Comment transférer des données d'un PDF vers Excel sans (trop de) douleur Trucs et astuces

La réutilisation dans un chiffrier Excel de certaines données provenant d'un tableau dans un article en PDF nécessite parfois beaucoup de temps et de sueurs, en plus des erreurs de saisies possibles. Mais êtes-vous sûr de ne pas souffrir inutilement?

Voici une procédure simple pour les utilisateurs, sous Windows.

- 1) Vous aurez besoin d'un éditeur de texte efficace (non non non, surtout pas Word ni même BlocNotes; vous voulez du texte ASCII tout propre, sans code caché). Nous vous recommandons Notepad2, un utilitaire gratuit qui a pour grand avantage de fonctionner sans installation, et ne touche donc pas au registre de votre système.
 - a. <u>http://www.flos-freeware.ch/notepad2.html</u>
 - b. Pour gagner du temps, installez-le dans la barre des tâches au bas de votre écran en glissant le fichier .exe sur la barre des tâches.



c. Attention! Si vous êtes un utilisateur Endnote, pour cette opération, vous devrez ouvrir le PDF hors du lecteur de PDF intégré dans Endnote. La version intégrée dans le logiciel Endnote ne copie pas les données correctement, allez savoir pourquoi... Pour ce faire, vous pouvez lancer le Reader d'Acrobat hors d'Endnote via le bouton Open File.

iroups	Tool	s١	Vindow Help					
d fr) mod			ि ■ ∞ ½ 企 ⊗	0 E) "	🕞 💀 💭 🚣 🕰 🗸 🕐 Quick Sea	rch 🔍 🔹 😞 Hide S	
	Search Options >							
(18065)			Any Field + PDE with Note -	ontains		•		
(1)			Any field of bit interfield of the					
	Q Auteur(s) Année Titre Publication							
(1)	۰	C	Montoya, Francisco G.; Alcayde, Alf	re 20	18	A fast method for identifying worldwide scien	Telematics and Informatics	
(16699)								
(102)								

2) Une fois le PDF ouvert dans Acrobat Reader, copiez le tableau en prenant soin de bien intégrer toutes les cellules.

Substances	Concentration (n = 4)	ns (ng g^{-1}) from	n 100 mg of v	vet tissue
	A	В	С	D
Liver tissues				
Venlafaxine	0.69 ± 0.14	0.16 ± 0.06	0.16 ± 0.02	n/d
Paroxetine	0.35 ± 0.10	0.15 ± 0.05	0.14 ± 0.03	n/d
S-citalopram	0.41 ± 0.19	n/d	0.06 ± 0.03	n/d
Norfluoxetine	0.63 ± 0.17	0.32 ± 0.10	n/d	n/d
Fluoxetine	0.20 ± 0.06	0.11 ± 0.05	n/d	n/d
Desmethylsertraline	10.29 ± 0.76	1.17 ± 0.28	0.06 ± 0.02	n/d
Amitriptyline	0.29 ± 0.07	n/d	n/d	n/d
Sertraline	0.29 ± 0.05	0.15 ± 0.04	0.15 ± 0.08	0.10 ± 0.0
Brain tissues				
Venlafaxine	0.43 ± 0.10	n/d	n/d	n/d
Paroxetine	0.19 ± 0.11	n/d	0.13 ± 0.07	n/d
S-citalopram	0.18 ± 0.11	n/d	0.06 ± 0.03	n/d
Norfluoxetine	n/d	n/d	n/d	n/d
Fluoxetine	0.08 ± 0.02	0.11 ± 0.05	n/d	n/d
Desmethylsertraline	2.08 ± 0.45	0.78 ± 0.40	0.06 ± 0.02	n/d
Amitriptyline	n/d	n/d	n/d	n/d
Sertraline	0.21 ± 0.08	0.16 ± 0.06	0.04 ± 0.01	<loq< td=""></loq<>
Muscle tissues				
Venlafaxine	0.08 ± 0.03	n/d	n/d	n/d
Paroxetine	<loq< td=""><td>n/d</td><td><loq< td=""><td>n/d</td></loq<></td></loq<>	n/d	<loq< td=""><td>n/d</td></loq<>	n/d
S-citalopram	n/d	n/d	n/d	n/d
Norfluoxetine	<loq.< td=""><td>n/d</td><td>n/d</td><td>n/d</td></loq.<>	n/d	n/d	n/d
Fluoxetine	0.09 ± 0.01	0.075 ± 0.03	n/d	n/d
Desmethylsertraline	0.42 ± 0.15	0.29 ± 0.12	0.06 ± 0.02	n/d
Amitriptyline	n/d	n/d	n/d	n/d
Sertraline	0.12 ± 0.03	0.093 ± 0.02	0.05 ± 0.02	<loo< td=""></loo<>

