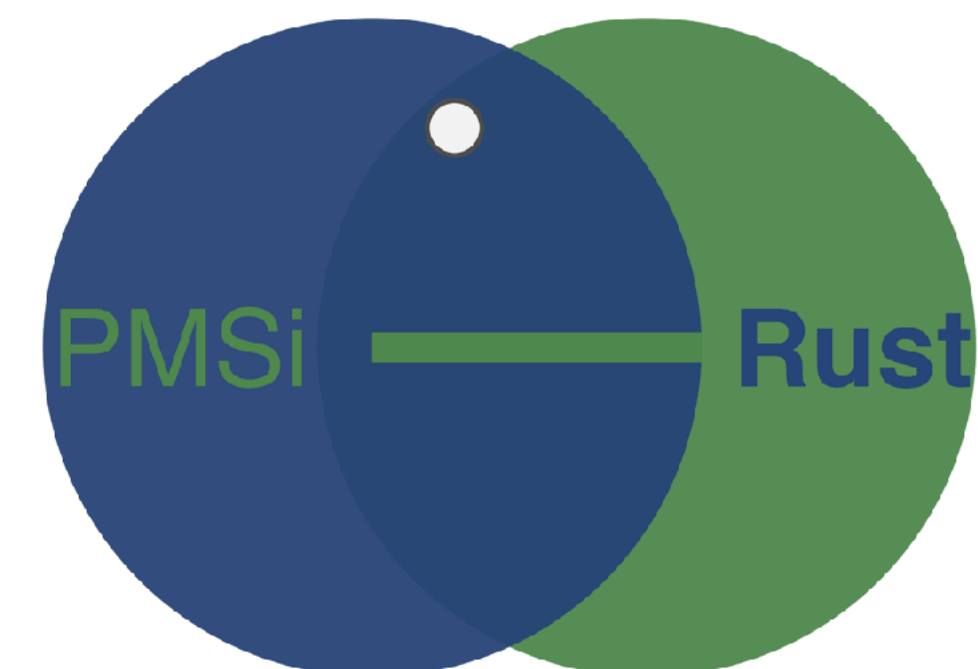
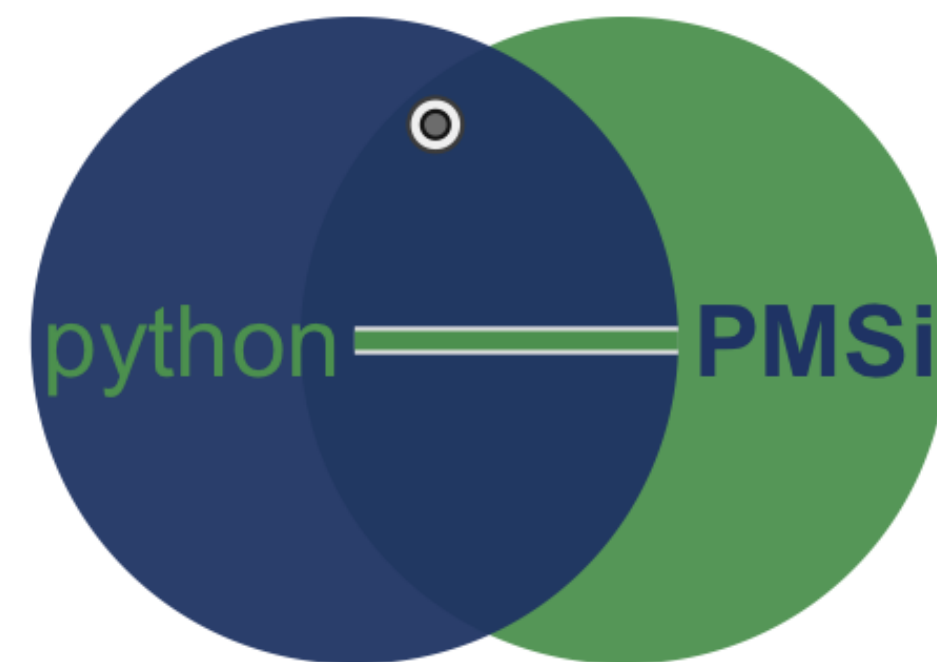
