Comment requêter les mesures de pluviométrie sur data.grandlyon.com ?

La documentation de data.grandlyon.com consacre un chapitre au service REST (WS). Ce service permet de télécharger des données via l'API de data.granlyon.com
https://rdata-grandlyon.readthedocs.io/fr/latest/services.html#services-rest-en-json

Les éléments ci-dessous illustrent les principes de cette documentation sur l'exemple des mesures de pluviométrie.

Dans l'onglet API de la fiche de données Mesures de pluviométrie de la Métropole de Lyon, le service WS (Web Service) est le seul disponible pour cette donnée. Nous allons nous intéresser à la construction des requêtes dans la partie "Créer une requête".

Comment est composée la requête ?

Conformément aux options sélectionnée (Format de sortie, Nombre d'enregistrements, Numéro du 1er enregistrement), le contructeur de requêtes (URL personnalisée) propose la requête par défaut suivante:
https://data.grandlyon.com/fr/datapusher/ws/timeseries/eau.pluviometrie_mesure/all.json?maxfeatures=100&start=1

La début de la requête 'https://data.grandlyon.com/fr/datapusher/ws/timeseries/' indique que nous allons utiliser le service WS (Web Service) de la base de données des séries temporelles 'timeseries'. Cela sera toujours identique pour toute donnée issue d'une série temporelle
Cette url est suivie du nom de la table qui est requêtée. Pour les mesures de pluviométrie, il s'agit de 'eau.pluviometrie_mesure'
La mention suivante 'all.json' indique que le résultat sera retourné au format JSon que nous pourrons lire dans notre navigateur internet. Le format CSV ('all.csv') est également disponible et permet de télécharger les résultats pour les exploiter sous Excel. Nous l'utiliserons plus tard pour récupérer les mesures.
Dans l'url, après le '?', on trouve les paramètres, séparés par des '&', que nous pouvons exploiter pour affiner la requête et les résultats renvoyés

Pour notre exemple sur la pluviométrie, le début de la requête ne variera pas pour pour l'instant: https://data.grandlyon.com/fr/datapusher/ws/timeseries/eau.pluviometrie_mesure/all.json?
Nous allons nous intéresser aux paramètres qui suivent

paramètres généraux

Il existe des paramètres généraux, qui sont partagées par toutes les données de data.grandlyon.com:

- 'maxfeatures' indique le nombre d'objets renvoyés. Pour tester, on peut le laisser à 100 mais il ne faut pas oublier de le modifier pour récupérer l'ensemble des résultats. Pour récupérer tous les résultats, on indique 'maxfeatures=-1'. Si maxfeatures n'est pas précisé, 1.000 résultats sont renvoyés par défaut. Pour tester notre requête, il est conseillé de laisser la valeur proposée 'maxfeatures=100'
- 'start' indique quel enregistrement de la requête est renvoyé en premier. Il sert à paginer les résultats (renvoyer les résultats 100 par 100, par ex). Pour récupérer tous les résultats qui nous intéressent, nous le laisserons à sa valeur par défaut 'start=1'
- 'sortby' permet de trier les résultats retournés, en indiquant le nom de l'attribut de tri. Par ex, on peut trier sur l'attribut qui marque l'horodatage ('sortby=horodate' si l'attribut s'appelle ainsi). Sans précision, la requête du constructeur renvoie les premiers enregistrements de la base de données (pas chronologiques, mais les premières mesures intégrées).

paramètres spécifiques au jeu de données

Nous pouvons ensuite préciser des paramètres spécifiques à notre jeu de données. Ces paramètres s'appuieront sur les attributs de notre donnée. Nos mesures contiennent les attributs suivants:

- 'identifiant', qui est numéro métier du site de pluviométrie
- 'horodate', qui est un horodatage (timestamptz) de la mesure. Nous pourrons interroger cette donnée en utilisant les dates au format ISO 'AAAA-MM-JJ' (par ex, '2024-05-31') ou les horodates au format ISO 'AAAA-MM-JJ HH:mm:ss' (par ex, '2022-04-01 05:00:00', voire '2022-04-01 05:00' les secondes étant optionnelles). Les horodates sont exprimées dans le fuseau horaires local (Paris)
- 'pluie_mm', qui représente la valeur de la mesure

Nous allons utiliser le nom de l'attribut et des opérateurs de comparaison adaptés. La forme générale du filtre sera __<opérateur>=<valeur à comparer> (entre attribut et opérateur, il y a 2 underscores _)

Par exemple, il est possible de demande les mesures d'un site particulier avec l'opérateur 'eq' (equal, égalité):
https://data.grandlyon.com/fr/datapusher/ws/timeseries/eau.pluviometrie_mesure/all.json?maxfeatures=100&start=1&identifiant__eq=21
La requête me retourne toujours 100 mesures, mais elles ne concernent que le site 21 comme demandé ci-dessus.

Voici la liste des opérateurs disponibles, avec des exemples de requêtes:

opérateur	signification	description	exemple
eq	equal	égalité	identifiant__eq=21
gt	greater than	strictement plus grand que	pluie_mm__gt=18
gte	greater than or equal	plus grand ou égal à	pluie_mm__gte=18
lt	lesser than	strictement plus petit que	horodate__lt=1986-01-03
lte	lesser than or equal	plus petit ou égal à	horodate__lte=1986-01-03
in	in	dans la liste (dont les éléments sont séparés par des virgules)	identifiant__in=1,2,3

Les opérateurs sont cumulables dans la même requête. Par exemple, si nous souhaitons les 100 premières mesures chronologiques du site 4 pour le mois d'avril 2022:
https://data.grandlyon.com/fr/datapusher/ws/timeseries/eau.pluviometrie_mesure/all.json?maxfeatures=100&start=1&identifiant__eq=4&horodate__gte=2022-04-01&horodate__lte=2022-04-30&sortby=horodate

Une fois satisfait de notre requête en JSon, nous souhaitons récupérer toutes les mesures du mois d'avril 2022 du site 4 en CSV, il suffit de modifier la requête ci-dessus, en remplaçant le

'all.json' par 'all.csv' et ne pas oublier de mettre le maxfeatures à -1:
https://data.grandlyon.com/fr/datapusher/ws/timeseries/eau.pluviometrie_mesure/all.csv?maxfeatures=-1&start=1&identifiant__eq=4&horodate__gte=2022-04-01&horodate__lte=2022-04-30&sortby=horodate