



1964 2014

BANQUE AFRICAINE DE DÉVELOPPEMENT 50 ANS AU SERVICE DE L'AFRIQUE  
AFRICAN DEVELOPMENT BANK 50 YEARS SERVING AFRICA



# SDMX

## Model d'information

Kamel Abdellaoui  
K.ABELLAOUI@afdb.org

Tunis 24- 25 Juin 2019



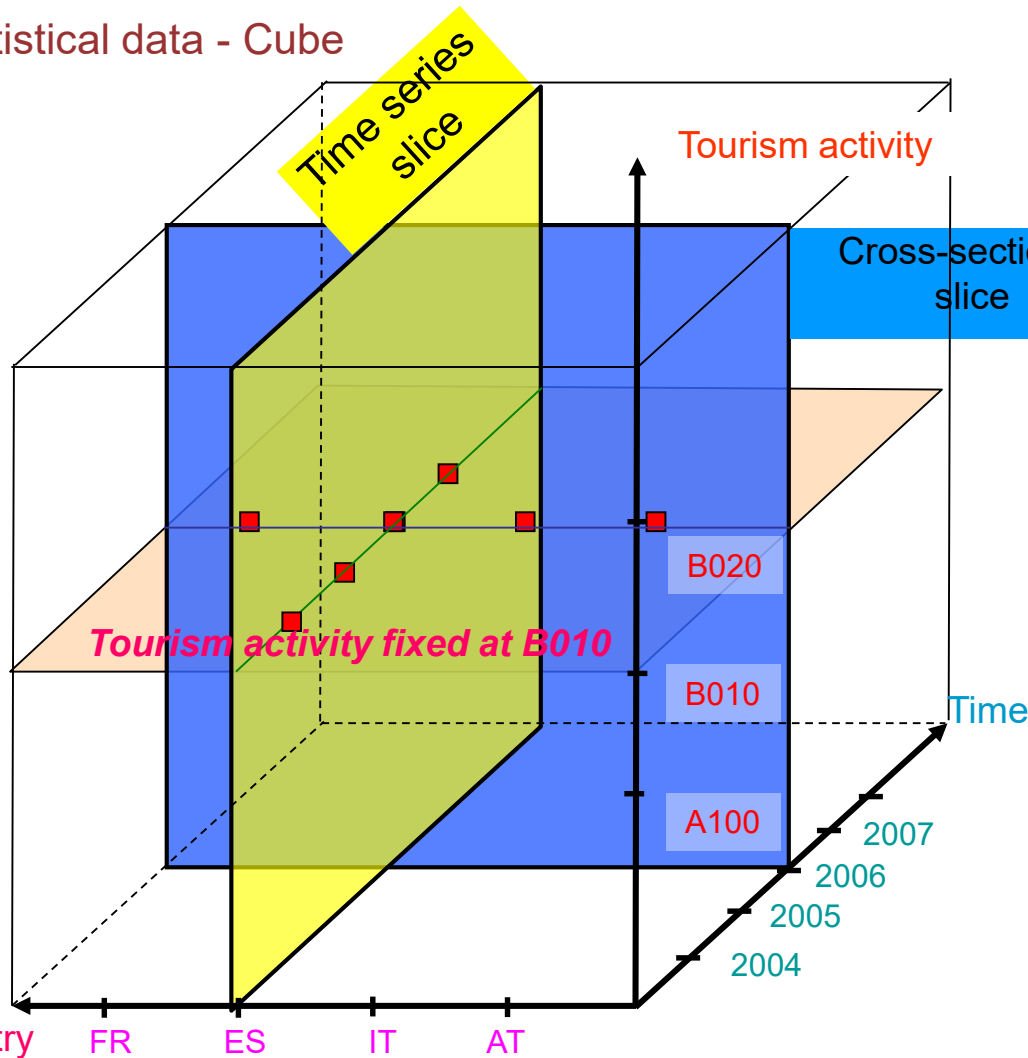
---

**“From Data Set”  
to  
“Data Structure Definition”**

# Different ways to represent data



Statistical data - Cube



Time series

+ UNIT		+ INDIC_TO	
NBR		A001	
+ ACTIVITY ▶		B010	
+ GEO ▶		ES	
TIME			
2004			1.257
2005			1.250
2006			1.216
2007			1.220

