

# OPEN DATA

## Les données ouvertes

Loire  
Atlantique

liberTIC

OPENDATA  
LOCALE

### Ateliers d'accompagnement à l'ouverture des données

Une action collective OpenData Locale

6/11 Sensibilisation à l'ouverture des données et à la réglementation

13/11 Atelier normalisation : théories et bonnes pratiques

20/11 Atelier normalisation : finalisation, améliorations et mises à jour

27/11 Atelier création de modèle

4/12 Atelier réutilisation des données

Nantes. Novembre 2017



# OPEN DATA

## Les données ouvertes

Loire  
Atlantique

liberTIC

OPENDATA  
LOCALE

- 1- L'ouverture du fichier de données
- 2- La qualification
- 3- Les corrections
- 4- Les ajouts de variables
- 5- L'enregistrement du CSV
- 6- Le nommage et les métadonnées

## ATELIER 3

### **Développement de compétences**

Savoir ouvrir et assurer la cohérence des données

Maîtriser la navigation dans les tables de données

Savoir standardiser les tables de données

Savoir filtrer et corriger les erreurs

Savoir enregistrer un fichier au format CSV

### **Création d'un CSV standardisé**

Loire  
Atlantique

liberTIC

OPENDATA  
LOCALE

# 1. L'ouverture du fichier

# Le fichier exercice

- Nom : PLF 2014 - Jaune - Associations subventionnées 2012.csv
- Source :  
<https://www.data.gouv.fr/fr/datasets/plf-jaune-associations-subventionnees/>
- Licence : Licence Ouverte / Open Licence
- Date de la dernière mise-à-jour : mardi 7 janvier 2014
- Version de l'exercice zippée à cette adresse :  
<http://www.2i2l.fr/spip.php?article472>

# L'ouverture du fichier

- Dézipper le fichier pour obtenir le CSV
- Nous utiliserons pour de multiples raisons le tableur Calc de la suite LibreOffice, surtout ici pour ses atouts dans les ouvertures et enregistrements de fichiers CSV
- Fichier → ouvrir → le fichier exercice
- Découvrir la fenêtre d'import de texte

# L'ouverture du fichier

Jeu de caractères  
Voir les problèmes  
d'accentuations

Séparateurs  
Voir les effets  
dans les données

Séparateurs  
de texte

Changer le format  
d'une colonne  
par un clic droit

Contrôler les décalages  
de colonnes

Si tout est ok,  
Valider !

Import de texte - [plf-2014-jaune-associations-subventionne-es-2012.csv]

**Importer**

Jeu de caractères : Europe occidentale (Windows-1252/WinLatin 1)

Langue : Par défaut - Français (France)

À partir de la ligne : 1

**Options de séparateur**

Largeur fixe  Séparé par

Tabulation  Virgule  Point-virgule  Espace  Autre

Fusionner les séparateurs

Séparateur de texte : "

**Autres options**

Champ entre guillemets comme texte  Détecer les nombres spéciaux

**Champs**

Type de colonne :

	Standard	Standard
1	Ministère	PROGRAMME
2	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e
3	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e
4	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e
5	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e
6	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e
7	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e
8	Ministère 01 - Affaires étrangères et européennes	105 - Action de la France e

Aide OK Annuler

# Le vocabulaire

Vocabulaires	Entêtes de colonne	Enregistrements en ligne
En tableur	Champs	données
En statistique	Variables	modalités
En logiciel	Paramètres	valeurs



# La découverte des données

- Il s'agit de faire un point sur tous les noms de champs.
- sélectionner la première ligne en cliquant sur le numéro 1 de la ligne
- copier la première ligne : Ctrl+c
- insérer une nouvelle feuille « variables », et se placer dans la cellule A1
- édition → collage spécial → transposer → ok

# La fiche explicative

- Nous pouvons voir les détails des variables et modalités dans le fichier qui accompagne la table de données :  
<https://www.data.gouv.fr/fr/datasets/plf-jaune-associations-subventionnees/>

