

# ANONYMISER SES DONNÉES : QUELQUES RESSOURCES

Dans certains cas, les données nominatives ne peuvent être diffusées en l'état. La réglementation en matière de protection des données personnelles est en effet stricte en France et plus globalement au sein de l'Union européenne. Comment anonymiser ses données de manière fiable, afin de pouvoir les diffuser ?

## DÉFINITIONS

Le site de la CNIL, [rubrique « anonymisation »](#), donne les définitions principales à connaître en la matière.

### Pseudonymiser n'est pas anonymiser

« La pseudonymisation est un traitement de données personnelles réalisé de manière à ce qu'on ne puisse plus attribuer les données relatives à une personne physique sans information supplémentaire.

Exemple : un nom devient un alias ou un numéro séquentiel. La pseudonymisation est réversible. » L'anonymisation, quant à elle, est définitive et ne permet pas d'identifier à posteriori les individus.

### Quelles étapes pour l'anonymisation ?

1. Identifier les informations à conserver selon leur pertinence ;
2. Supprimer les éléments d'identification directe, ainsi que les valeurs rares qui pourraient permettre une réidentification aisée des personnes ;
3. Distinguer les informations importantes des informations secondaires ou inutiles, c'est-à-dire pouvant être supprimées ;
4. Définir la finesse idéale et acceptable pour chaque information conservée.

### Selon la CNIL :

- ☑ La **randomisation** consiste à modifier les attributs dans un jeu de données de telle sorte qu'elles soient moins précises, tout en conservant la répartition globale. Cette technique permet de protéger le jeu de données du risque d'inférence. Mais attention ! Cette pratique est à proscrire, car elle consiste à introduire volontairement des erreurs.
- ☑ La **généralisation** consiste à modifier l'échelle des attributs des jeux de données, ou leur ordre de grandeur, afin de s'assurer qu'ils soient communs à un ensemble de personnes. Cette technique permet d'éviter l'individualisation d'un jeu de données. Elle limite également les possibles corrélations du jeu de données avec d'autres.

### Il faut donc absolument éviter :

- **l'individualisation** : il ne doit pas être possible d'isoler un individu dans le jeu de données.
- **la corrélation** : il ne doit pas être possible de relier entre eux des ensembles de données distincts concernant un même individu.
- **l'inférence** : il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

## QUELQUES LOGICIELS

### Amnesia, la solution open source et gratuite proposée par OpenAIRE

Amnesia permet d'effectuer ses recherches et de partager ses résultats en respectant les directives du Règlement général sur la protection des données (RGPD) grâce à des algorithmes d'anonymisation des données.

- Voir la présentation de l'outil sur [Youtube](#)
- Voir les modalités d'utilisation et le mode d'emploi sur le [site d'OpenAIRE](#)
- [Les tutoriels](#)

### Comparatif de cinq outils gratuits :

[ARX Data Anonymization Tool](#), [Amnesia](#), [μ-ARGUS](#), [sdcMicro](#) et [Anonimatron](#)

L'entreprise [Aircloak](#), spécialisée en anonymisation de données, propose sur son site un [comparatif de cinq outils gratuits](#), particulièrement adaptés à une opération unique d'anonymisation sur un jeu de données. Aircloak étant fournisseuse de solutions payantes, ce comparatif est à considérer avec précaution, mais il permet néanmoins d'avoir un panorama rapide.

### Comparatif de trois outils payants :

[Oracle Data Masking](#), [Arcad Software](#), [We Data/Octopize](#)

Le magazine DSIH, portant sur l'actualité des systèmes d'information hospitaliers et de la e-santé, propose un comparatif de ces trois logiciels. Voir à ce sujet les articles de Cédric Cartau ([partie 1](#), [partie 2](#)).

## POUR ALLER PLUS LOIN

- Nguyen, Benjamin. « [Techniques d'anonymisation](#) ». Statistique et société, Vol. 2, N° 4 décembre 2014, p. 43-50.
- Avis du Groupe de travail européen sur la protection des données, au sujet des [techniques d'anonymisation](#), 2014.
- Ben Fredj, Feten. [Méthode et outil d'anonymisation des données sensibles](#). Cryptographie et sécurité[cs.CR]. Conservatoire national des arts et métiers - CNAM ; Université de Sfax (Tunisie). Faculté des Sciences économiques et de gestion, 2017.
- Datastorm. [Déployer une procédure d'anonymisation : se poser les bonnes questions](#). Livre blanc, septembre 2020.