



# Institut de Santé Publique, d'Epidémiologie et de Développement (I.S.P.E.D)

## DIPLÔME UNIVERSITAIRE SCIENCES DE DONNEES EN SANTE PUBLIQUE

### 1 - Objectifs de la formation

L'objectif global du DU est de doter les étudiants des connaissances et des compétences sur les pré-requis pour mener à bien un projet de Sciences de Données en Santé Publique.

À l'issue du programme, les diplômés seront en mesure de contribuer activement à un projet de Sciences de Données pour la Santé Publique et aptes de poursuivre dans un programme de Master en Sciences de Données pour la Santé.

### 2 - Responsables

**Professeur Rodolphe Thiebaut** (PU-PH),

**Professeur Gayo Diallo** (PU)

ISPED – Université de Bordeaux

146 rue Léo Saignat

33076 Bordeaux cedex

05 57 57 13 93

[rodolphe.thiebaut@u-bordeaux.fr](mailto:rodolphe.thiebaut@u-bordeaux.fr)

[gayo.diallo@u-bordeaux.fr](mailto:gayo.diallo@u-bordeaux.fr)

### 3 - Organisation de la formation :

- **3-1 Capacité d'accueil :**

Le nombre minimum d'étudiants est fixé à **10** et le maximum à **30**.

- **3-2 Durée de la formation :**

La formation dure un an et à lieu tous les ans

- **3-3 Enseignement théorique :**

Le volume horaire global de la formation est de **180h**.

L'enseignement se fait exclusivement en distanciel.

L'enseignement est effectué en français.

La formation s'appuie sur 2 modules transversaux (module 1 et 2) et un triple socle en Épidémiologie (module 3), Statistiques (module 4), Informatique de Santé (module 5).

En fonction du cursus académique et/ou professionnel antérieur, les étudiants pourront être dispensés du suivi de certains modules, pour tout ou partie, et se verront remettre une attestation d'équivalence par l'équipe pédagogique. Une commission pédagogique se réunira pour statuer.

L'enseignement se déroule d'octobre à juillet de l'année universitaire.

- **3-4 Stage :**

Ce diplôme n'est pas ouvert au stage.

#### 4 - Conditions d'inscriptions

**Titres requis ou niveau :**

- Étudiants en santé publique ou sciences de la santé (médecine, pharmacie, biologie, etc.) souhaitant se spécialiser en analyse de données.
- Professionnels: Ingénieurs informaticiens ou statisticiens désirant acquérir des compétences en science des données en santé pour améliorer leur pratique.
- Diplômés en mathématiques, informatique ou statistiques intéressés par le domaine santé.

L'admission sera conditionnée à un processus de sélection par un jury composé des responsables de la formation et des intervenants qui procédera à une analyse individuelle des dossiers de candidature.

***Il conviendra de joindre au dossier de candidature : un curriculum vitae (3 pages maximum), une lettre de motivation et le diplôme requis pour accéder à la formation (à minima).***

#### 5 - Coût de la formation

En formation initiale : **650€/an**

En formation continue :

- Reprise d'études non financées : **1 250€ /an**
- Reprise d'études financées : **2 100€ /an**

*Auquel s'ajoutent les droits d'inscription correspondant à la base du droit d'inscription de Licence*

#### 6 - Contrôle des connaissances et conditions de validation :

- **6-1 : Nombre de sessions :**

Il est organisé 1 session d'examen par an.

- **6-2 : Nature des épreuves :**

L'examen consiste en :

- un contrôle continu en ligne pour chaque module sous forme de QCM ou de cas pratique et comptant pour 25% de la note générale,

- une épreuve écrite individuelle en présentiel comptant pour 75% de la note générale en fin d'année,
- L'assiduité aux cours est obligatoire et sera monitorée à travers les connexions sur la plateforme d'enseignement.

- **6-3 : Validation :**

Pour être déclaré admis, le candidat devra avoir suivi l'intégralité des modules auxquels il est inscrit et obtenu une moyenne générale  $\geq 10$ .

- **6-4 : Dispositions prévues en cas d'échec :**

Le redoublement est autorisé sous réserve d'une demande motivée et validée par l'équipe pédagogique du diplôme.

Les étudiants devront s'acquitter de nouveau des droits d'inscription.

