

## ○ Objectif

- Se familiariser avec les étapes de nettoyage et de préparation de données tabulaires au format Excel en vue de leur publication sur le portail Open Data.

## ○ Sommaire

- 10. Publication de données tabulaires (Excel)
- ... et exercices pratiques

# 10. Publication de données tabulaires: Utilisation d'Excel

- A partir d'un jeu de données « brut » au format Excel
  - Non directement exploitable par une machine
  - Au format large
  - Avec éventuellement des données à caractère personnel
- Publier le jeu de données sur le portail Open Data (de pré-production)
  - Données directement exploitables par une machine
  - Au format long
  - Valeurs standardisées
  - Données validées
  - Sans données à caractère personnel
  - Accompagnées de métadonnées

**=> En utilisant Excel**

- Population urbaine, données de la Banque Mondiale

<https://data.worldbank.org/indicator/SP.URB.TOTL?view=chart>

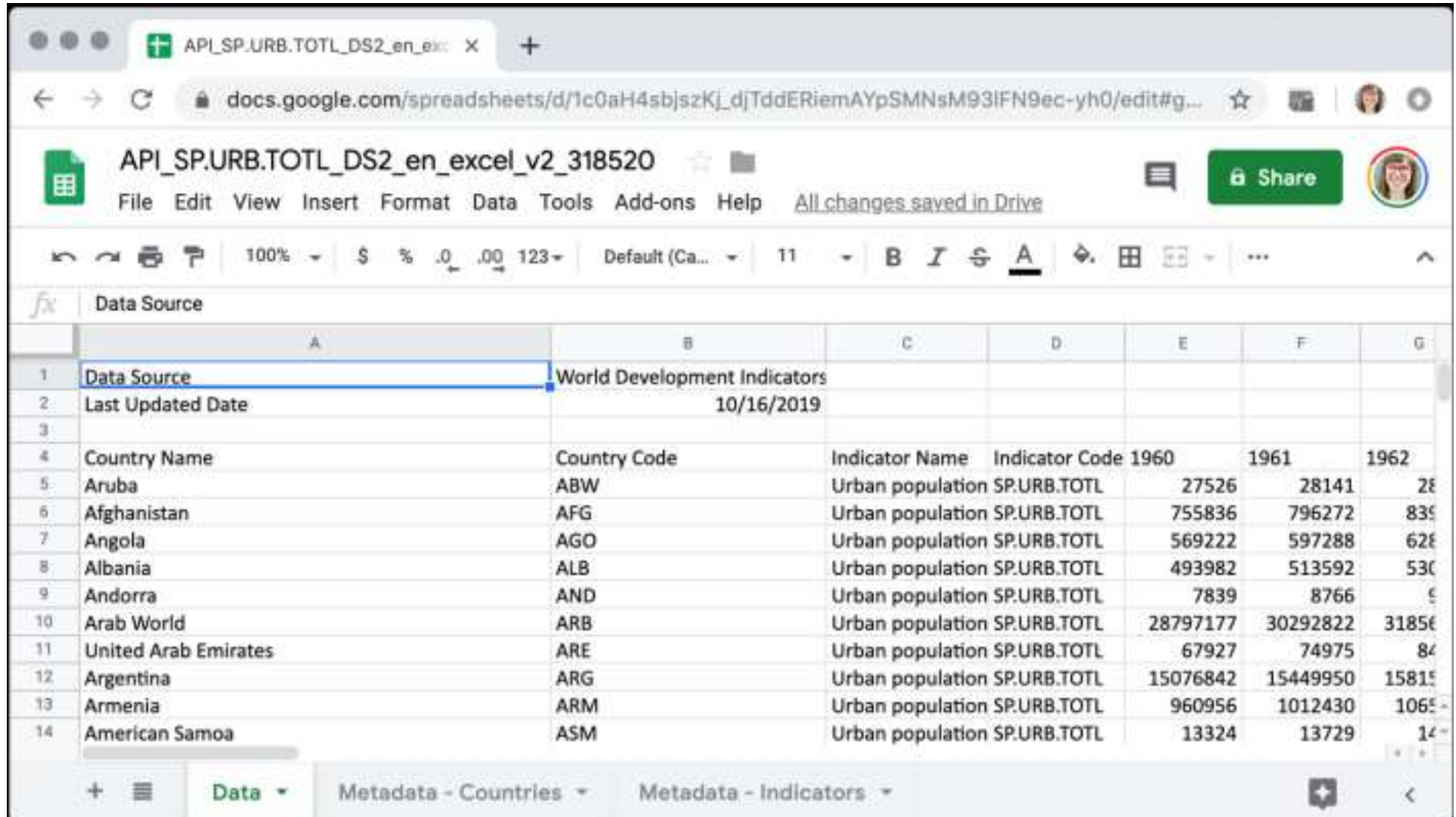
<http://api.worldbank.org/v2/en/indicator/SP.URB.TOTL?downloadformat=excel>

