

Qualité de Données, Data Wrangling

**ECTS : 2**

**Volume horaire : 21**

**Description du contenu de l'enseignement :**

Le cours a pour objectif d'introduire les caractéristiques principaux liés à la qualité des données ainsi que de présenter les algorithmes pour leur traitement.

La qualité de l'information a de graves répercussions sur l'efficience et l'efficacité des organisations et des entreprises. Par exemple, dans les organisations privées, telles que les sociétés de marketing ou les banques, il n'est pas surprenant d'avoir plusieurs registres de clients, mis à jour par différentes procédures organisationnelles, ce qui entraîne des informations contradictoires et en double. Ce cours vise à présenter les problèmes liés à la qualité de données, et de présenter les solutions qui peuvent être utilisées pour les résoudre.

En particulier, il a pour objectif d'examiner les trois points suivants :

- Dimensions de la qualité de données. Le terme « qualité de données » est assez vague et englobe plusieurs facettes. Dans cette partie, nous allons examiner les différents dimensions de la qualité de données, allant de l'exactitude, complétude, jusqu'à la consistance et la confiance.
- Algorithme et solutions théoriques. Dans cette partie nous allons examiner en détails quelques solutions théoriques proposées pour résoudre certains problèmes de qualité de données, en particulier celles liées à l'identification d'objet.
- Outils et solutions pratiques. Dans cette partie, nous allons utiliser des outils, et montrer comment ils peuvent être utilisés pour améliorer la qualité de données.

**Document susceptible de mise à jour - 11/02/2026**

**Université Paris Dauphine - PSL - Place du Maréchal de Lattre de Tassigny - 75775 PARIS Cedex 16**