



**United
Nations**

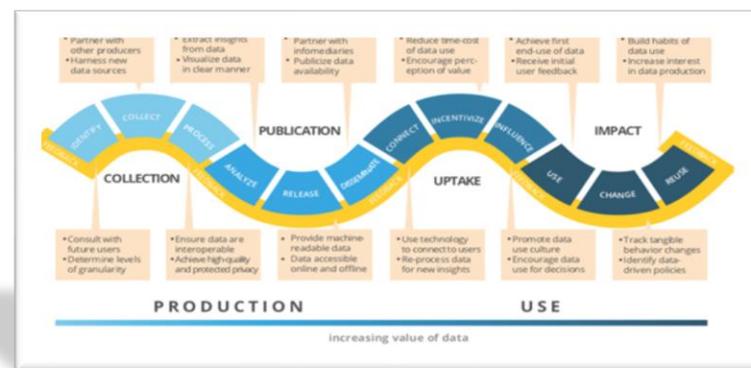
Department of Economic and Social Affairs
Statistics

Introducción a la modelación de datos usando el estándar SDMX

Interoperabilidad

- Capacidad para compartir, unir, analizar en conjunto, intercambiar y reutilizar sin fricciones los datos producidos por diferentes fuentes, en diferentes momentos, con el fin de generar información enriquecida para la mejor toma de decisiones
- La interoperabilidad es una característica de los datos de buena calidad y de los sistemas eficaces de gestión de datos

La interoperabilidad se debe entender a lo largo de toda la cadena de valor de los datos, desde su recolección hasta su utilización



Interoperabilidad y modelación de datos

- La interoperabilidad depende de las decisiones que se tomen sobre la modelación de datos y metadatos
- Un mismo contenido de información puede representarse de muy diversas maneras en diferentes sistemas y organizaciones
- Normalmente no existe una única forma “correcta” de representar la información
 - Algunas estructuras de datos son más apropiadas para la gestión de procesos transaccionales (p.ej., para la captura de datos de una encuesta, o el mantenimiento de una base de datos de registro civil)
 - Otras estructuras son más adecuadas para analizar y comunicar los datos a los usuarios (p.ej., para la creación de visualizaciones de datos para tableros digitales de monitoreo)

¿Qué es la modelación de datos?

- Un proceso que se enfoca en los siguientes aspectos:
 1. **Identificar** claramente y sin ambigüedades las cosas que un conjunto de datos debe capturar
 2. Seleccionar las propiedades claves que deben capturarse para **describir** con sentido (“significado”) esas cosas
 3. Decidir cómo esas cosas descritas se **relacionan** entre sí
 4. Decidir cómo **codificar formalmente** dicha información
- En esencia se trata del modelo entidad-relación que forma parte de la mayoría de los sistemas de gestión de base de datos modernos

Ejemplos

- El contenido de un conjunto de datos puede hacer referencia a **entidades** tales como “ciudad”, “persona” o “actividad”
- Estas entidades pueden ser descritas utilizando **atributos** como “nombre”, “edad”, o “industria”.
- En una aplicación dada, podría ser útil capturar las **relaciones** entre entidades y atributos. Por ejemplo:
 - El hecho que una o más personas vivan en una ciudad,
 - El hecho que una persona esté empleada en uno más tipos de actividad ...

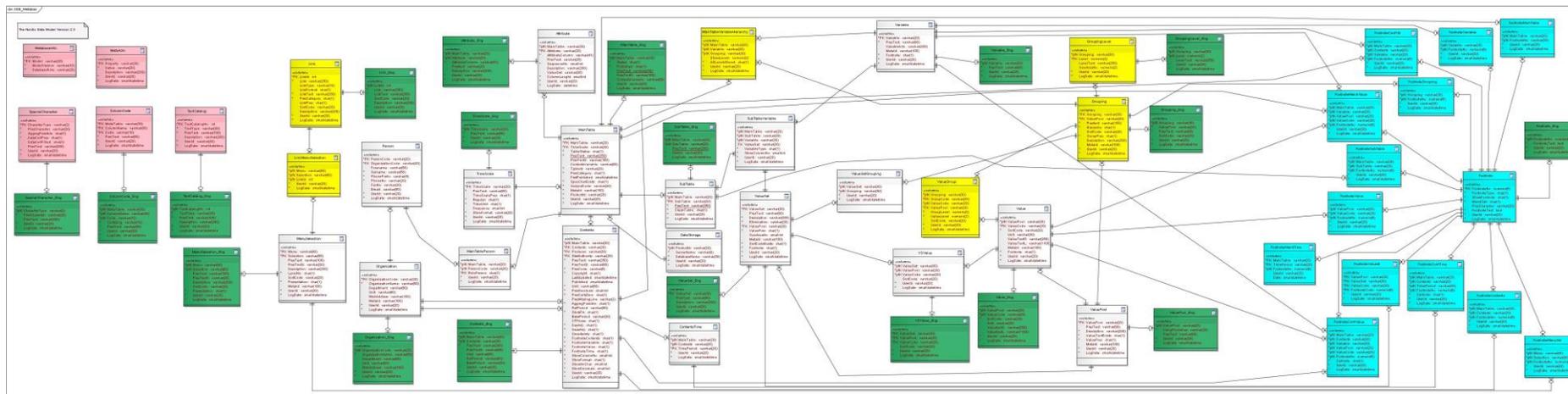
Modelos canónicos de datos y metadatos

- Son modelos que siguen **patrones estandarizados** específicos, que los hacen altamente **reutilizables** y que son propicios para el **intercambio** de datos.
- Proporcionan un **formato común** que puede ser utilizado para expresar diferentes conjuntos de datos
- Contribuyen a establecer una visión común de los **componentes** que forman parte de un conjunto de datos, y como estos se **relacionan** tanto entre si como con otros conjuntos de datos.
- Reducen el numero de transformaciones que los usuarios tienen que hacer por cuenta propia con el fin de **integrar** los datos de diferentes fuentes