- Mise en place
  - Examiner les données
  - Renommer le fichier
  - Garder une copie des données initiales
- Préparation des en-têtes
  - Supprimer toutes les « métadonnées » mélangées aux données
  - Séparer les cellules fusionnées et les en-têtes sur plusieurs lignes
  - Expliciter les unités dans les en-têtes, valeurs numériques propres
  - Contrôler la qualité des en-têtes
  - Figurer la première ligne et première colonne
  - Supprimer les colonnes et lignes inutiles

- Transformer les données
  - Supprimer les séparateurs de milliers
  - Standardiser le texte
  - Normaliser les dates
  - Rendre les valeurs atomiques
  - Ajouter des colonnes si nécessaire
  - Mettre les données au format long
- Valider les données
  - Ajouter des contraintes de validation
- Rédaction des métadonnées
  - Préparer les métadonnées descriptives
  - Détailler les métadonnées structurelles

- Rédaction des métadonnées
  - Préparer les métadonnées descriptives
  - Détailler les métadonnées structurelles
- Publication des données
  - Création du jeu de données sur le portail Open Data

# Examiner les données



API\_SP.URB.TOTL\_DS2\_en\_excel\_v2\_318520

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

100% \$ % .0 .00 123 Default (Ca... 11 B I S A

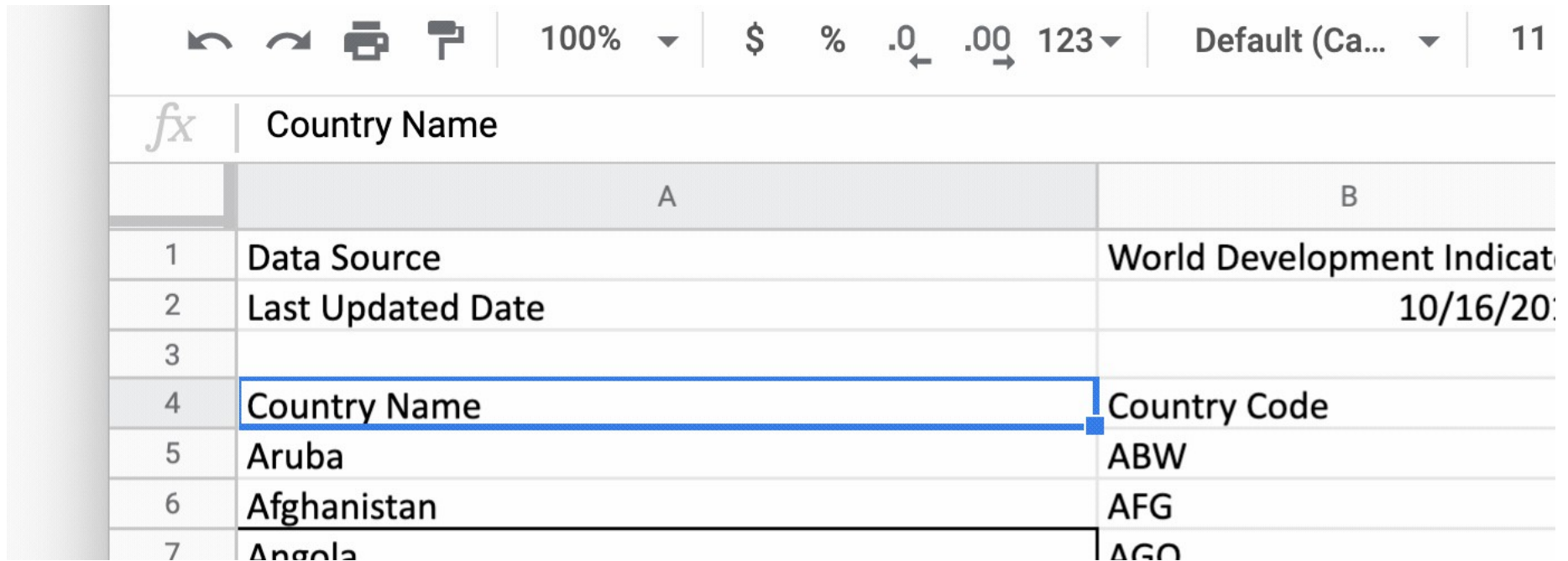
Data Source

	A	B	C	D	E	F	G
1	Data Source	World Development Indicators					
2	Last Updated Date	10/16/2019					