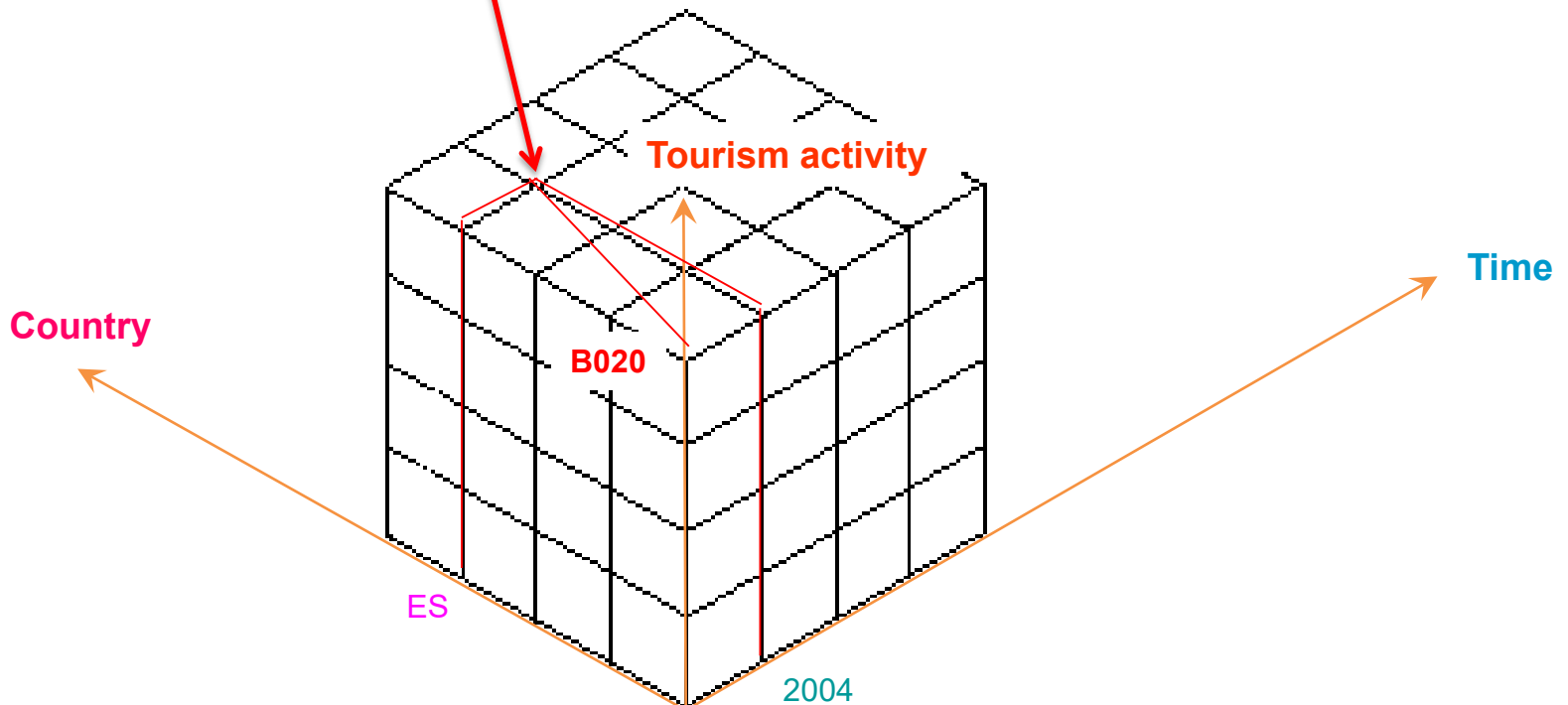
Cross-section for 2006

+ INDIC_TO		+ TIME	
A001		2004	
+ ACTIVITY ▶		B010	
+ GEO			
ES			1257
FR			8.289
IT			2.529
AT			546

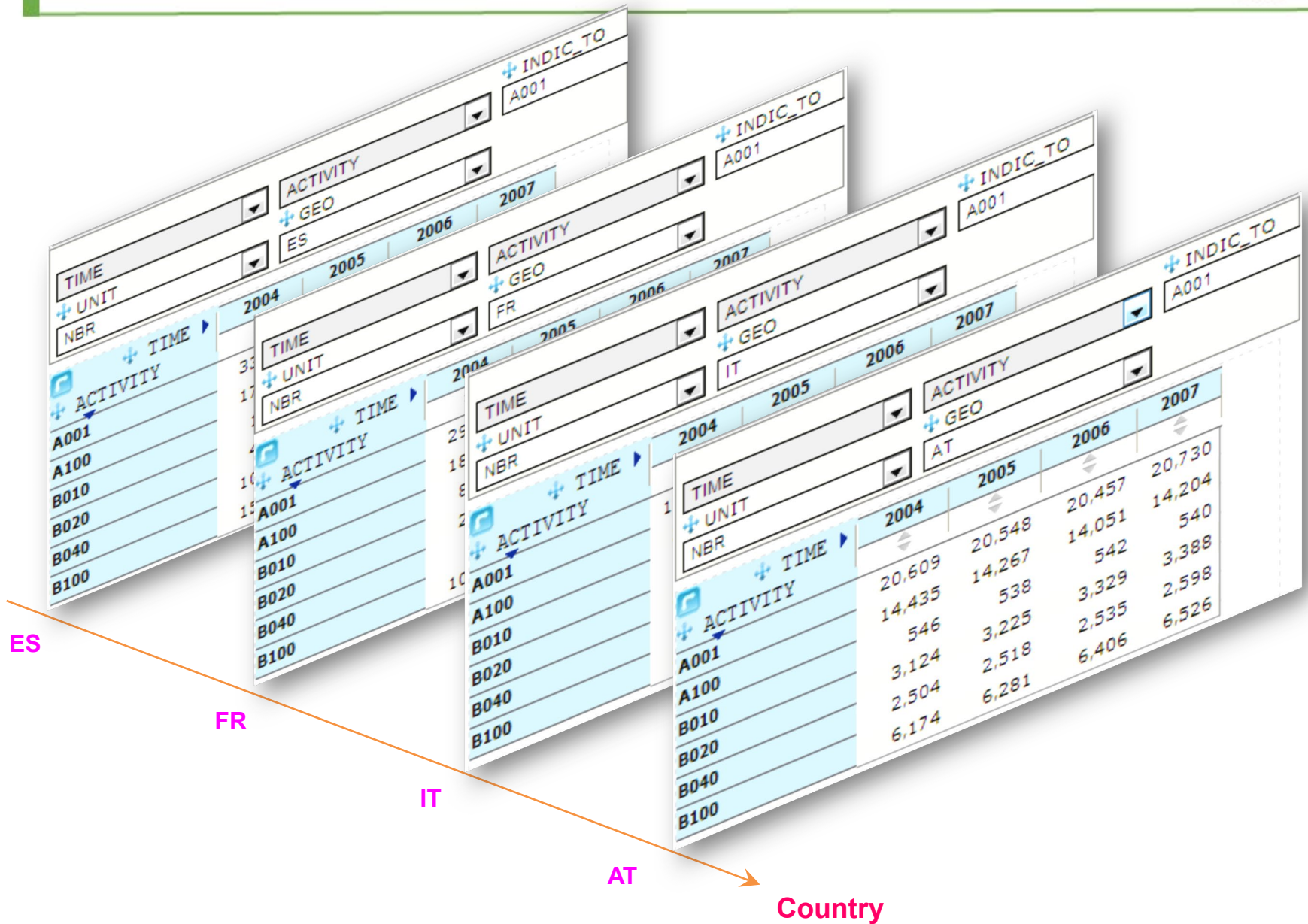
# Statistical data cube



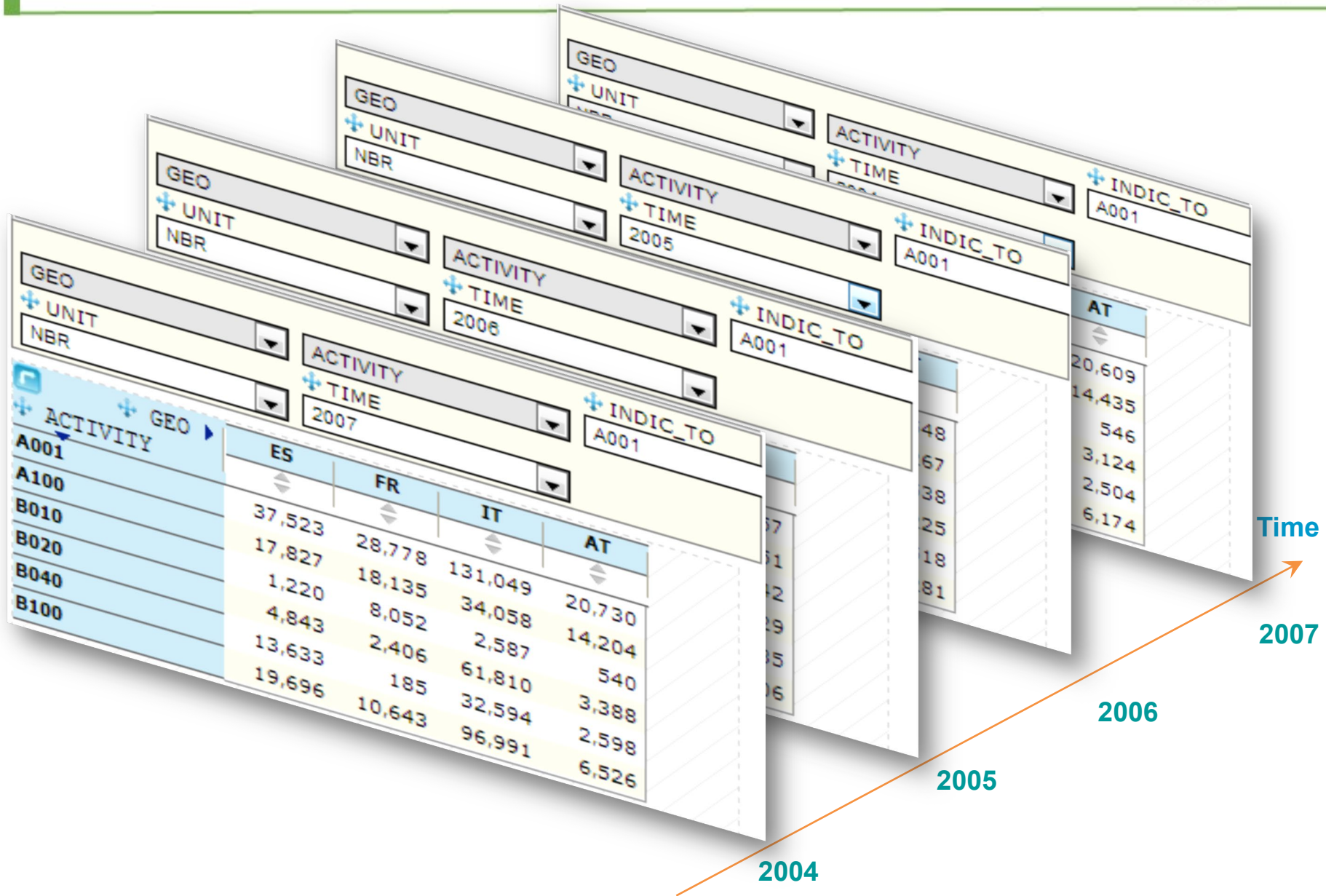
ACTIVITY	TIME	GEO		
+ UNIT	+ INDIC_TO			
NBR	A001			
+ ACTIVITY	B020	B020	B020	B020
+ TIME	2004	2005	2006	2007
+ GEO				
ES	4,405	4,552	4,524	4,843
FR	2,251	2,329	2,325	2,406
IT	56,586	68,385	68,376	61,810
AT	3,124	3,225	3,329	3,388