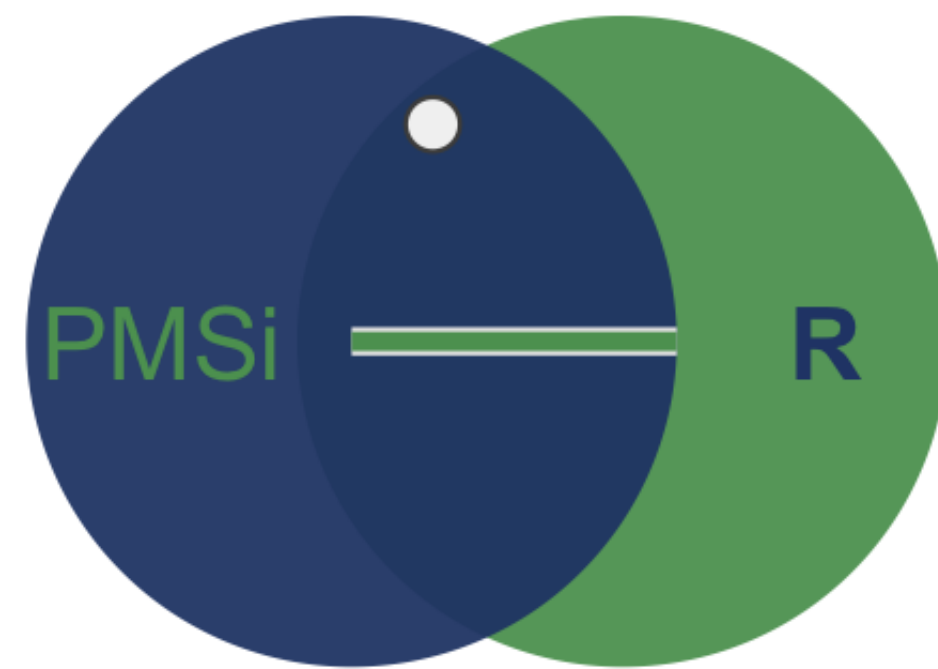


