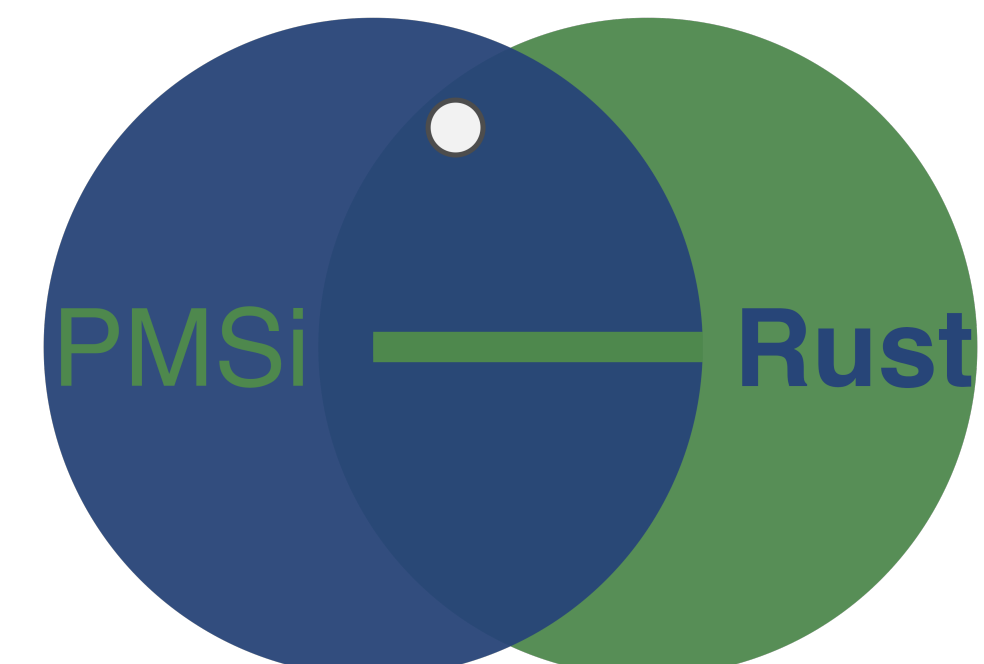
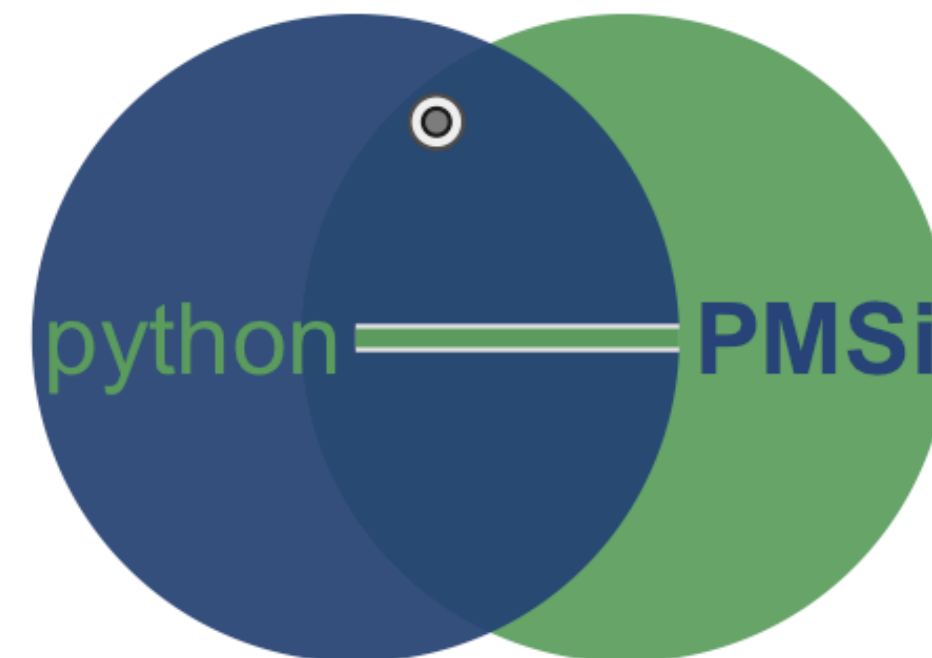
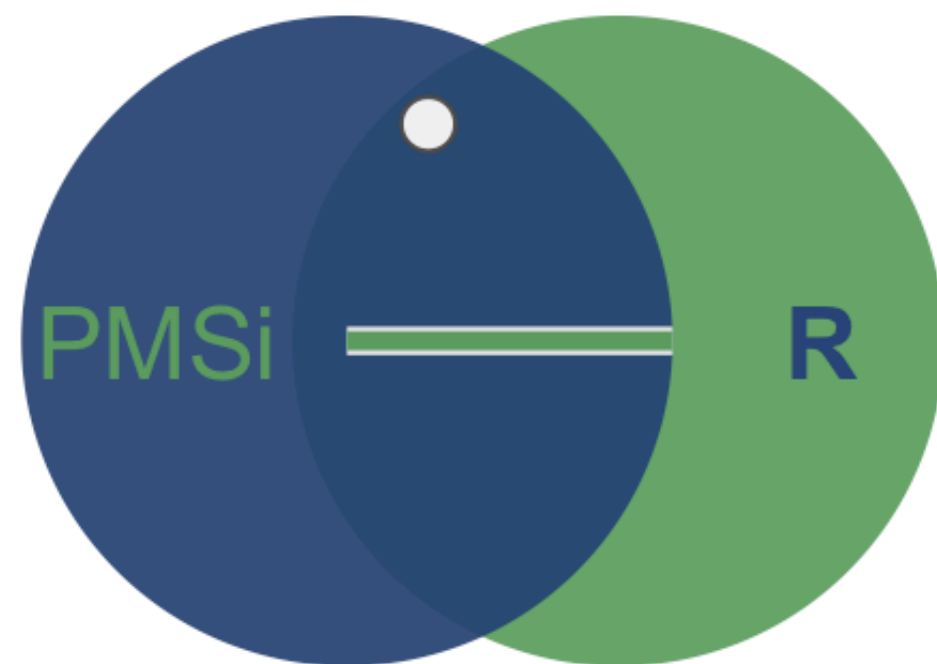


