1964 2014

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

sdmx
Statistical Data and Metadata eXchange

Modélisation des données

Kamel Abdellaoui
K.ABDELLAOUI@afdb.org

Tunis 24- 25 Juin 2019

Chiffres vs Données

Population by region, district, age groups and sex, 2010

Districts	All ages			0-14 years			15-64 years			65+ years		
	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female
All regions	**24,658,823**	**12,024,845**	**12,633,978**	**9,450,398**	**4,798,944**	**4,651,454**	**14,040,893**	**6,727,948**	**7,312,945**	**1,167,532**	**497,953**	**669,579**
Western	**2,376,021**	**1,187,774**	**1,188,247**	**926,514**	**470,537**	**455,977**	**1,359,590**	**676,710**	**682,880**	**89,917**	**40,527**	**49,390**
Jomoro	150,107	73,561	76,546	60,046	30,711	29,335	83,467	40,250	43,217	6,594	2,600	3,994
Ellembelle	87,501	42,317	45,184	34,465	17,397	17,068	48,730	23,301	25,429	4,306	1,619	2,687
Nzema East	60,828	29,947	30,881	24,960	12,802	12,158	33,575	16,265	17,310	2,293	880	1,413
Ahanta West	106,215	50,999	55,216	44,014	22,157	21,857	57,520	27,080	30,440	4,681	1,762	2,919
Sekondi Takoradi Metropolis	559,548	273,436	286,112	182,674	91,060	91,614	353,662	172,401	181,261	23,212	9,975	13,237
Kwesimintsim	*232,617*	*113,726*	*118,891*	*76,332*	*38,122*	*38,210*	*147,490*	*71,595*	*75,895*	*8,795*	*4,009*	*4,786*
Takoradi	*97,352*	*48,470*	*48,882*	*27,920*	*13,732*	*14,188*	*65,292*	*32,905*	*32,387*	*4,140*	*1,833*	*2,307*
Sekondi	*70,361*	*33,828*	*36,533*	*21,841*	*10,996*	*10,845*	*44,573*	*21,256*	*23,317*	*3,947*	*1,576*	*2,371*
Essikadu-Ketan	*159,218*	*77,412*	*81,806*	*56,581*	*28,210*	*28,371*	*96,307*	*46,645*	*49,662*	*6,330*	*2,557*	*3,773*
Shama	81,966	38,704	43,262	33,769	17,112	16,657	44,323	20,158	24,165	3,874	1,434	2,440
Mpohor-Wassa East	123,996	62,470	61,526	51,792	26,760	25,032	67,292	33,401	33,891	4,912	2,309	2,603
Tarkwa Nsuaem Municipal	90,477	46,662	43,815	34,464	17,447	17,017	53,314	27,929	25,385	2,699	1,286	1,413
Prestea/Huni Valley	159,304	80,493	78,811	64,965	32,879	32,086	89,098	45,111	43,987	5,241	2,503	2,738
Wassa Amenfi East	83,478	42,896	40,582	35,055	18,115	16,940	45,297	23,207	22,090	3,126	1,574	1,552
Wassa Amenfi West	161,166	83,227	77,939	67,552	34,797	32,755	88,404	45,732	42,672	5,210	2,698	2,512

Chiffres vs Données

- Les **Chiffres** par eux-mêmes sont sans signification.

- Les **données** doivent être correctement **décrites** pour être utilisables. Les descriptions permettent aux utilisateurs de savoir ce que les données représentent réellement.

Développement d'un modèle de données pour un échange en SDMX

- Similaire au développement d'une base de données relationnelle

- En SDMX, le modèle de données est représenté par un **Data Structure Definition**.
  - La «forme» du DSD est à peu près similaire à celle du schéma en étoile.

- Pour concevoir un DSD, nous devons d'abord trouver les *concepts* qui **identifient** et **décrivent** les données**.**

Développement d'un modèle de données pour un échange en SDMX

Concepts :

- “A unit of knowledge created by a unique combination of characteristics”*
- Chaque concept décrit quelque chose à propos des données.
- Les concepts doivent exprimer toutes les caractéristiques pertinentes des données.