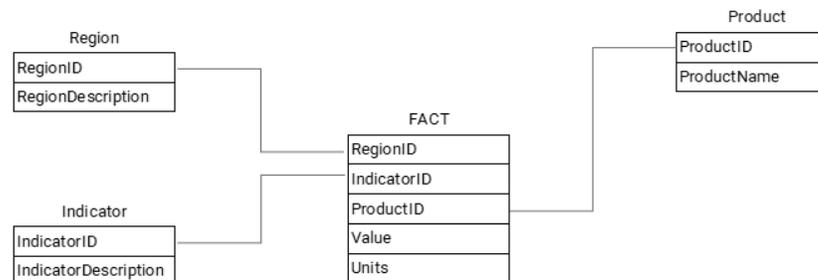
Simplificar las tareas de los usuarios

- El principio básico es **ocultar de la vista de los usuarios la complejidad interna de los modelos operacionales de datos**, de manera que ellos se puedan **concentrar en el uso** de los datos en lugar de tener que dedicar tiempo descifrar su estructura y significado

Modelo de datos operacional



Modelo de datos canónico



Esta labor de simplificación no es gratis

- Los proveedores de datos tienen que asumir la responsabilidad de tomar los datos de sus estructuras operativas originales y **transformarlos** a formatos consensuados para su presentación, diseminación y distribución.
- Esto implica la realización de operaciones de “extracción-transformación-carga” (ETL por sus siglas en inglés) sin que el usuario se percate de ello



El “cubo de datos” multidimensional

- Es un modelo simple que permite presentar en un solo cuadro todos los datos relevantes de una población bajo estudio
- Cada observación se caracteriza por:
 - **Mediciones:** Valores observados acerca de una o más variables de interés
 - **Dimensiones:** Un conjunto de características que identifican de manera única a cada observación y las distinguen una de otra
 - **Atributos:** Un conjunto de características adicionales que ayudan a describir e interpretar cada observación con mayor detalle

Dominios de las dimensiones, atributos y mediciones

- Cada dimensión, atributo o medida representa **un concepto**
- Los conceptos pueden tomar diversos valores en la practica:
 - Tomados de una lista de códigos (p.ej., “código CIIU”)
 - Conforme a un formato específico (p.ej., “aaaa-mm-dd” para fecha de nacimiento)
 - Dentro de un intervalo numérico específico (p.ej., “números entre 0 and 1” para el porcentaje de curules ocupadas por mujeres en el parlamento).
 - Valores de un tipo específico (p.ej., “texto”)

Importancia de la modelación de datos

- Un modelo de datos define el tipo de contenido que puede ser capturado y transmitido.
- Cualquier falla en el diseño del modelo de datos puede tener un efecto negativo sobre la capacidad para intercambiar datos
 - Conceptos faltantes (pérdida de información)
 - Conceptos con roles no adecuados (errores inesperados, tales como registros aparentemente duplicados)
 - Modelos sub-optimizados (p.ej., mensajes “muy pesados”)

Algunas consideraciones para la modelación de datos

- Los datos deben mantener su significado incluso fuera del contexto usual
- Los conceptos deben ser independientes uno del otro
- Se deben evitar dimensiones redundantes
- Las dimensiones deben ser relevantes para todas o para casi todas las posibles observaciones



**United
Nations**

Department of Economic and Social Affairs
Statistics

Cómo desarrollar un modelo de datos



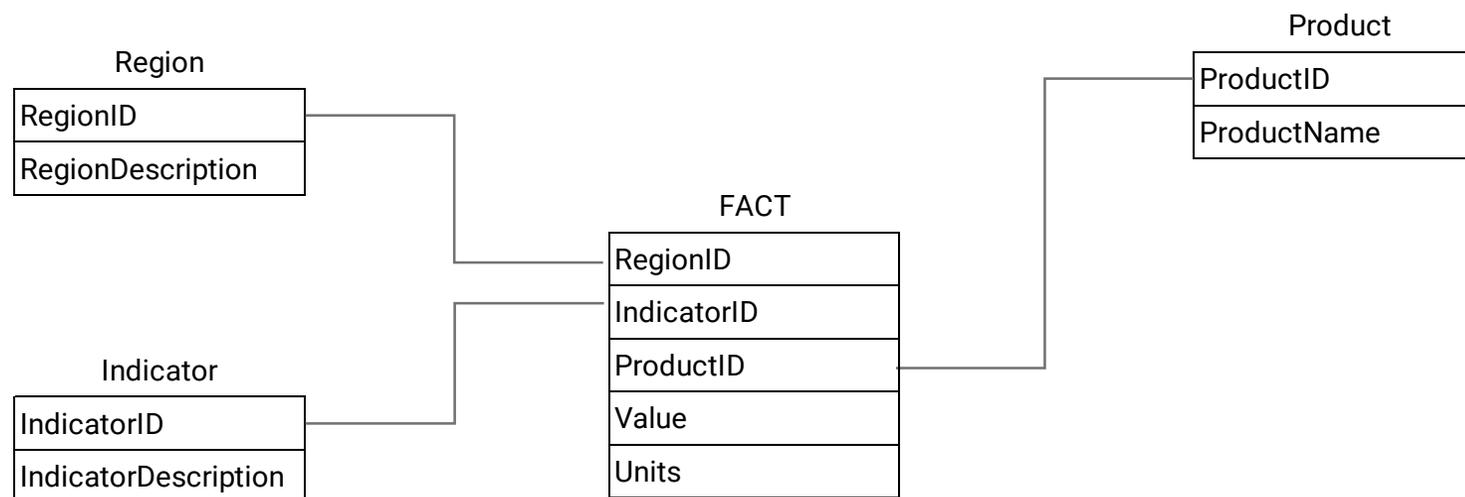
¿Números o datos?

