Réunion téléphonique avec les constructeurs : questionnements sur les fichiers de transmission des données 10s et QH.

# Fichiers de Données (10s et QH, voir plus)

Données QH :

À la base : besoin de rénover le fichier ISO pour permettre la transmission de données sur plus de 5 caractères.

Autre problématique : beaucoup de métadonnées pas nécessairement utiles dans le fichier ISO.

Données 10s : il faut prévoir un nouveau format de fichier pour transmettre l’ensemble des données 10s.

## Questions sur le format

Jusqu’où rénover le format du fichier ISO ?

* Seulement améliorer le format actuel (taille des champs, métadonnées…) ?
* Ou bien le rénover complètement et faire un nouveau format ?

Si nouveau format, pourquoi pas en créer un similaire pour les données 10s et QH ? (cela peut-il engendrer moins de développements ?)

Faut-il conserver le même format en entrée (prov. Stations) et en sortie (geodair) pour les fichiers de données QH dans le poste central ?

* Si oui : le fichier de données doit aussi bien contenir les données de tout un réseau que celles d’une station.
* Si non, beaucoup de métadonnées pourraient être gérées par le PC uniquement : (méthode, métadonnées du site, appareil…). Elles n’ont a priori pas besoin d’être dans le fichier entrant. Cependant dans ce cas le poste central doit être très rigoureux dans le suivi de ses configurations et donner l’accès à ces informations facilement.

Ce nouveau format de fichier pour les QH doit pouvoir être utilisé en import manuel sur le poste central. Quelles implications sur le contenu ?

Le format des fichiers de données :

* Avoir un contenu lisible (csv par exemple) et plus facilement exploitable par un opérateur.
* Un format plus évolutif que le ISO7168 : un séparateur plutôt qu’un nombre de caractères pour les données par exemple.
* Pour la transmission faut-il un format binaire ou une compression de fichiers ascii suffit ?
* Formats : csv, binaire ou xml json ou autre ?

## Questions sur le contenu

Pour les données 10s et QH :

* Limiter en durée ? 1 journée max par exemple
* Limiter en nombre de mesures par fichier ?

Quelle valeur est présente dans les fichiers :

* signal ?
* mesure physique ? (ppb)
* mesure de concentration ? (microg/m3)
* pour les QH : concentrations ; pour les 10s mesures : mesures physiques

Fichiers de données 10s :

* permettre le découpage du QH ?
* avis aasqa : pas besoin spécialement d’avoir les données 10s du QH en cours.

Questions sur le dimensionnement des fichiers.

## Autres

Nom (extension du fichier).

* Identique données QH et 10s ?
* faut-il continuer à distinguer le caractère brut ou validé des données dans l’extension du fichier ? (pas nécessaire a priori, ça ne concerne que les fichiers iso en sortie du poste central)

Attention il peut y avoir des cas où on importe des données déjà validées dans les postes centraux (polair)

* faut-il donner une indication sur le caractère complet ou pas de la journée dans l’extension du fichier ? (notion d’écrasement des données dans la base du poste central ?)

Proposition :

Arguments dans la commande de demande de fichier qui précisent la plage temporelle demandée à la minute près.

Le fichier de données contient au max la plage demandée à la min près ; il peut contenir moins si pour x raisons les mesures ne sont plus présentes en station ou ne le sont pas encore.

Définir comment sont spécifiées les mesures absentes dans les fichiers  pour distinguer :

* les cas où l’acquisition a échoué,
* où les mesures ne sont plus présentes dans la mémoire station,
* où la mesure n’était pas définie, pas active,
* …

Définir la manière dont le poste central doit gérer les ré-acquisitions, écrasement de données…

Comment seront gérés ces fichiers dans les stations ?

* Créés à la demande ?
* Toujours présents et complétés au fur et à mesure ?
* Quelle profondeur est possible ?
* Quelles règles d’effacement  en station ?

## Informations présentes

***Liste des informations présentes a minima pour une donnée 10s ou QH (on ne précise pas sous quelle forme sont indiquées ces informations) :***

* N° du site (paramètre site)
* N° de la mesure (ou code polluant + discriminant)
* unité
* indication sur la périodicité des données présentes dans le fichier : QH ou 10s
* Date
* Horodatage
* Valeur
* code qualité

(voir également la question de la précision)

en jaune : paramètres associés à chaque donnée

Dans ce cas de figure, très peu de métadonnées sont présentes, c’est au poste central de faire le lien entre les métadonnées station (nom du site, coordonnées…), les métadonnées mesure (polluant, appareil, méthode…) et les données à bancariser.

Le poste central doit être en mesure de suivre l’historique des changements de métadonnées : changement d’appareil, de méthode…

Sinon métadonnées mesures potentielles :

* code polluant,
* code discriminant,
* méthode
* appareil de mesure
  + n° de série
  + type
* gamme
* facteur de conversion, coefficients diverses
* hauteur prélèvement
* …
* Métadonnées du site

## Présentation des données

Formats de présentation des données :

* Par blocs de données comme le fichier iso actuel
* Par lignes (trop volumineux a priori)

Séparateur : ;

Suggestion de format par bloc

*Ligne d’entête du fichier :*

Les éléments constituant la ligne d’entête du fichier sont :

* le nombre de blocs de mesures présents dans le fichier,
* le n° du site de mesure,
* l’offset entre l’horodatage des données et l’heure TU au HHMMSS
* la périodicité des mesures affichées dans le fichier au format HHMMSS

Exemple :

Fichier de données QH contenant 3 mesures, pour le site de mesure 99005, l’heure de la station est à l’heure TU.

3|99005|0|001500<fin de ligne>

*Bloc de mesure :*

Les éléments constituant la ligne d’entête d’un bloc mesure sont :

* le numéro de la mesure en station,
* l’unité des données présentes dans le fichier,
* la date et horodatage de la première donnée présente dans le bloc au format AAMMJJHHMMSS
* le nombre de lignes de données pour le bloc.

Exemple :

Entête du bloc de la mesure n°10, mesurée en ppb, contenant une heure de données 10s à partir de 15h30 pour la journée du 12/11/2015 :

10|ppb|151112153000|360<fin de ligne>

Si problème de com :

* iseo : code D et 0
* fde : code N
* si station fde éteinte : pas de données, rien du tout dans le fichier ISO

quid si il manque des données au milieu de la journée ?