* Source: **Metadata Common Vocabulary**

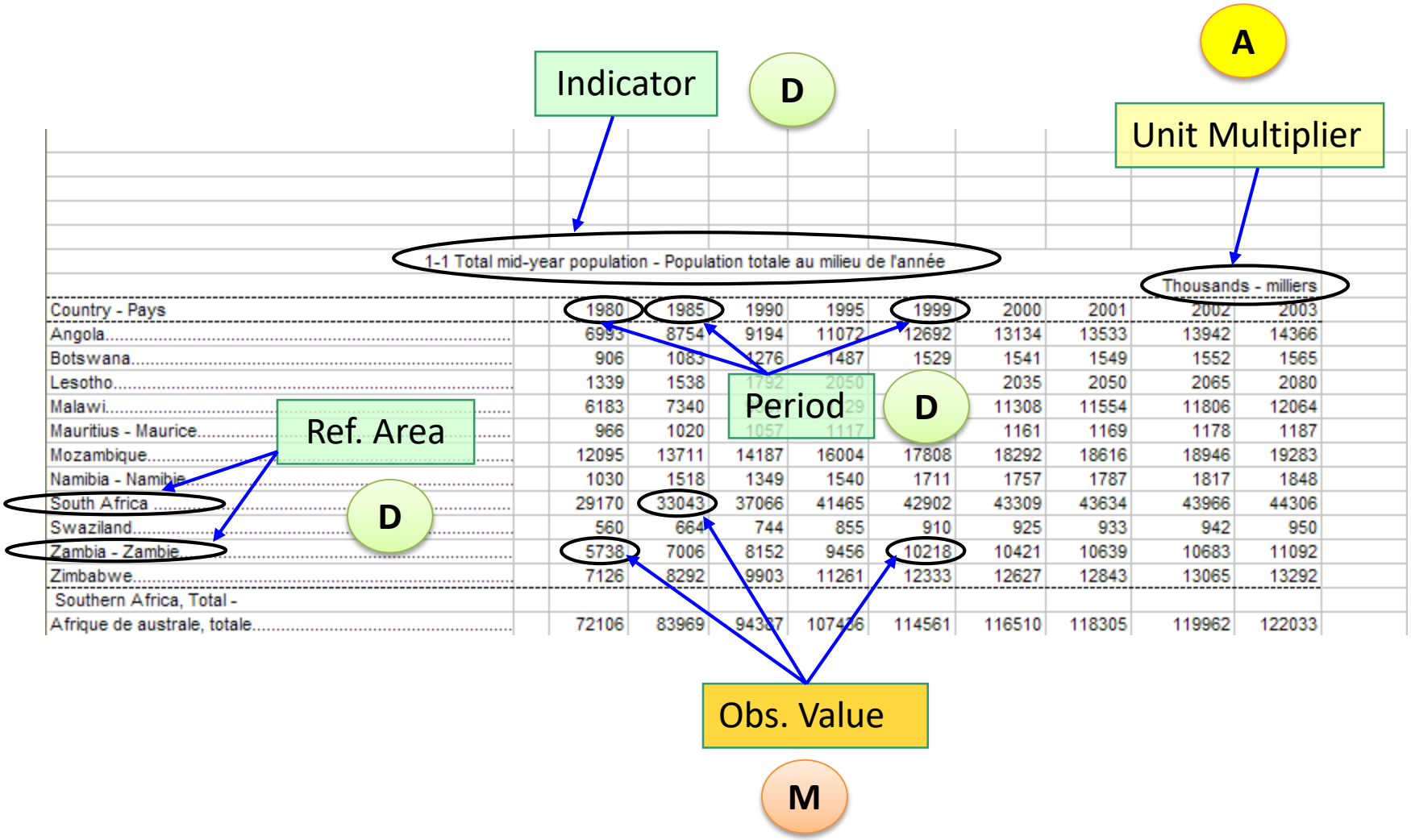
SDMX Data Structure Definition building blocks

Dimensions qui identifient la valeur d'observation

Attributes qui ajoute une métadonnée additionnelle à propos de la valeur d'observation

