

Echange des donnees et metadonnees statistiques (SDMX)

Convertisseur SDMX

Dany Ghafari, SDG and Environment Statistics Unit, UNEP

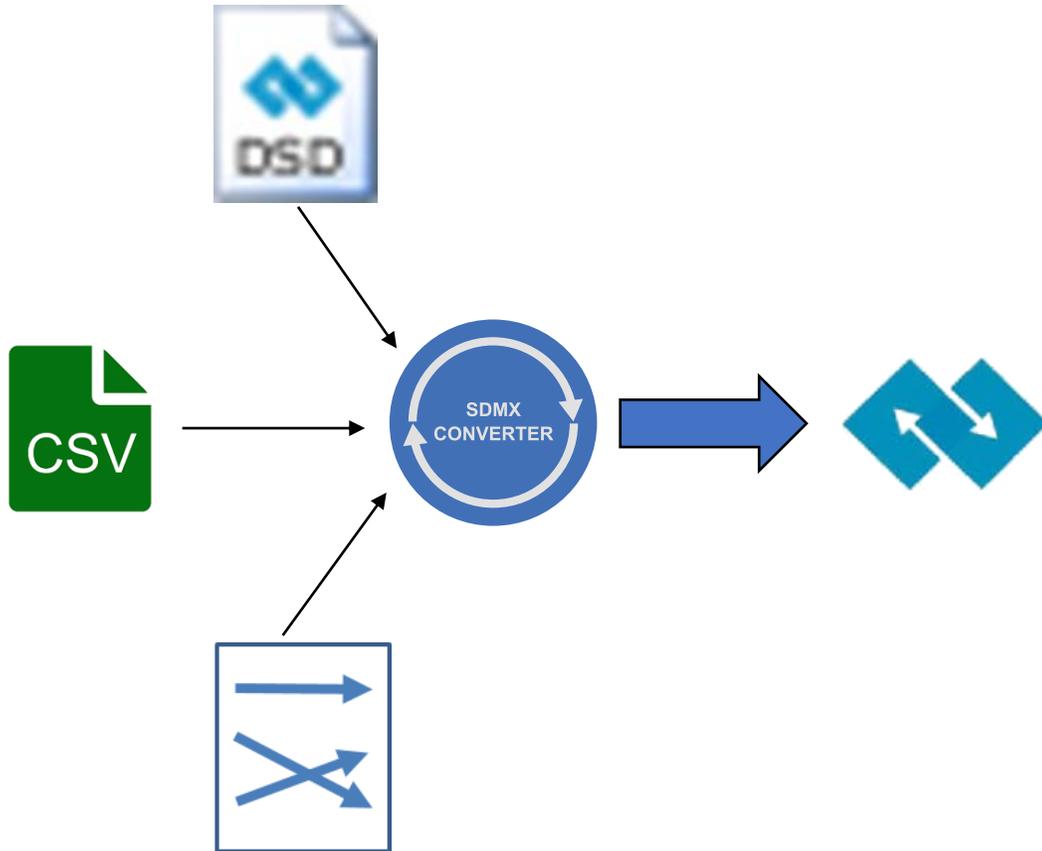
Convertisseur SDMX

- Logiciel développé par Eurostat
- Peut être utilisé pour convertir des données à partir d'une variété de formats en SDMX et vice versa
- Prend en charge CSV, Excel, DSPL et autres
- Peut être utilisé pour transformer....
 - Données non SDMX vers SDMX
 - SDMX vers non SDMX
 - Format SDMX vers un autre format SDMX

Convertisseur SDMX : Applications

- Le convertisseur SDMX est disponible en tant que:
 - Application de bureau avec une interface utilisateur graphique
 - Application en ligne de commande
 - Service Web
 - Bibliothèque Java
 - Application Web
- Télécharger à partir de https://circabc.europa.eu/ui/group/088149e5-0472-405b-839b-57d5970052cc/library/76a63fc2-3d22-42cc-85f2-4f6f30b4d8eb?p=1&n=10&sort=modified_DESC