t (A) 20% v/v effluent; (B) 20% v/v effluent treated with O₃; (C) 10% v/v effluent; (D) 10% v/v effluent treated with O :

3) Coller dans Notepad2.

) 💕 🔯 🗐 🖕 🔍 🖉 👗 🔝 👘 🔛 🗐	0. 🔾 🖃 📈 📭
1 Liver tissues	· · · · · · · · · · · · · · · · · · ·
2 Venlafaxine $0.69 \pm 0.14 \ 0.16 \pm 0.06 \ 0.$	$16 \pm 0.02 \text{ n/d}$
3 Paroxetine 0.35 ± 0.10 0.15 ± 0.05 0.1	$4 \pm 0.03 \text{ n/d}$
4 S-citalopram 0.41 ± 0.19 n/d 0.06 ± 0.	03 n/d
5 Norfluoxetine 0.63 ± 0.17 0.32 ± 0.10	n/d n/d
6 Fluoxetine 0.20 ± 0.06 0.11 ± 0.05 n/d	n/d
7 Desmethylsertraline 10.29 ± 0.76 1.17	± 0.28 0.06 ± 0.02 n/d
<pre>8 Amitriptyline 0.29 ± 0.07 n/d n/d n/d</pre>	
9 Sertraline 0.29 ± 0.05 0.15 ± 0.04 0.1	$5 \pm 0.08 \ 0.10 \pm 0.03$
10 Brain tissues	
11 Venlataxine 0.43 ± 0.10 n/d n/d n/d	
12 Paroxetine 0.19 \pm 0.11 n/d 0.13 \pm 0.0/	n/a
13 S-Citalopram 0.18 \pm 0.11 H/d 0.06 \pm 0.	03 n/a
15 Eluovetine 0.08 \pm 0.02 0.11 \pm 0.05 p/d	n/d
16 Desmethylsertraline 2 08 + 0 45 0 78 +	$0.40.0.06 \pm 0.02 \text{ n/d}$
17 Amitriptyline n/d n/d n/d n/d	
18 Sertraline 0.21 ± 0.08 0.16 ± 0.06 0.0	$4 \pm 0.01 < LOQ$
19 Muscle tissues	
20 Venlafaxine 0.08 ± 0.03 n/d n/d n/d	
21 Paroxetine <loq <loq="" d="" d<="" n="" td=""><td></td></loq>	
22 S-citalopram n/d n/d n/d n/d	
23 Norfluoxetine <loq d="" d<="" n="" td=""><td>91 /d</td></loq>	91 /d
24 Fluoxetine 0.09 ± 0.01 0.0/5 ± 0.03 n/	a n/a
25 Desmethy/Sertraine 0.42 ± 0.15 0.29 ±	0.12 0.06 ± 0.02 H/d
27 Sertraline 0.12 + 0.03 0.003 + 0.02 0	05 + 0.02 <1.00
27 Ser er ur me 0.12 2 0.03 0.033 2 0.02 0.	03 1 0.02 KLOQ

- 4) Pour qu'Excel puisse reconnaître les colonnes, il faut insérer un caractère qui servira de séparateur. Dans l'exemple, les colonnes sont séparées par un espace entre deux données, mais ce ne sera pas suffisant, car il y a aussi des espaces à l'intérieur des colonnes.
- 5) Sélectionner l'espace entre des colonnes.

