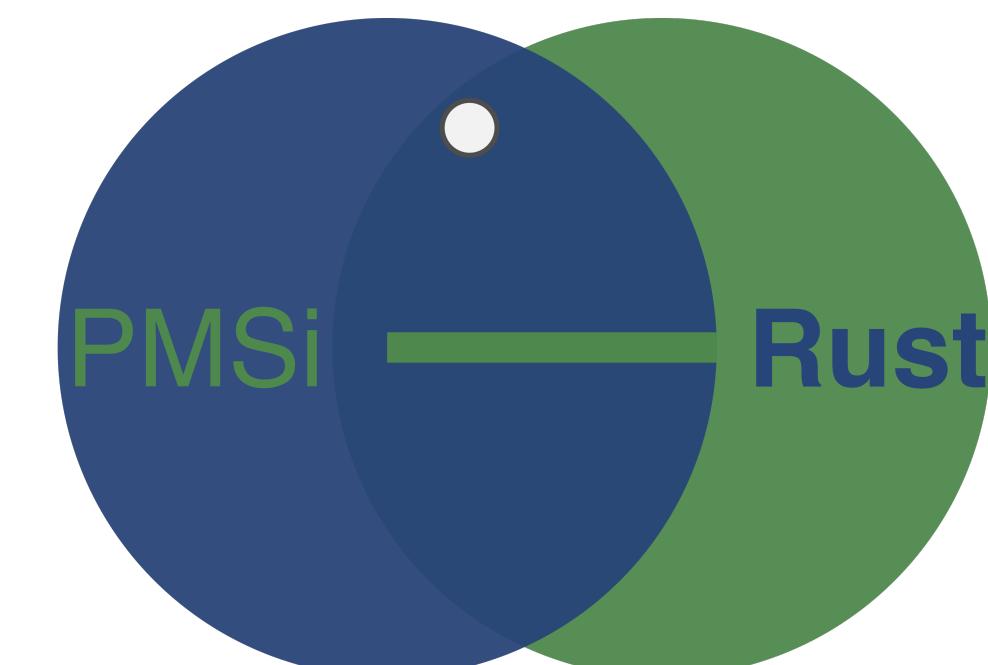
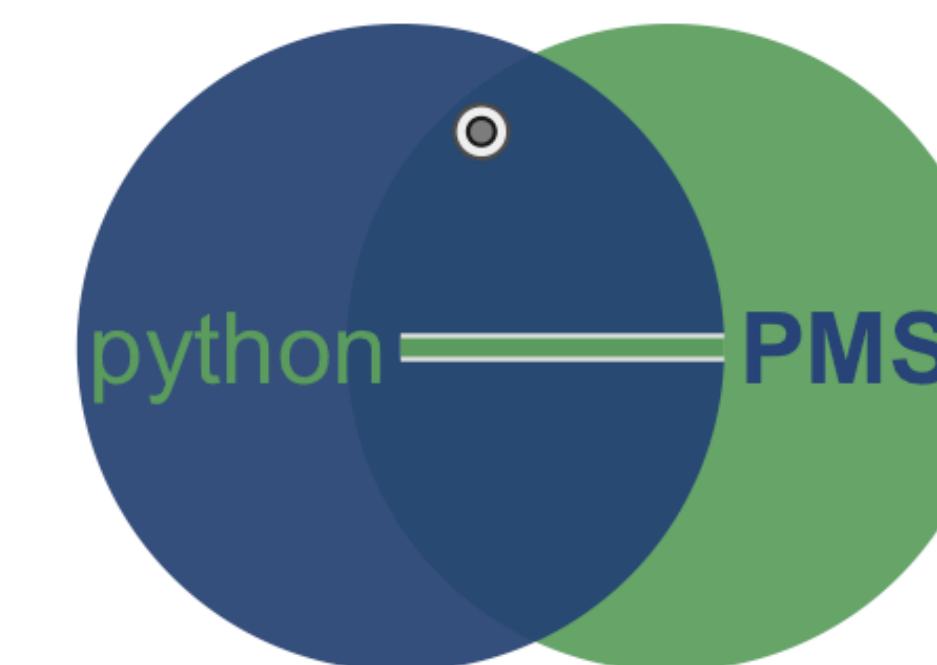