Conversion de données en SDMX



- Pour transformer des données en SDMX à l'aide de SDMX Convertisseur, vous avez besoin
 - Les données sources au format CSV, DSPL, Excel, etc.
 - Une définition de structure de données (DSD) selon laquelle le jeu de données SDMX sera structuré
 - mappages montrant comment les données source sont mappées aux concepts de la définition de la structure de données
 - Comme toujours lors de la configuration de l'échange SDMX, la configuration des mappages prend plus de temps et d'effort

Convertisseur SDMX

- Les données et les mappages peuvent être placés dans la même feuille de calcul
 - Les mappages peuvent également être fournis à partir d'un fichier externe
- Des informations supplémentaires peuvent être ajoutées pour faciliter la saisie des données.
 - Par exemple. listes de codes pour la validation et l'affichage des descriptions

Noms des feuilles de travail

- La feuille de calcul nommée **Paramètres** contient des mappages
 - Montre comment les cellules, les lignes et les colonnes correspondent aux dimensions et aux attributs de la DSD
- Les noms de feuille de calcul commençant par **VAL** sont ignorés
 - Peut être utilisé pour stocker des listes de codes ou d'autres informations auxiliaires
- Les noms des feuilles de calcul de transcodage doivent commencer par **Trans**
- Si une seule feuille de paramètres est utilisée comme ci-dessus, toutes les autres feuilles de calcul sont considérées comme des données et seront traitées
- Plusieurs feuilles de paramètres peuvent être utilisées si les feuilles de travail ont des mappages différents.
 - Dans ce cas, la feuille de calcul nommée **Parameter_mapping** doit indiquer quelles feuilles de calcul de paramètres s'appliquent à quelles feuilles de calcul de données
 - Les noms de feuille de calcul de paramètres doivent commencer par **Parameter**

Exercice 4 : Utilisation du convertisseur SDMX

- Utilisez SDMX Converter pour récupérer des données à partir d'une feuille de calcul pré-mappée et les convertir en SDMX

Mappages Excel: la feuille de calcul Paramètres

Concept

Mapping type

Cell where data starts

	A	B	C	D	E	F	G	H	I
1	Element	Type	PosType	Position				DataStart	H2
2	FREQ	DIM	FIX	A				NumColumns	1
3	REPORTING_TYPE	DIM	FIX	N				TranscodingSheet	Transcoding
4	SERIES	DIM	COLUMN	B					
5	REF_AREA	DIM	COLUMN	E					
6	TIME_PERIOD	DIM	COLUMN	G					
7	SEX	DIM	FIX	_T					
8	AGE	DIM	MIXED	COLUMN	P	FIX		_T	
9	URBANISATION	DIM	FIX	_T					
10	INCOME_WEALTH_QUANTILE	DIM	FIX	_T					
11	EDUCATION_LEV	DIM	FIX	_T					
12	OCCUPATION	DIM	FIX	T					

Position or value

Number of obs. columns

Concept Role

Feuille de calcul des mappages Excel

- **Elément**: nom du concept DSD
- **Type** : rôle du concept
 - **DIM** : cote
 - **ATT** : attribut
- **PosType** : Mappage ou type de position
- **Position** : position ou valeur
- **DataStart** : la première cellule contenant une observation
- **NumColumns** : nombre d'observations par ligne

Colonne PosType : mappage ou type de position

- Les types de mappage/position suivants sont pris en charge:
 - **CELL**
 - **ROW**
 - **COLUMN**
 - **FIX**
 - **OBS_LEVEL**
 - **MIXED**
 - **SKIP**

Type de mappage : CELLULE

- La valeur pour l'ensemble de la feuille de calcul est spécifiée dans la cellule fournie dans la colonne **Position**
- Par exemple. si la feuille de calcul ne doit contenir que des données pour un seul indicateur, son code peut être fourni dans une cellule.