# Statistical data – time series presentation



# STATISTICAL DATA– Cross-sectional



# From a statistical table to a Data Structure Definition: Tourism example (Concepts)



Number of establishments, bedrooms and bed-places - national - annual data

**FREQUENCY**

**TOURISM\_ACTIVITY**

**UNIT**

**TOURISM\_INDICATOR**

**TIME**

**COUNTRY**

**OBS\_STATUS**

Concept Identifier	Concept name
FREQUENCY	Frequency
COUNTRY	Country
TOURISM_INDICATOR	Tourism Indicator
TOURISM_ACTIVITY	Tourism Activity
TIME	Time Period
OBS_VALUE	Observation
UNIT	Unit
OBS_STATUS	Observation status

Last update: 05-03-2013

Table Customization [show](#)

INDIC\_TO  
A001 - A001 - Number of establishments

UNIT  
NBR - NBR - Number

ACTIVITY  
A100 - A100 - Hotels and similar establishments

TIME	2004	2005	2006	2007
GEO				
ES - Spain		e 17,607	18,304	17,827
FR - France		p 18,689		
IT - Italy		33,527		
AT - Austria	14,435	14,267		

Available flags:

b break in series p provisional c confidential  
e estimated  
f forecast  
i see metadata

# From a statistical table to a Data Structure Definition.

## Tourism example (Concept Scheme)



Concept Scheme:	
ID:	CS_TOURISM
Version:	1,0
Maintenance Agency:	ESTAT
Name:	(English) List of statistical concepts of tourism tables
	(French) Liste des concepts statistiques des tables tourisme

Concept Identifier	Concept name
FREQUENCY	Frequency
COUNTRY	Country
TOURISM_INDICATOR	Tourism Indicator
TOURISM_ACTIVITY	Tourism Activity
TIME	Time Period
OBS_VALUE	Observation
UNIT	Unit
OBS_STATUS	Observation status

**Concept Scheme:** List of concepts used in a data structure definitions or in a metadata structure definitions.



# From a statistical table to a Data Structure Definition.

## Tourism example (Codelists)



### Number of establishments, bedrooms and bed-places - national - annual data

Last update: 05-03-2013

Table Customization [show](#)

INDIC\_TO: A001 - A001 - Number of establishments

TIME: [dropdown] GEO: [dropdown]

UNIT: NBR - NBR - Number ACTIVITY: A100 - A100 - Hotels and similar establishments

TIME	2004	2005	2006	2007
GEO				
ES - Spain	17,402	17,607	18,304	17,827
FR - France	18,598	18,689	18,361	18,135
IT - Italy	33,518	33,527	33,768	34,058
AT - Austria	14,435	14,267	14,051	

- Available flags:
- b break in series
  - p provisional
  - c confidential
  - e estimated
  - r revised
  - n not significant
  - f forecast
  - s Eurostat estimate
  - z not applicable (incl. real zero)
  - i see metadata
  - u unreliable

Do we need codes for concept values?

# From a statistical table to a Data Structure Definition.

## Tourism example (Codelists)



INDIC\_TO

<input type="checkbox"/>	Code	Label
<input checked="" type="checkbox"/>	A001	Number of establishments
<input checked="" type="checkbox"/>	A002	Bedrooms
<input checked="" type="checkbox"/>	A003	Number of bed-places

CL\_TOUR\_INDICATOR

UNIT

<input type="checkbox"/>	Code	Label
<input checked="" type="checkbox"/>	NBR	Number

CL\_UNIT

Back to Select Data

ACTIVITY

<input type="checkbox"/>	Code	Label
<input checked="" type="checkbox"/>	A001	Collective tourist accommodation establishments
<input checked="" type="checkbox"/>	A100	Hotels and similar establishments
<input type="checkbox"/>	B040	accommodation n.e.s.
<input checked="" type="checkbox"/>	B100	Other collective accommodation

CL\_TOUR\_ACTIVITY

GEO

<input type="checkbox"/>	Code	Label
<input checked="" type="checkbox"/>	AT	Austria
<input checked="" type="checkbox"/>	ES	Spain
<input checked="" type="checkbox"/>	FR	France
<input checked="" type="checkbox"/>	IT	Italy

CL\_AREA

**Codelist:** A predefined list from which some statistical coded concepts take their values.

# From a statistical table to a Data Structure Definition.

## Tourism example (Codelists)



- Each Code List is defined uniquely by:
1. an ID,
  2. a maintenance agency,
  3. a version.

The name can be provided in several languages.