pmeasyr / pypmsi / pmsi-io

des communs en santé pour l'information  
médicale



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

# Introduction : Qui - quoi - où - qu'est-ce ?

- **Qui** : Guillaume Pressiat
- **Quoi** : statisticien
- **Où** : au département des données de santé du CHU de Brest
- DIM / Département d'*information médicale* : **Qu'est-ce ?**
  - *Données* qui résument l'activité et les typologies de soins réalisées dans un établissement de santé (actes, diagnostics, données démographiques, etc.), via le programme de médicalisation des systèmes d'information (PMSI)
  - Les DIM produisent, structurent et valident l'information médicale transmise à l'ATIH, aux ARS et à l'Assurance Maladie
  - Ces données alimentent les bases nationales de données de santé (PMSI, SNDS)
- Des données produites dans tous les hôpitaux du pays au même format

# Introduction : information médicale : qu'est-ce ?

## ● Exemples de codes diagnostics CIM-10

- E66 ~ obésité
- J45 ~ asthme
- C34 ~ cancer du poumon
- S72 ~ fracture du fémur
  
- E : métabolisme, endocrinien
- J : appareil respiratoire
- C : tumeurs
- S : fractures et traumatismes

## ● Exemples de codes actes CCAM

- NFKA006 : prothèse de genou
- NEKA020 : prothèse de hanche
  
- NF : genou
- NE : hanche
- K : retirer et remplacer
- A : par abord ouvert
  
- NFQK002 : radio du genou
  
- Q : observer
- K : par rayons X, radiographie

# Introduction : La donnée nationale est **obtenue** de manière locale

- Donnée / Obtenue | Data / Capta :
  - Dans les années 1990, les sociologues des sciences ont critiqué l'usage du mot « donnée » pour désigner des objets en réalité construits, arrachés au terrain au prix d'un temps et d'efforts parfois considérables. Le mot « obtenue » est notamment suggéré par Bruno Latour comme une alternative souhaitable.
  - Cf article de blog [ici](#)
- Perception : celui qui analyse la donnée ne voit plus forcément ce que voit celui qui l'a produite (*image : un producteur de blé dans une boulangerie ?*)
- La donnée concentre et cache la complexité des processus qui la produisent
- ▶ Intérêt de la pluri-disciplinarité autour des données (*on est tous dans le pétrin*)