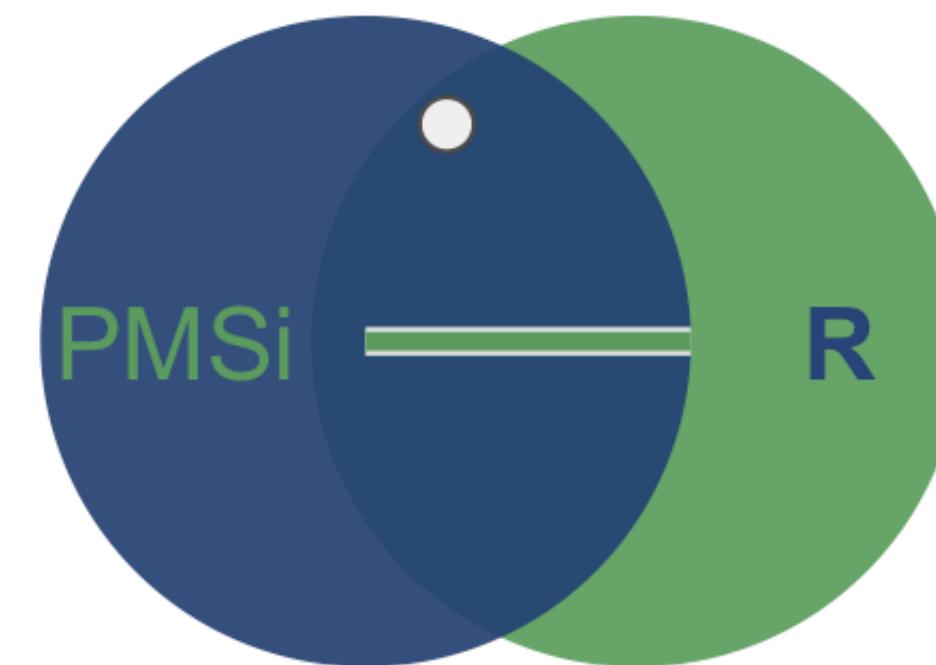
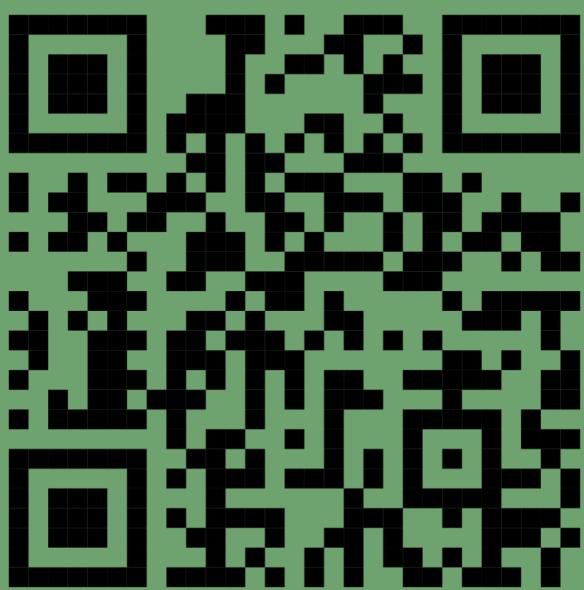


