

## DURÉE / LIEU

20 jours / 120 heures  
en présentiel à Rennes

- 1 cycle de 13 jours
- 1 cycle de 7 jours

Travail personnel : environ 150 heures

## TARIF

5 250 € net

## MODALITÉS DE CANDIDATURE

**Date limite de dépôt des candidatures :** 5 semaines avant le début de la formation

**Un CV** est à joindre à votre demande d'inscription

Votre demande d'inscription sera soumise au responsable pédagogique. Une confirmation d'inscription vous sera adressée au plus tard 3 semaines avant le début de la formation.

Les candidatures sont examinées par ordre d'arrivée.

Conditions générales de vente disponibles sur notre site.

## PARTENARIAT



Nos données  
au service  
de la Santé

## CONTACTS

**Une question concernant votre candidature ?**

[candidaturefc@ehesp.fr](mailto:candidaturefc@ehesp.fr)

02.99.02.25.00

**Qui contacter si vous êtes en situation de handicap ?**

Formation accessible aux personnes en situation de handicap avec possibilité d'aménagement de la formation

**Référent handicap EHESP, Sandrine VILALLON**

[referent-handicap@ehesp.fr](mailto:referent-handicap@ehesp.fr)

Ecole des hautes études en santé publique

15, avenue du Professeur-Léon-Bernard

CS 74312 - 35043 Rennes Cedex

Tél. + 33(0)2 99 02 22 00

[formation-continue.ehesp.fr](http://formation-continue.ehesp.fr)



Découvrir la formation  
puis accéder au  
formulaire de candidature



[formation-continue.ehesp.fr](http://formation-continue.ehesp.fr)



E H E S P

Formation  
continue

DIPLÔME D'ÉTABLISSEMENT (DE)

Traitement des données de santé : PMSI et SNDS

# SCIENCE DES DONNÉES PMSI

## Les +

Une formation en collaboration avec **L'ATIH**

Une **montée en compétences progressive et structurée** : du requêtage de base à l'analyse prédictive sur les données PMSI (MCO, SMR, HAD)

Une **maîtrise approfondie de R et SQL** sur la plateforme de L'ATIH

Une **double expertise** reconnue en data science et en exploitation des bases médico-administratives

Une maîtrise de **trois langages leader** de la data science (R, SQL et Python)

 20 jours / 120 heures  
2 cycles de 3 modules

 Rennes

 Mars à novembre 2026

# DIPLÔME D'ÉTABLISSEMENT - SCIENCE DES DONNÉES PMSI



## COMPÉTENCES VISÉES

- Exploiter efficacement les bases de données PMSI (MCO, SMR, HAD) dans une démarche analytique structurée
- Maîtriser les langages R, SQL et Python pour le requêtage, la manipulation et l'analyse de données hospitalières
- Construire et automatiser des requêtes SQL complexes dans R en environnement réel (plateforme ATIH)
- Élaborer des indicateurs d'activité ou de parcours et produire des analyses statistiques descriptives ou inférentielles
- Implémenter des modèles prédictifs sur les données PMSI à l'aide de méthodes de machine learning (classification, arbres, ...)
- Structurer un projet d'analyse PMSI du plan d'extraction au plan de restitution, avec traçabilité et reproductibilité (R Markdown, Quarto document, Notebooks)
- Mettre en œuvre des projets de data science hospitalière intégrant des données PMSI



## PUBLIC VISÉ

Toute personne souhaitant développer ou renforcer ses capacités à exploiter les bases PMSI à des fins d'analyse, de pilotage ou de recherche, notamment :

- Médecin DIM, médecin de santé publique, praticiens hospitaliers
- Techniciens de l'information médicale (TIM), statisticien, data-manager, ingénieur hospitalier
- Personnel ARS, Assurance Maladie, chercheurs travaillant sur les données médico-administratives



## RESPONSABLE PÉDAGOGIQUE

Dr Juste GOUNGOUNGA  
Professeur Département Méthodes quantitatives en santé publique, EHESP



## PRÉREQUIS

Connaissance du PMSI MCO



## PARCOURS PÉDAGOGIQUE

Le diplôme peut être réalisé en 1 ou 2 ans. Il doit inclure la réalisation des deux cycles.

Le cycle Traitement des données PMSI est dupliqué en deux sessions pour 2026.



## ÉVALUATION/ VALIDATION

- Recueil des attentes
- Auto-positionnement préalable
- Contrôle d'assuidité
- Recueil des appréciations

L'évaluation des acquis est réalisée à la fin de chaque cycle, et comprend :

- un QCM de validation des connaissances en fin de chaque module
- un projet individuel de traitement des données PMSI mobilisant les compétences acquises, de l'extraction à l'analyse prédictive



## MÉTHODES PÉDAGOGIQUES

- Accès aux ressources pédagogiques sur la plateforme « Campus »
- Exposé magistral, démonstration
- Etudes de cas
- Exercices d'application

Nombre maximal de participants : 12



## PERSPECTIVES

Renforcement des compétences en analyse et valorisation des bases PMSI via R, SQL via R et Python en vue d'évoluer vers des fonctions nécessitant des capacités d'exploitation des données PMSI ou données de santé en générale