## 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.
- Avantage : tous les DIM ont la même donnée de sortie (et tous les EDS cette même donnée d'entrée !), pour les quatre champs PMSI => **super pour partager des projets !**
- Les langages informatiques et l'algorithmique avant eux permettent de traiter l'information de manière rationnelle et reproductible.



Le langage R



Le langage  
Python



Le langage  
Rust

## Méthodes : Des données et des langages communs

- Des bibliothèques de fonctions écrites dans les langages R, Python et Rust 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

### Dépendances

- `pmeasyr` utilise la bibliothèque **`dplyr`**
- `pypmsi` et `pmsi_rust` utilisent la bibliothèque **`polars`**

# Méthodes : Des données et des langages communs

- 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)

## 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 (requête d'optimisation DIM, croisement PMSI et DPI, etc.)
  - 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'appropriier les données en local (autonomie, compréhension, technicité, règles d'exceptions)



# Résultats : Une méthode commune, deux exemples en image

```
R : Lire et requêter les données

library(pmeasyr)
library(dplyr, warn.conflicts = FALSE)
library(nomensland)
p ← noyau_pmeasyr(finess = '750712184',
                 annee = 2022,
                 mois = 12,
                 path = '~/Documents/data/mco/',
                 tolower_names = TRUE,
                 lib = FALSE)

rsa ← irsa(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 ad hoc
exemple_requete ← list(
  actes = c('QEFA003', 'QEFA005', 'QEFA010',
            'QEFA013', 'QEFA015', 'QEFA019',
            'QEFA020')
)

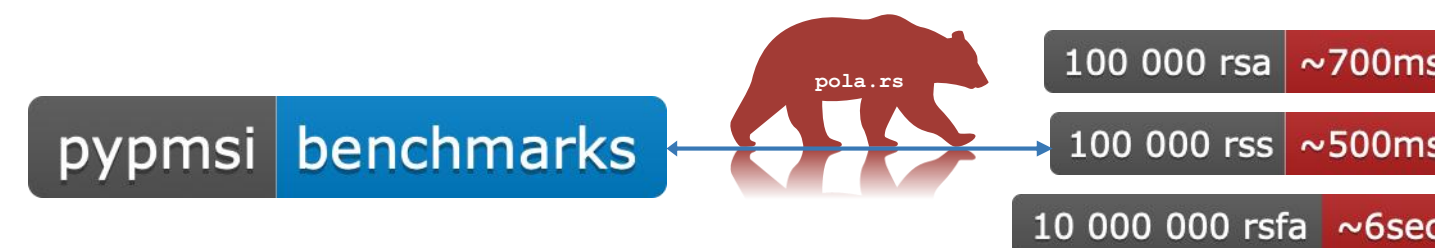
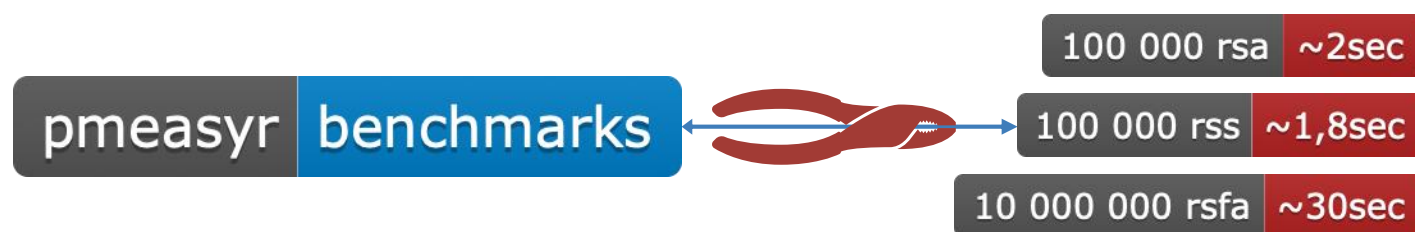
sejours_qefa ← requete(rsa, exemple_requete)
```

```
python : Lire les données et les convertir en fichier parquet

import polars as pl
import pypmsi as pm

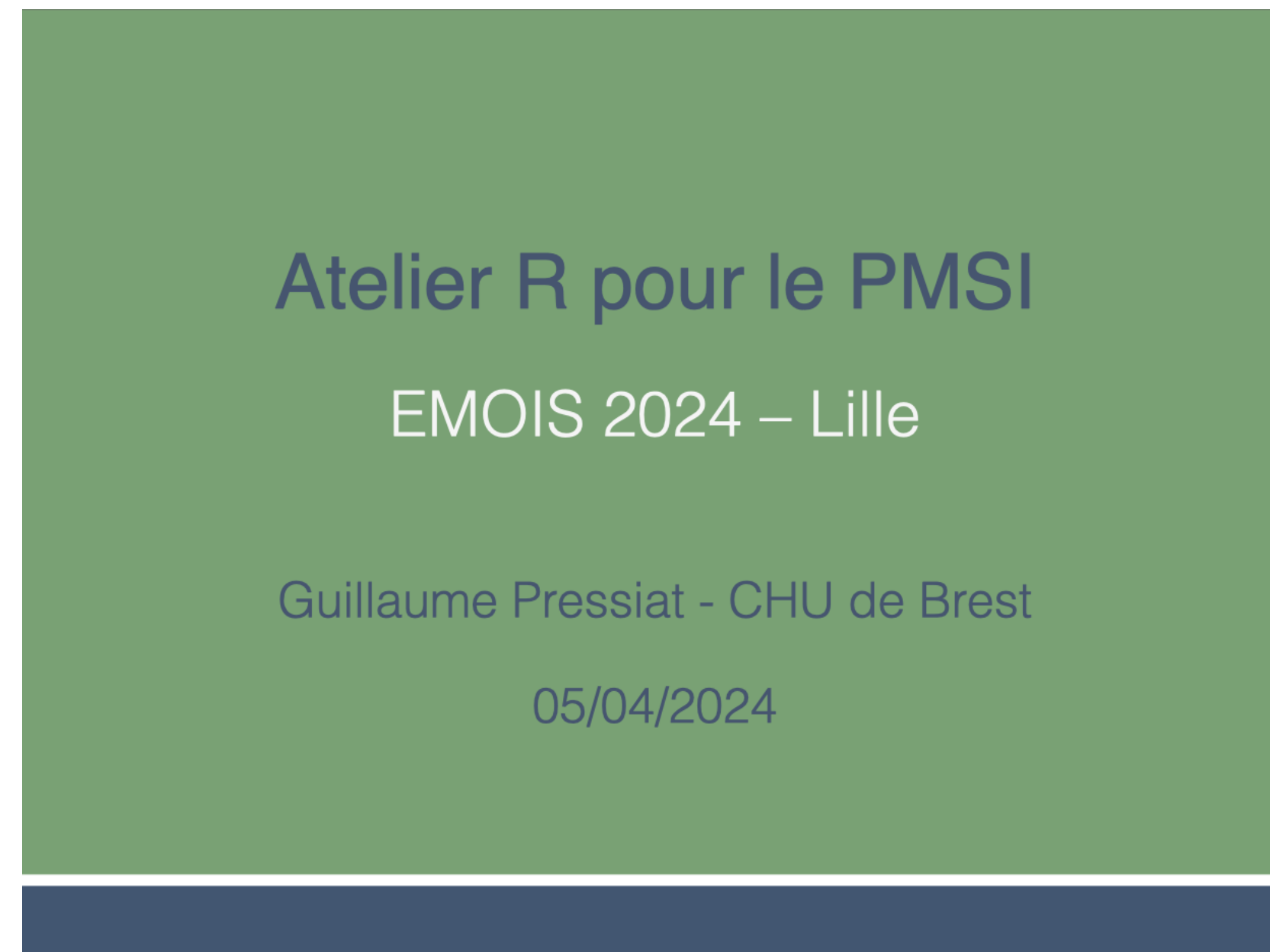
p = pm.noyau_pmsi(
    finess = '290000017',
    annee = 2021,
    mois = 12,
    path = '~/Documents/data/mco')

# Lire les données 2022
rsa = p.irsa(annee = 2022, typi = 4)
# rsa['rsa'] : Shape(141_901, 89)
#
# | nofiness | novrsa | cle_rsa | novrss | ... | ndas | na | ghm |
# |---|---|---|---|---|---|---|---|
# | str | str | str | str | ... | i32 | i32 | str |
# | 290000017 | 227 | ABCDEFGHI | 121 | ... | 10 | 2 | 23Z02Z |
# | 290000017 | 227 | ABCDEFGHI | 121 | ... | 113 | 90 | 10C024 |
# | ... | ... | ... | ... | ... | ... | ... | ... |
# | 290000017 | 227 | ABCDEFGHI | 121 | ... | 0 | 2 | 23M20T |
# | 290000017 | 227 | ABCDEFGHI | 121 | ... | 2 | 0 | 28Z07Z |
# | 290000017 | 227 | ABCDEFGHI | 121 | ... | 27 | 25 | 03M074 |
#
# écrire les données en fichier parquet
for key, tab in rsa.items():
    tab.write_parquet("~/Documents/data/parquet/rsa/" +
                     key + ".parquet")
```