# pmeasyr / pypmsi

des communs en santé pour l'information  
médicale des établissements  
*et plus si affinités*



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

**Liminaire** : La donnée nationale est **obtenue** de manière locale

**Liminaire** : La donnée nationale est obtenue de manière locale

- Donnée / Obtenue | Data / Capta

**Liminaire** : La donnée nationale est obtenue de manière locale

## ● Donnée / Obtenue | Data / Capta

- ❖ « La tentation de l'idéalisme vient peut-être du mot même de données qui décrit aussi mal que possible ce sur quoi s'appliquent les capacités cognitives ordinaires des érudits, des savants et des intellectuels. Il faudrait remplacer ce terme par celui, beaucoup plus réaliste, d'obtenues et parler par conséquent de bases d'obtenues, de sublata plutôt que de data. » - Latour, B., *Pensée retenue, pensée distribuée*, 2007

**Liminaire** : La donnée nationale est obtenue de manière locale

## ● Donnée / Obtenue | Data / Capta

- ❖ *« La tentation de l'idéalisme vient peut-être du mot même de données qui décrit aussi mal que possible ce sur quoi s'appliquent les capacités cognitives ordinaires des érudits, des savants et des intellectuels. Il faudrait remplacer ce terme par celui, beaucoup plus réaliste, d'obtenues et parler par conséquent de bases d'obtenues, de sublata plutôt que de data. » - Latour, B., Pensée retenue, pensée distribuée, 2007*
- ❖ *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.*

# Liminaire : La donnée nationale est obtenue de manière locale

## ● Donnée / Obtenue | Data / Capta

- ❖ « La tentation de l'idéalisme vient peut-être du mot même de données qui décrit aussi mal que possible ce sur quoi s'appliquent les capacités cognitives ordinaires des érudits, des savants et des intellectuels. Il faudrait remplacer ce terme par celui, beaucoup plus réaliste, d'obtenues et parler par conséquent de bases d'obtenues, de sublata plutôt que de data. » - Latour, B., *Pensée retenue, pensée distribuée*, 2007
- ❖ 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.
- ❖ « Croire que les données sont intrinsèquement quantitatives – évidentes, neutres sur le plan des valeurs et indépendantes de l'observateur – exclut la possibilité de les concevoir comme qualitatives, co-dépendamment constituées – en d'autres termes, de reconnaître que data sont des captas [...] Je suggère que nous repensions fondamentalement les données comme des captas en termes d'ambiguïté plutôt que de certitude, et que nous trouvions des moyens d'exprimer graphiquement la complexité de l'interprétation. » - Drucker J., *Humanities Approaches to Graphical Display*