# La fiche explicative

Nom du champ	Correspondance
Millésime	PLF et document budgétaire de rattachement
Type de Budget	Rattachement à un type de budget suivant la nomenclature : général / annexe / comptes péciaux
Ministère en 2011	Ministère de rattachement de la subvention pour l'année du versement
Programme en 2011	Programme de rattachement de la subvention
Association	Nom de l'association
SIREN	N° SIREN de l'association
Subvention 2011 en euros	Montant de la subvention
Département	Département de domiciliation de l'association
Ville	Ville de domiciliation de l'association
Objet de la subvention	Libellé de la subvention
Evaluation	champ libre indiquant si cette subvention fait l'objet d'une évaluation
Mission en 2011	Mission de rattachement pour l'année du versement

# Les raccourcis clavier

- CTRL + Origine, pour revenir en A1
- CTRL + Fin, pour voir le nombre de lignes et colonnes, découvrir la fin de la table de données
- CTRL + Flèche vers le bas, pour découvrir le peuplement des colonnes et vérifier qu'il n'exsiste pas trop de cellules vides, dans chacune des colonnes

## 2. La qualification

# La qualification des données

- Il s'agit de faire un inventaire détaillés des modalités saisies pour toutes les colonnes.
- Nous ne pouvons commencer un traitement et une analyse sans un inventaire critique des modalités en présence.
- Pour cela, nous utiliserons les filtres, voire des tris à plat (ou distributions de modalités).

# Le décret sur les subventions

- Il existe le décret 2017-779 du 5 mai 2017 relatif à l'accès sous forme électronique aux données essentielles des conventions de subvention.
- Le décret est accessible à cette adresse :  
<https://www.legifrance.gouv.fr/eli/decret/2017/5/5/PRMJ1636989D/jo/texte/fr>

# La comparaison des variables

## Les variables contenues dans le fichier en ligne

- MILLESIME
- LIB
- Ministère
- PROGRAMME
- ASSOCIATION
- SIREN
- DEPARTEMENT
- VILLE
- Subvention 2012
- OBJET
- EVALUATION
- Convention
- Éléments de la convention
- Mission

## Les variables proposées par l'association OpenData-France

- COLL\_NOM
- COLL\_SIRET
- COLL\_TYPE
- ACTE\_DATE
- ACTE\_REF
- ATTRIB\_NOM
- ATTRIB\_SIRET
- ATTRIB\_TYPE
- SUB\_OBJET
- SUB\_MONTANT
- SUB\_NATURE
- SUB\_DATE\_DEBUT
- SUB\_DATE\_FIN
- SUB\_VERSEMENT
- SUB\_DISPOSITIF
- RAE\_NUM
- NOTIF\_UE



# 3. Les corrections

# La correction des champs

- Renommer les champs existants,
- Écrire en majuscules ou minuscules, mais poser une constante : sélection et MAJ+F3
- Supprimer les accents,
- Remplacer les espaces par des tirets-bas
- Ajouter les champs qui manquent

# La correction des erreurs

- Erreurs de format
- Erreurs d'espace
- Erreurs textuelles
- Erreurs textuelles dans une longue série de modalités
- Erreurs de casse dans les colonnes
- Erreurs d'accentuation dans les colonnes
- Erreurs de l'information même !
- Erreurs oubliées : en voyez-vous d'autres ?

# Les erreurs de format

- Comment sont notés les départements dans la colonne G ? Quel est le souci ?
- Depuis le fichier d'origine le souci se corrige en imposant un format texte, à la colonne G, à l'ouverture du CSV, lors de l'importation du fichier.
- Autrement, vous pouvez corriger en imposant un format texte à la colonne et en faisant des recherches et remplacer de chaque numéro unique (ex : 1 → 01).

# Les erreurs d'espace

- Combien de modalités différentes dans la première colonne ? Jaune PLF 2014 - Subventions 2012
- Corrections en utilisant les filtres, et en écrasant les données erronées.

# Les erreurs textuelles

- En colonne B, on devrait trouver une liste de 3 modalités, il en existe davantage.
- Correction pour une erreur régulière :  
rechercher/remplacer → Ctrl+h
- Correction pour une erreur unique : filtre et copier/coller

# Les erreurs de phrases

- Par exemple en colonne C, avec les Ministères, comment voir si nous n'avons pas de faux-doublons dans les modalités ?
- Correction : réaliser un tri à plat sur l'ensemble de la variable.
- Nous vous laissons refaire l'exercice dans la colonne concernant le PROGRAMME.

