



Guillaume Pressiat, Département des Données de Santé, CHU de Brest



Introduction : La donnée nationale est obtenue de manière locale

Le PMSI constitue un système robuste qui permet aux ministères et aux agences du secteur de la santé de concevoir, d'adapter l'offre de soins et son financement.

Des moyens techniques conséquents sont mis en place au niveau national pour homogénéiser, structurer et collecter les données PMSI, les rassembler dans des bases de données (ATI, SNDS) accessibles au plus grand nombre.

Nous souhaitons aborder ces données avec une autre focale. D'où viennent-elles ? Qui en sont les producteurs ? De quels moyens disposent-ils ?

Nous proposons des outils communs de réutilisation des données PMSI pour les producteurs et autres acteurs locaux de la donnée dans les établissements.

Méthodes : Des données et des langages communs

Le format des données PMSI qui transitent mensuellement des établissements de santé vers les agences est un format ministériel commun. Nous nous appuyons sur la stabilité de ce format, arrêté annuellement : une forme d'interopérabilité des données.

Les langages informatiques et l'algorithmique avant eux permettent de traiter l'information de manière rationnelle et reproductible.

Des bibliothèques de fonctions écrites dans les langages R et Python ont été développées et partagées au niveau national pour lire les données et accéder à des méthodes et référentiels d'information médicale :

- Au DIM siège de l'AP-HP en 2017 pour R, au sein d'un groupe d'utilisateurs multi-établissements co-animé avec les médecins DIM
- Au DIM du CHU de Brest en 2023 pour Python, en lien avec le Centre de Données Cliniques (CDC)
- Au DIM du CHU de Brest fin 2023 pour Rust

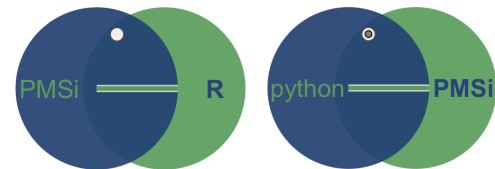
Le groupe de contributeurs s'est étoffé pour maintenir ces outils :

Auteurs : Guillaume Pressiat (CHU de Brest), Dr Namik Taright (AP-HP)

Contributeurs : Dr Blaise Tramier (CH du Pays d'Aix), Dr Joris Muller (CHRU de Strasbourg), Dr Alexandre Georges (HAD AP-HP), Dr Jan Chrusciel (CH Champagne Sud), Xavier Mouranche (AP-HP), Dr Rémi Flicoteaux (AP-HP)



Le langage R Le langage Python Le langage Rust



Dépendances : pmeasyr utilise la bibliothèque dplyr, pypmsi utilise la bibliothèque polars

Résultats : Une méthode commune

Ces bibliothèques permettent de partager une même approche pour :

- Lire les données : sans en questionner l'usage
- Partager des requêtes : pour partager les usages
- Trouver des atypies à corriger
- Valoriser l'activité comme e-PMSI
- Alimenter des entrepôts et lacs de données (décisionnel et recherche)
- Ancrer la donnée pour fiabiliser les résultats (reproductibilité)
- Partager des applications réutilisant ces données

Utiliser cette approche permet de mieux s'approprier les données en local (autonomie, compréhension, technicité, règles d'exceptions) y compris grâce à une documentation ouverte sur internet (voir le code à réponse rapide en haut à droite).

```
R : Lire et requêter les données
library(pmeasyr)
library(dplyr, warn.conflicts = FALSE)
library(dunnstam)
library(ggthemes)
p <- novau_pmeasyr(finess = "750712184",
  annee = 2022,
  mois = 12,
  path = "~/Documents/data/nco/",
  tolower_names = TRUE,
  lib = FALSE)

rsa <- lrs(p) %>% prepare_rsa

# chirurgie de l'obésité
chir_obesite <- get_all_listes("Chirurgie bariatrique")
sejours_chir_obesite <- lancer_requete(rsa, chir_obesite)

# recours exceptionnel
recours_exc <- get_all_listes("Recours Exceptionnel")
sejours_recours_exc <- lancer_requete(rsa, recours_exc)

# requête de base
exemple_requete <- list(
  actes = c("QEFAB03", "QEFAB05", "QEFAB10",
    "QEFAB13", "QEFAB15", "QEFAB19",
    "QEFAB28")
)

sejours_qef <- requete(rsa, exemple_requete)
```

```
python : lire les données et les convertir en fichier parquet
import polars as pl
import pyarrow as pa

p = pl.read_csv(
  files = "200000017",
  annee = 2021,
  mois = 12,
  path = "~/Documents/data/nco/"
)

# lire les données 2022
rsa = p.lrs(annee = 2022, typ1 = 4)
rsa[rsa[1] = Shape(141, 96)]

# tableau de données
| actifines | novras | cde_rsa | novras | rdas | n | dm |
|-----|-----|-----|-----|-----|-----|-----|
| 200000017 | 227 | ANCIEN FOR | 121 | 18 | 5 | 220027 |
| 200000017 | 227 | ANCIEN FOR | 121 | 118 | 98 | 180000 |
| 200000017 | 227 | ANCIEN FOR | 121 | 9 | 0 | 200007 |
| 200000017 | 227 | ANCIEN FOR | 121 | 6 | 6 | 240072 |
| 200000017 | 227 | ANCIEN FOR | 121 | 27 | 25 | 830074 |

# écrire les données en fichier parquet
for key, tab in rsa.items():
  tab.write_parquet(f"~/Documents/data/parquet/rsa/{key}.parquet")
```



Enquête utilisateurs (réalisée en août 2023) : 35 répondants dont :

- 75% médecins DIM / 25% statisticiens - techniciens
- 60% d'utilisateurs dans des CH, 25% en CHU et 15% en ESPIC
- 55% d'utilisateurs réguliers (au moins une fois par mois)

Quels usages ?

- 60% : indicateurs, analyses d'activités et requêtes diverses (files actives)
- 40% : amélioration du codage
- 33% : alimentation de bases de données



Conclusion : Du local au national

Les pratiques statistiques hétérogènes entre établissements et à l'échelle nationale conduisent à un éparpillement des connaissances, à des divergences de méthodes et de résultats, expliqués en partie par le recours fréquent à des logiciels payants. Pourtant ce sont bien les mêmes données qui sont étudiées, juste à des échelles différentes. Selon la loi pour une République numérique promulguée le 7 octobre 2016, il faudrait favoriser la « circulation des données et du savoir ».

La proposition portée par ces outils et leurs auteurs est de fédérer les établissements au-delà de la remontée des données en colligeant des méthodes et une documentation pour les métiers de la donnée PMSI à toutes les échelles du territoire (établissements, agences, ATI, SNDS).