* Untitled - Notepad2
File Edit View Settings ?
🗋 😂 📮 🖃 🗢 🕲 👗 🛍 👫 🔩 🖃 🔍
1 Liver tissues 2 Venlafaxine $[0.69 \pm 0.14 \ 0.16 \pm 0.06 \ 0.16$ 3 Paroxetine $0.35 \pm 0.10 \ 0.15 \pm 0.05 \ 0.14 \pm 0.50 \ 0.14 \pm 0.19 \ n/d \ 0.06 \pm 0.03$ 5 Norfluoxetine $0.63 \pm 0.17 \ 0.32 \pm 0.10 \ n/d$ 6 Fluoxetine $0.20 \pm 0.06 \ 0.11 \pm 0.05 \ n/d \ n/d$ 7 Desmethylsertraline $10.29 \pm 0.76 \ 1.17 \pm 0.50 \ n/d \ n/d \ n/d$

6) Dans le menu Edit => Replace (ou Ctrl-H).

File	dit view Settings	?	
	Undo	Ctrl+Z	👫 🥼 💀 🔇
1	Redo	Ctrl+Y	
2	Cut	Ctrl+X	$6 \pm 0.06 0.1$
4	Conv	Ctrl+C	d 0.06 ± 0.
5	Copy All	Alt+C	$.32 \pm 0.10$
7	Paste	Ctrl+V	0.76 1.17
8.	Swan	Ctrl+K	/d n/d n/d
10	Clear	Del	± 0.04 0.1
11	Clear All	Dei	n/d n/d
12	Clear Cliphoard		0.13 ± 0.07 d 0.06 ± 0.0
14	Select All	Ctrl+A	/d
15			0.45 0.78 ±
17.	Line	•	/d
18	Block	•	± 0.06 0.04
20	Brace	•	n/d n/d
21	Convert	•	d
23	Insert	•	n/d
24	Find	Ctrl+F	$5 \pm 0.03 \text{ n/m}$ 0.15 0.29 +
26	Save Find Text	Alt+F3	/d
27	Find Next	F3	B ± 0.02 0.
	Find Previous	Shift+F3	
	Select To Next	F2	
	Replace	Ctrl+H	
	Goto	Ctrl+G	

7) L'avantage de cette méthode est que Notepad a déjà défini la cible à remplacer.



8) Insérez un point-virgule comme caractère de remplacement.
 Attention! Le caractère choisi doit être absent des données à traiter. Confirmez (bouton Replace All)



9) Chaque espace présent dans les données a été remplacé par un point-virgule.



10) Toutefois, nos données comportaient aussi des espaces à l'intérieur des colonnes (dans le cas ici, les caractères plus ou moins (<u>±</u>) ainsi que des espaces entre certains mots tels que liver_tissues).

Substances	Concentrations $(ng g^{-1})$ from 100 mg of wet tissue $(n = 4)$				
	А	В	С	D	
Liver tissues					
veniaraxine	0.69 ± 0.14	0.16 ± 0.06	0.16 ± 0.02	n/d	
Paroxetine	0.35 ± 0.10	0.15 ± 0.05	0.14 ± 0.03	n/d	
S-citalopram	0.41 ± 0.19	n/d	0.06 ± 0.03	n/d	
Norfluoxetine	0.63 ± 0.17	0.32 ± 0.10	n/d	n/d	
Fluoxetine	0.20 ± 0.06	0.11 ± 0.05	n/d	n/d	
Desmethylsertraline	10.29 ± 0.76	1.17 ± 0.28	0.06 ± 0.02	n/d	
Amitriptyline	0.29 ± 0.07	n/d	n/d	n/d	
Sertraline	0.29 ± 0.05	0.15 ± 0.04	0.15 ± 0.08	0.10 ± 0.03	

11) Corrigez à la main les quelques données textuelles qui ne sont pas redondantes et donc ne peuvent s'automatiser.



12) Sélectionnez les chaînes de caractères redondantes qui peuvent être corrigés en lot.



13) Remplacez de la même manière que ci-haut (sans oublier de réinsérer les espaces si souhaité, comme ici avant et après le ±). Confirmez (bouton Replace All).

Attention! Cette opération de ménage des données peut parfois nécessiter plusieurs itérations selon le nombre de caractères à nettoyer.

Replace	Text		×
	Search String:		Find
	;±;	-	Replace
	Replace with:		Replace All
	±	•	In Selection
	Match and	Dec'humen erward	Close
	Match whole word only	Search un	
	Match beginning of word only	Close after replace	
	Regular expression search		
	Transform backslashes		

14) Les données sont maintenant prêtes à être importées dans Excel.



15) Sauvegardez le fichier avec l'extension .csv ou .txt (les deux fonctionnent).

