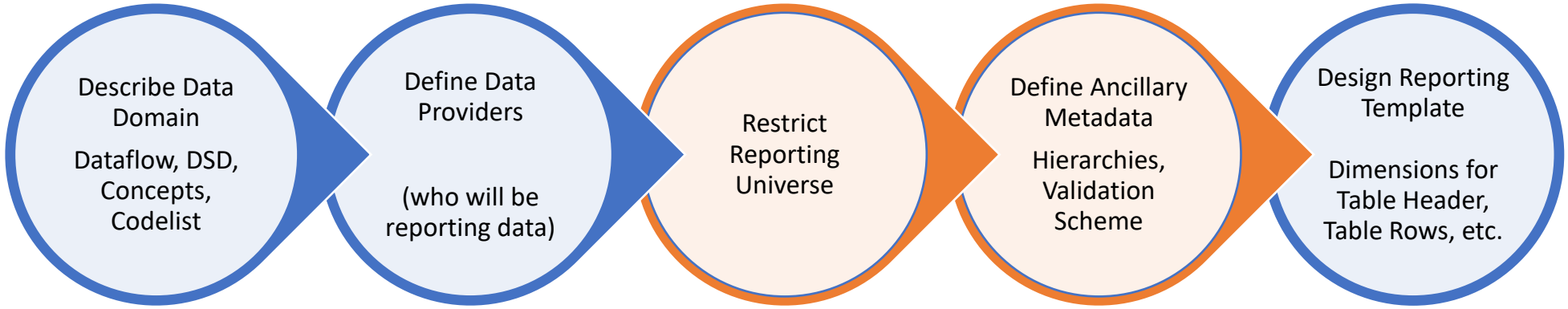
Worksheet	Rule Id	Dimension	Formula	Error Count
IMMUNISATION	DTP_D2_LTE_D1	VACCINE	DTP2<=DTP1	1
IMMUNISATION	DTP_D3_LTE_D2	VACCINE	DTP3<=DTP2	0
IMMUNISATION	HEPB_D2_LTE_D1	VACCINE	HEPB2<=HEPB1	0
IMMUNISATION	HEPB_D3_LTE_D2	VACCINE	HEPB3<=HEPB2	0
IMMUNISATION	HIB_D2_LTE_D1	VACCINE	HIB2<=HIB1	0
IMMUNISATION	HIB_D3_LTE_D2	VACCINE	HIB3<=HIB2	0