Measure qui est la valeur d'observation

Identification des Concepts



Country - Pays	1980	1985	1990	1995	1999	2000	2001	2002	2003
Angola	6993	8754	9194	11072	12692	13134	13533	13942	14366
Botswana	906	1083	1276	1487	1529	1541	1549	1552	1565
Lesotho	1339	1538	1792	2050		2035	2050	2065	2080
Malawi	6183	7340	8629			11308	11554	11806	12064
Mauritius - Maurice	966	1020	1057	1117		1161	1169	1178	1187
Mozambique	12095	13711	14187	16004	17808	18292	18616	18946	19283
Namibia - Namibie	1030	1518	1349	1540	1711	1757	1787	1817	1848
South Africa	29170	33043	37066	41465	42902	43309	43634	43966	44306
Swaziland	560	664	744	855	910	925	933	942	950
Zambia - Zambie	5738	7006	8152	9456	10218	10421	10639	10683	11092
Zimbabwe	7126	8292	9903	11261	12333	12627	12843	13065	13292
Southern Africa, Total - Afrique de australe, totale	72106	83969	94387	107486	114561	116510	118305	119962	122033

Dimension

- Lesquels des concepts sont utilisés pour identifier une observation?

Ref. Area	Period	Indicator	
Angola	2000	1-1 Total mid-year population -	13134

• Lorsque tous les 3 sont connus, nous pouvons localiser sans ambiguïté une observation dans le tableau.

1-1 Total mid-year population - Population totale au milieu de l'année

Thousands - milliers

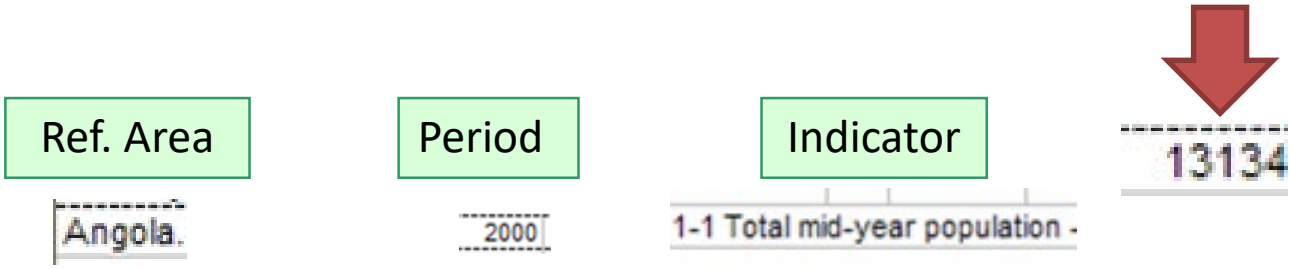
Country - Pays	1980	1985	1990	1995	1999	2000	2001	2002	2003
Angola	6993	8754	9194	11072	12692	13134	13533	13942	14366
Botswana	906	1083	1276	1487	1529	1541	1549	1552	1565
Lesotho	1339	1538	1792	2050	2037	2035	2050	2065	2080
Malawi	6183	7340	9667	11129	11270	11308	11554	11806	12064
Mauritius - Maurice	966	1020	1057	1117	1151	1161	1169	1178	1187
Mozambique	12095	13711	14187	16004	17808	18292	18616	18946	19283
Namibia - Namibie	1030	1518	1349	1540	1711	1757	1787	1817	1848
South Africa	29170	33043	37066	41465	42902	43309	43634	43966	44306
Swaziland	560	664	744	855	910	925	933	942	950
Zambia - Zambie	5738	7006	8152	9456	10218	10421	10639	10683	11092
Zimbabwe	7126	8292	9903	11261	12333	12627	12843	13065	13292
Southern Africa, Total - Afrique de australe, totale	72106	83969	94387	107436	114561	16510	118305	119962	122033

Dimensions Spéciales

- **TIME** dimension : fournit le temps d'observation. Si un DSD décrit des données de série chronologique, il doit avoir une dimension TIME.