P Ordinateur Disque local (C:)	IMG_04003PG	
Lecteur DVD RW (E:)	F-APS-2000025217-20171128.pdf 20180105 tot 4 III	
Nom du fichier exemple.csv Type : All files (*.*)]	•

16) Dans Excel, menu Données, Fichier texte.

X 🗉 🛏	2.6							
FICHIER	ACCUEI	INSERT	ON MISE	EN PAGE	FORMULES	DONNÉES	RÉVISION	AFI
Access	Web Fichie texte	er vutres	Connexions existantes	Actualiser tout +	Connexions	A↓ AAA	r Filtrer	Teffa Te Réaj
	Donné	es e <mark>s</mark> ternes			Connexions		Trier et f	iltrer
A1	*	\times	f _x					
	A	В	С	D	E	F	G	ł
1								
2								

17) Sélectionnez votre fichier de données nettoyées. Confirmez (bouton Importer)

🎉 🕨 Ordinateur	DATA (D:) entrepot		- 4,	Rechercher dans : en	trepot 🔎
 Nouveau de 	ossier)H •	1 0
cuments	A Man		Modifié le	Туре	Taille
ages	C exemple.csv		2018-02-16 13:44	Fichier CSV Micro	2 K
sique	al Comonorus		2018-01-05 14:23	Document texte	120 K
léos	EZproxyLogFile20171220.txt		2017-12-21 14:19	Document texte	1 641 K
gault Jean-Daniel	152017021211.csv		2017-12-15 11:02	Fichier CSV Micro	1 K
nateur	eprint_auteurstop10.csv		2017-08-31 16:49	Fichier CSV Micro	2 K
au	001.bt		2017-08-31 09:28	Document texte	1 K
LUVIAQ	jstor_DDA-INRS-lino-mod.txt		2017-08-22 14:49	Document texte	143 K
IGUILLE	overlap_titles.txt		2017-08-22 14:18	Document texte	4 K
ISTUWACH	unique_titles.txt		2017-08-22 14:18	Document texte	315 K
NUCESABLE	objects_without_portfolios.txt		2017-08-22 14:17	Document texte	4 K -
TINEEDINOOK	* 1	m			•
Nom de	fichie: exemple.csv				
			Outils 👻	Importer	Annuler

18) L'assistant d'importation de texte s'ouvre. Choisissez Délimité comme type de données d'origines. Confirmez (bouton Suivant).



19) Sélectionnez Point-virgule comme séparateurs (ou tout autre caractère que vous auriez utilisé à la place du point-virgule). Vous pouvez déjà voir dans l'aperçu des données que tout se présente bien.... Confirmez (bouton Suivant).

Assistant Importation Cette étape vous per l'aperçu ci-dessous.	on de texte - Étape 2 sur 3 ermet de choisir les séparateurs contenus dans vos données. Vous pouvez voir les changements	sur votre texte dans
Séparateurs Tabulation Point-virgule Espace Autge :	Interpréter des séparateurs identiques consécutifs comme uniques Identificateur de tegte :	
Liver tissues Venlafaxine Paroxetine S-citalopram Norfluoxetine	$\begin{array}{c} 0.69 \pm 0.14 & 0.16 \pm 0.06 & 0.16 \pm 0.02 & n/d \\ 0.35 \pm 0.10 & 0.15 \pm 0.05 & 0.14 \pm 0.03 & n/d \\ 0.41 \pm 0.19 & n/d & 0.06 \pm 0.03 & n/d \\ 0.63 \pm 0.17 & 0.32 \pm 0.10 & n/d & n/d \end{array}$	•
	Annuler < <u>P</u> récédent <u>S</u> uivant	> <u>T</u> erminer

20) Excel vous propose ensuite de choisir le format des données (Date, nombre, etc.). Vous pourrez aussi faire ce travail plus tard une fois les données dans Excel. Confirmez (bouton Terminer).