Débouchés possibles :

- Data analyst hospitalier (DIM, ARS, ...)
- Chargé d'études en santé publique, data manager en santé
- Consultant en exploitation de données de santé
- Chercheur exploitant les bases médico-administratives

## MODULE 2 - REQUÊTAGE ET ANALYSE INTER-CHAMPS PMSI SUR LA PLATEFORME ATIH VIA R

### Contenu de formation

- Présentation du contenu des tables des champs SMR et HAD
- Filtres spécifiques dans les deux champs
- Requêtes PMSI SMR et HAD de type Rbase, dplyr avec R
- Construction de fonctions et de manipulation d'objets R
- Organisation avancée de l'environnement de travail sur RStudio
- Débogage des fonctions R et profiling de code R
- Requêtes de synthèse : les statistiques médicales par champ et par patients
- Utilisation de nomenclatures générales et PMSI des différents champs
- Graphes d'indicateurs d'activités PMSI
- Automatisation d'analyse : Rmarkdown, Quarto document, RShiny

5 jours • 1<sup>er</sup> au 5 juin 2026 OU 19 au 23 octobre 2026

## MODULE 3 - EXPLOITATION AVANCÉE DES DONNÉES PMSI PAR LA DATA SCIENCE AVEC R

### Contenu de formation

- **Machine Learning : méthodologie générale**
- **Le workflow R de l'apprentissage supervisé**
  - Construction du jeu de données et préparation des données
  - Comparaison des modèles de classification supervisée ou de régression (Gradient Boosting, Random Forest, etc)
  - Paramétrage et entraînement du modèle
  - Évaluation du modèle
  - Diagnostique, gestion du sur-apprentissage ou du sous-apprentissage, cross-validation
- **Introduction à l'apprentissage non-supervisé**
  - Cas d'utilisation sur les données médico-administratives (PMSI, ScanSanté...)

3 jours • 24 au 26 juin 2026 OU 25 au 27 novembre 2026

## CYCLE PYTHON POUR L'ANALYSE DES DONNÉES DE SANTÉ ET LA DATA SCIENCE

### MODULE 1 - INTRODUCTION À PYTHON AVEC DES DONNÉES PMSI

#### Contenu de formation

- Introduction à Python (distributions, installations, interfaces shell, notebook, IDE)
- Présentation d'Anaconda et Jupyter lab
- Base du langage : instruction, variable, scripts
- Les objets de collection de python : dict, list, set, tuple
- Instruction des contrôles : conditions, boucles, fonctions
- Présentation de Programmation Orienté Object : Classe, attribut, méthode, module, package. Importation des modules/packages
- Calcul scientifique avec NumPy : vecteur et matrice
- Data management avec pandas (DataFrame et Serie, import/export, indexation, filtering, traitement des manquants)
- Cas d'utilisation sur les données du PMSI MCO

2 jours • 12 au 13 octobre 2026

### MODULE 2 - PYTHON POUR L'ANALYSE DES DONNÉES

#### Contenu de formation

- Data management avec pandas : agrégation, jointure, crosstab...
- Data visualisation avec matplotlib et seaborn
- Utilisation des méthodes d'analyses statistiques classiques : T-tests + Anova, Chi2, PCA, régression logistique
- Librairie scikit-learn pour l'apprentissage statistique (les modèles usuels de régression, régression logistique et arbres de décision...)
- Cas d'utilisation sur les données du PMSI MCO

3 jours • 14 au 16 octobre 2026

### MODULE 3 - MACHINE LEARNING APPLIQUÉ AUX DONNÉES PMSI AVEC PYTHON

#### Contenu de formation

- **Machine Learning : méthodologie générale**
- **Le pipeline de l'apprentissage supervisé**
  - Construction du jeu de données et préparation des données
  - Comparaison des modèles de classification supervisée ou de régression (Gradient Boosting, Random Forest...)
  - Paramétrage et entraînement du modèle
  - Evaluation du modèle
  - Diagnostic, gestion du sur-apprentissage ou du sous-apprentissage, Cross-Validation
- **Introduction à l'apprentissage non supervisé**
- **Cas d'utilisation sur les données PMSI**

2 jours • 23 au 24 novembre 2026

## PROGRAMME ET MODULES

### CYCLE TRAITEMENT DES DONNÉES PMSI AVEC R ET SQL

#### MODULE 1 - PLATEFORME DES DONNÉES HOSPITALIÈRES ET TECHNIQUES DE BASE DE REQUÊTAGE VIA RSTUDIO

##### Contenu de formation

- Présentation de la plateforme des données hospitalières de l'ATIH contenant les bases des 4 champs PMSI
- Notions de bases de données
- Architecture de la base PMSI MCO mise en place par l'ATIH
- Notions de base du langage SQL et R
- Prise en main du requêtage sur la plateforme des données hospitalières à l'aide de scripts SQL et R depuis RStudio
- Exercices d'extraction et d'analyse de données à partir des tables du PMSI MCO sur la plateforme des données hospitalières

5 jours • 16 au 20 mars 2026 OU 7 au 11 septembre 2026