

ATELIER N°2

DIFFUSER SANS DIVULGUER : L'ART DE PARTAGER LES DONNÉES CONFIDENTIELLES

Clara Baudry (), Julien Jamme (**)*

() Insee, Département des méthodes statistiques - Experte gestion statistique de la confidentialité*

*(**) Insee, Département des méthodes statistiques - Expert gestion statistique de la confidentialité*

Pour assurer la confidentialité de nos données, impossible de se retrancher derrière une unique solution valable en toute circonstance. En revanche, partager une démarche méthodologique commune, cela est possible ! **Cet atelier a ainsi pour ambition de fournir aux participants un cadre conceptuel et technique permettant d'interroger chaque nouvelle situation avec les moyens adéquats : concepts, méthodes, mesures.**

Diffuser des informations issues de données confidentielles nécessite un arbitrage entre le souci de protéger la vie privée des individus et le secret des affaires des entreprises d'une part, et le souhait de mettre à disposition des informations utiles auprès du public d'autre part.

En s'appuyant sur des cas concrets rencontrés dans la statistique publique française (diffusion des QPV) ou étrangère (diffusion du Censur 2021), l'atelier poursuivra principalement trois objectifs :

1. présenter les **concepts** clés de la protection de données statistiques,
2. détailler les moyens de mener un arbitrage, objectivement fondé, entre le **risque** et l'**utilité**,
3. familiariser l'auditeur avec les **différentes familles de méthodes** afin de trouver le meilleur des compromis possibles. L'atelier abordera à cette occasion les avantages comparatifs des **méthodes de blanchiment** et des **méthodes perturbatrices**.

L'atelier s'attachera à présenter les différents types de risques de divulgation (identification, attribut, inférence, différenciation) et les principaux scénarios d'attaque auxquels peuvent être confrontées les données statistiques que nous publions auprès du public.

Le principe des méthodes telles que les clés aléatoires, le swapping ou la génération de données synthétiques sera expliqué et les méthodes comparées entre elles. Les avantages et les inconvénients d'utiliser des méthodes perturbatrices en lieu et place des méthodes suppressives seront notamment discutés.

Pré requis

Certaines parties de l'atelier feront appel à des connaissances de base en probabilités.

Durée : 3 heures.