# Les erreurs de casse

- Dans la colonne du nom de l'association, vous en avez en minuscules.
- Correction en sélectionnant la colonne et en changeant la casse.
- Attention, terrain glissant, ne pas dénaturer les données !



# Les erreurs d'accentuation

- Avec les exemples THEATRE et THÉÂTRE ou FEDERATION et FÉDÉRATION, etc.
- Correction pour une erreur régulière : rechercher/remplacer → Ctrl+h
- Attention, terrain glissant, ne pas dénaturer les données !

# Les erreurs d'information

- Plus compliqué à trouver...
- nous avons ici le cas de l'identifiant SIREN, les premiers qui apparaissent dans le filtre font 8 chiffres au lieu de 9.

# Les erreurs oubliées

- En voyez-vous d'autres ?

# 4. Les ajouts

# L'ajout de variables

- Ajout d'une date de début de subvention
  - Ajout d'une date de fin de subvention

# L'ajout de la date de début

- Nom de la variable : SUB\_DATE\_DEBUT
- Les dates se posent en respectant un format international ISO-8601 : AAAA-MM-JJ
- Par exemple : 2017-09-18
- Ce format est reconnu par le tableur et vous pouvez incrémenter les cellules.

# L'ajout de la date de fin

- Nom de la variable : SUB\_DATE\_FIN
- Les dates se posent en respectant un format international ISO-8601 : AAAA-MM-JJ
- Par exemple : 2017-09-18
- Ce format est reconnu par le tableur et vous pouvez incrémenter les cellules.

