

# Données budgétaires des municipalités locales du Cameroun

Ce document est un guide pratique à l'insertion des données budgétaires locales dans OpenSpending.

Il est basé sur les différents types de problèmes rencontrés lors la conversion des données durant la mise en place du site [cameroon.openspending.org](http://cameroon.openspending.org).

Ce guide ne présente pas une liste exhaustive des instructions à suivre pour formater les données pour OpenSpending, mais récapitule les problèmes fréquemment rencontrés avec les données.

## Pourquoi avons-nous besoin de données correctement formatées ?

La base de données OpenSpending nécessite des données structurées et lisibles par un ordinateur.

### Qu'appelle-t-on des données lisibles par un ordinateur ?

Des données lisibles par un ordinateur sont des données qu'un programme peut facilement trier, filtrer et analyser.

Les formats les plus courants de fichiers lisibles par des machines sont Excel, CSV (Comma Separated Values) et XML. OpenSpending fonctionne avec des fichiers CSV, qui peuvent facilement être exportées à partir d'Excel ou d'autres tableurs. Regardez dans les options d'exportation du programme que vous utilisez.

Dans Excel : Fichier -> Enregistrer sous -> Format : «CSV (séparateur : point-virgule) (.csv)»;

Dans Google Docs : Fichier> Télécharger comme ... > CSV.

## Pourquoi ne peut-on pas utiliser un tableau dans un document Word ou un fichier PDF avec des «données structurées» ?

Les programmes comme Word sont conçus pour mettre en page des documents de manière adaptée et visuellement agréable à l'oeil humain. Néanmoins, ces fichiers ne sont pas lisibles par une machine de la même manière : les tableaux à l'intérieur ne peuvent pas être analysés, triés, filtrés ou exportés facilement dans une autre application comme OpenSpending.

Il est bien souvent possible de convertir des données non structurées en données structurées, par exemple, en copiant/collant un tableau issu d'un document Word dans Excel.

Certaines erreurs peuvent être introduites dans le processus, il sera alors nécessaire d'effectuer du nettoyage dans les données, processus qui sera détaillé dans ce document.

Dans d'autres cas, en particulier avec les fichiers PDF, un copier/coller peut échouer complètement, une application spécifique telle que Acrobat Professional, ABBYY FineReader ou PDF2HTML sera alors nécessaire.

## Structurer les données pour OpenSpending

En plus de s'assurer que le format de fichier de données soit lisible par une machine, les données doivent également être structurées de manière spécifique afin d'être chargés dans OpenSpending.

OpenSpending est très flexible en ce qui concerne la structure, par exemple, l'ordre des colonnes n'a pas d'importance, mais les contenus des colonnes doivent être normalisés.

En se basant sur les données des municipalités locales du Cameroun, nous avons élaboré un modèle simple afin qu'il serve de modèle aux futures conversions de données. Ce fichier contient des exemples de données du conseil de Tignère et un modèle de données.

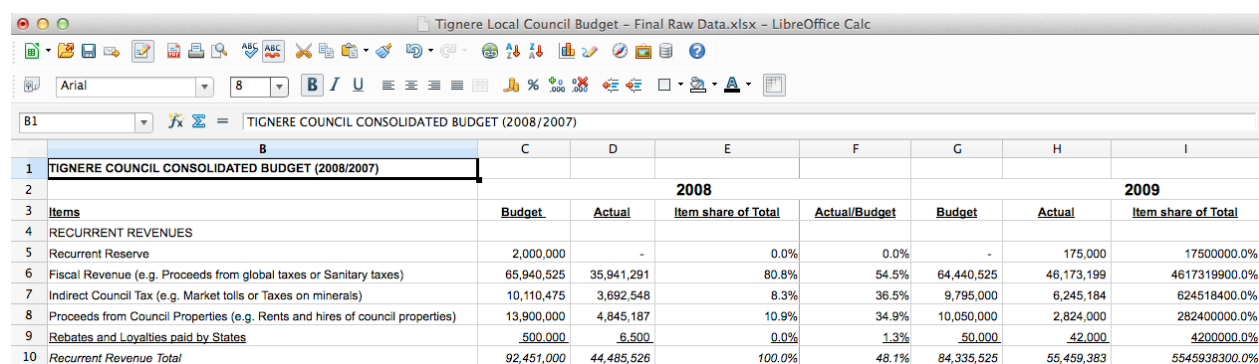
Voici quelques lignes directrices qui vous aideront à mettre en forme des données pour une utilisation dans OpenSpending.

