

Modèle de données pivot Theia/OZCAR

Format des fichiers de données

Avant-propos :

L'objectif de ce document est de décrire le format du fichier de données associé à l'objet **result** du modèle de données pivot.

Cela s'applique aux données de séries temporelles de type numérique (EnumDataTypes=Numeric) ou textuel (EnumDataTypes=Text)

Format d'un fichier de données de séries temporelles :

Chaque fichier contient la série temporelle d'une variable acquise par une station.

Le format de fichier est un format ASCII avec un séparateur « ; ». Le nom de fichier porte l'extension .txt.

Le début du fichier contient des métadonnées sur 4 lignes d'entêtes commençant par #:

```
#Date_of_extraction; « The date of the extraction in UTC and ISO8601 » ;\n
#Observation_ID; « the ID of the observation of the metadata JSON file » ;\n
#Dataset_title; « The title of the dataset » ;\n
#Variable_name; « The name of the observed property of the observation » ;\n
```

NB: L'id de l'observation indiqué dans l'entete doit correspondre à la valeur du champ **observationId** indiqué dans le **fichier de métadonnées .json**

Le reste du fichier contient les données de séries temporelles.

Chaque ligne représente un enregistrement. Chaque ligne est composée de 7 colonnes obligatoires suivies ou non de colonnes optionnelles. La première ligne de la série de données numériques correspond à l'intitulé de ces colonnes.

```
dateBeg;dateEnd;latitude;longitude;altitude;value;qualityFlags;[liste de valeur additionnelles
séparées par des « ; »];\n
2005-12-31T06:00:00Z;2006-01-01T06:00:00Z;13.5933;2.6544;152;-9999.9;21100.0;11;1,a,25;\n
2006-01-01T06:00:00Z;2006-01-02T06:00:00Z;13.5933;2.6544;152;-9999.9;21100.0;11;1,a;\n
2006-01-02T06:00:00Z;2006-01-03T06:00:00Z;13.5933;2.6544;152;-9999.9;21100.0;11;;
```

Chaque colonne est séparée par des « ; ». Le séparateur des décimal est un « . ».

Les éléments des colonnes `dateBeg` et `dateEnd` sont des dates UTC exprimées au format ISO 8601 « YYYY-MM-DDThh:mm:ssZ ».

- Pour les mesures instantanée, seul le champ `dateEnd` est rempli, `dateBeg` est laissé vide.
- Les éléments des colonnes `longitude`, `latitude` correspondent aux coordonnées de la mesure dans le référentiel de coordonnées WGS84 (ESPG:4326) avec une préférence d'au moins quatre chiffres après la virgule.
- Les éléments de la colonne `altitude` correspondent à l'altitude absolue de la mesure par rapport au niveau de la mer. Cette altitude est décrite en mètre.
- Les éléments de la colonne `value` sont les valeurs numériques de la variable considérée.
- Les éléments de la colonne `qualityFlags` correspondent aux drapeaux de qualité pour la chaque mesure. La signification de ces drapeaux de qualités est décrite dans le format pivot de métadonnées. Plusieurs drapeaux de qualité peuvent être levés pour une mesure. Ces différents drapeaux sont séparés par des « | ». Si des mesures de l'observation n'ont pas de drapeau de qualité, les éléments de la colonne de drapeaux de qualité correspondants seront décrits par une chaîne de caractère vide séparé par des « ; ».
- Les colonnes suivantes correspondent à des valeurs additionnelles aux mesures de l'observation (exemples : incertitudes, erreurs, paramètres d'instrumentations...). Ces valeurs additionnelles sont décrites dans le format pivot de métadonnées pour chaque observation. Le titre d'une colonne de valeur additionnelle est identique à celui décrit dans le format pivot de métadonnées. Les éléments des colonnes de valeurs additionnelles sont des éléments de texte libre. Lorsqu'une colonne de valeur additionnelles est déclarée pour une observation, la valeur additionnelle est décrite pour chaque mesure de l'observation. Si la valeur additionnelle n'est pas décrite pour des mesures de l'observation, les éléments de la colonne de valeur additionnelle correspondants seront décrits par une chaîne de caractère vide séparé par des « ; ».

Format des noms des fichiers de données .txt et archive ZIP

Format des noms de fichiers de données .txt

Chaque fichier .txt regroupant les données de séries temporelles d'une variable à une position donnée est nommé par l'id de l'observation déclaré dans le fichier de métadonnées .json : `observationId.txt`

Ex : `CATC_OBS_CE.Sap_Odc_1.txt`

Format des noms de l'archive .zip regroupant tous les fichiers données .txt d'un jeu de donnée

Tous les fichiers de données .txt des observations d'un même jeu de données sont regroupés dans un fichier ZIP nommé de la manière suivante :

« `identifiant 4 lettres du producteur de données` »_DAT_« `identifiant pérenne du jeu de données fourni par le producteur` ».zip

Ex : `CATC_DAT_CE.Sap_Odc.zip`