

## DURÉE / LIEU

22 jours (132 heures)

9 modules de 2 à 4 jours à **Rennes**

**Présentiel** : 1 cycle de 16 jours (6 modules)  
+ 1 cycle de 6 jours (3 modules)

**Travail personnel** : environ 150 heures

## TARIF

5250 € net

Sous réserve de validation par les instances de l'EHESP

## MODALITÉS DE CANDIDATURE

- **Date limite de candidature** : le 31/01/2025
- **Un CV** est à joindre à votre demande d'inscription. En fonction de votre parcours professionnel, le responsable pédagogique pourra vous contacter pour un entretien.

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

Conditions générales de vente disponibles sur notre site.

## PARTENARIAT



### 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 télécharger le **bulletin de candidature**



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



EHESP

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
- Un diplôme avec une **double expertise** données de santé et Datascience
- La pratique de **trois langages leader** de la data science (Python, R et SQL)
- L'appropriation des **algorithmes de Machine Learning**



1 cycle de 6 modules  
de 2 à 4 jours

1 cycle de 3 modules  
de 2 jours

**Modules ouverts à la carte**



Rennes



22 jours

**Mars à  
novembre 2025**

# D.E. SCIENCE DES DONNÉES PMSI



## COMPÉTENCES VISÉES

- Exploiter efficacement les données PMSI
- Maîtriser des langages, outils et méthodes phares de la data science : Python, R, SAS SQL, machine learning
- Répondre à une demande d'étude en choisissant les outils et les bases pertinentes
- Piloter des projets hospitaliers axés sur l'exploitation des données PMSI



## PUBLIC VISÉ

Toute personne souhaitant exploiter les données issues du PMSI via des langages, outils et méthodes de la data science tels que Python, R, SAS SQL, SAS Enterprise Guide et les méthodes de machine learning : médecin DIM, médecin de santé publique, statisticien, data-manager, ingénieur hospitalier, personnel ARS, Assurance Maladie, chercheurs, bureau d'études



## PARCOURS PÉDAGOGIQUE

Recueil des attentes et des besoins des apprenants

Auto-positionnement préalable à l'entrée en formation : questionnaire

Ce diplôme est structuré autour de neuf modules permettant d'aborder les principales thématiques de données en santé et les méthodes de data science.

Parcours sur une ou deux années



## MÉTHODES PÉDAGOGIQUES

Accès aux ressources pédagogiques sur la plateforme « Campus »

Exposé magistral, démonstration

Etudes de cas

Exercices d'application

**Nombre maximum de participants : 10**



## PRÉREQUIS

Connaissance du PMSI MCO

## CONDITIONS

### PREALABLES POUR SUIVRE LA FORMATION

Disposer d'un ordinateur et d'une connexion à haut débit

Maîtriser les outils à distance

## PROGRAMME ET MODULES

### CYCLE TRAITEMENT DES DONNÉES PMSI

#### MODULE 1

#### PLATEFORME DES DONNÉES HOSPITALIÈRES ET TECHNIQUES DE BASE DE REQUÊTAGE

##### Contenu de formation

- Présentation de la plateforme de l'ATIH contenant des bases PMSI
- Notions de base de données
- Architecture de la base PMSI MCO mise en place par l'ATIH
- Prise en main du logiciel SAS Enterprise Guide
- Exercices d'analyse de données à partir d'une base PMSI MCO

**4 jours • 24 au 27 mars 2025**

#### MODULE 3

#### SAS PROGRAMMATION SQL ET TECHNIQUES DE REQUÊTAGE AVANCÉES

##### Contenu de formation

- Présentation de la plateforme de l'ATIH contenant des bases PMSI
- Architecture des bases PMSI MCO-SSR mise en place par l'ATIH
- Logiciel SAS Enterprise Guide
- Programmation SQL SAS (join, Subqueries, SQL procedure, SET Operators, Advanced PROC SQL Features)
- Exercices et études de cas sur le chaînage des séjours à partir des données PMSI MCO-SSR via programmation SQL