Equlaity Check Summary

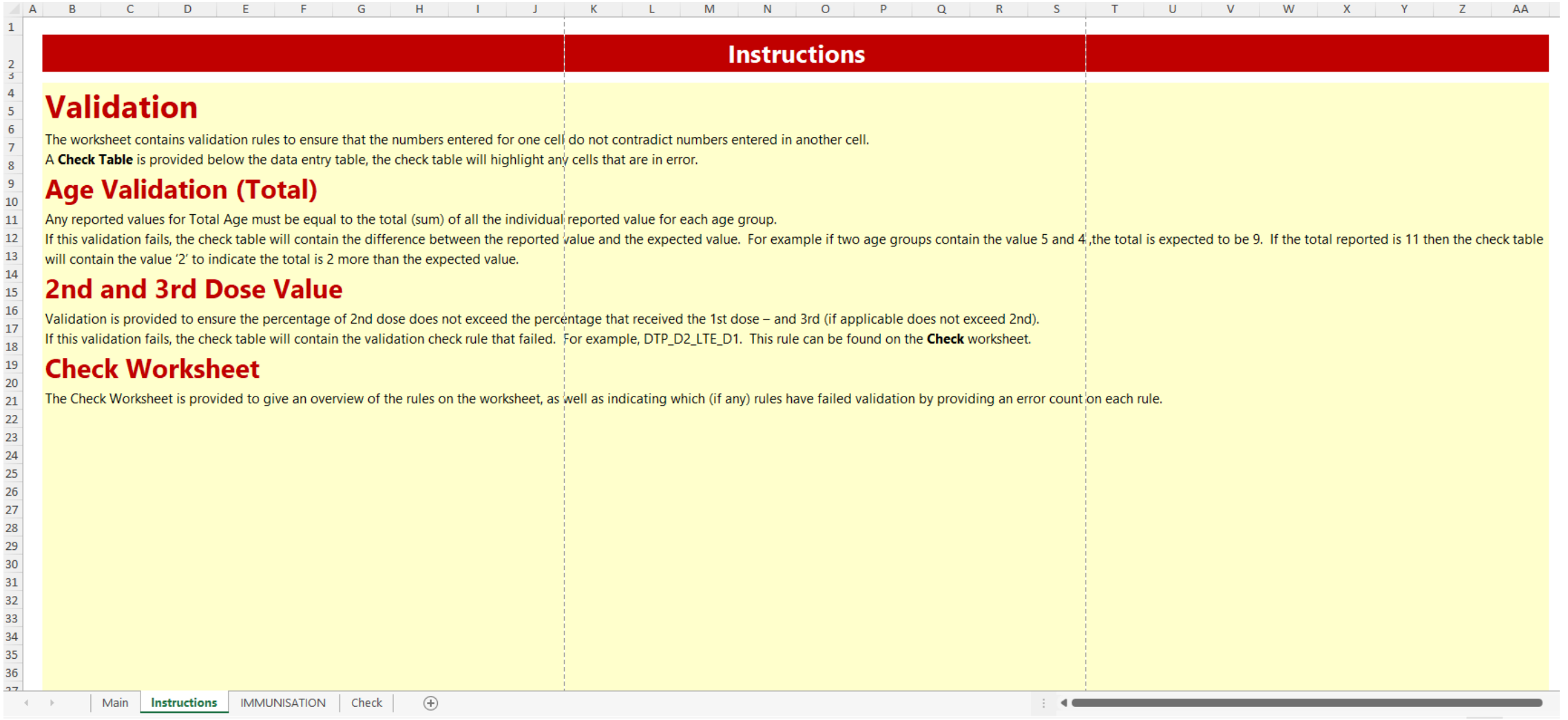
Equality check failures result in the difference between the reported value and the expected value being displayed in the Check Table. Zero values indicate there is no difference between the reported value and the expected value. This summary shows the maximum difference, for each worksheet that define equality checks.

Worksheet	Max Difference
IMMUNISATION	1

Reporting Template – Template Design updated Workflow



Features – Instruction Sheet



Instructions

Validation

The worksheet contains validation rules to ensure that the numbers entered for one cell do not contradict numbers entered in another cell.
A **Check Table** is provided below the data entry table, the check table will highlight any cells that are in error.

Age Validation (Total)

Any reported values for Total Age must be equal to the total (sum) of all the individual reported value for each age group.
If this validation fails, the check table will contain the difference between the reported value and the expected value. For example if two age groups contain the value 5 and 4 ,the total is expected to be 9. If the total reported is 11 then the check table will contain the value '2' to indicate the total is 2 more than the expected value.

2nd and 3rd Dose Value

Validation is provided to ensure the percentage of 2nd dose does not exceed the percentage that received the 1st dose – and 3rd (if applicable does not exceed 2nd).
If this validation fails, the check table will contain the validation check rule that failed. For example, DTP_D2_LTE_D1. This rule can be found on the **Check** worksheet.

Check Worksheet

The Check Worksheet is provided to give an overview of the rules on the worksheet, as well as indicating which (if any) rules have failed validation by providing an error count on each rule.

Main | Instructions | IMMUNISATION | Check

References

FMR Docker 10 min quick start	https://www.sdmx.io/resources/containers/fmr-docker-mysql/
Download FMR	https://www.sdmx.io/resources/download/
FMR product page	https://www.sdmx.io/tools/fmr/
FMR Wiki – general reference	https://fmrwiki.sdmxcloud.org/Main_Page
Excel Reporting Template reference	https://fmrwiki.sdmxcloud.org/Excel_Reporting_Template

Convert Report Template Data to another data format
https://fmrwiki.sdmxcloud.org/Data_Transformation_Web_Service

Asynchronous Data Conversion and Data Validation
https://fmrwiki.sdmxcloud.org/Asynchronous_Data_Validation_and_Transformation_Web_Service