Code lists	
<b>ID:</b>	CL_AREA
<b>Version:</b>	1,0
<b>Maintenance Agency:</b>	ESTAT
<b>Name:</b>	(English) List of géographical ISO codes
Code ID	Name (English)
AT	Austria
BE	Belgium
DE	Germany
ES	Spain
FR	France
IT	Italy
PT	Portugal

+	GEO
ES - Spain	
FR - France	
IT - Italy	
AT - Austria	

# From a statistical table to a Data Structure Definition.

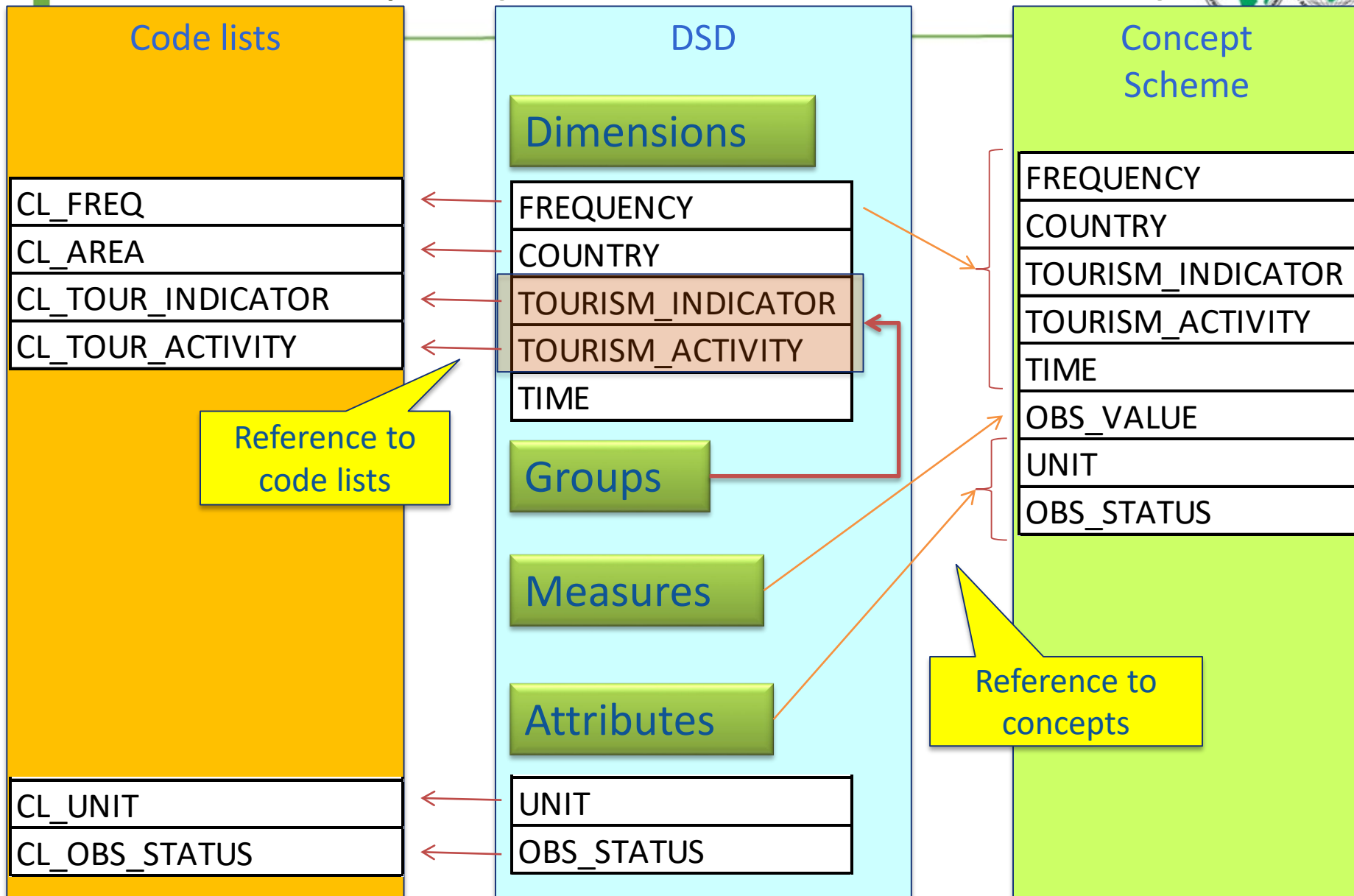
## Tourism example (ConceptScheme & Codelists)



<b>Concept Scheme:</b>	
<b>ID:</b>	CS_TOURISM
<b>Version:</b>	1.0
<b>Maintenance Agency:</b>	ESTAT
<b>Name:</b>	(English) List of statistical concepts of tourism tables
	(French) Liste des concepts statistiques des tables tourisme

<b>Concept Identifier</b>	<b>Concept name</b>	<b>Code list</b>
FREQUENCY	Frequency	CL_FREQ
COUNTRY	Country	CL_AREA
TOURISM_INDICATOR	Tourism Indicator	CL_TOUR_INDICATOR
TOURISM_ACTIVITY	Tourism Activity	CL_TOUR_ACTIVITY
TIME	Time Period	
OBS_VALUE	Observation	
UNIT	Unit	CL_UNIT
OBS_STATUS	Observation status	CL_OBS_STATUS

# Tourism example (DSD element and references)



# Tourism example (Data Structure Definition)



Number of establishments, bedrooms and bed-places - national - annual data

FREQUENCY

Last update: 05-03-2013

Table Customization [show](#)

TIME	2004	2005	2006	2007
ES - Spain		e 17,607	18,304	17,827
FR - France		p 18,689	18,361	18,135
IT - Italy		33,527		
AT - Austria	14,435	14,267	14,267	

TOURISM\_ACTIVITY

UNIT

TOURISM\_INDICATOR

TIME

COUNTRY

OBS\_VALUE

Available flags:

b break in series p provisional c confidential  
e estimated  
f forecast  
i see metadata

OBS\_STATUS

DIMENSIONS ATTRIBUTES MEASURES

**FREQUENCY**

# Number of establishments, bedrooms and bed-places - national - annual data

Last update: 05-03-2013

Table Customization [show](#)

**TOURISM\_ACTIVITY**

**UNIT**

**TOURISM\_INDICATOR**

**TIME**

**COUNTRY**

**OBS\_VALUE**

**OBS\_STATUS**

	2004	2005	2006	2007
ES - Spain		17,607	18,304	17,827
FR - France		18,689	18,361	18,135
IT - Italy		33,527		
AT - Austria	14,435	14,267	14,267	

Available flags:

- b break in series
- p provisional
- c confidential
- e estimated
- f forecast
- i see metadata

Role	Dimension or attribute name	Identifier	Attachement level	Code list
Dimension	Frequency	FREQ		CL_FREQ
Dimension	Country	COUNTRY		CL_AREA
Dimension	Tourism Indicator	TOUR_INDICATOR		CL_TOUR_IND
Dimension	Tourism activity	TOURISM_ACTIVITY		CL_TOUR_ACT
Dimension	Period	TIME		
Attribute	Unit	UNIT	Dataset	CL_UNIT
Attribute	Observation status	OBS_STATUS	Observation	CL_OBS_STATUS
Measure	Observation value	OBS_VALUE		

# Tourism example (Data Structure Definition)



<b>Data Structure Definition:</b>	
<b>ID:</b>	TOURISM_A
<b>Version:</b>	1.0
<b>Maintenance Agency:</b>	ESTAT
<b>Name:</b>	(English) Data structure definition for table Number of establishments, bedrooms and bed-places
	(French) DSD pour la table Nombre d'établissements, chambres et lits

Role	Dimension or attribute name	Identifier	Attachement level	Code list
Dimension	Frequency	FREQ		CL_FREQ
Dimension	Country	COUNTRY		CL_AREA
Dimension	Tourism Indicator	TOUR_INDICATOR		CL_TOUR_IND
Dimension	Tourism activity	TOURISM_ACTIVITY		CL_TOUR_ACT
Dimension	Period	TIME		
Attribute	Unit	UNIT	Dataset	CL_UNIT
Attribute	Observation status	OBS_STATUS	Observation	CL_OBS_STATUS
Measure	Observation value	OBS_VALUE		