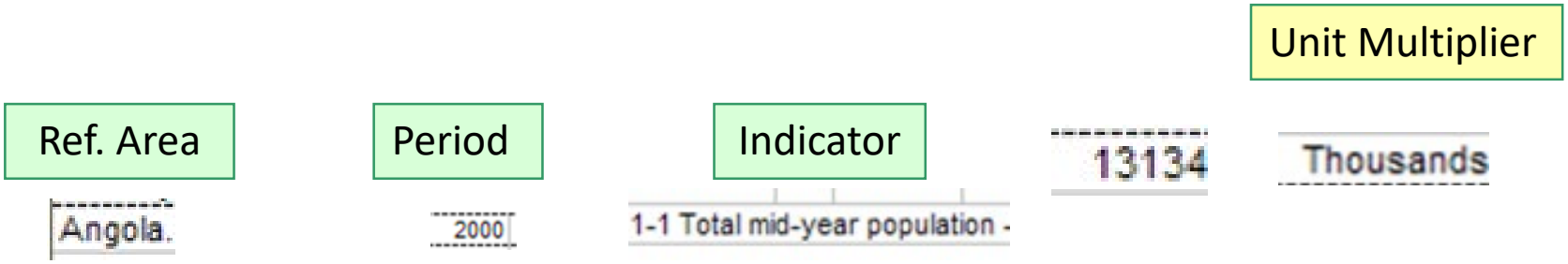
- **FREQUENCY** dimension : décrit l'intervalle entre les observations. S'il existe une dimension TIME, une autre dimension doit être marquée comme dimension FREQUENCY.

Primary Measure



- Observation Value représente un concept décrivant les valeurs réelles transmises.
- En SDMX, ce concept est nommé **Primary Measure**.
- Primary Measure est généralement représenté par le concept **OBS_VALUE**.

Attribute



- Dans notre exemple, **Unit Multiplier** représente des informations supplémentaires sur les observations.
- Ce concept n'est pas utilisé pour identifier une série ou une observation.
- En SDMX ces concepts sont appelés **attributes**.

Exercice1 : Identifier concepts and rôles

MDG Literacy rates of 15-24 years old, both sexes, percentage						Last updated: 02 Jul 2012
Country	**1991**	**2000**	**2002**	**2005**	**2006**	**2010**
Thailand		98.0		98.1		
Uganda	69.8		80.8		84.1	87.4

Literacy rates of 15-24 years old, men, percentage						Last updated: 02 Jul 2012
Country	**1991**	**2000**	**2002**	**2005**	**2006**	**2010**
Thailand		98.1		98.2		
Uganda	77.2		86.0		87.3	89.6

Literacy rates of 15-24 years old, women, percentage						Last updated: 02 Jul 2012
Country	**1991**	**2000**	**2002**	**2005**	**2006**	**2010**
Thailand		97.8		97.9		
Uganda	63.1		76.2		81.1	85.5

Exercice1 : Identify concepts and roles

- Identify concepts in the table
- Mark each concept as:
  - Dimension
  - Time Dimension
  - Primary Measure (i.e. observation value)
  - Attribute

Dimension ou Attribute?

- Le choix du rôle d'un concept a de profondes implications sur la structure des données.
- Concepts that identify data, should be made dimensions. Concepts that provide additional information about data, should be made attributes.

- Si un concept est une dimension, il est possible d'avoir des séries chronologiques qui ne diffèrent que par la valeur de ce concept.
- E.g. if Unit of Measure is a dimension, it is possible to have separate series for “T” and “T/HA” or, more controversially, “KG” and “T”

Dimension ou Attribute? (2)

Cambodia		
Fixed and Mobile telephone subscriptions	2013	20.6 million
Fixed and Mobile telephone subscriptions	2012	19.7 million
Fixed and Mobile telephone subscriptions	2013	140.9 per 100 pop.

Unit of measure en tant que dimension…

Ref.Area (D)	Indicator (D)	Time Period (D)	Unit (D)	Unit Mult. (A)	Obs. Value (M)
Cambodia	Fixed and Mobile telephone subscriptions	2013	Number	Millions	20.6
Cambodia	Fixed and Mobile telephone subscriptions	2012	Number	Millions	19.7
Cambodia	Fixed and Mobile telephone subscriptions	2013	Per 100 pop.	Units	140.9

Dimension ou Attribute? (3)

Unit of measure en tant que attribute…

Violation!