	C	D	E
	15.1.1 Forest area as a proportion of total land area		
1		SERIES	AG_LND_FRST I
2	<i>(in percents)</i>		
3	Items	REF_AREA	2011
4	Kyrgyz Republic	KG	5.6
5	Batken oblast	KG05	9.8
6	Djalal-Abad oblast	KG03	4.1
7	Ysyk-Kul oblast	KG02	3.2
8	Naryn oblast	KG04	3.1
9	Osh oblast	KG06	6.4
10	Talas oblast	KG07	4.5
11	Chui oblast	KG08	2.2
12	Approved by the Decree of the Government of the Kyrgyz Republic dated July 26, 2011 No. 407		

Type de mappage : LIGNE

- Les valeurs du concept sont stockées dans la ligne spécifiée dans la colonne **Position**

	C	D	E	F	G	H	I
1	1.1.1 Proportion of population living below the international poverty line (1.9 USD)						
2	<i>(as a per cent to number of population)</i>						
3		SERIES:	SI_POV_DAY	AGE:	_T	UNIT	PT
4		URBANISATION:	_T	EDU.LEV	T	UNIT	M. 0
5	Items	REF_AREA	2007	2008	2009	2010	2011
6	Kyrgyz Republic	KG	0.13	0.09	0.29	0.28	0.01
7	Batken oblast	KG05	0.00	0.27	0.00	0.21	0.00
8	Djalal-Abad oblast	KG03	0.20	0.00	0.00	0.22	0.00
9	Ysyk-Kul oblast	KG02	0.05	0.34	0.58	0.11	0.00
10	Naryn oblast	KG04	0.13	0.63	0.03	3.46	0.31
11	Osh oblast	KG06	0.33	0.03	0.00	0.09	0.00

Type de mappage : COLONNE

- Les valeurs du concept sont stockées dans la colonne spécifiée dans la colonne **Position**

	C	D	E	F	G	H	I
1	1.1.1 Proportion of population living below the international poverty line (1.9 USD)						
2	<i>(as a per cent to number of population)</i>						
3		SERIES: SI_POV_DAY	AGE: _T			UNIT PT	
4		URBANISATION: _T	EDU.LEV _T			UNIT M.	0
5	Items	REF_AREA	2007	2008	2009	2010	2011
6	Kyrgyz Republic	KG	0.13	0.09	0.29	0.28	0.01
7	Batken oblast	KG05	0.00	0.27	0.00	0.21	0.00
8	Djalal-Abad oblast	KG03	0.20	0.00	0.00	0.22	0.00
9	Ysyk-Kul oblast	KG02	0.05	0.34	0.58	0.11	0.00
10	Naryn oblast	KG04	0.13	0.63	0.03	3.46	0.31
11	Osh oblast	KG06	0.33	0.03	0.00	0.09	0.00

Type de mappage : COLONNE (2)

- Également utilisé avec la représentation basée sur les enregistrements, lorsque chaque ligne contient un enregistrement ou une observation, et chaque colonne contient des valeurs pour un concept.

	A	B	D	E	F	G	H	I	J	K	L	M	N
1	M49 Code	Series Code	Indicator R	Country	Disaggre	Year	IMR	Age Group	Unit	Nature	Footnote	Source Det	Time Detail
2	4	SH_DYN_IM	3.2.1	Afghanistan	BOTHSEX	2000	90.8	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
3	8	SH_DYN_IM	3.2.1	Albania	BOTHSEX	2000	23.1	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
4	12	SH_DYN_IM	3.2.1	Algeria	BOTHSEX	2000	33.9	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
5	20	SH_DYN_IM	3.2.1	Andorra	BOTHSEX	2000	4	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
6	24	SH_DYN_IM	3.2.1	Angola	BOTHSEX	2000	122.9	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
7	28	SH_DYN_IM	3.2.1	Antigua and	BOTHSEX	2000	13.1	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
8	32	SH_DYN_IM	3.2.1	Argentina	BOTHSEX	2000	17.3	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
9	51	SH_DYN_IM	3.2.1	Armenia	BOTHSEX	2000	26.6	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
10	36	SH_DYN_IM	3.2.1	Australia	BOTHSEX	2000	5.1	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
11	40	SH_DYN_IM	3.2.1	Austria	BOTHSEX	2000	4.6	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
12	2	SH_DYN_IM	3.2.1	Azerbaijan	BOTHSEX	2000	50.8	<1Y	PER_1000_LIVE	NA		Source: Uni	2000