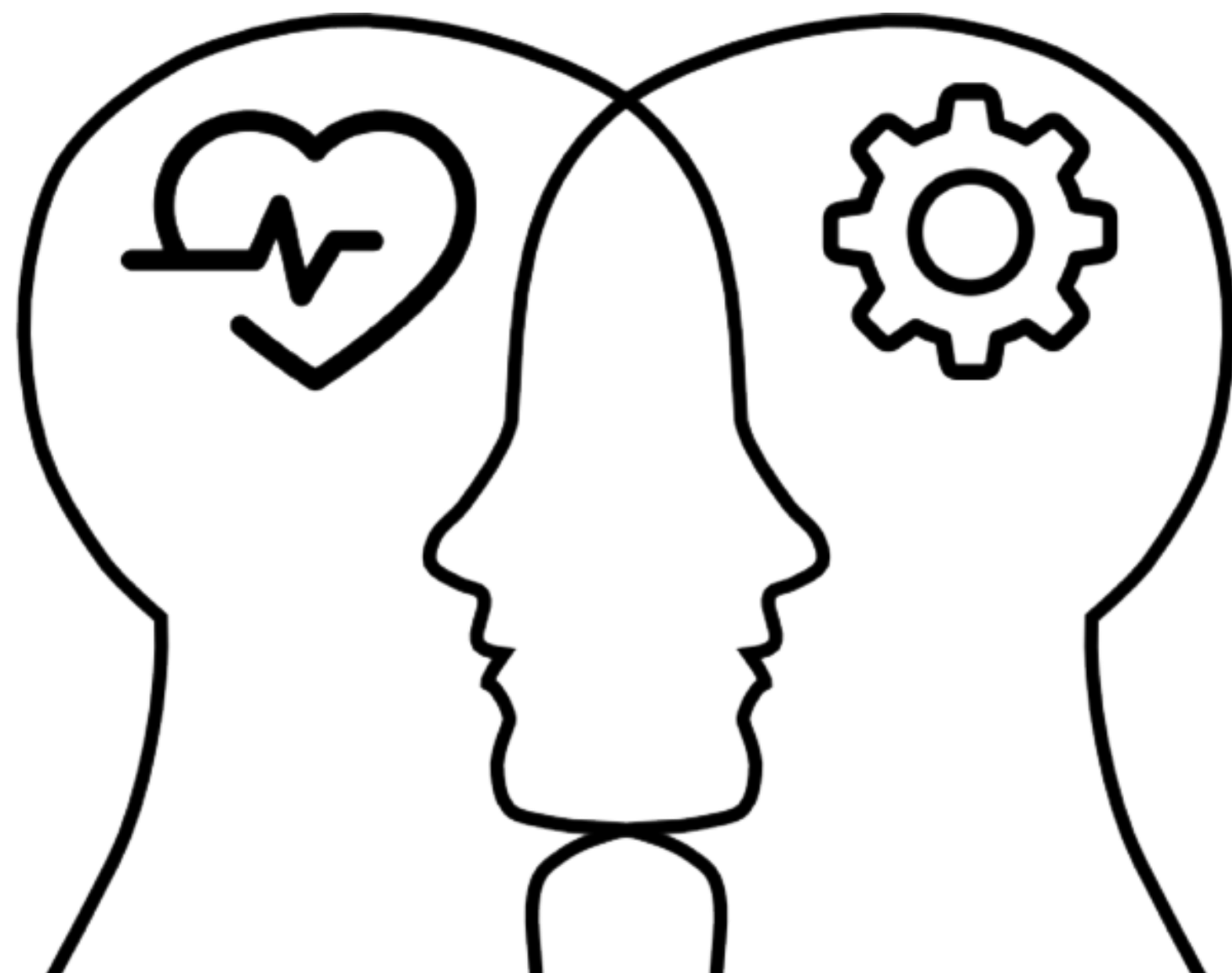
# Atelier au congrès EMOIS 2024

- Atelier R pour le PMSI lors du congrès EMOIS 2024 : reproduire des tableaux e-PMSI avec R et exemples de requêtes avancées : <https://guillaumepressiat.github.io/emois-2024/>



# Datathon InterHop 2024

- Datathon **InterHop** 2024 : indicateurs périnatalité avec OMOP, SQL et Python (ou R)
- <https://interhop.org/2024/08/31/maternite>
- InterHop : Association qui promeut l'interopérabilité et l'open source en santé
- ETL PMSI vers OMOP Common Data Model réalisé sur la plate-forme data science d'InterHop
- Production d'un premier dashboard synthétique d'indicateurs