1.1 Proportion of population below \$1 (PPP) per day													
Series	1990	1992	1994	1996	1998	1999	2000	2002	2006	2007	2008	2009	2011
Rwanda													
MDG Population below \$1 (PPP) per day, percentage Last updated: 02 Jul 2012							74.6 ^{1,3}		72.1 ^{1,3}				63.2 ^{1,3}
State of Palestine													
MDG Population below \$1 (PPP) per day, percentage Last updated: 02 Jul 2012										0.4 ^{1,2,3}		0.0 ^{1,2,3}	
Thailand													
MDG Population below \$1 (PPP) per day, percentage Last updated: 02 Jul 2012	11.6 ^{1,3}	8.6 ^{1,3}	4.1 ^{1,3}	2.5 ^{1,3}	2.1 ^{1,3}	3.2 ^{1,3}	3.0 ^{1,3}	1.6 ^{1,3}	1.0 ^{1,3}		0.4 ^{1,3}	0.4 ^{1,3}	

- Un conjunto de números no tiene significado por sí mismo
- Para ser útil, tiene que ser descrito de manera adecuada
- Al añadir descriptores, permitimos que usuarios sepan qué representan estos números. Los números pasan entonces a ser datos.

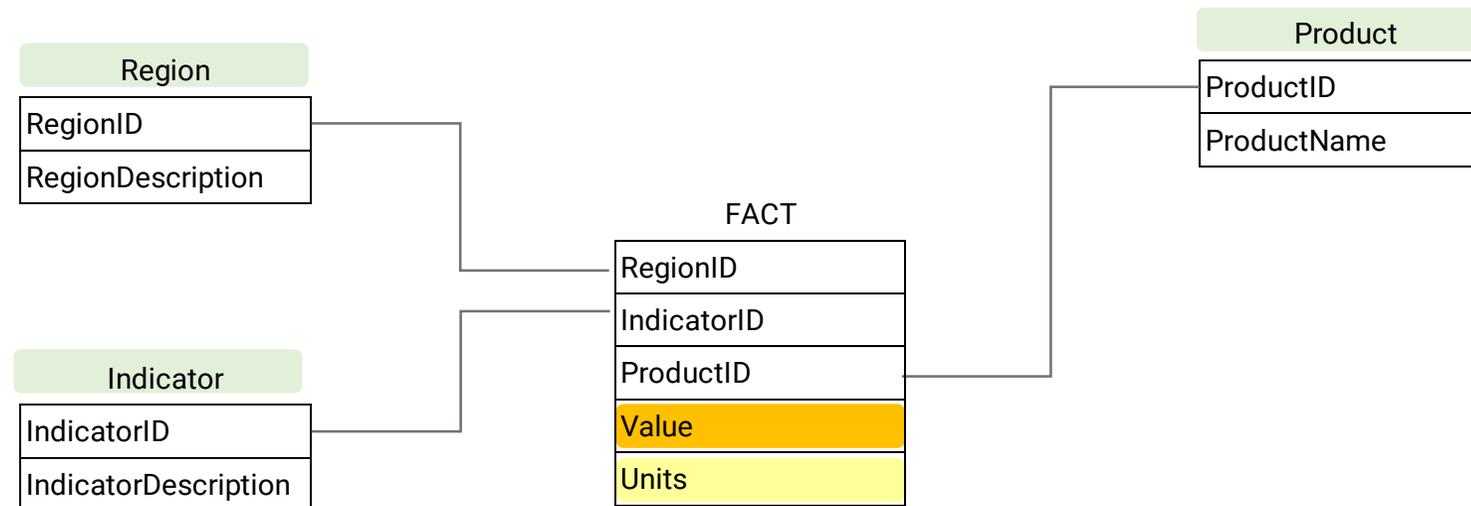
Desarrollar un modelo de datos para fines de diseminación e transferencia

- La tarea es similar al desarrollo de una base de datos relacional
- En SDMX, un modelo de datos se representa a través de una “**Definición de Estructura de Datos**” (DSD, por sus siglas en inglés)
- La forma de una DSD en SDMX se parece a un “esquema de estrella”



Conceptos

- Para diseñar una DSD, primero tenemos que identificar los **conceptos** que describen todas las características relevantes de nuestro conjunto de datos



Identificación de conceptos

	1-1 Total mid-year population - Population totale au milieu de l'année										
Country - Pays	1980	1985	1990	1995	1999	2000	2001	Thousands - milliers		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	1057	1117	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	116510	118305	119962	122033		

Dimensiones

- ¿Qué conceptos se utilizan para identificar una observación (es decir, para distinguir una observación de otra)?
 - Indicador
 - Área de referencia
 - Periodo
- Cuando conocemos estos 3 conceptos, podemos localizar sin ambigüedad la observación correspondiente en el cuadro.
- En SDMX, estos conceptos se conocen como **dimensiones**.
 - Una dimensión es similar a lo que en bases de datos se llama “llave primaria”

Medición primaria

- El valor observado es un concepto que describe las mediciones del fenómeno en cuestión.
- En SDMX, este concepto se conoce como **medición primaria**.
- La medición primaria se representa usualmente como **OBS_VALUE**.

Atributos

- En nuestro ejemplo, **el multiplicador de las unidades** representa información adicional acerca de las observaciones.
- Este concepto no se necesita para identificar una observación específica (ni para distinguir una observación de otra)
- A tales conceptos se les conoce en SDMX bajo el nombre de **atributos**.

¿Dimensión o atributo?