Ref.Area	Indicator	Time Period	Unit	Unit Mult.	Obs. Value
Cambodia	Fixed and Mobile telephone subscriptions	2013	Number	Millions	20.6
Cambodia	Fixed and Mobile telephone subscriptions	2012	Number	Millions	19.7
Cambodia	Fixed and Mobile telephone subscriptions	2013	Per 100 pop.	Units	140.9

- L'ensemble de données ci-dessus n'est pas valide: doublons des clés
- Les deux valeurs ci-dessus ne sont différentes que par leurs attributs

Dimension or Attribute? (4)

"Unit of measure" en tant que attribute avec changement de l'indicateur …

Ref.Area	Indicator	Time Period	Unit	Unit Mult.	Obs. Value
Cambodia	Fixed and Mobile telephone subscriptions	2013	Number	Millions	20.6
Cambodia	Fixed and Mobile telephone subscriptions	2012	Number	Millions	19.7
Cambodia	Fixed and Mobile telephone subscriptions per 100 population	2013	Per 100 pop.	Units	140.9

- Maintenant, il n'y a pas de violation car chaque ligne a une clé de série unique
- Le concept d'unité est tou

Représentation

- Lorsque les données sont transférées, ses la description des ses concepts doivent avoir des valeurs valides.

- A concept peut être :
  - Codifié (Coded)
  - Non codifié avec format (Un-coded with format)
  - Non codifié non formaté (Un-coded free text)

Code

- “A language-independent set of letters, numbers or symbols that represent a concept whose meaning is described in a natural language.”
- A sequence of characters that can be associated with a descriptions in any number of languages.
  - Descriptions can be updated without disrupting mappings or other components of data exchange.

Code List

- Une liste prédéfinie à partir de laquelle certains concepts statistiques codés prennent leurs valeurs. "
- Une liste de codes énumère toutes les valeurs possibles pour un concept ou un ensemble de concepts
  - Liste de code de sexe
  - Liste de codes de pays
  - Liste de codes d'indicateurs, etc.

Code List: Some Examples

Code	Description
SI_POV_DAY1	Population below international poverty line (1.1.1)
SI_POV_EMP1	Employed population below international poverty line (1.1.1)
SI_POV_NAHC	Population below national poverty line (1.2.1)
SI_COV_BENFTS	Population covered by at least one social protection floor/system (1.3.1)
SI_COV_CHLD	Children covered by social protection (1.3.1)
SI_COV_DISAB	Population with severe disabilities collecting disability social protection benefits (1.3.1)
SI_COV_LMKT	Population covered by labour market programs (1.3.1)
SI_COV_MATNL	Mothers receiving maternity benefits and benefits for newborns (1.3.1)
SI_COV_PENSN	Population above retirement age receiving a pension (1.3.1)

Code	Description (EN)	Description (FR)
_T	Total or no breakdown by education level	Total ou aucune ventilation par niveau de s
ISCED11_0	Early childhood education	Education de la petite enfance
ISCED11_01	Early childhood educational development	Développement éducatif de la petite enfan
ISCED11_02	Pre-primary education	Enseignement préprimaire
ISCED11_1	Primary education	Enseignement primaire
ISCED11_10	Primary education	Enseignement primaire

Code	Description
1	World
2	Africa (M49)
4	Afghanistan
5	South America (M49)
8	Albania
9	Oceania (M49)
10	Antarctica
11	Western Africa (M49)
12	Algeria

Un-coded Concepts

- Peut être en texte libre: tout texte valide peut être utilisé comme valeur pour le concept.
  - Footnote , commentaire…
- Peut avoir un format spécifique
  - Code postal : 5 digits

Représentation des concepts en SDMX

- **Dimensions** doivent être codés ou avoir leur format spécifié.
  - Free text n'est pas autorisé.

- **Attributes** peut être codé ou non codé; le format peut éventuellement être spécifié.

Exercice 2: Représentation

- Suite à l'exercice1 , déterminez la représentation de chaque concept
  - Codé, formaté, free-text

- Développer des listes de codes et des formats pour vos concepts
  - Utilisez n'importe quelle approche pour vos codes

MERCI

A.AIH@AFDB.ORG