**Liminaire** : La donnée nationale est obtenue de manière locale

**Liminaire** : La donnée nationale est obtenue de manière locale

- La donnée et le kaléïdoscope : l'endroit de nos subjectivités



**Liminaire** : La donnée nationale est obtenue de manière locale

- La donnée et le kaléïdoscope : l'endroit de nos subjectivités
- Perception : celui qui regarde la donnée ne voit plus forcément ce que voit celui qui l'a produite

**Liminaire** : La donnée nationale est obtenue de manière locale

- La donnée et le kaléïdoscope : l'endroit de nos subjectivités
- Perception : celui qui regarde la donnée ne voit plus forcément ce que voit celui qui l'a produite
- La donnée concentre et cache la complexité des processus qui la produisent  
=> intérêt de la pluri-disciplinarité autour des données

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

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

- Contexte : le Programme de Médicalisation des Systèmes d'Information

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

- Contexte : le Programme de Médicalisation des Systèmes d'Information
  - CIM-10, CCAM, GHM, CSARR, GME, données médico-administratives, etc., **figées**

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

- Contexte : le Programme de Médicalisation des Systèmes d'Information
  - CIM-10, CCAM, GHM, CSARR, GME, données médico-administratives, etc., **figées**
- 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.

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

- Contexte : le Programme de Médicalisation des Systèmes d'Information
  - CIM-10, CCAM, GHM, CSARR, GME, données médico-administratives, etc., **figées**
- 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 (ATIH, SNDS) accessibles au plus grand nombre.



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

- Contexte : le Programme de Médicalisation des Systèmes d'Information
  - CIM-10, CCAM, GHM, CSARR, GME, données médico-administratives, etc., **figées**
- 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 (ATIH, 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 ?

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

- Contexte : le Programme de Médicalisation des Systèmes d'Information
  - CIM-10, CCAM, GHM, CSARR, GME, données médico-administratives, etc., **figées**
- 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 (ATIH, 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

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

- Contexte : le Programme de Médicalisation des Systèmes d'Information
  - CIM-10, CCAM, GHM, CSARR, GME, données médico-administratives, etc., **figées**
- 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 (ATIH, 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
- **les DIM et les EDS en particulier !**

# Méthodes : Des données et des langages communs

## 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.

## 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 EDS ont la même donnée d'entrée, pour les quatre champs PMSI => super pour partager des projets !**

## 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 EDS ont la même donnée d'entrée, pour les quatre champs PMSI => super pour partager des projets !**
- **Le lien EDS - DIM : un échange fructueux**



## 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 EDS ont la même donnée d'entrée, pour les quatre champs PMSI => super pour partager des projets !**
- **Le lien EDS - DIM : un échange fructueux**
- Les langages informatiques et l'algorithmique avant eux permettent de traiter l'information de manière rationnelle et reproductible.

## 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 EDS ont la même donnée d'entrée, pour les quatre champs PMSI => super pour partager des projets !**
- **Le lien EDS - DIM : un échange fructueux**
- Les langages informatiques et l'algorithmique avant eux permettent de traiter l'information de manière rationnelle et reproductible.

## 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 EDS ont la même donnée d'entrée, pour les quatre champs PMSI => super pour partager des projets !**
- **Le lien EDS - DIM : un échange fructueux**
- 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

## Méthodes : Des données et des langages communs

- 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 :

## Méthodes : Des données et des langages communs

- 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

## Méthodes : Des données et des langages communs

- 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)**



## Méthodes : Des données et des langages communs

- 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

## Méthodes : Des données et des langages communs

- 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

### Dépendances

- 👁 `pmeasyr` utilise la bibliothèque **`dplyr`**
- 👁 `pypmsi` et `pmsi_rust` utilisent la bibliothèque **`pola.rs`**

# 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

## Résultats : Une méthode commune

- Ces bibliothèques permettent de partager une même approche pour :