- Los conceptos que son necesarios para **identificar datos individuales**, son considerados **dimensiones**
- Los conceptos que proporcionan **información adicional**, son considerados **atributos**

Dimensiones especiales

- El **TIEMPO** es una dimensión que especifica el momento o el intervalo de tiempo al que una observación hace referencia.
- El **AREA DE REFERENCIA** es una dimensión que describe la ubicación a el area geográfica a la que una observación hace referencia (p.ej., país, región, ciudad, ...)

→ **Todo sucede en un momento (o periodo) específico del tiempo**

→ **Todo sucede en algún lugar**

Ejemplo

Target number	Title of target	Indicator number	Title of Indicator	Unit of measurement	Disaggregation	2015	2016	2017	2018	Responsible ministry, agency
Goal 3. Ensure healthy lives and promote well-being for all at all ages										
3.1	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births	3.1.2	Proportion of births attended by skilled health personnel	Percent	Hemmesi	99.9	100.0	99.9	99.9	Ministry of Health and Medical Industry
					Including regions:					
					Ashgabat	99.8	100.0	99.9	99.9	
					Ahal velayat	99.9	99.9	99.9	99.9	
					Balkan velayat	100.0	100.0	99.9	99.9	
					Dashgouz velayat	100.0	100.0	100.0	100.0	
					Lebal velayat	99.9	99.9	99.9	99.9	
					Mary velayat	100.0	100.0	100.0	100.0	

Identificación de conceptos

Objetivo

Periodo de tiempo

Agencia responsable

Target number	Title of target	Indicator number	Title of Indicator	Unit of measurement	Disaggregation	2015	2016	2017	2018	responsible ministry, agency
Goal 3. Ensure healthy lives and promote well-being for all at all ages										
3.1	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births	3.1.2	Proportion of births attended by skilled health personnel	Percent	Hemmesi	99.9	100.0	99.9	99.9	Ministry of Health and Medical Industry
					Including regions:					
					Ashgabat	99.8	100.0	99.9	99.9	
					Ahal velayat	99.9	99.9	99.9	99.9	
					Balkan velayat	100.0	100.0	99.9	99.9	
					Dashgouz velayat	100.0	100.0	100.0	100.0	
					Lebal velayat	99.9	99.9	99.9	99.9	
					Mary velayat	100.0	100.0	100.0	100.0	

Meta

Indicador

Unidad de medición

Área de referencia

Valor observado

Identificación de conceptos

Target number	Title of target	Indicator number	Title of Indicator	Unit of measurement	Disaggregation	2015	2016	2017	2018	Responsible ministry, agency
Goal 3. Ensure healthy lives and promote well-being for all at all ages										
3.1	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births	3.1.2	Proportion of births attended by skilled health personnel	Percent	Hemmesi	99.9	100.0	99.9	99.9	Ministry of Health and Medical Industry
					Including regions:					
					Ashgabat	99.8	100.0	99.9	99.9	
					Ahal velayat	99.9	99.9	99.9	99.9	
					Balkan velayat	100.0	100.0	99.9	99.9	
					Dashgouz velayat	100.0	100.0	100.0	100.0	
					Lebal velayat	99.9	99.9	99.9	99.9	
					Mary velayat	100.0	100.0	100.0	100.0	

Ejercicio 1: Identificación de conceptos

- Identifique los conceptos que se encuentran en el cuadro de datos
- Marque cada concepto como:
 - Dimensión
 - Medición primaria
 - Atributo
- Identifique la dimensión correspondiente al tiempo (periodo de referencia)
- Identifique la dimensión correspondiente al lugar geográfico (área de referencia)

Representación

- Cuando los datos se transmiten de un sistema a otro, los conceptos que describen cada observación deben tener valores “permitidos”
- Un concepto puede estar
 - Codificado
 - No codificado, pero restringido a valores que siguen un formato especificado
 - No codificado – en formato de texto libre

Representación de conceptos en SDMX

- Las **dimensiones** tienen que estar codificadas o seguir un formato específico.
→ No se permite texto libre para las dimensiones.
- Los **atributos** pueden estar codificados o no; los atributos no codificados pueden o no requerir un formato específico.

Codigos

- Combinaciones de letras, números y otros símbolos **independientes de cualquier idioma**
 - **Representan los diferentes valores que un concepto puede asumir**, cuyo significado puede estar descrito en cualquier numero de idiomas
- La utilización de códigos permite actualizar las descripciones en lenguaje natural correspondientes a los valores que un concepto puede asumir, sin tener que alterar los flujos de datos establecidos con base en la estructura de datos previamente acordada

Listados de códigos

- Se tratan de listados predefinidos (diccionarios) que contienen todos los posibles valores que un concepto codificado puede asumir
- Por ejemplo:
 - Para el concepto “Sexo”: { M, F, _T }
 - Para el concepto “Área geográfica”: Clasificación M49 o ISO-3 de códigos de país
 - Para el concepto “Indicador”: El catálogo de los códigos que representan a cada uno de los indicadores ODS
- Los listados de códigos:
 - Se mantienen independientemente de los datos que hacen uso de ellos
 - Deben armonizarse entre los diferentes actores que participan en el intercambio

Listados de códigos: algunos ejemplos

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 enfance
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

Conceptos no codificados

- Pueden ser texto libre: Pueden utilizar cualquier texto arbitrario
 - P.ej.: Nota al pie, Fuente, ...
- Pueden requerir valores con un formato específico
 - Código postal: 5 dígitos
 - Fecha: aaaa-mm-dd

DSD para los ODS: algunos retos en la codificación

- El uso inconsistente de codificaciones puede generar aparentes registros duplicados y otras anomalías
- Es necesario desarrollar manuales de codificación y asegurar su aplicación (por ejemplo, a través de restricciones del contenido)
- Se utiliza un código único para los casos en que no haya desglose (p.ej., para “Total” y “No aplica”) – esto simplifica la codificación

DSD para los ODS: algunos retos en la codificación

- Debido a que la DSD cubre indicadores de naturaleza muy diversa, no siempre resulta obvio la forma correcta de codificar algunas dimensiones
- Por ejemplo, ¿cuál debería ser el código para la dimensión SEX cuando el indicador es “nacimientos atendidos por personal calificado”?:
 - ¿No aplica? ¿Total? ¿Femenino?

