

*Réunion de Co-construction normalisation
Normalisation d'un protocole d'échange de données en topologie fédérée
Projet QUOALISE

*2020-10-29

14h - 15h30

*Participants

- Cyril - Projet SEN
- Elias - Projet SEN
- Cédric - Projet SEN
- Guillaume Zin - plug-in Domoticz Linky
- François Bodin - IRISA

*Proposition d'ordre du jour

- Formalisme des données et métadonnées
 - RUDI
 - Domoticz
 - Quoalise
- Questions RUDI :
 - Quelle organisation pour produire le document que François nous a fourni la dernière fois
 - Séparation métadonnées unitaires / catalogue
 - Est ce qu'une entrée de catalogue n'est pas une donnée ? cf. XMPP qui semble se baser le plus possible sur la réutilisation d'un même formalisme
 - Les métadonnées se voient attribuer un identifiant unique et persistant. Pourquoi, comment ?
 - Les métadonnées restent accessibles même quand les données ne le sont plus ?
 - Références pour inspiration ?

*Formalisme RUDI, en développement

- deux niveaux de meta

+ unitaire: pour chaque mesure, resultat, caractériser le contexte

+ description de l'ensemble: origine, schéma, format)

* => catalogue RUDI = description, indexable

- Identifiant global (DOI)
- Identifiant local (optionel)
- Provenance
 - entité
 - date production
 - date révision
 - type (flux, fichier)
- fréquence de rafraichissement
- description informelle
- conditions d'accès
 - confidentialité
 - licences
 - habilitations / droits d'accès
- modèle économique
- Anonymisation

- si données perso mises à disposition de manière anonymisée

- Conditions d'usage
- Variantes existantes
- Couverture temporelle
- Couverture géographique
- Volume
 - Pas mal de discussion autour de ce sujet
 - Besoin de connaître la quantité avant de la demander
- Schémas de données utilisés
- Mots clés
 - Ne normalise pas les requêtes
 - Fixer une centaine de termes spécifiques vs tous les mots du dictionnaire
- Thèmes
- Sous-thèmes
- Langues utilisées au niveau de la restitution des données
 - ex Keolis est aussi ailleurs
- Formats disponibles
- Connecteurs (Déjà une liste définie ?)
- Statut stockage
- Consentement
- Règles de gestion du cycle de vie du jeu de données

*Formalisme Domoticz ?

*Pile technique

basé sur Sqlite

API Plugin assez neuve, Guillaume étant à l'origine de la demande de l'ouverture

Domoticz n'est pas prévu du tout pour loader une série de 24h d'un coup (plutôt en temps réel)

Le plugin actuel est en "boucle ouverte", il n'a pas accès aux données qu'il a remonté précédemment...

Données potentielles très diverses : Vent, pression, humidité, conso, eau etc.

*Pour une courbe affichée

- Vue par heure (par exemple sur les 7 derniers jours)
- Vue par jour, une conso par jour
- Devices de différents types avec tables spécifiques
- Pour données de conso, utilisation d'un device « multimeter »
- Déjà trop de devices dans Domoticz pour en créer un nouveau de type "linky"

L'idéal serait une vue de la personne qui souscrit la tarification avec les données

*Etat des lieux "énergie" dans Domoticz :

Deux types de compteur d'énergie

- heure pleine / heure creuse
- usage / index

Aucun n'est idéal, il faudrait agréger plusieurs compteurs (mais non prévu par défaut).

*Formalisme Quotalise, en développement, exemple courbe de charge

<https://pastebin.com/emSkmqjn>

XML, parce que XMPP

Ressemble aux document de travail de François sur la qualité de l'air (projet AQMO)

Station de collecte qui peut être intéressante, avec multi-capteurs.