3							
4	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962
5	Aruba	ABW	Urban population	SP.URB.TOTL	27526	28141	28141
6	Afghanistan	AFG	Urban population	SP.URB.TOTL	755836	796272	836272
7	Angola	AGO	Urban population	SP.URB.TOTL	569222	597288	628288
8	Albania	ALB	Urban population	SP.URB.TOTL	493982	513592	530592
9	Andorra	AND	Urban population	SP.URB.TOTL	7839	8766	9566
10	Arab World	ARB	Urban population	SP.URB.TOTL	28797177	30292822	31856822
11	United Arab Emirates	ARE	Urban population	SP.URB.TOTL	67927	74975	84975
12	Argentina	ARG	Urban population	SP.URB.TOTL	15076842	15449950	15815950
13	Armenia	ARM	Urban population	SP.URB.TOTL	960956	1012430	1065430
14	American Samoa	ASM	Urban population	SP.URB.TOTL	13324	13729	14729

+ Data Metadata - Countries Metadata - Indicators



# Préparer les en-têtes



100% | \$ % .0 .00 123 | Default (Ca... | 11

	A	B
1	Data Source	World Development Indicat
2	Last Updated Date	10/16/20
3		
4	Country Name	Country Code
5	Aruba	ABW
6	Afghanistan	AFG
7	Angola	AGO

Avant

Afghanistan		Angola	
Population	Urban population	Population	Urban population

Après

Afghanistan Population	Afgahnistan Urban population	Angola Population	Angola Urban population
---------------------------	---------------------------------	----------------------	----------------------------

Afghanistan	1960	755836
	1961	796272
	1962	839385
	1963	885228
	1964	934135
	1965	986074