Ejercicio 2: Representación

- Utilizando el modelo anterior, determine la forma de representación de cada concepto
 - Codificada, siguiendo un formato predeterminado, o en texto libre
- Desarrolle los listados de códigos y especifique los formatos que acompañen a los conceptos que Vd. Identifico como “codificados” o “formateados”
 - No existe ninguna restricción en el tipo de códigos o formatos que Vd. puede utilizar. Invente sus propias reglas de codificación/formato; puede basarse en experiencias previas o simplemente en cualquier enfoque pragmático.

Recapitulando

- Conceptos:

- Medición primaria [OBS_VALUE]
- Dimensiones [**SERIES, REF_AREA, TIME_PERIOD**, SEX, AGE, EDU_LEVEL...]
- Atributos [**UNIT_MEASURE, UNIT_MULT, OBS_STATUS, FREQ, NATURE, SOURCE...**]

Desgloses



Recapitulando

REPRESENTACION	DIMENSION	ATRIBUTO
Codificado	✓	✓
Con formato	✓	✓
Texto libre	x	✓



**United
Nations**

Department of Economic and Social Affairs
Statistics

El modelo de información SDMX

Metadatos estructurales

¿Qué?

- Identificadores y descriptores (dimensiones y atributos)

¿Dónde?

- Incorporados en los datos

¿Para qué?

- Para identificar, buscar, explorar, vincular, filtrar, visualizar...

Metadatos de referencia

- Solo descriptores (atributos)
- Pueden almacenarse e intercambiarse independientemente de los datos
- Para entender, interpretar, evaluar calidad, utilizar adecuadamente...

Definición de estructura de datos (DSD)

- Representa un modelo de datos utilizado en el intercambio de datos
- Especifica una estructura común y estandarizada para múltiples conjuntos de datos en un dominio específico
- Una DSD contiene:
 - El listado de **conceptos** que son relevantes para el conjunto de datos
 - La **estructura dimensional** (descripción del rol que juega cada concepto como dimensión, atributo o medición)
 - Los **listados de códigos** utilizados para representar valores específicos que pueden asumir las dimensiones y ciertos atributos en el conjunto de datos
 - **Grupos** que describen colecciones de observaciones en el conjunto de datos (p.ej., series de tiempo o secciones transversales dentro de un panel espacio-temporal)

Series de tiempo

- Un conjunto ordenado de observaciones de la misma variable, tomadas en diferentes puntos del tiempo
- Las observaciones que pertenecen a una misma serie de tiempo solo difieren en su dimensión TEMPORAL:
 - Los valores de todas las demás dimensiones son idénticos.
 - Puede haber atributos a nivel de serie de tiempo que son comunes para toda la serie
 - Los atributos a nivel observación pueden diferir entre los elementos de una serie de tiempo.

Serie de tiempo: ejemplo

1.1 Proportion of population below \$1 (PPP) per day													
Series	1990	1992	1994	1996	1998	1999	2000	2002	2006	2007	2008	2009	2011
Rwanda													
MDG Population below \$1 (PPP) per day, percentage Last updated: 02 Jul 2012							74.6 ^{1,3}		72.1 ^{1,3}				63.2 ^{1,3}
State of Palestine													
MDG Population below \$1 (PPP) per day, percentage Last updated: 02 Jul 2012										0.4 ^{1,2,3}		0.0 ^{1,2,3}	
Thailand													
MDG Population below \$1 (PPP) per day, percentage Last updated: 02 Jul 2012	11.6 ^{1,3}	8.6 ^{1,3}	4.1 ^{1,3}	2.5 ^{1,3}	2.1 ^{1,3}	3.2 ^{1,3}	3.0 ^{1,3}	1.6 ^{1,3}	1.0 ^{1,3}		0.4 ^{1,3}	0.4 ^{1,3}	
1.2 Poverty gap ratio													
Series	1990	1992	1994	1996	1998	1999	2000	2002	2006	2007	2008	2009	2011
Rwanda													
MDG Poverty gap ratio at \$1 a day (PPP), percentage Last updated: 02 Jul 2012							36.9 ^{1,3}		34.8 ^{1,3}				26.6 ^{1,3}
State of Palestine													
MDG Poverty gap ratio at \$1 a day (PPP), percentage Last updated: 02 Jul 2012										0.1 ^{1,2,3}		0.0 ^{1,2,3}	
Thailand													
MDG Poverty gap ratio at \$1 a day (PPP), percentage Last updated: 02 Jul 2012	2.4 ^{1,3}	1.6 ^{1,3}	0.7 ^{1,3}	0.4 ^{1,3}	0.3 ^{1,3}	0.5 ^{1,3}	0.5 ^{1,3}	0.3 ^{1,3}	0.2 ^{1,3}		0.0 ^{1,3}	0.1 ^{1,3}	
Footnotes													
1 Based on nominal per capita consumption averages and distributions estimated from household survey data.													
2 Based on Purchasing Power Parity (PPP) dollars imputed using regression.													
3 Source: http://research.worldbank.org/PovcalNet/index.htm													

Datos en corte transversal

- Un conjunto de observaciones que hacen referencia al mismo periodo, pero toman diferentes valores en una o mas dimensiones no temporales (tales como área geográfica, sexo, edad, tipo de actividad...)
- Por ejemplo, todos los datos en una encuesta o censo usualmente se refieren a un periodo especifico en el tiempo, y las tabulaciones se organizan conforme a dimensiones tales como ubicación geográfica, sexo, grupo de edad, industria, etc.