### 1. OpenSpending nécessite qu'il y ait une ligne d'en-tête dans le fichier. C'est ce que le logiciel va utiliser pour identifier les noms de vos colonnes.

Par exemple, ce document présentant les données du budget Tignère comportait initialement 3 rangées d'en-tête, avant le nettoyage des données.

Or il est nécessaire que la ligne d'en-tête soit la première ligne dans le fichier car toutes les autres sont traitées comme des lignes de données.

*Le fichier original :*



TIGNERE COUNCIL CONSOLIDATED BUDGET (2008/2007)							
B	C	D	E	F	G	H	I
1	TIGNERE COUNCIL CONSOLIDATED BUDGET (2008/2007)						
2			2008				2009
3	Items	Budget	Actual	Item share of Total	Actual/Budget	Budget	Actual
4	RECURRENT REVENUES						
5	Recurrent Reserve	2,000,000	-	0.0%	0.0%	-	175,000
6	Fiscal Revenue (e.g. Proceeds from global taxes or Sanitary taxes)	65,940,525	35,941,291	80.8%	54.5%	64,440,525	46,173,199
7	Indirect Council Tax (e.g. Market tolls or Taxes on minerals)	10,110,475	3,692,548	8.3%	36.5%	9,795,000	6,245,184
8	Proceeds from Council Properties (e.g. Rents and hires of council properties)	13,900,000	4,845,187	10.9%	34.9%	10,050,000	2,824,000
9	Rebates and Loyalties paid by States	500,000	6,500	0.0%	1.3%	50,000	42,000
10	Recurrent Revenue Total	92,451,000	44,485,526	100.0%	48.1%	84,335,525	55,459,383

*Le fichier retravaillé :*

Cameroon Local Council Budget Data Template ☆

File Edit View Insert Format Data Tools Help All changes saved

Lucy Chambers ▾

Share

No other viewers

	A	B	C	D	E	F	G	H	I	J
	Head-account	Head-Account Description	Sub-account	Sub-account Description	Year	Reporting Type	Amount	Revenue/Expenditure	Recurrent/Investment	Council Name
1	110.xxx	Uncategorized	110.101	Reserves for overheads	2008	Budget	2000000	REVENUE	RECURRENT	TIGNERE COUNCIL
2	110.xxx	Uncategorized	110.101	Reserves for overheads	2008	Actual	0	REVENUE	RECURRENT	TIGNERE COUNCIL
3	110.xxx	Uncategorized	110.101	Reserves for overheads	2009	Budget	0	REVENUE	RECURRENT	TIGNERE COUNCIL

Vous remarquerez plusieurs changements :

- La première rangée de données d'origine qui était le titre du document a été supprimée. Vous aurez la possibilité de donner un titre à votre ensemble de données lors de la phase de chargement dans OpenSpending.
- Les lignes ne doivent contenir qu'un seul type d'information, soit une ligne par budget (ou «fait»). Dans le fichier original, on peut voir que chaque ligne contient les données budgétaires prévisionnelles et les dépenses réelles, et ce pour plusieurs années. Dans la version corrigée, chaque ligne ne contient les données que d'une seule année, et les différents intitulés ont été reportés dans les colonnes. Par exemple, il est maintenant possible de filtrer la colonne H «recettes / dépenses» pour afficher uniquement les postes qui de recettes. Il en est de même pour d'autres critères tels que le type de déclaration.

*Notez que le formatage de données pour OpenSpending implique souvent la création de nombreuses lignes en plus par rapport au document original.*

- Les colonnes de pourcentage n'ont pas été reportées car OpenSpending génère ce calcul automatiquement.

A noter : les données peuvent se présenter aussi bien en français qu'en anglais.

