

# CONFIDENTIALITÉ STATISTIQUE

Maxime Beauté<sup>1</sup>, Clément Guillo<sup>2</sup>

<sup>1</sup> *Division recueil et traitement de l'information, Département des méthodes statistiques, Insee, 88 Avenue Verdier, 92120 Montrouge, [maxime.beaute@insee.fr](mailto:maxime.beaute@insee.fr)*

<sup>2</sup> *Division méthodes et référentiels géographiques, Département des méthodes statistiques, Insee, 88 Avenue Verdier, 92120 Montrouge, [clement.guillo@insee.fr](mailto:clement.guillo@insee.fr)*

Outre le respect des règlements nationaux et européens, la confidentialité des données permet de s'assurer de la confiance des répondants aux enquêtes. Cette confiance influe positivement sur les taux de réponse et favorise donc la production de statistiques de qualité.

Anonymiser des données diffusées publiquement passe d'une manière ou d'une autre par une réduction de l'information diffusée, autrement dit par une perte d'information. Dans ce domaine, l'enjeu est de permettre l'élaboration de statistiques les plus précises possibles, tout en empêchant la divulgation d'informations sensibles au niveau individuel.

Les méthodes de gestion du secret statistique sont nombreuses. Elles diffèrent en particulier selon la nature des données à traiter : tableaux de données agrégées ou données individuelles. Choisir une méthode revient à déterminer comment minimiser la perte d'information nécessaire pour garantir le secret : il peut s'agir de masquer une partie des données, de les restructurer, d'y ajouter du bruit, etc. Choisir les paramètres d'une méthode revient à fixer cet équilibre entre quantité d'information diffusée et risque de divulgation d'informations confidentielles.

Pour protéger une information, outre le fait d'anonymiser les données elles-mêmes, il est également possible de restreindre l'accès à ces données.

À la fin de l'atelier, les participants seront en mesure de connaître :

- les enjeux liés à la confidentialité statistique ;
- le cadre législatif en vigueur ;
- l'approche consistant à contrôler l'accès aux données en guise de protection ;
- un panorama des méthodes de gestion de la confidentialité les plus utilisées :
  - sur des tableaux de données agrégées ;
  - sur des bases de données individuelles ;
  - sur des données géographiques carroyées.

## Notes biographiques

Maxime Beauté est expert sur les questions de confidentialité statistique à l'Insee. Il participe aux réunions de réflexion au sujet de la confidentialité dans des groupes de travail européens. Il réalise également une veille méthodologique sur la gestion du secret dans les données tabulées et individuelles. Il forme et sensibilise les agents du Service statistique public aux enjeux liés à la confidentialité statistique. Enfin, il participe aux travaux de protection de tableaux de données avant leur diffusion pour différents services de l'Insee.

Clément Guillo encadre une partie des travaux méthodologiques portant sur les statistiques, l'analyse spatiale et la confidentialité à l'Insee. Il a participé à plusieurs travaux portant sur la production de données géographiques dont le géoréférencement des logements de la source fiscale. Par ailleurs, il participe aux réunions de réflexions, notamment sur les questions de confidentialité, autour de l'intégration de statistiques dans les données géographiques à l'Insee et dans des groupes de travail européens.