Vista de series de tiempo vs. vista de corte transversal

2.1 Net enrolment ratio in primary education

	2009	2010	2011
Morocco			
Total net enrolment ratio in primary education, both sexes		94.1	96.2
Total net enrolment ratio in primary education, boys		95	96.8
Total net enrolment ratio in primary education, girls		93.3	95.6
State of Palestine			
Total net enrolment ratio in primary education, both sexes	88.2	89.2	
Total net enrolment ratio in primary education, boys	88.2	89.8	
Total net enrolment ratio in primary education, girls	88.2	88.5	
Uganda			
Total net enrolment ratio in primary education, both sexes	94.2	91	
Total net enrolment ratio in primary education, boys	93.1	89.7	
Total net enrolment ratio in primary education, girls	95.3	92.3	

Las dimensiones “país” y “sexo” se escogieron como dimensiones de corte transversal.

Nótese que la dimensión temporal aún esta presente (pero solo asume un valor fijo)

2.1 Net enrolment ratio in primary education

2010

	Total	Boys	Girls
Morocco	94.1	95	93.3
State of Palestine	89.2	89.8	88.5
Uganda	91	89.7	92.3

Llaves en SDMX

- La **llave de una serie** identifica de manera única a una serie de tiempo
 - Consiste en todas las dimensiones excepto **TIEMPO**
- La **llave de un grupo** identifica de manera única a un grupo
 - Consiste en una o mas dimensiones (las cuales pueden o no incluir la dimensión **TIME**)

Conjunto de datos

- Una colección de datos lógicamente relacionados entre si, y que comparten una estructura común
- Puede entenderse como una colección de una o mas series de tiempo y/o uno o mas cortes transversales
- Un “conjunto de datos” sirve, por ejemplo, como un contenedor para intercambiar múltiples series de tiempo (p.ej., en un mensaje SDMX)

Ejercicio 3:

Codificación de series de tiempo

- Utilizando el cuadro del ejercicio anterior, identifique cada serie de tiempo a través de sus “llaves”
- Para cada serie de tiempo, proporcione los valores que corresponden a su llave.

Mensajes SDMX

- Todo tipo de información relacionada con SDMX se intercambia en documentos denominados “mensajes”
- Un mensaje SDMX puede ser enviado en diferentes formatos (p.ej., XML, JSON, CSV)
- Existen diferentes tipo de mensajes SDMX para diferentes fines:
 - El mensaje de tipo **Structure** se utiliza para transmitir información estructural, tal como una DSD, MSD, esquema conceptual, etc.
 - **GenericData**, **StructureSpecificData**, y otros mensajes se utilizan para transmitir datos.
- Un mensaje SDMX en formato XML se conoce como mensaje SDMX-ML.



La definición de estructura de datos ODS



La definición de estructura de datos (DSD) para los ODS

- Desarrollada por el Grupo de Trabajo sobre SDMX para los indicadores ODS, establecido por el IAEG-SDGs in abril de 2016
- La primera versión fue publicada oficialmente el 14 de junio de 2019

📄 SDG DSD Matrix Version 1.0

📄 Global SDG DSD v1.0

📄 Guidelines for the Global DSD for SDGs

<https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/>

La definición de estructura de datos (DSD) para los ODS

Artefact/object	ID	Agency	Version
SDG DSD	SDG	IAEG-SDGs	1.0
Concept Scheme	SDG_CONCEPTS	IAEG-SDGs	1.0

Concept Name	Role	Type	Code List
FREQ	Dimension	Code	CL_FREQ
REPORTING_TYPE	Dimension	Code	CL_REPORTING_TYPE
SERIES	Dimension	Code	CL_SERIES
REF_AREA	Dimension	Code	CL_AREA
SEX	Dimension	Code	CL_SEX
AGE	Dimension	Code	CL_AGE
URBANISATION	Dimension	Code	CL_URBANISATION
INCOME_WEALTH_QUANTILE	Dimension	Code	CL_QUANTILE
EDUCATION_LEV	Dimension	Code	CL_EDUCATION_LEV
OCCUPATION	Dimension	Code	CL_OCCUPATION
CUST_BREAKDOWN	Dimension	Code	CL_CUST_BREAKDOWN
COMPOSITE_BREAKDOWN	Dimension	Code	CL_COMP_BREAKDOWN
DISABILITY_STATUS	Dimension	Code	CL_DISABILITY
ACTIVITY	Dimension	Code	CL_ACTIVITY
PRODUCT	Dimension	Code	CL_PRODUCT
TIME_PERIOD	Dimension	Date/Time	
OBS_VALUE	Measure	Double	
UNIT_MULT	Attribute (mandatory, obs)	Code	CL_UNIT_MULT
UNIT_MEASURE	Attribute (mandatory, ts)	Code	CL_UNIT_MEASURE
OBS_STATUS	Attribute (mandatory, obs)	Code	CL_OBS_STATUS
NATURE	Attribute (optional, obs)	Code	CL_NATURE
TIME_DETAIL	Attribute (optional, obs)	String	
COMMENT_OBS	Attribute (optional, obs)	String	
COMMENT_TS	Attribute (optional, ts)	String	
TIME_COVERAGE	Attribute (optional, obs)	Date/time	
UPPER_BOUND	Attribute (optional, obs)	Double	
LOWER_BOUND	Attribute (optional, obs)	Double	
BASE_PER	Attribute (optional, obs)	Date/time	
SOURCE_DETAIL	Attribute (optional, obs)	String	
GEO_INFO_URL	Attribute (optional, ts)	String	
GEO_INFO_TYPE	Attribute (optional, ts)	Code	CL_GEO_INFO_TYPE
DATA_LAST_UPDATE	Attribute (optional, ts)	Date/time	

- 16 Dimensiones
- 15 Atributos
- 20 listados de códigos

La definición de estructura de datos (DSD) para los ODS

- Una sola DSD se utiliza para todos los indicadores ODS
- Como cubre indicadores de naturaleza muy diversa, no todas sus dimensiones son aplicables para todos los casos. Por ejemplo:
 - Edad (AGE) no es aplicable para el indicador “Área del terreno cubierto por bosque”
 - El valor **_T** (desglose no aplicable o no disponible) se utiliza en estos casos.