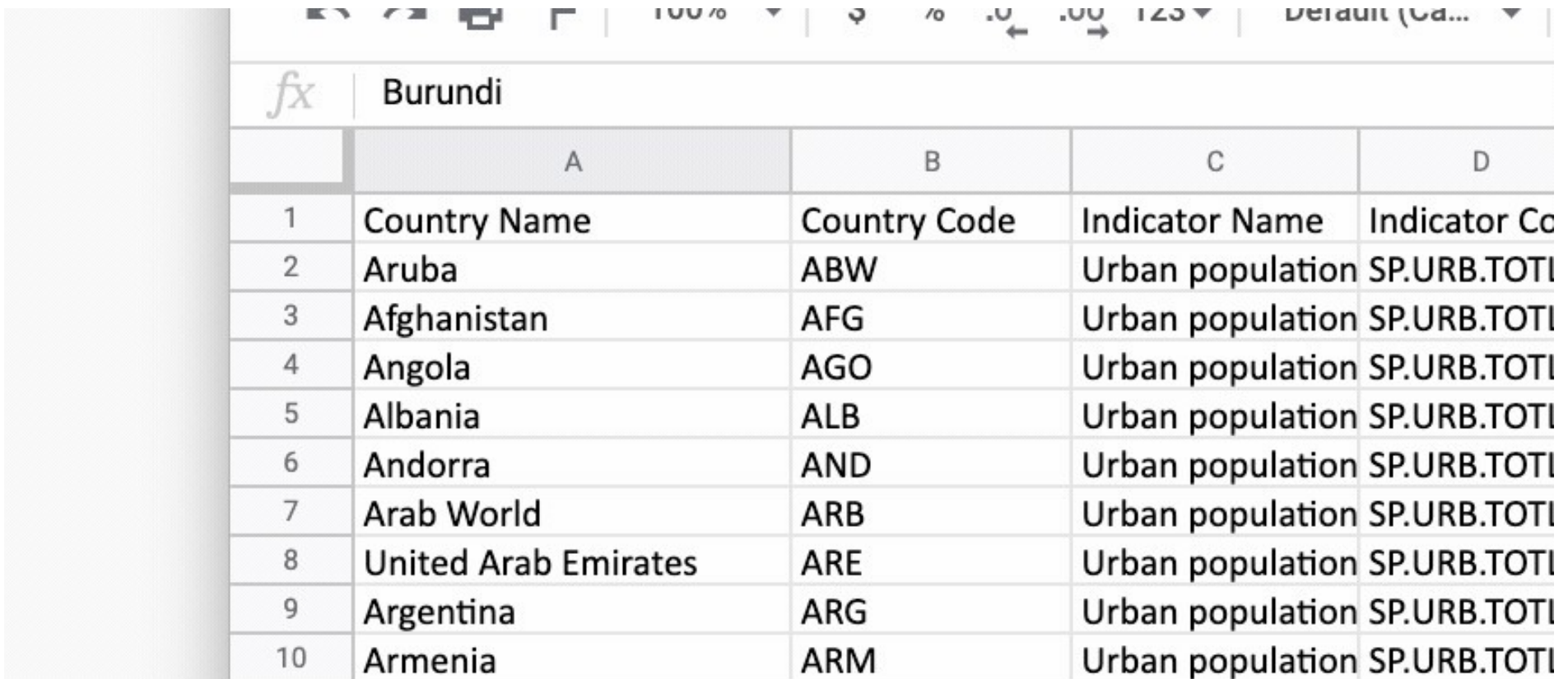
## Avant

	fake GDP	fake share
Afghanistan	293 Euro	46.48%
Angola	383 Euro	29.40%
Albania	919 Euro	19.01%
Andorra	294 Euro	31.07%

## Après

	fake GDP in Euro	fake share in %
Afghanistan	293	46.48
Angola	383	29.40
Albania	919	19.01
Andorra	294	31.07

# Figurer les volets d'en-tête



The screenshot shows an Excel spreadsheet with a table. The table has four columns labeled A, B, C, and D. The first row is a header row with the following content: Country Name, Country Code, Indicator Name, and Indicator Co. The subsequent rows contain data for various countries, including Aruba, Afghanistan, Angola, Albania, Andorra, Arab World, United Arab Emirates, Argentina, and Armenia. The indicator name for all these countries is 'Urban population' and the indicator code is 'SP.URB.TOTI'.

	A	B	C	D
1	Country Name	Country Code	Indicator Name	Indicator Co
2	Aruba	ABW	Urban population	SP.URB.TOTI
3	Afghanistan	AFG	Urban population	SP.URB.TOTI
4	Angola	AGO	Urban population	SP.URB.TOTI
5	Albania	ALB	Urban population	SP.URB.TOTI
6	Andorra	AND	Urban population	SP.URB.TOTI
7	Arab World	ARB	Urban population	SP.URB.TOTI
8	United Arab Emirates	ARE	Urban population	SP.URB.TOTI
9	Argentina	ARG	Urban population	SP.URB.TOTI
10	Armenia	ARM	Urban population	SP.URB.TOTI

## Avant

755,836	796 272
569,222	597 288
493,982	513 592
7,839	8 766
28,797,177	30 292 822
67,927	74 975

## Après

755836	796272
569222	597288
493982	513592
7839	8766
28797177	30292822
67927	74975

# Jour 2 - FIN