# Tourism DSD: summary



## DIMENSIONS

DIMENSIONS											
Position in Key	CONCEPT					REPRESENTATION				Dimension Type	XS Level
	ID	Name	CONCEPT SCHEME			CODE LIST			TEXT		
			ID	VER	AGENCY	ID	VER	AGENCY			
1	FREQ	Frequency	TOUR_CONCEPTS	1,0	ESTAT	CL_FREQ	1,0	ESTAT		Frequency	Section
2	COUNTRY	Country	TOUR_CONCEPTS	1,0	ESTAT	CL_AREA	1,0	ESTAT		Measure	Observation
3	TOURISM_INDICATOR	Tourism indicator	TOUR_CONCEPTS	1,0	ESTAT	CL_TOUR_IND	1,0	ESTAT			Section
4	TOURISM_ACTIVITY	Toursim activity	TOUR_CONCEPTS	1,0	ESTAT	CL_TOUR_ACT	1,0	ESTAT			Section
TIME	TIME	Period	TOUR_CONCEPTS	1,0	ESTAT				Integer	Time	Section

## MEASURES

MEASURES											
Type	CONCEPT					REPRESENTATION				Measure Dimension	Code
	ID	Name	CONCEPT SCHEME			CODE LIST			TEXT		
			ID	VER	AGENCY	ID	VER	AGENCY			
Primary	OBS_VALUE	Observation Value	TOUR_CONCEPTS	1,0	ESTAT				Decimal	N/A	N/A

## ATTRIBUTES

ATTRIBUTES											
Attach-ment Level	CONCEPT					REPRESENTATION				Assignment Status	XS Level
	ID	Name	CONCEPT SCHEME			CODE LIST			TEXT		
			ID	VER	AGENCY	ID	VER	AGENCY			
Observation	OBS_STATUS	Observation status	TOUR_CONCEPTS	1,0	ESTAT	CL_OBS_STATUS	1,0	ESTAT		Conditional	Observation
Dataset	UNIT	UNIT	TOUR_CONCEPTS	1,0	ESTAT	CL_UNIT	1,0	ESTAT		Mandatory	Dataset

# Examples of concrete classes

# Example of flat Codelist



CodeList	
<b>id</b>	CL_FREQ
<b>Name (en)</b>	Frequency
<b>Name (es)</b>	Frecuencia
<b>Description (en)</b>	it provides a list of values indicating the "frequency" of the data (e.g. monthly) and, thus, indirectly, also implying the type of "time reference" that could be used for identifying the data with respect time.
<b>Description (es)</b>	Se proporciona una lista de valores que indican la "frecuencia" de los datos (por ejemplo, mensual) y, por tanto, indirectamente, también define el tipo de "tiempo de referencia" que podría ser utilizado para la identificación de los datos con respecto tiempo.
<b>version</b>	1.0
<b>agencyID</b>	SDMX

Each Code List is defined uniquely by:

1. an ID,
2. a maintenance agency,
3. a version.

The name and description can be provided in several languages.

Each item is defined by an ID,

The name and description can be provided in several languages.

id		Name	Description
A	en	Annual	Annual
	es	Anual	Anual
M	en	Monthly	Monthly
	es	Mensual	Mensual
Q	en	Quarterly	Quarterly
	es	Trimestral	Trimestral
S	en	Half-yearly, semestre	Half-yearly, semestre
	es	Semestral	Semestral
W	en	Weekly	Weekly
	es	Semanal	Semanal

# Example of flat codelist with simple hierarchy



## CL\_NUTS

id	Description	Parent id
BE2	VLAAMS GEWEST	
BE3	REGIONE WALLONE	
BE31	Prov. Brabant Wallon	BE3
BE32	Prov. Hainaut	BE3
BE321	Ath	BE32
BE322	Charleroi	BE32
BE323	Mons	BE32
BE324	Mouscron	BE32
BE33	Prov. Lieja	BE3
BE34	Prov. Luxemburgo (B)	BE3
BE35	Prov. Namur	BE3

```
-BE2
-BE3
  |-BE31
  |-BE32
    |-BE321
    |-BE322
    |-BE323
    |-BE324
  |-BE33
  |-BE34
  |-BE35
```

# Example of Category Scheme



CategoryScheme	
id	DDB_21
Name (en)	Dissemination Database
Name (es)	Base de Datos de Difusión
Description (en)	Common dissemination category scheme <u>thta</u> can comprise high-level categories such as finance, economics, health, tourism, transport, demography
Description (es)	Esquema de categoría más habituales para la difusión de los datos que puede comprender muchas categorías de alto nivel como finanzas, economía, salud, turismo, transporte, demografía
version	1.0
agencyID	ESTAT

□ **Category Schema:**  
A hierarchical division of categories into groups based on common characteristics.

CATEGORIES				
id	lang	Name	Description	Parent category id
DEMO	en	Demography	Demography	
	es	Demografía y población	Demografía y población	
STS	en	Short Term Statistics	Short Term Statistics	
	es	Estadísticas Económicas a Corto Plazo	Estadísticas Económicas a Corto Plazo	
PROD	en	Production	Production	STS
	es	Producción	Producción	

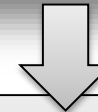
+ [DEMO] Demography  
+ [STS] Short term Statistics  
|- [PROD]: Production

# Example of Dataflow



Dataflow	
id	STS_PROD_MANU
Name (en)	Indices of manufacturing without coffee threshing
Name (es)	Indices de la actividad manufacturera sin trilla
Description (en)	Statistical concepts for the manufacturing industry
Description (es)	Indices of production and actual sales and employment by manufacturing activity with threshing
version	1.0
agencyID	ESTAT

CategoryScheme	
id	DDB_21
Name (en)	Dissemination Database
Name (es)	Base de Datos de Difusión



+ Demography  
 + Short term Statistics  
 + Production  
 - Indices Of Manufacturing....

Referenced Data Structure Definition	
id	PROD_MANUFACT_TS
version	1.0
agencyID	ESTAT

**Data Flow:**  
 A structure which describes, categorises and constrains the allowable content of a collection of data

# Data Structure Definition, Dataflow, Dataset



DSD

DIMENSIONS													
Position in Key	CONCEPT				REPRESENTATION				Dimension Type				
	ID	Name	CONCEPT SCHEME	CODELIST	ID	VER	AGENCY	TEXT FORMAT					
1	FREQ	Frequency	CS_NA	1.2	ESTAT	CL_FREQ	2.0	SDMX	Frequency				
2	ADJUSTMENT	Adjustment indicator	CS_NA	1.2	ESTAT	CL_ADJUSTMENT	1.1	ESTAT					
3	REF_AREA	Reference area	CS_NA	1.2	ESTAT	CL_AREA	1.2	SDMX					
4	COUNTERPART_AREA	Counterpart area	CS_NA	1.2	ESTAT	CL_AREA	1.2	SDMX					
5	REF_SECTOR	Reporting institutional sector	CS_NA	1.2	ESTAT	CL_SECTOR	1.2	ESTAT					