Assistant Importation de texte - Éta	pe 3 sur 3	22 8
Cette étape vous permet de sélection Format des données en colonne Standard Tegte Date : AMJ Colonne non distribuée Aperçu de données	nner chaque colonne et de dé L'option Standard convertit valeurs en texte.	înir le format des données. es valeurs numériques en nombres, les dates en dates et les autres Ayancé
Standard Standard Liver tissues 0.69 ± 0.14 Venlafaxine 0.35 ± 0.10 S-citalopram 0.41 ± 0.19 Norflucxetine 0.63 ± 0.17	Standard Standard 0.16 ± 0.06 0.16 ± 0. 0.15 ± 0.05 0.14 ± 0. n/d 0.06 ± 0. 0.32 ± 0.10 n/d	Standard
		Annuler < <u>P</u> récédent Suivant > <u>Terminer</u>

21) Excel vous demande à partir de quelle cellule vous voulez importer les données en tableau. Confirmez (bouton OK).

A1 - : >	$< \sqrt{f_x}$								
AA	3 C	D	E	F	G				
1	Importer des donn	iées			22 9				
3	Sélectionnez la mé	éthode d'affici	hage de ces dou	nnées dans v	otre classeur				
4	O Tablea	u	lage ac ces aoi	inities dans v	oure classear.				
5	Rapport de tableau croisé dynamique								
6	Graphique croisé dynamique								
7	📑 🔘 Ne créer que la connexion								
8	Intérer les données dans								
9	Feuille de calcul <u>e</u> xistante :								
10	=SAS1								
11	Nouvelle feuille de calcul								
12	Ajouter ces do	nnées au <u>m</u> od	èle d <u>e données</u>	;					
13									
14	P <u>r</u> opriétés			ОК	Annuler				
15	<u> </u>								

22) Et les données sont importées.

1	А	В	С	D	E	F
1	Liver tissues					
2	Venlafaxine	0.69 ± 0.14	0.16 ± 0.06	0.16 ± 0.02	n/d	
3	Paroxetine	0.35 ± 0.10	0.15 ± 0.05	0.14 ± 0.03	n/d	
4	S-citalopram	0.41 ± 0.19	n/d	0.06 ± 0.03	n/d	
5	Norfluoxetine	0.63 ± 0.17	0.32 ± 0.10	n/d	n/d	
6	Fluoxetine	0.20 ± 0.06	0.11 ± 0.05	n/d	n/d	
7	Desmethylsertraline	10.29 ± 0.76	1.17 ± 0.28	0.06 ± 0.02	n/d	
8	Amitriptyline	0.29 ± 0.07	n/d	n/d	n/d	
9	Sertraline	0.29 ± 0.05	0.15 ± 0.04	0.15 ± 0.08	0.10 ± 0.03	
10	Brain tissues					
11	Venlafaxine	0.43 ± 0.10	n/d	n/d	n/d	
12	Paroxetine	0.19 ± 0.11	n/d	0.13 ± 0.07	n/d	
13	S-citalopram	0.18 ± 0.11	n/d	0.06 ± 0.03	n/d	
14	Norfluoxetine	n/d	n/d	n/d	n/d	
15	Fluoxetine	0.08 ± 0.02	0.11 ± 0.05	n/d	n/d	
16	Desmethylsertraline	2.08 ± 0.45	0.78 ± 0.40	0.06 ± 0.02	n/d	
17	Amitriptyline	n/d	n/d	n/d	n/d	
18	Sertraline	0.21 ± 0.08	0.16 ± 0.06	0.04 ± 0.01	<loq< td=""><td></td></loq<>	
19	Muscle tissues					
20	Venlafaxine	0.08 ± 0.03	n/d	n/d	n/d	
21	Paroxetine	<loq< td=""><td>n/d</td><td><loq< td=""><td>n/d</td><td></td></loq<></td></loq<>	n/d	<loq< td=""><td>n/d</td><td></td></loq<>	n/d	
22	S-citalopram	n/d	n/d	n/d	n/d	
23	Norfluoxetine	<loq< td=""><td>n/d</td><td>n/d</td><td>n/d</td><td></td></loq<>	n/d	n/d	n/d	
24	Fluoxetine	0.09 ± 0.01	0.075 ± 0.03	n/d	n/d	
25	Desmethylsertraline	0.42 ± 0.15	0.29 ± 0.12	0.06 ± 0.02	n/d	
26	Amitriptyline	n/d	n/d	n/d	n/d	
27	Sertraline	0.12 ± 0.03	0.093 ± 0.02	0.05 ± 0.02	<loq< td=""><td></td></loq<>	
28						

23) Relaxez, faites-vous un thé au jasmin, la vie est belle...