Theia/OZCAR pivot data model

Data file format

Foreword :

The objective of this document is to describe the format of the data file associated with the result object of the pivot data model.

It applies on time series data of numerical type (EnumDataTypes=Numeric) or textual type (EnumDataTypes=Text).

Format of a time series data file :

Each file contain the time series data of one variable acquired by a station.

The file format is an ASCII format with a";" separator. The file name has the extension.txt.

The beginning of the file contains metadata on 4 header lines starting with #:

```
#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: T**he observation id in the header** must match the value of the obser **observationId** field specified in the **.json metadata file.**

The rest of the file contains time series data.

Each line represents one record. Each row is composed of 7 mandatory columns followed or not by optional columns. The first line of the numerical data series corresponds to the title of these columns.

```
dateBeg;dateEnd;latitude;longitude;altitude;value;qualityFlags;[list of additional values separated
by « ; »];\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;;
```

Each column is separated by";". The decimal separator is a". "».

The elements of the `dateBeg` and `dateEnd` columns are UTC dates expressed in ISO 8601 format "`YYYY-MM-DDThh:mm:ssZ`".

- For instantaneous measurements, only the `dateEnd` field is filled in, `dateBeg` is left blank.
- The elements of the `longitude,latitude` columns correspond to the coordinates of the measurement in the coordinate reference frame WGS84 (ESPG:4326) with a preference of at least four decimal places.
- The elements in the `altitude` column correspond to the absolute altitude of the measurement with respect to sea level. This altitude is described in metres.
- The elements of the `value` column are the numerical values of the variable considered.
- The elements in the `qualityFlags` column correspond to the quality flags for each measurement. The meaning of these quality flags is described in the metadata pivot format. Several quality flags can be raised for one measurement. These different flags are separated by "|". If observation measurements do not have a quality flag, the elements of the corresponding quality flag column will be described by an empty character string separated by";".
- The following columns correspond to values additional to the observation measurements (examples: uncertainties, errors, instrumentation parameters, etc.). These additional values are described in the pivot metadata format for each observation. The title of an additional value column is the same as that described in the metadata pivot format. The elements of the additional value columns are free text elements. When an additional value column is reported for an observation, the additional value is described for each observation measure. If the additional value is not described for observation measurements, the corresponding elements of the additional value column will be described by an empty string separated by";".

Format of .txt data file names and ZIP archive names

Format of .txt data file names

Each.txt file containing the time series data of a variable at a given position is named by the observation id (see: description_fields_JSON_v1.0.docx) declared in the.json file: observationId.txt

Ex: CATC_OBS_CE.Sap_Odc_1.txt

Format of the .zip archive grouping all the .txt data files of a dataset

All data files . txt of observations from the same dataset are grouped together in a ZIP file named as follows:

`"4 letters of the data producer ID "_DAT_" permanent identifier of the data set provided by the producer".zip`

`Ex: CATC_DAT_CE.Sap_Odc.zip`