Utilise GeoJSON pour la localisation

Métadonnées

- Iq XMPP
 - Source de la donnée (entité)
 - Destinataire de la donnée (entité)
- Qualise
 - Métadonnées => Correspond presque au catalogue RUDI?
 - Identification capteur
 - Entité (marque / fabricant / exploitant + id local) : Ex. Enedis + PRM
 - Type (compteur électrique, TODO liste à définir)
 - Mesure
 - Grandeur physique : ex. puissance électrique (W), TODO liste à définir
 - Nom métier : ex. courbe de charge, libre et ouvert
 - problème de cohérence, de contraintes
 - si fermé, il faut quelqu'un qui gère le truc
 - Type d'aggregation : moyenne, cumul, instantané TODO liste à définir
 - Interval de mesure
 - Données (SENML)
 - timestamp
 - valeur
 - unité (répété, aussi dans métadonnée de manière plus expressive)
 - identifiant unique (répété, aussi dans métadonnée de manière plus expressive)

*La situation côté RUDI

Parmi la pile technique des projets menés par l'IRISA à Rennes :

- MongoDB pour le stockage de séries temporelles
 - en type "cache"
 - Déjà quelques soucis de scaling
 - Attention à la licence contaminante
- "Irods" qui fonctionne en fédération basé sur postgres pour le stockage à froid
- Pas mal de Linux et LXC
- ESP8266
- Remontée de donnée par 4G et Lora, WiFi pour maintenance uniquement.

(mise en place d'une typologie "noeud producteur de données" dans le projet RUDI)

*Quelle organisation de travail pour atteindre cette production (le doc de métadonnées) ?

Suivi d'un ~process de normalisation :

version de base produite par une personne, Commentaire sur le Wiki, puis Réunion qui entérine ça

Confronté aux pratiques de GeoBretagne et du SIG Rennes Métropole (basés sur INSPIRE).

Travail avec Keolis par exemple :

- débats d'ordre technique, sur la complexité juridique ou les performance

- nombre de transactions obligatoire
- Partenaire de RUDI, donc impliqué de fait
- Producteurs de données intéressés par la démarche

Il reste une phase de négociation en 2nd volet : champ facultatif ou obligatoire.

- DOI obligatoire ? identifiant payant (similaire à NDD), enregistrement nécessaire > arbitrage de RUDI = ne pas mettre ce champ obligatoire.

Il s'agit des métadonnées du "jeu de données" présent dans le catalogue (et non pas de chaque donnée ensuite, qui elles ne sont pas normalisées ici).

Utile pour la publication

Déploiement d'un proxy par machine virtuelle

*Inspiration et références ?

Presque pas, plutôt de l'expérience et de l'échange en se rattrapant ce tout ce qui est déjà standard, et SIG

Géoportail

IFREMER par exemple sea data net qui gère vraiment bien la

- plein de producteurs de données (très hétérogènes)
- systèmes de normalisation très intéressant
 - choix des nom
 - notion de fédération
 - présentation
 - mais pas de données personnelles

SeaDataNet: <https://www.seadatanet.org/>

https://www.dropbox.com/s/453kfrgiazsrnbb/SDN_IRISA_20190326.pdf?dl=0

<http://aqmo.irisa.fr/>

<https://www.datashield.ac.uk/> (a vérifier)

*Qualité de service pour les opérateurs

Certains producteurs ont des obligations (ex. Keolis, c'est leur contrat), d'autres non, sont responsables de leur production

Des clauses d'engagement sont proposées, le producteur se positionne dedans

Enjeu des quotas d'accès aux données : souvent dépassés, il faut donc mettre une limitation (par exemple un seuil de gratuité). C'est le cas pour Domoticz sur DataConnect en ce moment.

*Gestion des Consentements

Éventuellement reprise de l'outil "Pulsar" de l'ULAQ au Canada : <https://pulsar.ca> ~15 homme-années

Pour les traitements/anonymisations :

Plugin de "R" - DataShield <http://opaldoc.obiba.org/en/latest/r-user-guide/datashield.html> /

<https://www.datashield.ac.uk/>

*Perspectives

Une fois l'API de RUDI stabilisée (d'ici quelques mois), il serait possible de tester la publication de données issues de Quotalise dans le catalogue RUDI.

À priori pour l'instant pas de soucis à lier une ressource quotalise si possible sous forme d'url