## Résultats : Une méthode commune

- Ces bibliothèques permettent de partager une même approche pour :
  - **Lire les données : sans en questionner l'usage**



## 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

## 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

## 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

## 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)**

## 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)

## 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

## 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)

# 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)

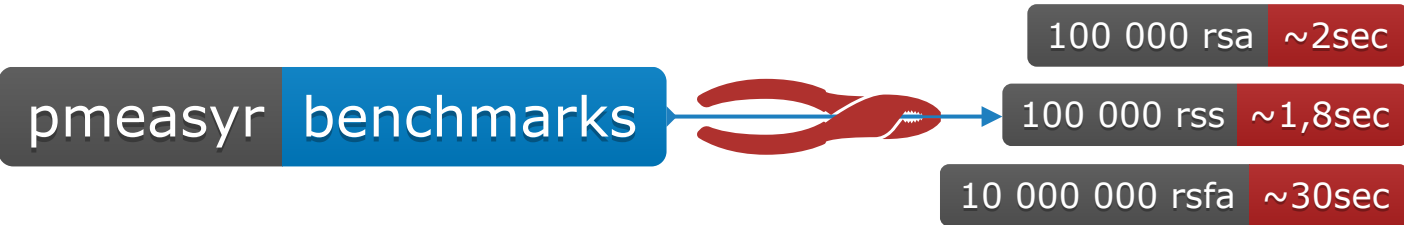
rsa <- irsa(p) %>% prepare_rsa

# 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

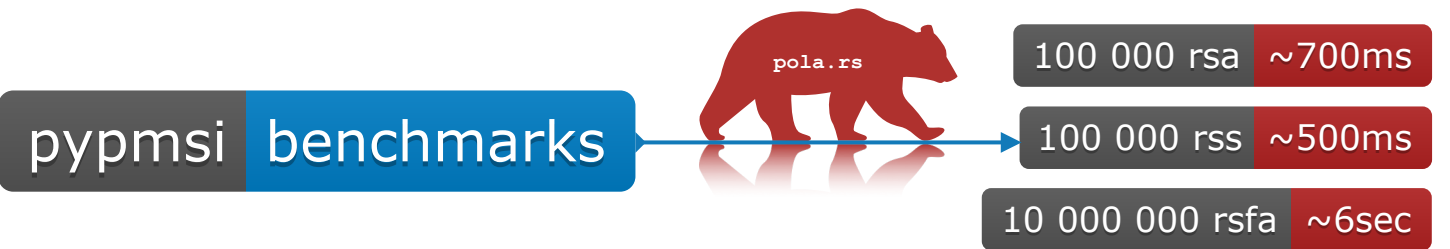
import polars as pl
import pypmsi as pm

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")
```





**Conclusion** : Du local au national

## 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* ».

## 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, ATIH, SNDS).

# **Discussion** : avantages et limites de la donnée PMSI

# Discussion : avantages et limites de la donnée PMSI

- Avantages

## **Discussion** : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité

## Discussion : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale

## **Discussion** : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale
  - Des données communes à tous les établissements de santé



## **Discussion** : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale
  - Des données communes à tous les établissements de santé
- Limites

## Discussion : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale
  - Des données communes à tous les établissements de santé
- Limites
  - Une donnée moins détaillée que celle du SIH

## Discussion : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale
  - Des données communes à tous les établissements de santé
- Limites
  - Une donnée moins détaillée que celle du SIH
  - Pas de détail par UF

## Discussion : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale
  - Des données communes à tous les établissements de santé
- Limites
  - Une donnée moins détaillée que celle du SIH
  - Pas de détail par UF
  - Pas toutes les données : pas de documents

# Discussion : avantages et limites de la donnée PMSI

- Avantages
  - Des données figées : facilite la reproductibilité
  - Des données cohérentes avec la base nationale
  - Des données communes à tous les établissements de santé
- Limites
  - Une donnée moins détaillée que celle du SIH
  - Pas de détail par UF
  - Pas toutes les données : pas de documents
  - Pas de détails sur les intervenants

# Enquête utilisateurs

## Enquête utilisateurs

# Enquête utilisateurs

**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)

# Enquête utilisateurs

**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**



merci de votre attention

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