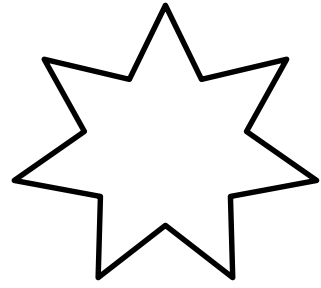
MEASURES												
TYPE	CONCEPT				REPRESENTATION				MEASURE DIMENSION	CODE		
	ID	Name	CONCEPT SCHEME	CODELIST	ID	VER	AGENCY	TEXT FORMAT				
Primary	OBS_VALUE	Observation value	CS_NA	1.2	ESTAT				N/A	N/A		

ATTRIBUTES												
ATTACHMENT LEVEL	CONCEPT				REPRESENTATION				ATTRIBUTE	ASSIGNMENT STATUS		
	ID	Name	CONCEPT SCHEME	CODELIST	ID	VER	AGENCY	TEXT FORMAT				
Observation	OBS_STATUS	Observation status	CS_NA	1.2	ESTAT	CL_OBS_STATUS	1.1	SDMX			M	
Observation	CONF_STATUS	Confidentiality status	CS_NA	1.2	ESTAT	CL_CONF_STATUS	1.0	SDMX			M	
Observation	COMMENT_OBS	Comments to the observation value	CS_NA	1.2	ESTAT						C	

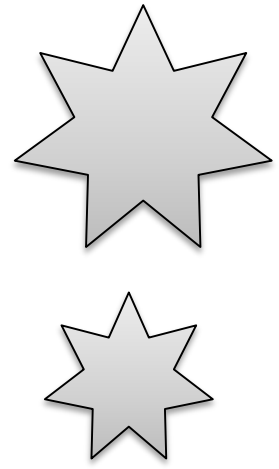
describes

Dataflow



Constraints on dimensions

Dataset



DIMENSIONS														
Position in Key	CONCEPT				REPRESENTATION				Dimension Type	IS level				
	ID	Name	CONCEPT SCHEME	CODELIST	ID	VER	AGENCY	TEXT						
1	FREQ	Frequency	TOUR_CONCEPTS	1.0	ESTAT	CL_FREQ	1.0	ESTAT	Frequency	Section				
2	COUNTRY	Country	TOUR_CONCEPTS	1.0	ESTAT	CL_AREA	1.0	ESTAT	Measure	Observation				
3	TOURISM_INDICATOR	Tourism indicator	TOUR_CONCEPTS	1.0	ESTAT	CL_TOUR_IND	1.0	ESTAT		Section				
4	TOURISM_ACTIVITY	Tourism activity	TOUR_CONCEPTS	1.0	ESTAT	CL_TOUR_ACT	1.0	ESTAT		Section				
Time	TIME	Period	TOUR_CONCEPTS	1.0	ESTAT				Integer	Time				

MEASURES												
Type	CONCEPT				REPRESENTATION				Measure Dimension	Code		
	ID	Name	CONCEPT SCHEME	CODELIST	ID	VER	AGENCY	TEXT				
Primary	OBS_VALUE	Observation value	TOUR_CONCEPTS	1.0	ESTAT				Decimal	N/A	N/A	
Cross-sectional	ES	Spain	TOUR_CONCEPTS	1.0	ESTAT	CL_AREA	1.0	ESTAT	COUNTRY	ES		
Cross-sectional	FR	France	TOUR_CONCEPTS	1.0	ESTAT	CL_AREA	1.0	ESTAT	COUNTRY	FR		
Cross-sectional	AT	Austria	TOUR_CONCEPTS	1.0	ESTAT	CL_AREA	1.0	ESTAT	COUNTRY	AT		
Cross-sectional	IT	Italy	TOUR_CONCEPTS	1.0	ESTAT	CL_AREA	1.0	ESTAT	COUNTRY	IT		

ATTRIBUTES													
Attach-ment level	CONCEPT				REPRESENTATION				Attribute	Assignment Status	IS level		
	ID	Name	CONCEPT SCHEME	CODELIST	ID	VER	AGENCY	TEXT					
Observation	OBS_STATUS	Observation status	TOUR_CONCEPTS	1.0	ESTAT	CL_OBS_STATUS	1.0	ESTAT			Conditional	Observation	
Dataset	UNIT	Unit	TOUR_CONCEPTS	1.0	ESTAT	CL_UNIT	1.0	ESTAT			Numerosity	Dataset	

