

Accéder à des données directement ou indirectement nominatives

Guide pour les rustres et les malappris

Laurent Toulemon, version du 20 juin 2017

L'accès aux données non anonymes pour la recherche est maintenant simple, facile et rapide, à condition de savoir comment s'y prendre. Cette note est destinée à vous aider.

Vérifier que les données « ouvertes » ne suffisent pas à vos recherches

L'Insee met sur son site de très nombreux fichiers individuels du recensement. Voir aussi sur le réseau Quetelet de très nombreux fichiers d'enquête. N'hésitez pas à contacter le producteur pour vérifier que les données ouvertes ne vous suffisent pas.

Le catalogue des sources non anonymes et disponibles pour la recherche figure sur le site du CASD. La liste des producteurs et des contacts figure sur la page « procédure » du site du comité du secret statistique.

Faire une demande au Comité du secret statistique

Récupérer sur le site du comité du secret le formulaire de demande (regarder au passage le schéma de la procédure dans le document PDF en haut de page). Consulter également les FAQ à droite de la page.

Remplissez le formulaire sobrement : pas besoin d'en faire des tartines, mais soyez précis sur l'objectif de votre demande et pourquoi il vous faut des données confidentielles.

Contactez le producteur des données pour avoir son avis et sa validation avant d'envoyer le formulaire au comité du secret.

Après un avis favorable du comité du secret

il faut signer le ou les engagements de confidentialité (comité du secret, éventuellement ministère des finances), faire une demande à la Cnil ; pour cela, contactez le correspondant informatique et liberté (CIL) de votre institution, normalement habilité à vous donner l'autorisation.

Ensuite (ou parallèlement) se faire enrôler au CASD si les données sont accessibles par ce canal (ce qui n'est pas toujours impératif).

L'accès est payant (tarif disponible sur le site du CASD) : coût fixe, location d'une SD-Box si vous n'en disposez pas, abonnements individuels. Prix dégressif : compter 3 000 euros par an pour un utilisateur (2 600 sans SD-Box, il y a divers tarifs). Si vous vous incrustez dans un projet déjà existant comme Big_Stat le coût marginal peut être ramené à 240 euros par an (ou 660 avec une SD-Box). Pour cela il faut impérativement que votre recherche s'inscrive fortement dans la problématique du projet (me contacter), la philosophie générale étant plutôt de séparer les projets de différentes équipes sur différents sujets. La liste des projets figure sur le site du CASD, mais sans le nom de principal demandeur, pour le moment.

Contacts et liens internet

Le recensement à l'Insee

Informations sur le recensement de la population sur le site de l'Insee (les fichiers existent pour toutes les années depuis 2004, mais il est un peu compliqué de retrouver les adresses. Je conseille d'aller sur le site de l'Insee et de rechercher « Logements, individus, activité, mobilités scolaires et professionnelles, migrations résidentielle » :

- <https://www.insee.fr/fr/metadonnees/source/s1321> (documentation)
- <https://www.insee.fr/fr/information/2409289> présentation des fichiers
- <https://www.insee.fr/fr/statistiques/2409559> fichiers de 2013 l'accès aux fichiers

Réseau Quetelet

- <http://www.reseau-quetelet.cnrs.fr>

Comité du secret statistique

- http://www.cnis.fr/cms/Accueil/activites/trois_comites/Comite_du_secret_statistique la page d'accueil
- http://www.cnis.fr/cms/Accueil/activites/trois_comites/Comite_du_secret_statistique/Les_procedures_daces_aux_donnees_confidentielles la page décrivant les procédures, avec un document PDF décrivant la procédure et le lien vers le formulaire
- comite-secret@cnis.fr l'adresse de contact

Centre d'accès sécurisé aux données (CASD)

- <https://casd.eu>
- <https://casd.eu/fr/donnees> la page qui liste les sources disponibles (au CASD ou ailleurs)
- <https://casd.eu/fr/tous-les-projets> la liste des projets

Last but not least

Ce texte est produit dans le cadre du projet Big_Stat, voir http://big_stat.site.ined.fr. Il est sujet à mise à jour et corrections. Merci d'envoyer vos retours, compléments et corrections éventuels à toulemon@ined.fr et giulia.ferrari@ined.fr