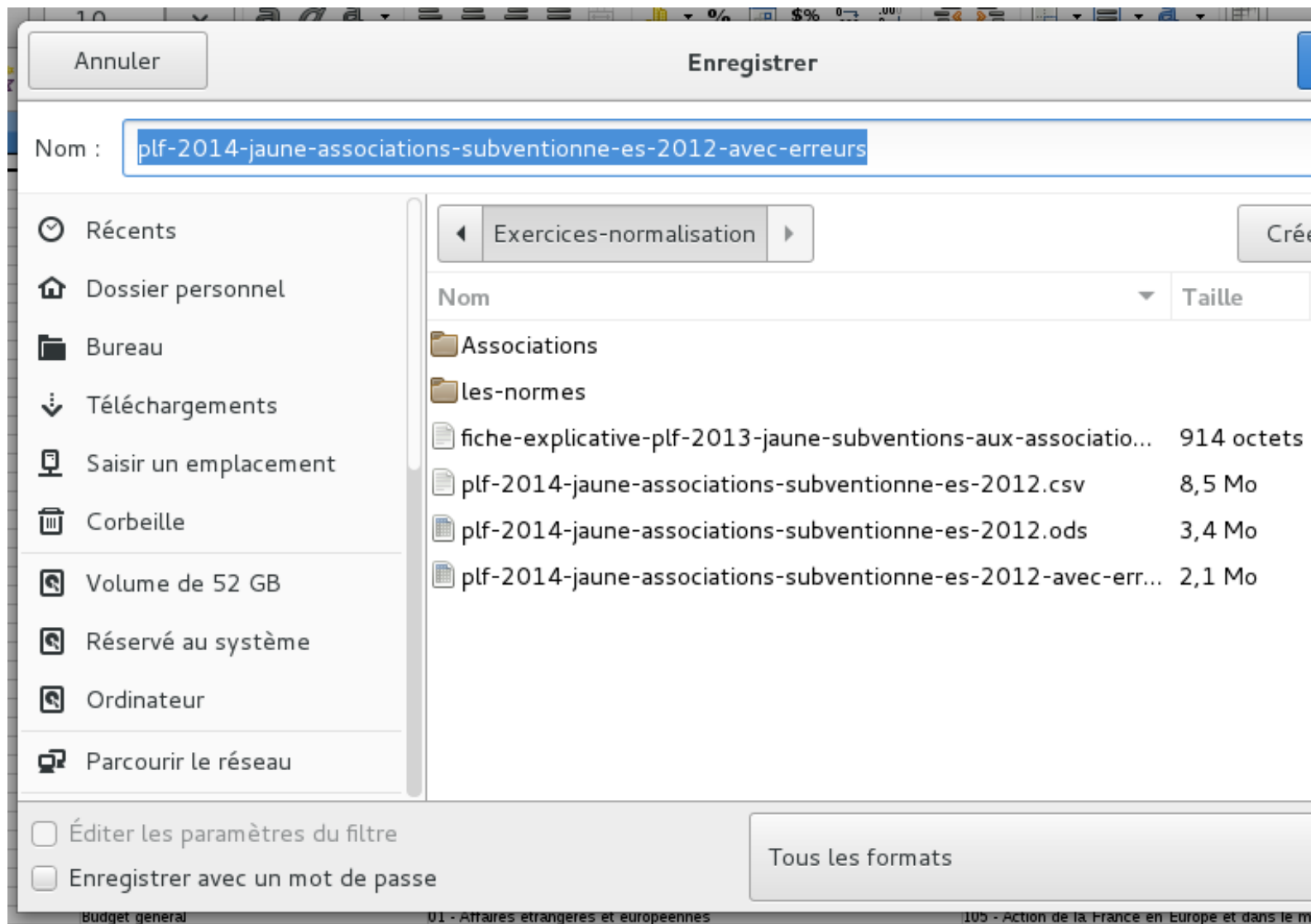
# 5. Le CSV



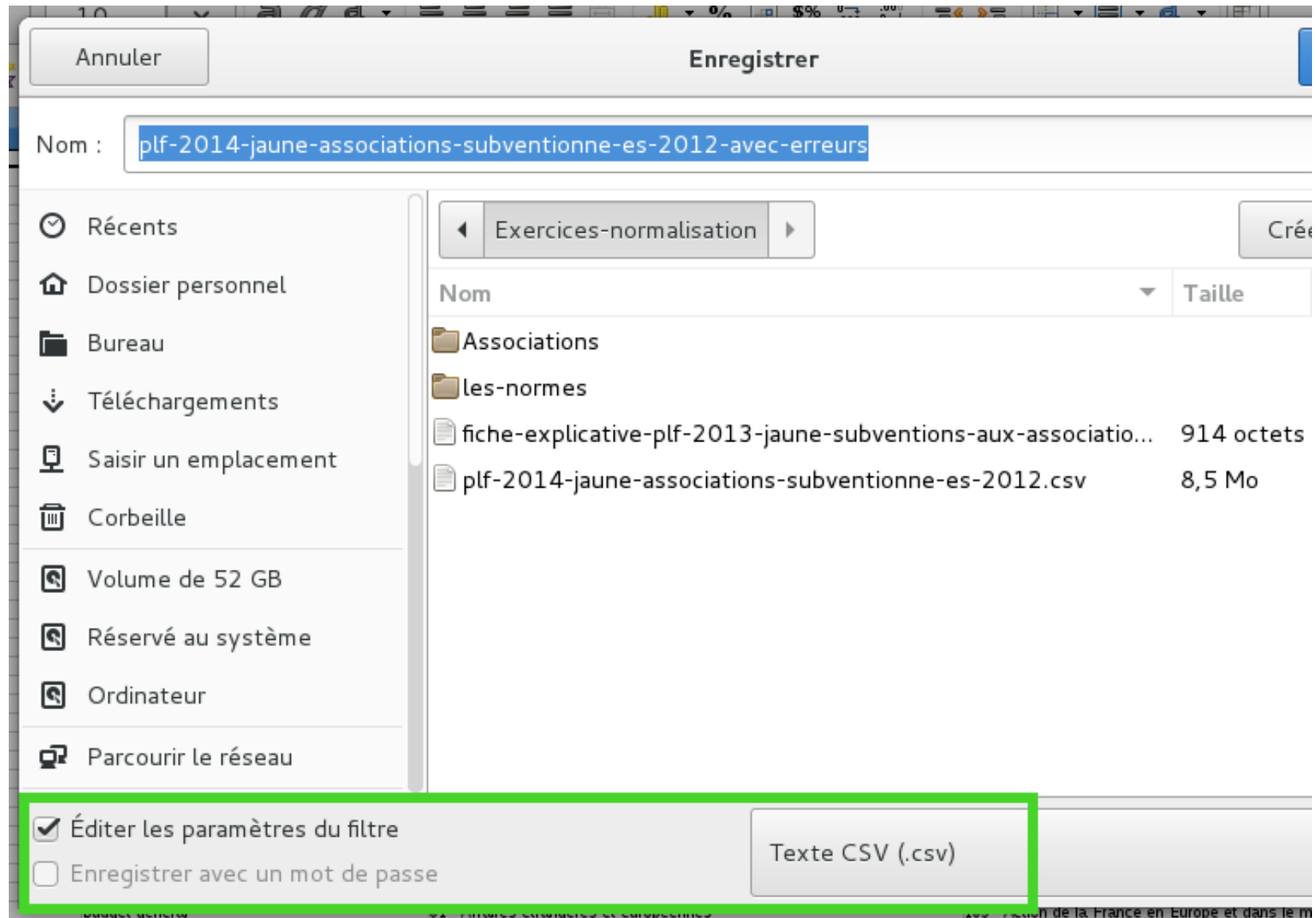
# CSV, les recommandations ODF

- Jeu de caractères en UTF-8
- Les données comportent leurs entêtes en première ligne
- Format CSV, séparé par des virgules (système anglais)
- Les données textuelles entre guillemets
- Les décimales séparées par un point (système anglais)
- Le retour lignes, en mode CRLF (système Windows !)
- Le CSV est vérifié
- Le CSV est documenté