## 7 - Délivrance du diplôme :

Après obtention, le Diplôme est remis à l'étudiant :

- par voie postale après une demande écrite
- en main propre sur présentation d'une pièce d'identité à l'UGANC

Création :

Conseil ISPED du 26/03/2026

Conseil du Collège Sciences de la santé du 19/05/2026

(Version 1)

## **PROGRAMME**

### **Objectif(s) de la formation**

L'objectif du diplôme, dispensé en français et à distance, est donc de former 30 étudiants par an (150 en 2030 à l'instar des DU de l'ISPED).

L'enjeu est d'initier les étudiants aux fondamentaux des sciences de données en santé publique, avec des compétences de base en statistiques, épidémiologie et informatique de santé.

L'enjeu est également de comprendre les enjeux éthiques, légaux et sociaux liés à l'utilisation des données de santé. La formation se fera en lien étroit avec des entreprises du secteur, dont le partenaire du projet Santé Numérique, IQVIA. Elles fourniront éventuellement des cas d'usages et des jeux de données pour les mises en œuvre pratiques.

À l'issue des cinq années du financement « Compétences et métiers d'avenir » (CMA), la formation sera suffisamment stabilisée.

Le DU pourra s'obtenir en validant l'ensemble des modules de la formation, par leur suivi ou l'obtention d'équivalences.

### **Compétences acquises**

L'approche choisie doit permettre aux apprenants de :

- Savoir utiliser et comprendre la littérature scientifique
- Savoir élaborer une question de recherche (méthodologie du "problème d'abord")
- Connaître les bases des méthodologies statistiques
- Connaître les principes fondamentaux de la science épidémiologique
- Avoir une connaissance pratique de la législation actuelle en matière de données
- Connaître les bases de la conception d'une base de données en santé
- Connaître les bases de la programmation et structure de données
- Connaître les outils numériques fondamentaux pour les sciences de données : shell, networks & Docker
- Connaître les outils pour une science des données communicable et reproductible

### Calendrier :

Le programme détaillé des enseignements sera communiqué ultérieurement, en amont du début des enseignements, par le secrétariat universitaire.

Matériel pédagogique : L'ensemble du matériel pédagogique sera transmis aux étudiants via la plateforme d'enseignement en ligne de l'IspeD.

### Accompagnement

Sur le plan pédagogique, les apprenants bénéficient de compléments d'informations par le biais d'un forum d'échange apprenant-enseignant, via Internet, auquel les enseignants répondent chaque semaine.

Type	COURS	Responsables
Modules Transverses	<b>Module 1 : Outils numériques pour les sciences de données (30 heures)</b>	
	Introduction aux outils numériques pour les sciences de données	Marina boudin
	Principes Linux, Shell et scripts	
	Gestion de versions et de codes, Maven	
	Principes des conteneurs et cas d'application avec Docker	
	Les langages YAML et Mark Down	
	<b>Module 2 : Aspects éthiques, réglementaires et recherche biblio (20h)</b>	
	Grands principes éthiques pour les sciences de données en santé	Guillaume Verdy
	Aspects réglementaires sur les données de santé et RGPD	
Les Hub Nationaux de Données , le SNDS et l'Espace Européen des Données de Santé		
Principes de Recherche Bibliographique et analyse de la littérature scientifique		
Principes pour une étude reproductible		
	<b>Module 3 : Aspects fondamentaux en Épidémiologie (40h)</b>	
Modules Socles	Introduction à l'Épidémiologie	Guillaume Verdy
	Les principes et les modalités de mise en œuvre des principaux schémas d'étude en épidémiologie	
	Le principe de la surveillance épidémiologique	
	Les protocoles d'études épidémiologie et traitement des biais	
	Les Grands principes des études de cohortes	
	<b>Module 4 : Fondamentaux en Biostatistiques (45 heures)</b>	
	Introduction aux statistiques en santé	Pierre Joly
	Définition de populations d'études	
	Estimation de confiance et analyse de variance	
	Méthodes de Régressions linéaires	
	Analyses multivariées	
	Mise en œuvre avec le langage R	
	<b>Module 5 : Fondamentaux en Informatique de Santé (45 heures)</b>	
	Principes algorithmiques et structure de données	Gayo Diallo
Les entrées/sorties et traitement de données hétérogènes (textes, tabulaires, etc)		
Modélisation de bases de données et mise en œuvre		
Programmation structurée et orientée objet		
Terminologies biomédicales		
Mise en œuvre avec Python		