

RFC 4180 : Common Format and MIME Type for Comma-Separated Values (CSV) Files

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 9 mai 2007. Dernière mise à jour le 22 août 2017

Date de publication du RFC : Octobre 2005

<https://www.bortzmeyer.org/4180.html>

Curieusement, bien que très utilisé, le format CSV n'avait jamais fait l'objet d'une spécification officielle. Voilà qui est fait.

CSV est un format simple et cela explique donc que le RFC soit très court. Pour enregistrer le type MIME `text/csv`, l'IETF avait besoin d'un standard stable de CSV, qu'est notre RFC. Normaliser après coup n'est jamais facile et on trouve dans le RFC quelques perles, qui reflètent l'état de CSV dans le monde réel. Par exemple, « *"The last record in the file may or may not have an ending line break"* » ou bien « *"There maybe an optional header line appearing as the first line of the file"* ».

Le format CSV étant très simple, il est trivial de produire du CSV à partir de n'importe quelle source de données. Mais, la plupart du temps, il n'y a pas d'effort particulier à faire, beaucoup de logiciels savent déjà exporter en CSV. Par exemple, avec PostgreSQL :

```
psql -c "COPY $TABLE TO STDOUT WITH CSV" $DATABASE
```

Et, si on veut l'en-tête facultatif :

```
psql -c "COPY $TABLE TO STDOUT WITH CSV HEADER" $DATABASE
```

Autre problème liée à la normalisation tardive : des questions comme l'échappement des caractères sont laissées dans le vague. Mais la prime revient, comme souvent, à un logiciel Microsoft, en l'occurrence Excel, qui, dans sa version française, exige un point-virgule et pas une virgule comme séparateur ! Il ne peut donc pas lire le CSV « standard ».

Il existe des bibliothèques pour manipuler du CSV facilement dans tous les langages de programmation. En Python, je recommande le module csv standard <<http://docs.python.org/library/csv.html>>. En C, il existe une libcsv <<https://sourceforge.net/projects/libcsv/>>.

Sur la difficulté à lire et écrire du CSV en pratique, on peut lire l'excellent « Alors comme ça tu veux faire du CSV? » <<http://sebsauvage.net/wiki/doku.php?id=csv>> » (en anglais, « *So You Want To Write Your Own CSV code?* ») <<https://tburette.github.io/blog/2014/05/25/so-you-want-to-write-your-own-csv-code/>>.)

Sur le Web, un examen de quelques fichiers CSV montre que peu de serveurs HTTP prennent la peine d'étiqueter avec le type correct. On trouve plus souvent `text/plain`. Un exemple de fichier CSV en ligne (et servi avec le type correct) est sur ce blog, en (en ligne sur <https://www.bortzmeyer.org/files/simple-http-transfer.csv>). C'était un fichier pcap (en ligne sur <https://www.bortzmeyer.org/files/simple-http-transfer.pcap>)), transformé en CSV grâce à tshark (commande `tshark -r simple-http-transfer.pcap -T fields -E separator=, -e ipv6.src -e tcp.srcport -e ipv6.dst -e tcp.dstport -e tcp.flags -e frame.len` ce qui fait que le fichier CSV comprend donc l'adresse IP source, le port source, l'adresse IP de destination, le port de destination, les booléens TCP et la longueur du paquet). Autre exemple correct, sur <<http://data.gouv.fr/>>, par exemple avec <<http://static.data.gouv.fr/e9/751955c5dd1f2953cb1296e64efc4fde.csv>> (c'est une liste de monuments historiques <<http://www.data.gouv.fr/fr/dataset/liste-des-immeu>). Si le fichier CSV comprend l'en-tête optionnel mentionné plus haut (ce qui est le cas de celui de `data.gouv.fr` que je viens de mentionner), le type MIME devrait en théorie se terminer par un paramètre, `; header=present`. En pratique, je ne l'ai jamais vu.

La section 5 du RFC, portant sur la sécurité, prétend que CSV ne contient que des données, pas de code, et ne pose donc pas de problèmes de sécurité à lire. C'est en fait excessivement optimiste avec Excel ou Google Spreadsheets <<http://georgemauer.net/2017/10/07/csv-injection.html>>.

Une des preuves de l'utilité de CSV est que mon article Transformer du XML en CSV <<https://www.bortzmeyer.org/xml-to-csv.html>> est un des plus populaires, grâce aux nombreuses requêtes « Comment je transforme du XML en CSV? » sur les moteurs de recherche.