# pmeasyr / pypmsi / pmsi-io

des communs en santé pour l'information  
médicale



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

# Introduction : Des communs, pourquoi ?

## Introduction : Des communs, pourquoi ?

- Le monde des DIM en deux composantes :

## Introduction : Des communs, pourquoi ?

- Le monde des DIM en deux composantes :
  - un réseau de cerveaux

# Introduction : Des communs, pourquoi ?

- Le monde des DIM en deux composantes :

- un réseau de cerveaux
  - Abimes
  - CoCoA
  - Blogs
  - CNIM
  - etc.

# Introduction : Des communs, pourquoi ?

- Le monde des DIM en deux composantes :
  - un réseau de cerveaux
    - Abimes
    - CoCoA
    - Blogs
    - CNIM
    - etc.
  - un archipel de systèmes d'informations

# Introduction : Des communs, pourquoi ?

- Le monde des DIM en deux composantes :
  - un réseau de cerveaux
    - Abimes
    - CoCoA
    - Blogs
    - CNIM
    - etc.
  - un archipel de systèmes d'informations
    - Facile d'être à la merci des éditeurs
    - Difficile de partager des requêtes et de s'entraider

## Introduction : Des communs, pourquoi ?

- Le monde des DIM en deux composantes :
  - un réseau de cerveaux
    - Abimes
    - CoCoA
    - Blogs
    - CNIM
    - etc.
  - un archipel de systèmes d'informations
    - Facile d'être à la merci des éditeurs
    - Difficile de partager des requêtes et de s'entraider
- Difficile de mettre en commun en pratique, chaque DIM aime faire sa propre cuisine

# Introduction : Des communs, pourquoi ?

# Introduction : Des communs, pourquoi ?

- Le contexte national des données de santé

- Le contexte national des données de santé

- La donnée PMSI est très utile / utilisée au national (ATIH, SNDS, HDH)
- les entrepôts de données de santé présentent de nouvelles promesses
- sans vraiment que les DIM soient forcément associés ?

- Le contexte national des données de santé
  - La donnée PMSI est très utile / utilisée au national (ATIH, SNDS, HDH)
  - les entrepôts de données de santé présentent de nouvelles promesses
  - sans vraiment que les DIM soient forcément associés ?
- Les DIM ont pourtant bien contribué à collecter de la donnée !

- Le contexte national des données de santé
  - La donnée PMSI est très utile / utilisée au national (ATIH, SNDS, HDH)
  - les entrepôts de données de santé présentent de nouvelles promesses
  - sans vraiment que les DIM soient forcément associés ?
- Les DIM ont pourtant bien contribué à collecter de la donnée !
- Quelle maille trouver pour rendre « interopérables » les DIM d'établissements entre eux, donner de la visibilité à leurs travaux et expertises, attirer de nouvelles compétences « data science », connecter leurs connaissances à d'autres acteurs ?

## Introduction : Des communs, pourquoi ?

- Le contexte national des données de santé
  - La donnée PMSI est très utile / utilisée au national (ATIH, SNDS, HDH)
  - les entrepôts de données de santé présentent de nouvelles promesses
  - sans vraiment que les DIM soient forcément associés ?
- Les DIM ont pourtant bien contribué à collecter de la donnée !
- Quelle maille trouver pour rendre « interopérables » les DIM d'établissements entre eux, donner de la visibilité à leurs travaux et expertises, attirer de nouvelles compétences « data science », connecter leurs connaissances à d'autres acteurs ?
- Mettre en commun pour être plus forts

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

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

- Donnée / Obtenu | Data / Capta

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

## ● Donnée / Obtenu | 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.

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

- **Donnée / Obtenu | 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.

- **Perception : celui qui regarde la donnée ne voit plus forcément ce que voit celui qui l'a produite**

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

- Donnée / Obtenu | 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.

- 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

# 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 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 !**

## 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'algorithme 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 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'algorithme 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 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'algorithme avant eux permettent de traiter l'information de manière rationnelle et reproductible.



# Méthodes : Des données et des langages communs

## 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 :

## 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

## 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)

## 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

## 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 **polaris**

# Méthodes : Des données et des langages communs

## 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 (requête d'optimisation DIM, croisement PMSI et DPI, etc.)

## 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

## 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)

## 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)

## 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

## 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
  - Partager la gestion des rejets de Druides

## 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
  - Partager la gestion des rejets de Druides
- 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'obésite
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)
```

pmeasyr benchmarks



100 000 rsa	~2sec
100 000 rss	~1,8sec
10 000 000 rsfa	~30sec

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

import polars as pl
import pypmsi as pm

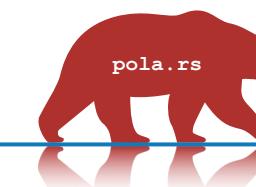
p = pm.noyau_pmsi(
    finess = '290000017',
    annee = 2021,
    mois = 12,
    path = '~/Documents/data/mco')

# 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	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")
```

pypmsi benchmarks



100 000 rsa	~700ms
100 000 rss	~500ms
10 000 000 rsfa	~6sec

# Cas pratique

## Cas pratique

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

## Cas pratique

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



## **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

## **Discussion** : avantages et limites

- Avantages

## **Discussion** : avantages et limites

- Avantages

- Des données figées : facilite la reproductibilité

## **Discussion** : avantages et limites

- Avantages

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

## **Discussion** : avantages et limites

### ● Avantages

- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé

## Discussion : avantages et limites

### ● Avantages

- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé
- => **Un socle commun de dialogue pour un réseau de DIM**