Type de mappage : FIX

- La valeur fixe pour l'ensemble du jeu de données est stockée dans la colonne **Position** et n'apparaît pas dans la feuille de calcul de données
 - Par exemple. si l'on s'attend à ce que les données soient toujours annuelles, la fréquence peut être codée pour l'ensemble de la feuille de calcul

	A	B	C	D
1	Element	Type	PosType	Position
2	FREQ	DIM	FIX	A

Type de mappage : MIXTE

- La valeur du concept est conditionnelle
- Peut être utilisé pour fournir une valeur par défaut

	A	B	C	D	E	F	G
1	Element	Type	PosType	Position		DataStart	K12
2	FREQ	DIM	FIX	A		NumColumns	1
3	REPORTING_TYPE	DIM	CELL	B4			
4	SERIES	DIM	COLUMN		2		
5	REF_AREA	DIM	MIXED	CELL	B3	FIX	TH
6	TIME_PERIOD	DIM	COLUMN		4		

- « Utilisez la cellule B3 pour le concept REF_AREA. Si la cellule est vide, utilisez la valeur fixe TH »

Type de mappage : OBS_LEVEL

- Peut être utilisé dans pour spécifier les attributs attachés au niveau de l'observation par rapport à la cellule contenant l'observation.
 - Par exemple, lorsque chaque ligne contient plusieurs observations et leurs attributs.

	A	B	C	D	E	F	G
37	OBS_EDP_WBB	ATT	SKIP				
38	OBS_STATUS	ATT	MIXED	OBS_LEVEL 1 CELL			H14
39	REF_PERIOD_DETAIL	ATT	CELL	B4			
40	REF_YEAR_PRICE	ATT	CELL	H6			

- « Pour l'attribut OBS_STATUS, utilisez la cellule située à 1 colonne à droite de la cellule contenant la valeur d'observation. Si cette cellule est vide, utilisez la valeur de la cellule H14. »

Type de mappage : SAUTER

- La valeur du concept n'est pas mappée
- Ne peut être utilisé qu'avec des attributs facultatifs

18	NATURE	ATT	FIX	C				
19	SOURCE_DETAIL	ATT	SKIP					
20	COMMENT_OBS	ATT	SKIP					

Utilisation des DSD et des flux de données

- SDMX Converter peut utiliser un DSD ou un flux de données pour valider ou convertir des ensembles de données
- L'utilisation de flux de données est généralement plus stricte car les contraintes de contenu sont généralement attachées aux flux de données
 - Par exemple, un ensemble de données SDG national structurellement valide réussira la validation par rapport au SDG DSD et au flux de données mondial du pays DF_SDG_GLC, mais échouera à la validation par rapport au flux de données mondial harmonisé DF_SDG_GLH
 - En outre, un ensemble de données SDG national structurellement valide échouera à la validation par rapport au flux de données du pays DF_SDG_GLC s'il contient des relations non valides entre les dimensions, par ex. Sexe féminin avec indicateur « Superficie terrestre couverte par la forêt ».

Utilisation des registres SDMX

- SDMX Converter peut charger des structures à partir d'un fichier ou d'un registre
- Si la structure est chargée à partir d'un fichier, elle doit être complète avec toutes les références requises.
 - Au moins, le fichier doit contenir un DSD, son Concept Scheme et toutes les listes de codes de référence
- Si la structure est chargée à partir d'un registre, toutes les structures référencées seront utilisées automatiquement
- Les adresses des registres couramment utilisés sont fournies par défaut. L'adresse de tout service Web conforme au registre peut être saisie manuellement.

Exercice 5 : Conversion de données à l'aide de flux de données et de contraintes de contenu

- Comparez la validation par rapport au DSD ODD et la validation par rapport aux flux de données ODD.

Exercice 6 : Mapper un fichier Excel

- Mappez une feuille de calcul sur le SDG DSD et utilisez SDMX Converter pour récupérer et convertir les données

Thank you!

Dany Ghafari, SDG and Environment Statistics Unit, UNEP

United Nations Avenue, Gigiri
PO Box 30552 – 00100 GPO Nairobi, Kenya

www.unep.org