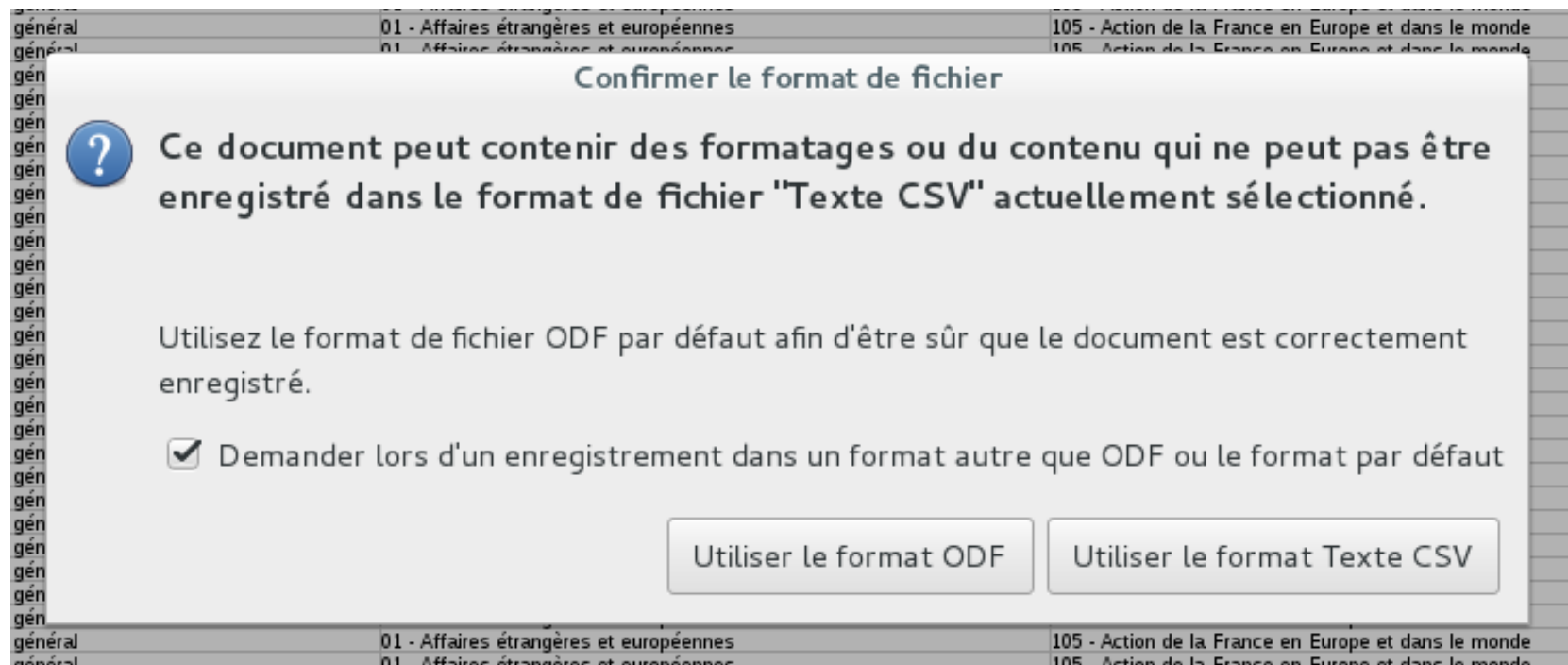
# CSV, l'enregistrement



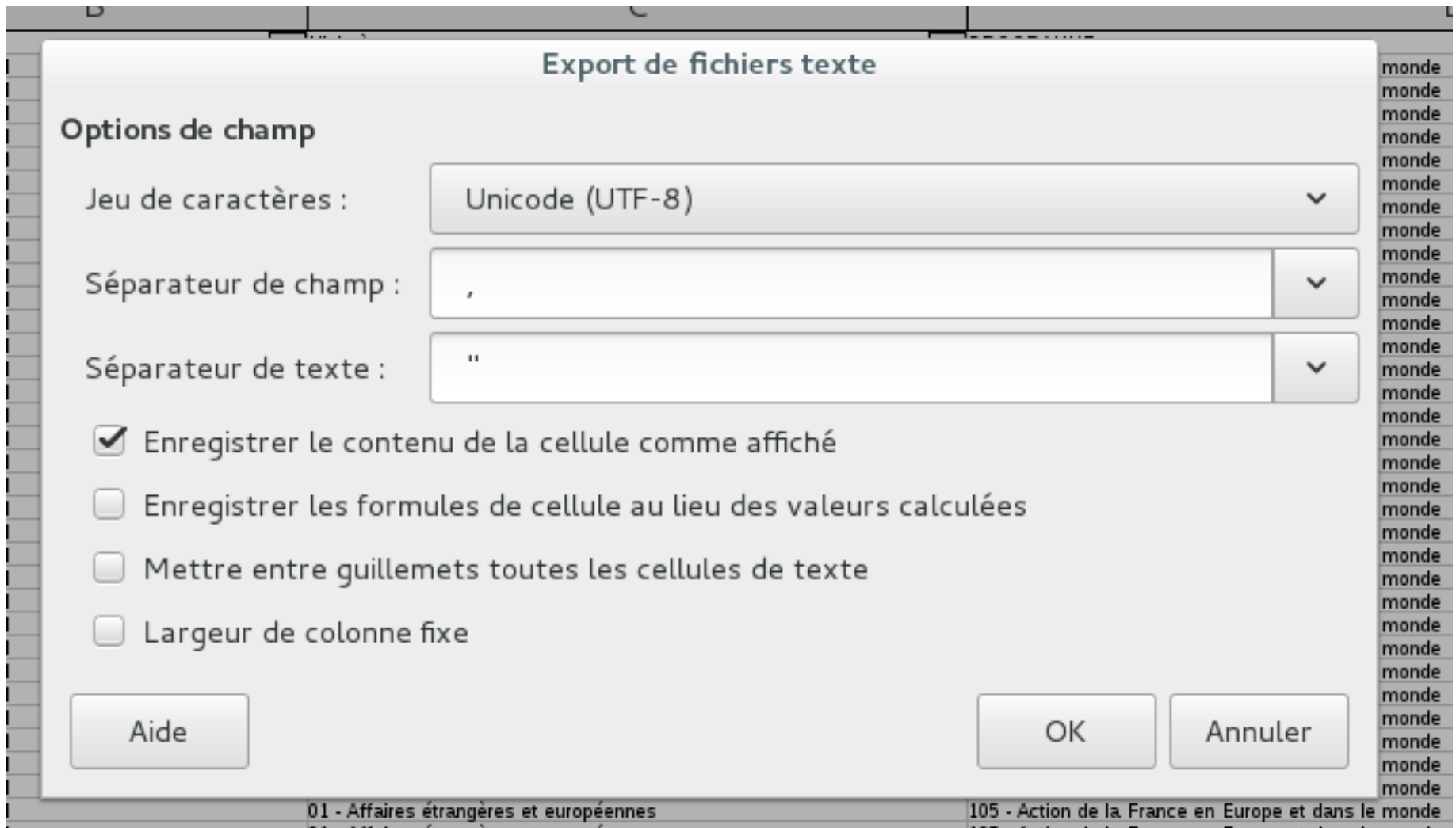
# CSV, éditer les paramètres








# CSV, l'avertissement



# CSV, définir les paramètres



# CSV, attention au poids du fichier

	fiche-explicative-plf-2013-jaune-subventions-aux-associations.csv	914 octets	Texte
	plf-2014-jaune-associations-subventionne-es-2012.csv	8,5 Mo	Texte
	plf-2014-jaune-associations-subventionne-es-2012.ods	3,4 Mo	Feuille de calcul
	plf-2014-jaune-associations-subventionne-es-2012-avec-erreurs.csv	9,1 Mo	Texte
	plf-2014-jaune-associations-subventionne-es-2012-avec-erreurs.ods	2,1 Mo	Feuille de calcul

# CSV, le test en ligne

The screenshot shows the CSV Lint website interface. At the top, there is a navigation bar with 'CSV Lint' and 'ALPHA' on the left, and 'About', 'Recent validations', 'Recent schemas', and 'API' on the right. Below the navigation bar is a header with the text 'Check your CSV files with CSVLint'. The main content area is divided into two columns. The left column contains introductory text: 'CSV looks easy, but it can be hard to make a CSV file that other people can read easily.', 'CSVLint helps you to check that your CSV file is readable. And you can use it to check whether it contains the columns and types of values that it should.', 'Just enter the location of the file you want to check, or upload it. If you have a schema which describes the contents of the CSV file, you can also give its URL or upload it.', and 'CSVLint currently only supports validation of delimiter-separated values (dsv) files. It is also possible to upload a .zip file of (dsv) files. Read more...'. The right column contains a form with the following elements: a section titled 'Enter a link to your CSV:' with a text input field labeled 'Enter URL' and a blue '+' button; a note stating 'Submitted urls are recorded in a public list of validation reports. If you want to validate private data then upload a file from