

## Passerelle - Bug #14404

### connecteur tableur : perfs déplorables sur un fichier ods

21 décembre 2016 10:53 - Frédéric Péters

<b>Statut:</b> Nouveau	<b>Début:</b> 21 décembre 2016
<b>Priorité:</b> Normal	<b>Echéance:</b>
<b>Assigné à:</b>	<b>% réalisé:</b> 0%
<b>Catégorie:</b>	<b>Temps estimé:</b> 0:00 heure
<b>Version cible:</b>	<b>Planning:</b>
<b>Patch proposed:</b> Non	
<b>Description</b>	
Avec le fichier en pièce jointe, sur tournai, le endpoint de base /data prenait 8,5 secondes. En mettant le fichier en CSV, ça prend 62 millisecondes.	
<b>Demandes liées:</b>	
Lié à Passerelle - Development #11260: mise en base des données du CSV	<b>Fermé</b> 07 juin 2016
Lié à Publik - Development #44020: développer notre petit module de lecture ods	<b>Solution proposée</b> 24 juin 2020
Dupliqué par Passerelle - Support #30282: csvdatasource: pyexcel-ods lent à l...	<b>Rejeté</b> 31 janvier 2019

### Historique

#### #1 - 21 décembre 2016 11:00 - Frédéric Péters

En local,

```
>>> def f():
...     t0 = time.time()
...     get_data_ods('/tmp/usage_des_actes_modif.ods')
...     print time.time() - t0
...
>>> f()
10.8324098587
```

Ce fichier fait 4 Ko, 23 lignes, 3 colonnes.

#### #2 - 21 décembre 2016 11:20 - Frédéric Péters

- *Sujet changé de connecteur tableur : perfs déplorables sur de l'ods à connecteur tableur : perfs déplorables sur un fichier ods*

La détection du nombre de colonnes/lignes se fait sur cette base :

```
<table:table-row table:number-rows-repeated="1048553" table:style-name="ro1">
  <table:table-cell table:number-columns-repeated="16384"/>
</table:table-row>
```

et badaboum.

#### #3 - 21 décembre 2016 11:25 - Jean-Baptiste Jaillet

il regarde chaque colonne vide puis dans chaque colonne il regarde chaque cellule ? (si je comprends bien). Et il regarde celle qui sont pas vide...

#### #4 - 21 décembre 2016 11:34 - Frédéric Péters

- *Lié à Development #11260: mise en base des données du CSV ajouté*

#### #5 - 21 décembre 2016 15:14 - Frédéric Péters

Sans passer par [#11260](#), un truc qui doit pouvoir être fait c'est charger le fichier en mode "stream", limiter le nombre de colonnes à celles où le connecteur a défini des identifiants, s'arrêter dès qu'une ligne est totalement vide.

#### #6 - 22 décembre 2016 13:39 - Benjamin Dauvergne

Le plus simple je pense c'est de convertir au chargement vers un format unique, par exemple CSV/UTF-8, mais si on sait écrire de l'ODS ça pourrait ça. À voir si on pourrait quand même garder le fichier original pour quand l'utilisateur veut le télécharger (quelqu'un qui utilise XLS n'a peut être pas de quoi éditer correctement du CSV/UTF-8 ou de l'ODS).

**#7 - 22 juin 2017 08:58 - Frédéric Péters**

- Fichier Communes-GRC.ods ajouté

Autre fichier pareillement gros, rencontré dans #17100.

**#8 - 24 juin 2018 11:21 - Frédéric Péters**

- Fichier meyzieu\_base\_nautique\_groupes\_b2n8fps.ods ajouté

Pour aller au-delà des fichiers déclarant par erreur des milliers de lignes ou colonnes, ce fichier (via #24726 ) (au premier regard) ne me semble pas exposer ces soucis. Mais décompressé son content.xml fait 3,6 Mo, ce qui suffirait à être trop lent. Ajouté à pyexcel qui est bien peu rapide, parser le fichier prend genre 5 secondes (dans mon test en local).

Bref, sur la proposition de Benjamin, produire un CSV "cache", ça serait bienvenu.

**#9 - 10 février 2019 14:05 - Frédéric Péters**

- Dupliqué par Support #30282: csvdatasource: pyexcel-ods lent à la lecture d'un ods de 217 lignes ajouté

**#11 - 17 juin 2020 14:49 - Frédéric Péters**

- Lié à Development #44020: développer notre petit module de lecture ods ajouté

**Fichiers**

---

usage_des_actes_modif.ods	4,19 ko	21 décembre 2016	Frédéric Péters
Communes-GRC.ods	18,4 ko	22 juin 2017	Frédéric Péters
meyzieu_base_nautique_groupes_b2n8fps.ods	80,1 ko	24 juin 2018	Frédéric Péters