

Bases de données, SNDS enrichi et parcours patient par clustering – BDD-PPC

Du 1er au 3 juin 2026 (3 jours)
9h-12h / 13h-16h

Lieu des cours : Isped, Campus Carreire
université de Bordeaux

Frais de formation :

Inscription individuelle : 360 €

Inscription institutionnelle : 900 €

Responsables :

- Fleur MOUGIN (responsable), professeure des universités en informatique, Isped, université de Bordeaux
- Amandine GOUVERNEUR (responsable), IQVIA, Senior Consultant
- Romane LE GOFF, IQVIA, doctorante CIFRE, équipe AHeAD, BPH U1219 Inserm
- Jade VADEL, IQVIA, Statisticienne Real World Solutions

Objectifs

- › Connaître les bases de données (BDD), les différents types de données et l'intérêt de leur appariement
- › Mettre en œuvre un appariement
- › S'initier au clustering

Pré-requis

- › Langage R et SAS
- › **Public cible** : professionnels, chercheurs, enseignants, étudiants (Master 2, doctorat) dans le champ sont les bienvenus

Programme

- › Dans le cadre du projet CAP Santé Numérique, lauréat de l'AMI « Compétences et métiers d'avenir » (CMA) et porté par l'Université de Bordeaux, plusieurs programmes de formation sont proposés pour les étudiants et professionnels issus de cursus scientifiques. Le présent module s'inscrit dans cette démarche et a été défini en concertation entre l'Isped et IQVIA. Le contenu est le suivant :

1) Enrichissement des BDD SNDS-données open source

- › Présentation des différentes bases de données disponibles (non exhaustif) ainsi que des avantages et des inconvénients de chacune
- › Présentation de l'intérêt de l'appariement des bases de données avec le SNDS
- › Contraintes réglementaires et éthiques (CNIL, RGPD, pseudonymisation)
- › Ouverture sur des entrepôts autorisés appariés au SNDS : HEH, P4DP, Biogroup

2) Clustering pour identifier les parcours des patients (trajectoires de traitement)

- › Introduction au clustering (concepts fondamentaux, algorithmes classiques, choix du nombre de clusters)
- › Etude de parcours de soins/traitements : découpage temporel via l'analyse de séquences d'états et application de l'Optimal Matching pour calculer la distance (similarité) entre séquences avant clustering

Mise en pratique avec R et SAS