your computer, using the Browse button below.'; a section titled 'Or upload a file:' with a blue button labeled 'Parcourir...' and the text 'Aucun fichier sélectionné.'; a checkbox labeled 'Add optional schema (in .json format)'; a text input field labeled 'Enter optional schema URL'; another 'Or upload a file:' section with a blue button labeled 'Parcourir...' and the text 'Aucun fichier sélectionné.'; and a large green button labeled 'Validate'.

- <http://csvlint.io/> (By Open Data Institute – Londres - <https://theodi.org/labs>)
- Si vous indiquez une adresse Web, une URL en ligne : votre fichier doit être anonymisé,
- Si vous parcourez votre disque dur pour chercher un CSV : attention donc de ne pas déposer de fichiers non-anonymisés

# 6. Le nommage et les métadonnées



# L'appellation du fichier

Le nom du fichier est figé : pérenne dans le temps  
il doit contenir :

- Le numéro du territoire, le nom de l'auteur, de la collectivité
- La description du contenu
- Des séparateurs avec des tirets bas, underscore
- La date de sa mise à jour ou son millésime

# Les informations sur le fichier

- Titre du jeu de données
- Description et usages
- Description de chaque champs
- Date de création
- Fréquence de mise à jour
- Couverture spatiale et temporelle
- Source, URL en ligne
- Licence utilisée
- Producteur des données

d'autres informations peuvent être placées dans un fichier texte ou bien sur la page web de l'espace de téléchargement.

# Contacts



Département de Loire-  
Atlantique  
Mission participation citoyenne  
et usages numériques  
02 40 99 89 22

[numerique.loire-  
atlantique.fr/ouvrevosdonnees](http://numerique.loire-atlantique.fr/ouvrevosdonnees)

[usages.numeriques@loire-  
atlantique.fr](mailto:usages.numeriques@loire-atlantique.fr)



Association Libertic  
Twitter @libertic  
[www.association-libertic.org](http://www.association-libertic.org)  
[contact@association-libertic.org](mailto:contact@association-libertic.org)