## Indicateurs périnat à partir d'OMOP



Code ▾

Code

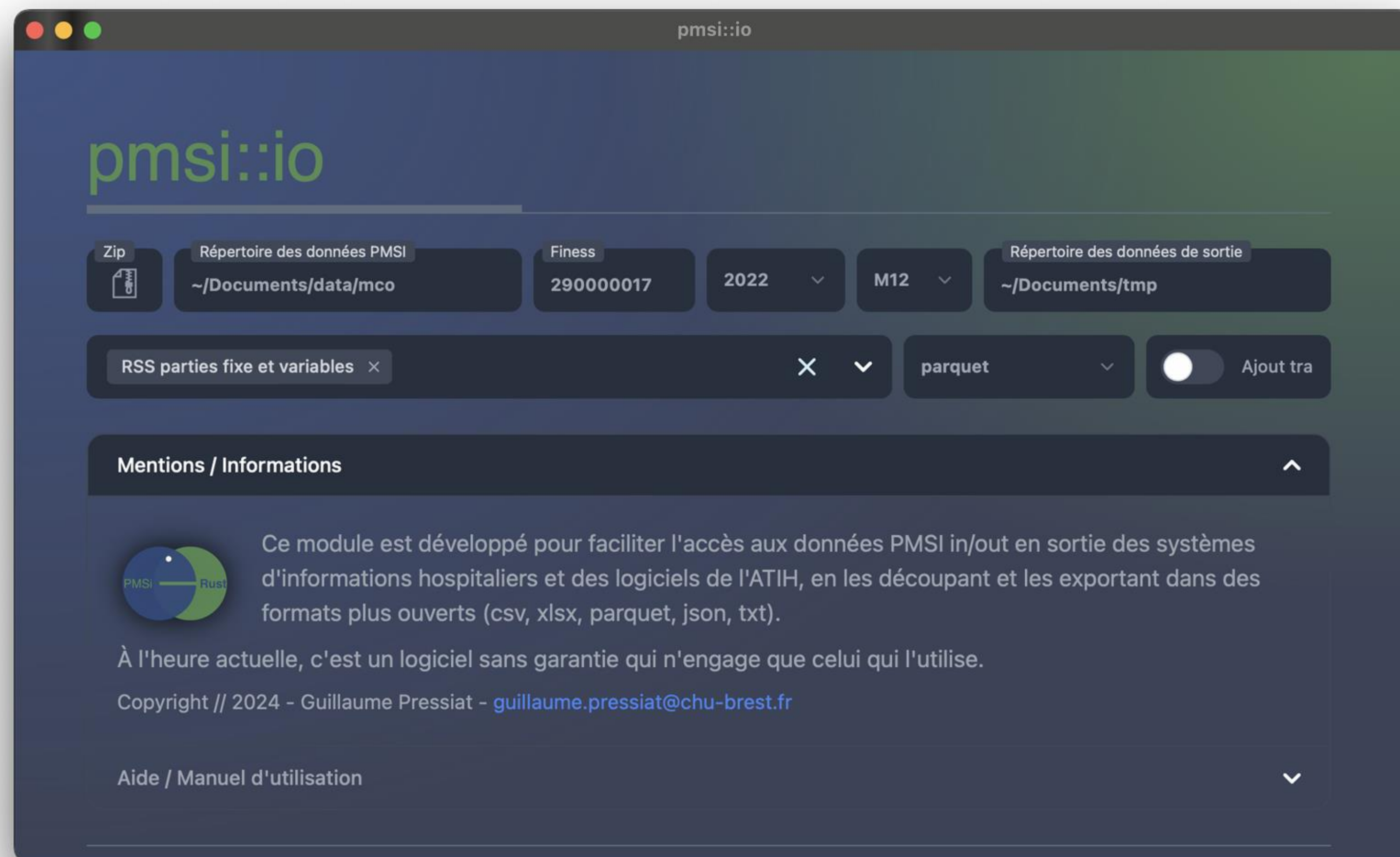
Code

Code

Code

# Nouveauté 2024 : pmsi::io

- Logiciel libre
- Objectif : Rendre accessible la donnée aux professionnels DIM qui ne maîtrisent pas la programmation
- Fonctionnalités : Lire et convertir les données PMSI in/out dans des formats plus ouverts (csv, parquet, xlsx, txt)
- Aspect technique : codé en Rust avec une bibliothèque attenante
- <https://github.com/GuillaumePressiat/pmsi-io>



# Enquête utilisateurs

## Enquête utilisateurs (réalisée en août 2023) :

- 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

- **pmsi-io : plus de 200 téléchargements**

- **Avantages**

- Les hôpitaux partagent le même format de données : autant partager les outils autour de ces données
- Utiliser R et Python au quotidien permet de diversifier les composantes de son métier
- Effet de synergie dans les équipes, pour recruter
- Écrire du code participe à s'émanciper de contraintes et à avoir un espace de liberté, presque sans limites !

- **Limites à dépasser**

- Chaque individu est unique et a sa propre manière de chercher à être « libre » dans son travail
- Les contraintes des structures empêchent parfois cette liberté (isolement, continuité de services, formation)
- La donnée et les algorithmes peuvent être communs et interopérables mais l'humain reste un être complexe politiquement

## 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. 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 de santé à toutes les échelles du territoire (établissements, agences, ATIH, SNDS).
- Comment faire réseau pour relier les acteurs de la donnée de santé à l'échelle du territoire ? Comment trouver un cadre socio-technique, pluri-disciplinaire, tenant compte de la réalité des contextes métiers, qui puisse relier producteurs et utilisateurs de données : s'inspirer d'un concept comme le data-mesh ?

merci de votre attention

- Pour plus d'informations, vous pouvez vous rendre sur cette page web
- <https://guillaumepressiat.github.io/>