## Discussion : avantages et limites

- Avantages

- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé
- => **Un socle commun de dialogue pour un réseau de DIM**

- Limites

## Discussion : avantages et limites

- Avantages

- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé
- => **Un socle commun de dialogue pour un réseau de DIM**

- Limites

- Une donnée moins détaillée que celle du SIH

## Discussion : avantages et limites

- Avantages

- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé
- => **Un socle commun de dialogue pour un réseau de DIM**

- Limites

- Une donnée moins détaillée que celle du SIH
- Pas de détail par UF

## Discussion : avantages et limites

- Avantages

- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé
- => **Un socle commun de dialogue pour un réseau de DIM**

- 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

- Avantages

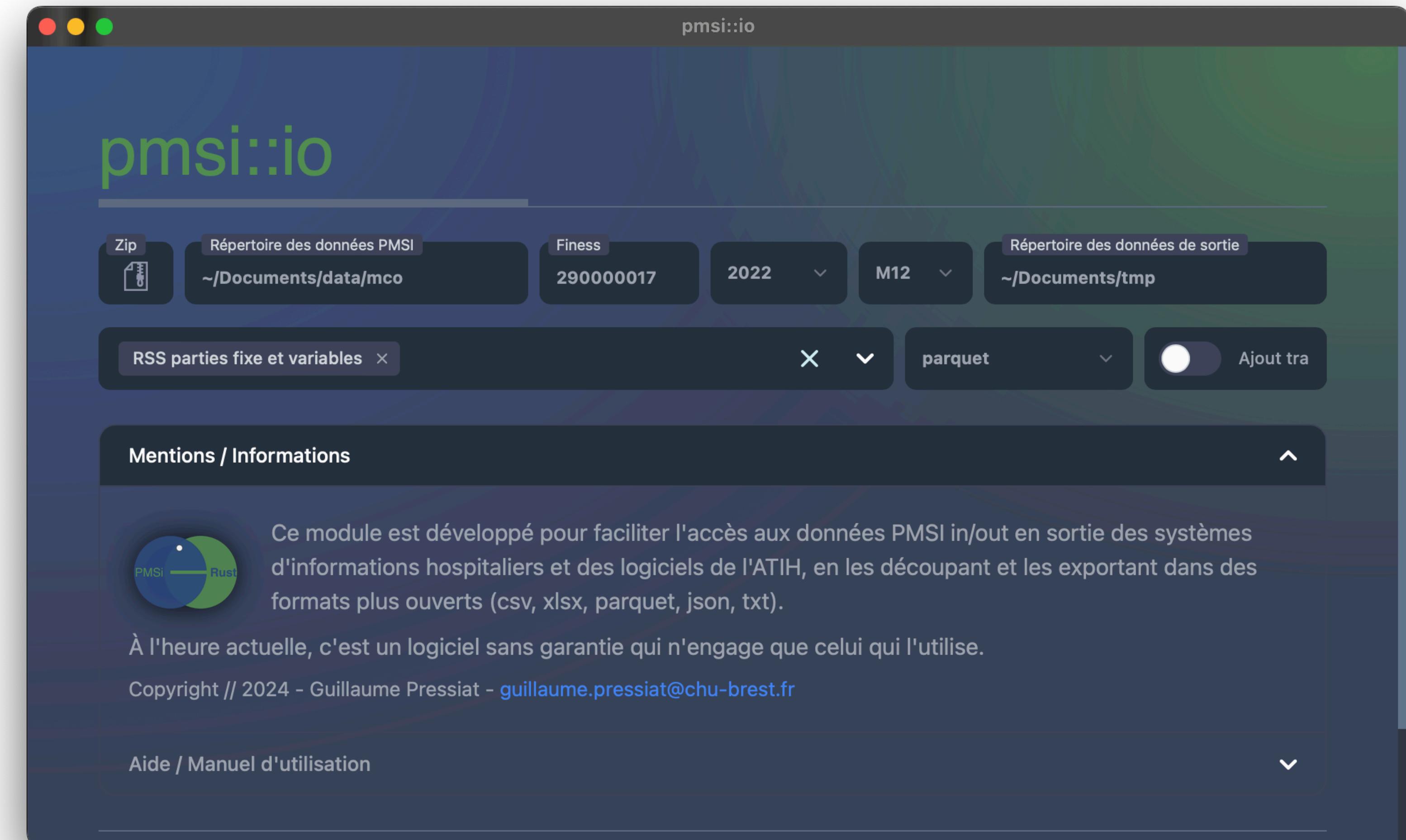
- Des données figées : facilite la reproductibilité
- Des données cohérentes avec la base nationale
- Un format de données commun à tous les établissements de santé
- => **Un socle commun de dialogue pour un réseau de DIM**

