



## **GOVERNANCE DE LA DONNEE – EXPLORATION DES METHODES ET OUTILS D'ANONYMISATION DES DONNEES**

L'accumulation de différents types de données complexifie leur partage. Qu'elles se trouvent dans un data lake / data hub ou bien dans des applications métiers, les données sont soumises à des réglementations et à des niveaux de confidentialité restreints. L'anonymisation est le processus de modification de contenu ou de structure de la donnée pour retirer les informations pouvant identifier personnellement un individu. La donnée est un actif important au sein de l'entreprise et au-delà dans les échanges avec les partenaires. Il est donc primordial de maîtriser les techniques qui permettent d'anonymiser, de privatiser les données afin de les partager sans devoir se soucier de leur caractère personnel.

### **Objectif du stage**

L'objectif du stage est de lister, tester et évaluer les outils clés en main et les solutions offrant des fonctionnalités d'anonymisation. Il s'agira de comparer les méthodes d'anonymisation de la donnée, d'analyser celles d'industrialisation de l'anonymisation au sein d'un SI et étudier les meilleures façons d'anonymiser tout en gardant la valeur pour un Data Scientist. Enfin, les façons de reconstruire de la donnée anonymisée seront analysées. Le stage a aussi pour objectif d'implémenter différentes méthodes d'anonymisation sur des données afin de les cartographier dans le cadre d'une étude comparative.

### **Travaux à réaliser**

Intégré pour la durée de votre stage à la Taskforce Data, vous serez amené, sous la supervision d'un consultant, à réaliser les travaux suivants :

- Rédiger une note de cadrage du sujet de stage, comprenant :
  - Votre compréhension du stage et ses objectifs
  - La définition d'un cas d'usage
  - Un planning prévisionnel et les jalons clés
- Définir les critères d'évaluation des solutions
- Faire un état de l'art des méthodes d'anonymisation
- Analyser les solutions clés en main ainsi que leur industrialisation dans un SI
- Implémenter des méthodes d'anonymisation avec ces solutions
- Définir des "plan d'anonymisation" expliquant comment anonymiser une donnée sans perdre sa valeur
- Implémenter des algorithmes d'anonymisation
- Analyser les méthodes de reconstruction d'une donnée déjà anonymisée :
  - Via pseudo anonymisation : Valider la possibilité d'un retour arrière lorsqu'elle est prévue
  - Via croisement de données : Valider l'impossibilité de reconstruire la donnée
- Enrichir le séminaire sur la gouvernance de la donnée de Nexworld

En parallèle, vous participerez à des missions de conseil, en intégrant une équipe de consultants Nexworld, placée sous la responsabilité d'un directeur de mission.

### **Outils et technologies (non exhaustif)**

- BigID, Egnyte, G2 Track, OneTrust, Box Zones, Aptible, Privatar
- Anonymisation, Pseudo-anonymisation, k-anonymity
- Java, Python

### **Profil**

Il est à noter que le stage est ouvert aux profils plus ou moins techniques. Un profil plus technique pourra développer les méthodes d'anonymisation, alors qu'un profil plus conseil pourra analyser les besoins réglementaires et le métier d'anonymisation, avec un axe RGPD par exemple.