## 2. Les colonnes d'un fichier doivent être homogènes. Cela signifie :

- Aucune ligne ou cellule ne doit être vide. Les données importées dans OpenSpending peuvent être parfois dé-normalisées : il peut y avoir des références aux feuilles de codes externes / données de base mais chaque ligne doit contenir toutes les informations nécessaires pour construire l'objet qui en résulte. Les colonnes (en particulier les classifications) doivent comporter une valeur pour chaque ligne, elles ne seront pas automatiquement remplies.
- Ne pas inclure de données pré-agrégées (comme par exemple des sous-totaux) dans les données. OpenSpending s'occupe de les calculer automatiquement.

*Exemple :*

## BUDGET COMMUNAL COUNCIL BUDGET

Première partie / Part one

Recettes / Revenue

TITRE <i>Tire</i>	CHAPITRE ARTICLE PARAGRAPHE (CLASSES)  <i>Head, Subhead paragraph</i>	NATURE DE LA RECETTE  <i>Nature of revenue</i>	RECETTES EXECUTES ET PORTEES AU DERNIER COMPTE ADMINISTRATIVE  <i>Executed expenditure recorded in the previous administrative account</i>	PREVISIONS DE L'EXERCICE PRECEDENT  <i>Revenue estimates on the previous financial year</i>	REVISIONS DE RECETTE DE L'EXERCICE VOTEES PAR LE CONSEIL MUNICIPAL  <i>Revenue estimates voted by the municipal council</i>	PREVISION DE RECETTE RETENUES PAR L'AUTORITE DE TUTELLE  <i>Revenue estimates approved by the supervisory authority</i>	OBSERVATION  <i>Observations</i>
	720 117	Produits de concession dans les cimetières <i>Cemeteries fees</i>					
	720 119	Produits des prestations de services fournies <i>Proceeds of services rendered</i>					
	720 120	Recettes des services <i>Revenue from services</i>	67 000	5 000 000	13 000 000		
	720 121	Autre produits d'exploitation du domaine <i>Other revenue form the exploitation</i>					
		TOTAL DU CHAPITRES 7.2 <i>TOTAL DU CHAPTER</i>	67 000	5 000 000	16 000 000		
	73	RISTOURNES SUR PRODUITS ET REDEVANCES CONSENTIES PAR L'ETAT <i>Rebates and royalties granted by the state</i>					
	730						

Commune d'Arrondissement de Yaoundé II

11/37

Comme vous pouvez le voir dans l'exemple ci-dessus :

- De nombreuses cellules sont vides. Elle doivent être soit remplies (par des données ou par «0»), soit supprimées.
- Il y a un total pré-agrégé (en gris), cette ligne doit être supprimée.

Nous verrons comment transformer les données et éviter quelques erreurs courantes dans les tutoriaux à venir.

### 3. Il doit y avoir une colonne (ou une combinaison de plusieurs colonnes) qui constitue un «identifiant unique»

OpenSpending a été construit dans le but de pouvoir mettre à jour des entrées dans la base de données à tout moment. Cela signifie qu'il est nécessaire de pouvoir attribuer un identifiant unique à chaque ligne de données, identifiant qu'OpenSpending va pouvoir utiliser pour déterminer si une ligne existante doit être mise à jour ou si il est nécessaire d'en créer une nouvelle.

La meilleure façon de le faire est simplement d'ajouter une colonne factice au jeu de donnée dans lequel vous mettez un nombre qui augmente pour chaque ligne.

Vous pouvez le faire dans Excel en :

- Tapant les chiffres dans les deux premières lignes,
- Sélectionnant et glissant les deux cellules vers le coin inférieur droit de la cellule pour prolonger la série.

### Autres point à vérifier lors de la préparation des données pour OpenSpending

- Les montants doivent être présentés en format purement numérique.

Ne pas utiliser de virgules. Le point (".") doit uniquement être utilisé comme séparateur décimal.

- Les montants doivent être les montants réels :

Montant
1657000

Plutôt que

Montant (en milliers)
1657

Les transformations que nous avons effectuées sur les données sont relativement simples et les jeux de données fournis pour ce projet jusqu'à présent sont relativement petits. Cela signifie qu'il est parfaitement possible de transformer les données en les modifiant manuellement ou en utilisant seulement les fonctions basiques des feuilles de calcul.