Dimensión: frecuencia (FREQ)

- Indica el nivel de recurrencia con el que ocurren las observaciones en el tiempo (p.ej., mensual, anual, bianual, etc.)
- La DSD para ODS actualmente solo permite datos con frecuencia anual
- En casos en que la frecuencia no sea anual, (p.ej., promedio para periodos de dos años) es necesario indicar los detalles en el atributo TIME_DETAIL.

Dimensión: tipo de reporte (REPORTING_TYPE)

- Se utiliza para distinguir entre datos reportados a nivel nacional, regional o global
- Los países utilizan el valor **N** (reporte nacional)
- Las organizaciones regionales utilizan el valor **R** (reporte regional)
- Las agencias internacionales utilizan el valor **G** (reporte global)

Dimensión: serie (SERIES)

- Se utiliza para representar “sub-indicadores”
 - Un mismo indicador ODS puede tener múltiples series
 - No confundir esta dimensión con las series de tiempo SDMX (cada “serie” puede tener una o más “series de tiempo”, es decir, múltiples desgloses con observaciones a lo largo del tiempo)
- Ejemplo: El indicador 5.5.1, “Proporción de curules ocupados por mujeres en (a) parlamentos nacionales y (b) gobiernos locales” tiene actualmente 4 series en la base de datos global para indicadores ODS:
 - SG_GEN_PARL Proporción de curules ocupados por mujeres en parlamentos nacionales
 - SG_GEN_PARLN Número de curules ocupados por mujeres en parlamentos nacionales
 - SG_GEN_PARLNT Proporción de curules en parlamentos nacionales
 - SG_GEN_LOCG Proporción de curules ocupados por mujeres en gobiernos locales

Dimensión: área de referencia geográfica (REF_AREA)

- País o área geográfica con la que se relaciona el fenómeno estadístico medido
- Se tiene contemplado que los países reporten valores a nivel nacional, pero estos pueden también extender el listado de códigos para incluir desglose geográfico sub-nacional

Dimensión: sexo (SEX)

- Condición de genero masculino o femenino. Esta dimensión se aplica solo si los datos pueden desglosarse por sexo.
- El código **_T** se utiliza cuando el desglose por sexo no es aplicable o no esta disponible
- En el caso de algunos indicadores de genero se fija el valor **F**
 - Por ejemplo, para la serie *Proporción de curules ocupados por mujeres en parlamentos nacionales* el valor de la dimensión sexo es siempre “F”

Dimensión: edad (AGE)

- Edad - or rango de edades – de los individuos en la población a la que se refiere una observación
- El código **_T** se utiliza cuando el desglose por edad no es aplicable
- Algunos indicadores corresponden por definición a un grupo de edades específico (por ejemplo, mortalidad infantil entre niños menores de 5 años de edad)

Dimensión: áreas urbanas / rurales (URBANISATION)

- Esta dimensión tiene 3 códigos:
 - _T (Total)
 - _U (Áreas urbanas)
 - _R (Áreas rurales)
- El código _T se utiliza cuando el desglose por urbanización no es aplicable

Dimensión: cuantil de ingreso o riqueza (INCOME_WEALTH_QUANTILE)

- Se utiliza actualmente para desgloses por quintil de ingreso o riqueza
- En el futuro podrá extenderse para cubrir deciles, percentiles, etc.
- El código **_T** se utiliza cuando el desglose por quintil de riqueza o ingreso no es aplicable

Dimensión: Nivel de educación (EDUCATION_LEV)

- Nivel educativo más alto que una persona ha completado satisfactoriamente
- La lista de códigos incluye las categorías de la clasificación internacional para estadísticas de la educación (ISCED11 y ISCED97), así como códigos para categorías especiales relevantes para ODS
- El código _T se utiliza cuando el desglose por nivel de educación no es aplicable

Dimensión: ocupación (OCCUPATION)

- Tipo de trabajo o empleo de un individuo que realiza una serie de tareas o deberes laborales
- El listado de códigos incluye las clasificaciones de la OIT (ISCO-08, ISCO-98, ISCO-68)
- El código _T se utiliza cuando el desglose por ocupación no es aplicable

Dimensión: Estatus de discapacidad (DISABILITY STATUS)

- Se utiliza para desglosar los indicadores ODS por tipo de discapacidad
- Actualmente solo se utiliza para distinguir entre personas con alguna discapacidad y personas sin ninguna discapacidad
- El código _T se utiliza cuando el desglose por estatus de discapacidad no es aplicable

Dimensión: actividad económica (ACTIVITY)

- Desglose por tipo de actividad económica con base en el tipo de bienes y servicios producidos
- El listado de códigos incluye las categorías CIU.
- El código _T se utiliza cuando el desglose por tipo de actividad económica no es aplicable

Dimensión: tipo de producto (PRODUCT)

- Código de producto
- Combina valores específicos para los ODS tomados de varias clasificaciones, incluyendo: CPC, flujos materiales, y otros productos no estandar
- El código _T se utiliza cuando el desglose por tipo de producto no es aplicable

Dimensión: Desglose customizado (CUST_BREAKDOWN)

- Se trata de una dimensión especial que facilita desgloses no estándar, principalmente en el contexto de datos nacionales
- Actualmente el listado de códigos se encuentra vacío, pero en el futuro será utilizado con códigos genéricos (e.g. CODE1, CODE2, etc), cuyo significado será asignado por los proveedores de datos nacionales
- Debe ser utilizado en combinación con el atributo CUST_BREAKDOWN_LB, el cual transmite la descripción del código customizado.
- El código **_T** se utiliza cuando el desglose customizado no es aplicable