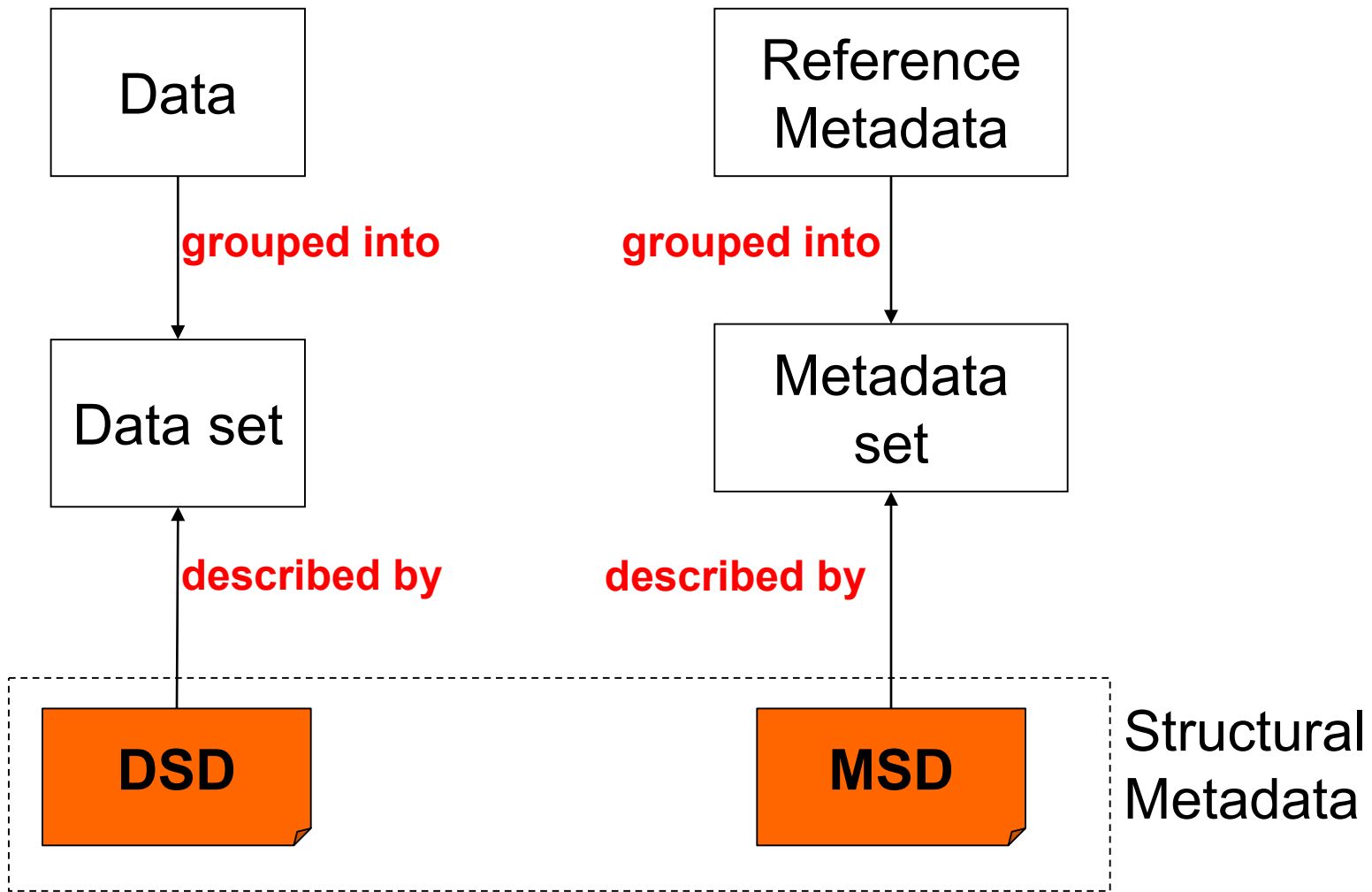
describes



Constraints on dimensions

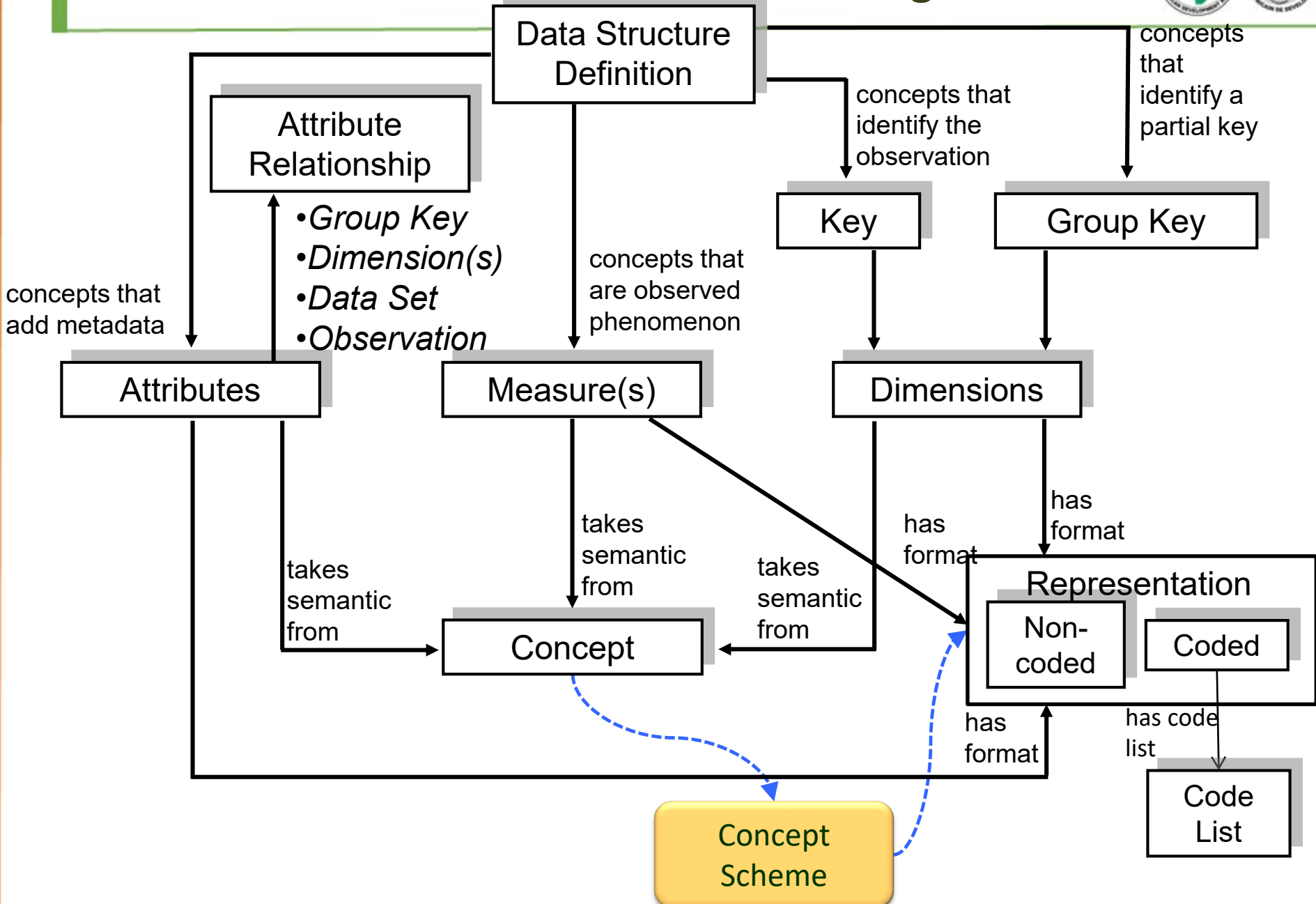


# Data Structure Definition, Dataflow, Dataset

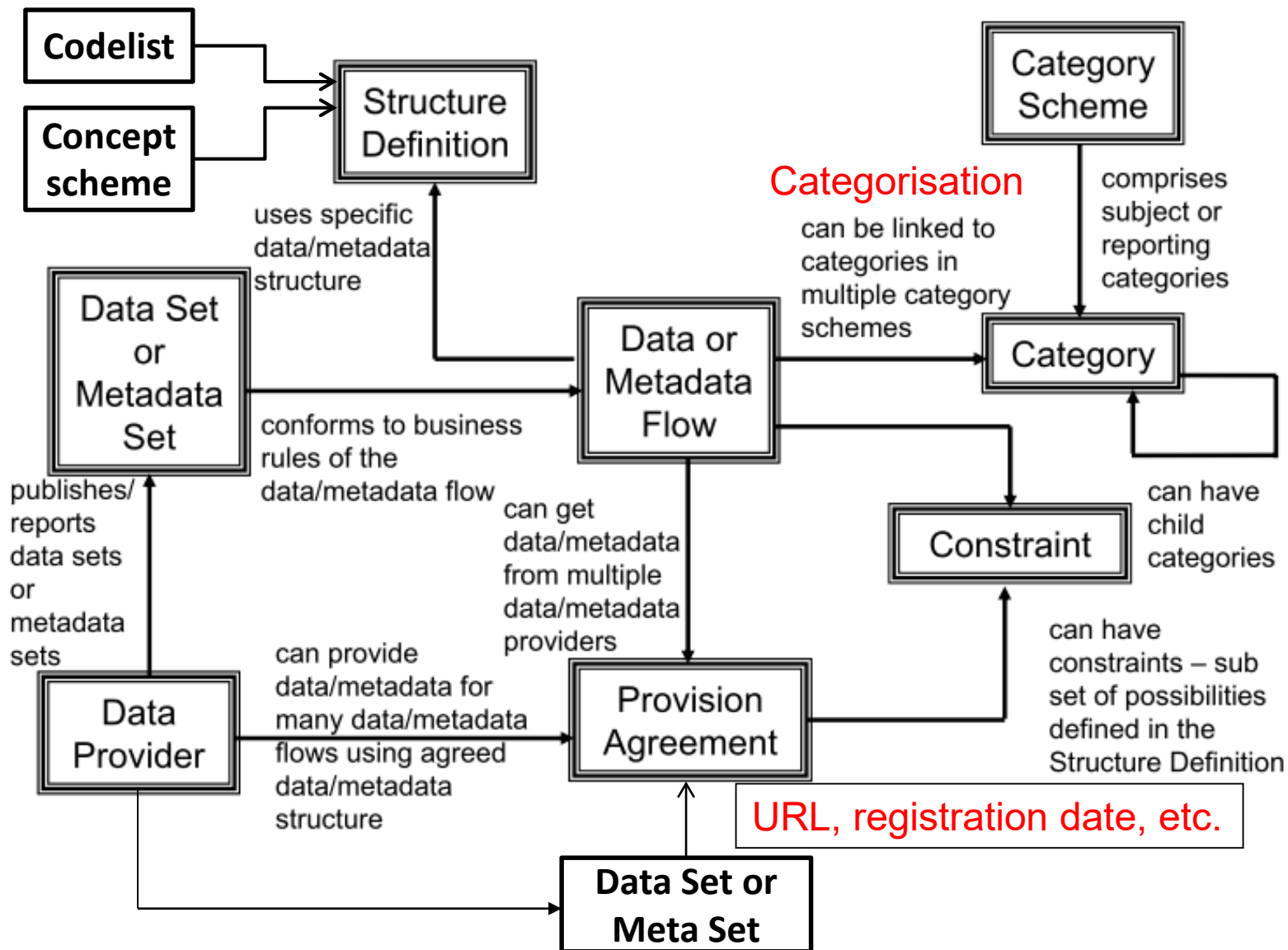




# SDMX Data Structure Definition building blocks



# High level schematic of the SDMX Information Model





- ❑ **Concept Schema**: List of concepts used in a data structure definitions or in a metadata structure definitions.
  
- ❑ **Codelist**: A predefined list from which some statistical coded concepts take their values.
  
- ❑ **Category Schema**: A hierarchical division of categories into groups based on common characteristics.
  
- ❑ **AgencyScheme, DataProviderScheme, DataConsumerScheme, OrganisationUnitScheme** (which all inherit from the abstract class OrganisationScheme)

# Information Model – Structure concrete classes (1/2)

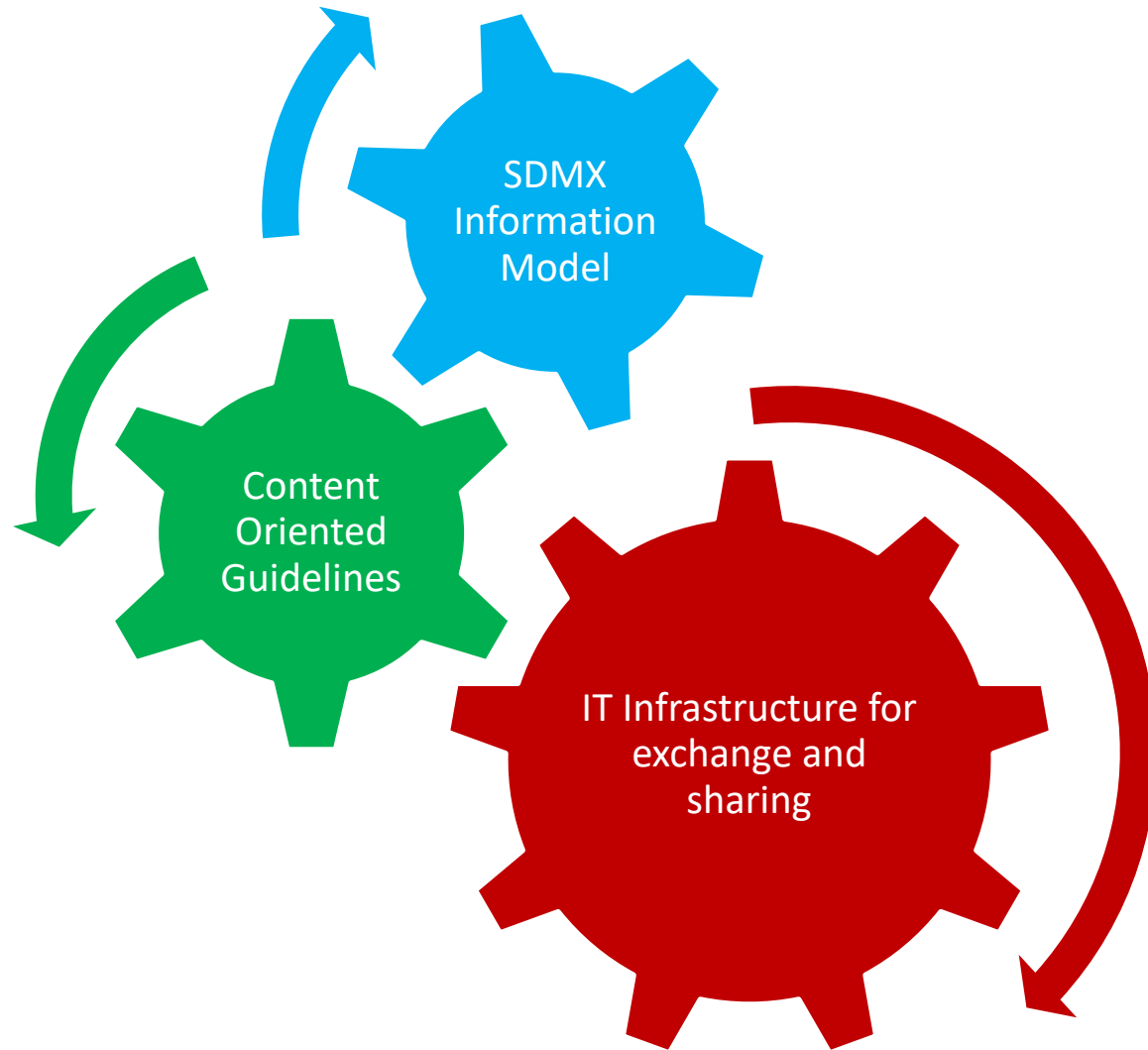


- ❑ **Dataset**: A collection of data reported or disseminated according to a dataflow
- ❑ **Data Structure Definition (DSD)**: set of structural metadata associated to a dataset, which include information about how concepts are associated with the measures, the dimensions and the attributes of a data cube; along with information about the representation of data and related descriptive metadata
- ❑ **Data Flow**: A structure which describes, categorises and constrains the allowable content of a collection of data
- ❑ **Provisional agreement**: The set of information which describes the way in which data sets and metadata sets are provided by a data provider
- ❑ **Constraints**: a construct that is used to define set of Code Lists or series key in order to apply validation or to assist visualisation.



- ❑ **Hierarchical Codelist:** specifies the organisation of the codes in one or more hierarchies, but does not define the codes themselves
- ❑ **Metadata Structure Definition:** a collection of metadata concepts, their structure and usage when used to collect, report or disseminate reference metadata
- ❑ **Metadata Set:** any organized collection of reference metadata
- ❑ **Metadata Flow:** a structure (without any metadata) of a flow of reference metadata, that providers will provide for different reference periods

# SDMX components



# THANK YOU

[A.AIH@AFDB.ORG](mailto:A.AIH@AFDB.ORG)