Cependant, il existe des outils permettant de transformer rapidement les données et contribuer ainsi à réduire les erreurs introduites dans les données, telles que les fautes de frappe.

Cette liste n'est pas exhaustive, mais vise les erreurs les plus fréquentes et les modifications nécessaires. Les étapes préliminaires sur la façon de corriger d'autres erreurs courantes dans les données ont été présentées dans [cette page web sur OpenSpending](#).

## Récapitulatif des point à vérifier avant l'export

- Il n'y a qu'une seule colonne d'en-tête
  - Chaque ligne ne contient qu'un seul type d'information ( Each row contains only one logical type of information (par exemple pas de données concernant des années différentes sur la même ligne)
  - Il n'y a pas de ligne ou de cellule vide
  - Il n'y a pas de total pré-calculé
  - Il y a une colonne / combinaison de colonnes qui constitue un identifiant unique
  - Exécutez une vérification orthographique. OpenSpending regroupe entre eux les objets similaires, donc par exemple écrire «Ministère des Affaires sociales» et «Ministere des Affaires sociales» - reviendrait à écrire deux entités distinctes.
- Il est également important de vérifier que la casse soit harmonisée.
- Dans les colonnes comportant des données numériques, il n'y a pas de séparateur de chiffres tels que les virgules
  - Les données dans les colonnes numériques sont les données réelles (plutôt que des données en milliers par exemple.)
  - Les dates sont au format JJ-MM-AAAA ou MM-AAAA ou AAAA.
  - Le fichier ne contient pas d'informations personnelles tels que des numéros de téléphone ou adresses.

**Vous êtes maintenant prêt à exporter votre fichier au format CSV et à le soumettre.**

**A NOTER : N'oubliez pas d'enregistrer votre fichier avec encodage UTF-8 afin de préserver les caractères spéciaux et les signes diacritiques.**

*Comment exporter votre fichier en tant que CSV*

Dans Google Docs : Fichier> Publier sur le Web> [Sélectionner les feuilles à publier]> Republier maintenant> Sous «Obtenir un lien vers les données publiées », sélectionnez' CSV 'dans le menu déroulant puis sélectionnez vos feuilles.

Dans Excel: Fichier> Enregistrer sous> CSV (séparateur : point-virgule) (Dans certaines versions : «CSV délimité par des virgules»)

Dans Numbers : Fichier> Exporter> CSV

Dans Libre Office (anciennement OpenOffice ou StarOffice) : Fichier> Enregistrer sous .. > Dans le champ "Type de fichier" menu : "Texte CSV (.csv)."

## **Récapitulatif des point à vérifier après l'export**

Après avoir exporté ouvrir votre fichier à vérifier :

- Les trémas et caractères spéciaux s'affichent pas correctement dans votre nouveau fichier
- Pour les données venant des fichiers Excel, assurez-vous que toutes les macros dans le document ont été correctement transformées en valeurs statiques par la fonction exportateur CSV. Si votre feuille de calcul Excel dispose d'une protection d'exportation, vous devrez peut-être trouver un outil pour la contourner.

## **Mettre le fichier en ligne**

Vous devez maintenant mettre en ligne vos données afin que OpenSpending sache où les récupérer. Il ya plusieurs façons de le faire, par exemple dans un dossier public dans Dropbox, ou les exporter directement à partir de Google Docs (Fichier> Publier sur le Web> Republier maintenant> [Dans «Obtenir un lien vers des données publiées 'article] Sélectionnez CSV et la feuille correcte de vos données> Copier le lien de l'encadré).

Nous avons utilisé DataHub, un répertoire en ligne pour les données, qui est une solution rapide et facile. Voici une vidéo ci-dessous expliquant comment procéder :

<https://vimeo.com/43720463>

L'idéal est que vous mettiez en ligne à la fois le fichier d'origine (avant les modifications) et la version modifiée.

Cela permet de les comparer facilement en cas de problèmes ou de questions concernant la provenance des données. Vous pouvez ajouter les deux jeux de données en tant que

ressources dans le même dossier sur DataHub, de sorte à indiquer qu'ils se réfèrent à la même chose