

RFC 8943 : Concise Binary Object Representation (CBOR) Tags for Date

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 21 novembre 2020

Date de publication du RFC : Novembre 2020

<https://www.bortzmeyer.org/8943.html>

Le format de données CBOR, normalisé dans le RFC 8949¹, possède un certain nombre de types de données de base, et un mécanisme d'extensions, les étiquettes ("*tags*"). Ce RFC spécifie deux nouvelles étiquettes, pour indiquer des dates.

Le RFC 8949 déclarait déjà deux types pour les estampilles temporelles, l'étiquette 0 pour une chaîne de caractères dont le contenu est une estampille au format du RFC 3339 (avec la date et l'heure), et l'étiquette 1 pour une estampille sous la forme d'un nombre de secondes depuis l'"*epoch*". Dans ce nouveau RFC sont ajoutées deux étiquettes, 100 pour un entier qui va stocker le nombre de jours depuis l'"*epoch*", et 1004 pour une date seule (sans heure) au format du RFC 3339. L'"*epoch*" est celle de la norme Posix 1 / IEEE Standard 1003.1, le 1 janvier 1970. Dans les deux cas, comme on ne stocke pas l'heure, des considérations comme le fuseau horaire ou les secondes intercalaires sont inutiles. Quant au calendrier utilisé, c'est le grégorien.

Dans ce calendrier, John Lennon (je reprends l'exemple du RFC...) est né le 9 octobre 1940 et mort le 8 décembre 1980. (Les dates utilisées dans ce RFC n'incluent pas l'heure.) Pour la première étiquette, 100, qui indique le nombre de jours depuis l'"*epoch*", l'auteur d'"*I Am the Walrus*" est né le -10676. C'est un nombre négatif puisque l'"*epoch*" utilisée est le 1 janvier 1970 <<https://www.bortzmeyer.org/epoch-50.html>>, après sa naissance. Lennon est mort le 3994. Pour le cas de la deuxième étiquette, 1004, il est né le 1940-10-09 et mort le 1980-12-08, suivant le format du RFC 3339. Le jour (lundi, mardi, mercredi...) est explicitement non mentionné, si on en a besoin, il faut le recalculer.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc8949.txt>

Les deux formats, en nombre de jours depuis l'epoch, et en RFC 3339 ont l'avantage que les comparaisons de date sont triviales, une simple comparaison d'entiers dans le premier cas, de chaînes de caractères dans le suivant, suffit.

Petit piège des dates indiquées sans l'heure, un même événement peut survenir à deux dates différentes selon le fuseau horaire. Ainsi, une vidéoconférence qui a lieu, à Tokyo, le 12 octobre à 10h00 sera considérée par les habitants d'Honolulu comme se tenant le 11 octobre à 15h00.

Les deux étiquettes ont été enregistrées à l'IANA <<https://www.iana.org/assignments/cbor-tags/cbor-tags.xml#tags>>. La valeur 100 pour la première a été choisie car 100 est le code ASCII de 'd' (pour "date").

Si vous voulez un fichier CBOR utilisant ces deux étiquettes, vous pouvez appeler le service <https://www.bortzmeyer.org/apps/date-in-cbor> qui vous renvoie un tableau avec les quatre façons de servir une date en CBOR, les deux standards du RFC 8949, et les deux de notre RFC. Ici, on utilise le programme `read-cbor` (en ligne sur <https://www.bortzmeyer.org/files/read-cbor.go>) pour afficher plus joliment :

```
% wget -q -O - https://www.bortzmeyer.org/apps/date-in-cbor | ./read-cbor -
Array of 5 items
...
  Tag 0
    String of length 20: 2020-11-21T06:44:33Z
  Tag 1
    Unsigned integer 1605941073
  Tag 100
    Unsigned integer 18587
  Tag 1004
    String of length 10: 2020-11-21
```