**2 jours • 12 au 13 juin 2025**

#### MODULE 2

#### SAS ENTERPRISE GUIDE ET TECHNIQUES DE REQUÊTAGE INTERMÉDIAIRES

##### Contenu de formation

- Architecture des bases PMSI MCO-SSR mise en place par l'ATIH
- Logiciel SAS Enterprise Guide
- Requête via SAS Guide
- Etude de cas 1 : Analyse de l'activité d'un GHT
- Etude de cas 2 : Parcours patient MCO-SSR
- Règles de bon usage

**2 jours • 10 au 11 juin 2025**

#### MODULE 4

#### PLATEFORME DES DONNÉES ATIH VIA R - INITIATION

##### Contenu de formation

- De SAS à R depuis le portail ATIH
- Requêtes PMSI MCO de type Rbase avec R
- Requêtes simple de prise en main de R : notions de programmation R
- Organisation de l'environnement de travail
- Utilisation des fonctions R fournies par des packages R
- Requêtes de synthèse : les statistiques médicales par groupe (GHM, DA, GP, CAS...), diagnostic, acte, case-mix par établissement, taux de recours, Taux d'attractivité et fuite
- Utilisation de nomenclatures et de jointure
- Graphes d'indicateurs d'activités PMSI à l'aide de la fonction plot()
- Études de cas : Statistique descriptive sur la prise en charge d'infarctus du myocarde non compliqués

**2 jours • 1er au 2 septembre 2025**

#### MODULE 5 PLATEFORME DES DONNÉES ATIH VIA R - INTERMÉDIAIRE

##### Contenu de formation

- De SAS à R depuis le portail ATIH
  - Description de l'architecture du pmsi SMR
  - Requêtes PMSI MCO-SSR de type SQL avec R : Tidyverse
  - Requêtes intermédiaires : programmation avancée R
  - Requêtes de synthèse : les statistiques médicales par groupe (GHM, DA, GP, CAS...), diagnostic, acte, case-mix par établissement, taux de recours, part de marché, Taux d'attractivité et fuite, statistiques des GHT ; tests de comparaisons d'indicateurs de l'activité hospitalière.
- Utilisation de nomenclatures et de jointure
- Graphes d'indicateurs d'activités PMSI à l'aide des fonctions du package ggplot et packages dérivés
- Concept et types de chaînage
- Études de cas :
  - Prise en charge d'infarctus du myocarde non compliqués ;
  - Chaînage de patient provenant des champ d'activité MCO-SMR,
  - Chaînage pluriannuelle
- Règles de bon usage

**3 jours • 3 au 5 septembre 2025**

#### MODULE 6 UTILISATION DE LA DATASCIENCE POUR EXPLOITER LES DONNÉES PMSI

##### 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...)
  - 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 • 26 au 28 novembre 2025**

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

#### MODULE 1 - PYTHON : INITIATION

##### 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 médico-administratives

**2 jours • 6 au 7 octobre 2025**

#### MODULE 2 - PYTHON : INTERMÉDIAIRE

##### 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 médico-administratives (PMSI, ScanSanté...)

**2 jours • 8 au 9 octobre 2025**



#### Responsable pédagogique Dr Juste GOUNGOUNGA

Professeur Département Méthodes quantitatives en santé publique, EHESP

Cette formation peut également vous intéresser :

#### DIPLÔME D'ÉTABLISSEMENT

• SCIENCE DES DONNÉES EN SANTÉ PUBLIQUE

#### MODULE 3 - PYTHON : AVANCÉ LA DATA SCIENCE ET LE MACHINE LEARNING

##### Contenu de formation

- Machine Learning : méthodologie générale
- Le pipeline de l'apprentissage supervisé via la librairie Scikit-Learn
  - 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 médico-administratives (PMSI, ScanSanté)

**2 jours • 24 au 25 novembre 2025**



#### PERSPECTIVES

Renforcer votre expertise en exploitation des données PMSI ou données de santé en générale à des fins stratégiques  
Expert data hospitalier, data manager / data analyst en santé

Chargé d'études statistiques en santé



#### ÉVALUATION/ VALIDATION

Recueil des attentes  
Auto-positionnement  
Recueil des appréciations  
Contrôle d'assuétude  
L'obtention de ce diplôme repose sur la réalisation et la validation de projets individuels d'analyse de données

Plus d'informations sur  
formation-continue.ehesp.fr

MàJ 02/07/2024