- 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**

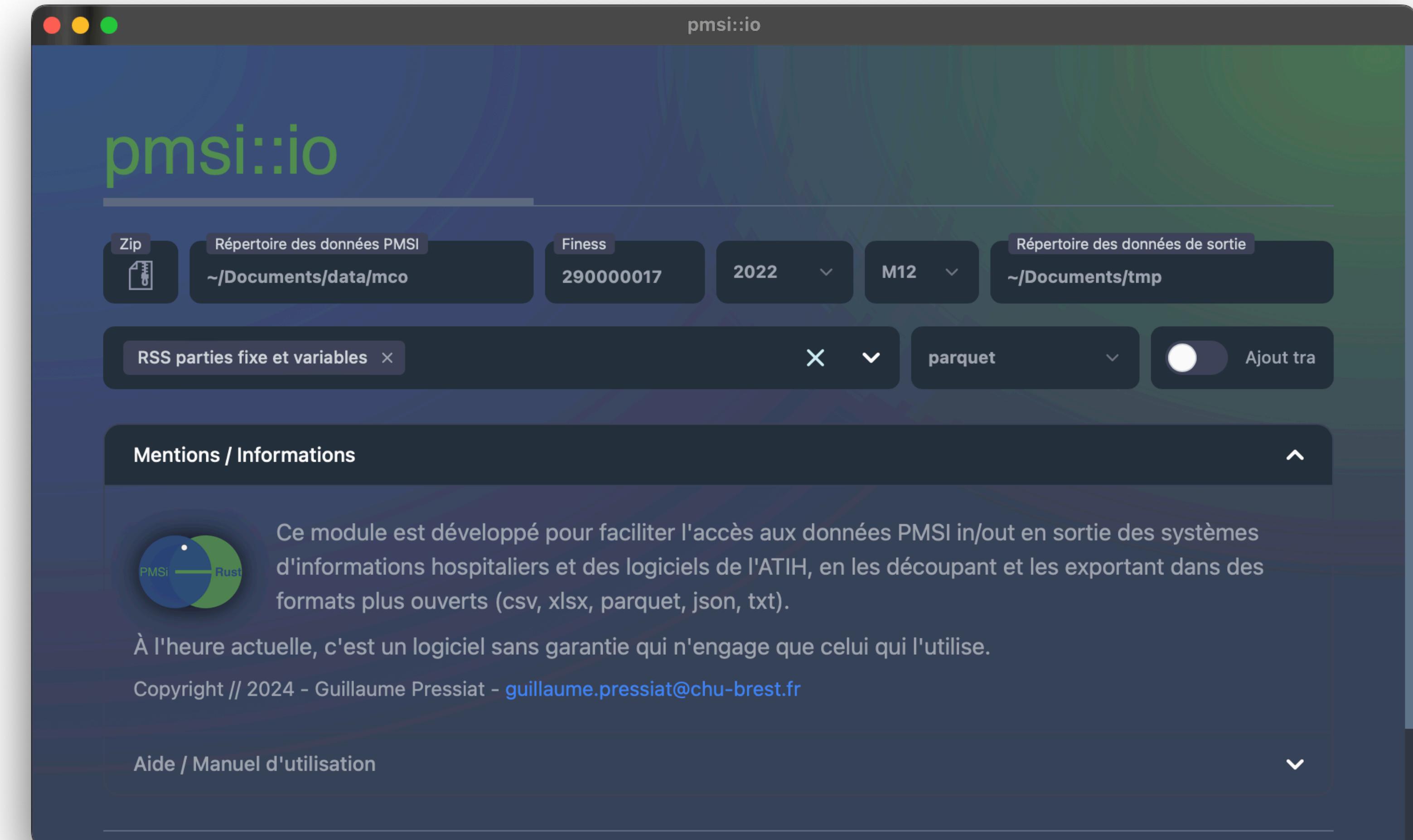
# Nouveauté 2024 : pmsi::io

# Nouveauté 2024 : pmsi::io



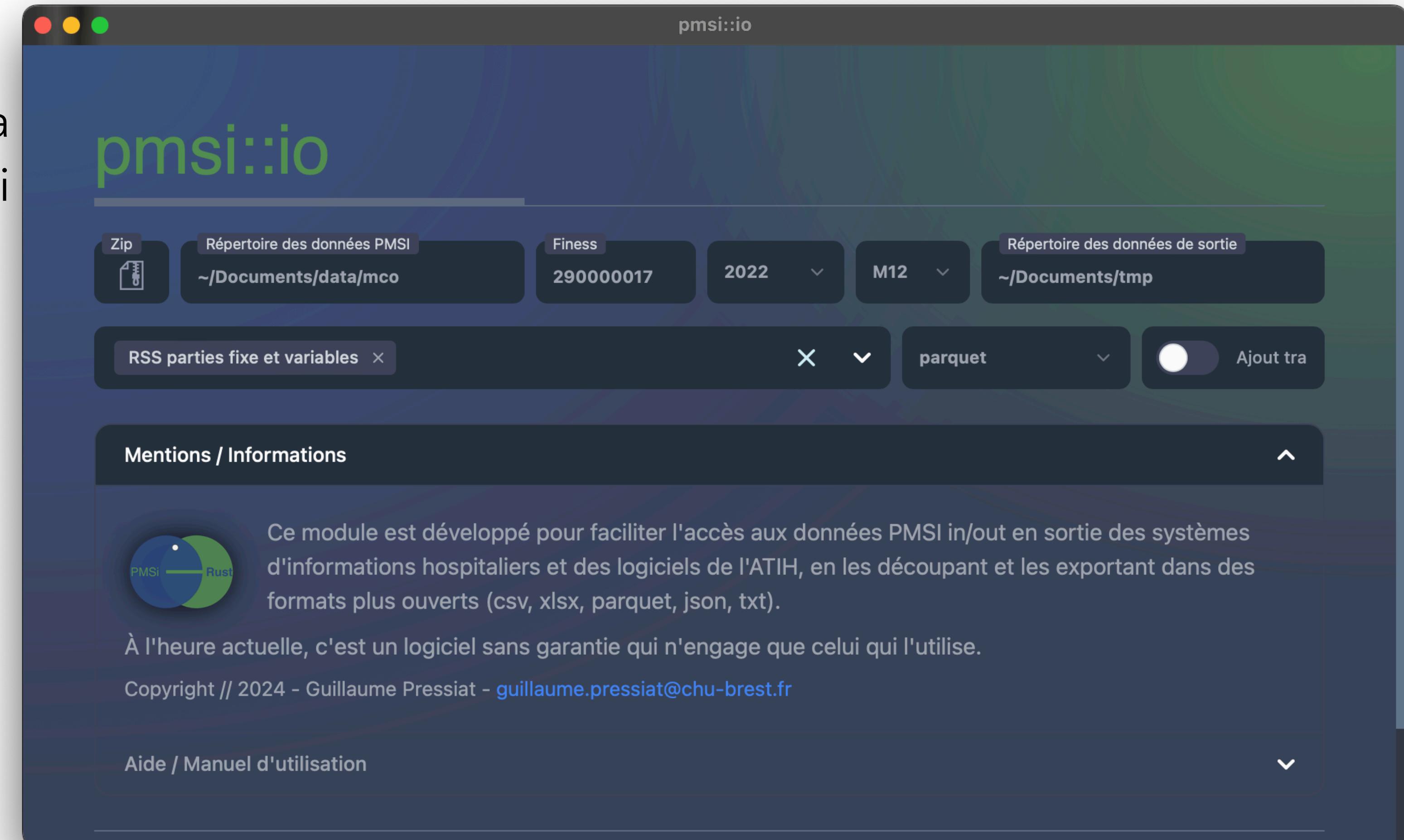
# Nouveauté 2024 : pmsi::io

- **Logiciel libre**



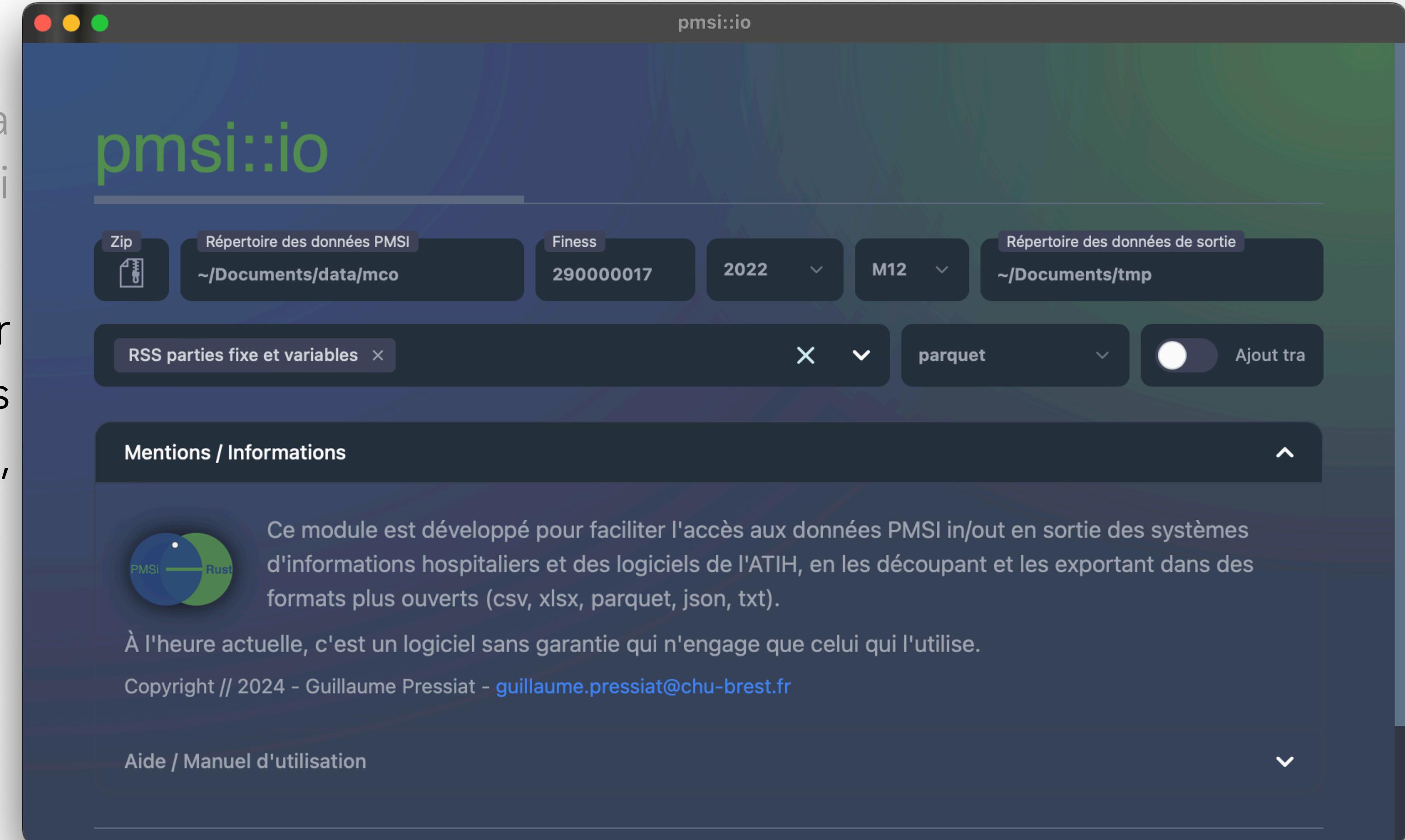
# Nouveauté 2024 : pmsi::io

- **Logiciel libre**
- **Objectif** : Rendre accessible la donnée aux professionnels DIM qui ne maîtrisent pas la programmation



# 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)



# 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)



# 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



# 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

**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**

# 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**

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

merci de votre attention

- Pour plus d'informations, vous pouvez vous rendre sur cette page web
- <https://guillaumepressiat.github.io/>
- Événement Meetup HDH à venir le 7 novembre 2024