Dimensión: desglose compuesto (COMPOSITE_BREAKDOWN)

- Se trata de una dimensión mixta que representa varios listados de códigos que han sido consolidados en uno solo
 - Por ejemplo, desglose por organización internacional, tipo de riesgo de desastre, etc.
- Se utiliza para desgloses que solo son aplicables para un número muy reducido de indicadores, de manera que se evite el uso de demasiadas dimensiones
- El código **_T** se utiliza cuando el desglose compuesto no es aplicable

Dimensión temporal: TIME_PERIOD

- Periodo de tiempo al que corresponde cada observación
- La convención para los ODS es que siempre se incluya el concepto TIME_PERIOD como el data del año de referencia, expresado en formato de 4 dígitos
- Cualquier información adicional se debe incluir en el atributo TIME_DETAIL
- Información estructurada sobre el periodo de tiempo se puede incluir en el atributo TIME_COVERAGE.

Medición primaria: valor observado (OBS_VALUE)

- Se utiliza para capturar el valor de la variable numerica de interes

Atributo: Estatus de la observación (OBS_STATUS)

- Se utiliza para capturar información sobre la calidad de un valor o para indicar un valor inusual o faltante
 - Por ejemplo, se puede utilizar para indicar la existencia de un quiebre en la serie de tiempo
- Es un atributo de uso obligatorio

Atributo: multiplicador de la unidad de medición (UNIT_MULT)

- Es el exponente base 10 tal que al multiplicar el valor de la observación por $10^{\text{UNIT_MULT}}$ se obtiene como resultado un valor expresado en la unidad de medición simple
- Si el valor observado corresponde a “millones”, el multiplicador de la unidad de medición es 6; si esta en trillones, es 9, etc. En los casos en que las observaciones se encuentran expresadas en unidades simples, el valor del multiplicador es 0.
- Es un atributo de uso obligatorio

Atributo: Unidad de medición (UNIT_MEASURE)

- Unidad en la que se expresan los valores de los datos
- En algunos casos no resulta obvio cual es la unidad de medida correcta.
- Es un atributo de uso obligatorio, codificado

Atributo: detalles del periodo de tiempo (TIME_DETAIL)

- Este atributo se utiliza para proporcionar detalles sobre el intervalo de tiempo que cubre cada observación (p.ej., para el periodo '2001-2003', la dimensión TIME_PERIOD sería 2002, pero el atributo TIME_DETAIL sería 2001-2003).
- Es un atributo opcional en formato de texto libre

Atributo: cobertura temporal (TIME_COVERAGE)

- Se utiliza para incluir la representación ISO8601 del intervalo de tiempo al cual se refiere cada observación
- Mientras que la dimensión TIME_PERIOD siempre se debe expresar como un año en cuatro dígitos, y el atributo TIME_DETAIL se expresa en formato de texto libre, el atributo TIME_COVERAGE puede ser utilizado para proporcionar el intervalo exacto en un formato estructurado estándar
- Se trata de un atributo de uso opcional, que sigue un formato predefinido

Atributo: periodo base (BASE_PER)

- Periodo que se utiliza como base para el caso que el valor observado sea un numero índice, o como referencia para el caso que se trate de un valor monetario en precios constantes (p.ej., cuando el valor observado se expresa en dólares constantes a precios de 2005)
- Cuando se utiliza este atributo, se espera que su valor sea siempre un año específico
- Es un atributo opcional de observaciones individuales.

Atributo: naturaleza del dato (NATURE)

- Proporciona información sobre la naturaleza del proceso que genera los valores observados
- Permite distinguir si un dato ha sido producido y diseminado como dato oficial de un país, o si ha sido estimado o ajustado por alguna otra entidad (p.ej., agencias internacionales).
- Usualmente se asigna el código C (“country data”) para los valores reportados a nivel nacional
- Se trata de un atributo opcional para las observaciones individuales

Atributo: Detalles de la fuente de datos (SOURCE_DETAIL)

- Proporciona información adicional sobre la fuente de datos, por ejemplo, si se trata de datos generados a partir de una encuesta en particular.
- Es un atributo opcional en formato de texto libre

Atributos: cotas superior e inferior (UPPER_BOUND y LOWER_BOUND)

- En los casos que el valor observado represente un estimado puntual, estos dos atributos pueden utilizarse para indicar las cotas superior e inferior de dicho estimado
 - En el caso de los ODM, era necesario crear diferentes códigos de series
- Se trata de un atributo opcional

Atributos: Comentarios/Notas (COMMENT_OBS y COMMENT_TS)

- Información adicional sobre observaciones individuales o sobre series de tiempo individuales (por ejemplo, notas aclaratorias explicando como se ha calculado o estimado una observación, u otros detalles a considerar para evaluar la calidad del dato o su comparabilidad con otros datos)
- El atributo COMMENT_OBS se utiliza para comentarios relacionados con observaciones individuales; el atributo COMMENT_TS se utiliza para comentarios sobre series de tiempo completa.
- Ambos son atributos opcionales

Atributo: referencia geoespacial (GEO_INFO_URL)

- Proporciona la dirección web de un archivo geo-informático.
- Se utiliza en combinación con GEO_INFO_TYPE.
- Es un atributo opcional a nivel de series de tiempo

Atributo: tipo de referencia geoespacial (GEO_INFO_TYPE)

- Especifica el tipo de archivo o servicio geoespacial referenciado en el atributo GEO_INFO_URL.
- Es un atributo opcional a nivel de series de tiempo



**United
Nations**

DESA
Statistics Division

Gracias.

<http://unstats.un.org/sdgs/>

Paso de Excel a SDMX