

Direct or indirect access to nominative data

Guide for Hicks and Dummies

Laurent Toulemon, version of June 20, 2017

Access to non-anonymous data for research is now simple, easy and fast, provided you know how to do it. This note is intended to help you.

Verify that "open" data is not enough for your research

INSEE has a very large number of individual census files on its site. The réseau Quetelet website also has numerous survey files. Do not hesitate to contact the producer to make sure that the open data is not enough for you.

The catalog of non-anonymous sources available for research can be found on the CASD website. The list of producers and contacts is on the "procedure" page of the Statistical Confidentiality Committee website.

Apply to the Committee on Statistical Confidentiality

Retrieve the application form from the confidentiality committee's website (look at the diagram of the procedure in the PDF document at the top of the page). Also see the FAQs on the right side of the page.

Be brief and to the point when you fill out the form: no need to make a meal of it, but be precise about the purpose of your request and why you need confidential data.

Contact the data producer for advice and validation before sending the form to the confidentiality committee.

After a favorable opinion from the confidentiality committee

Sign the confidentiality agreement(s) (confidentiality committee, possibly the Ministry of Finance), make a request to the CNIL. To do so, contact the data protection correspondent (CIL, *correspondant informatiques et libertés*) of your institution, who is normally accredited to give you the authorization.

Then (or at the same time) enroll in the CASD if the data is accessible via this channel (which is not always necessary).

Access is fee-paying (rate available on the CASD website): standing charge, rental of an SD-Box if you do not have one, individual subscriptions. Economies of scale: count 3,000 euros per year for a single user (2,600 without SD-Box, there are various tariffs). If you worm your way into an existing project like Big_Stat the marginal cost can be reduced to 240 euros per year (or 660 with an SD-Box). To do this, your research must be closely linked to the project objectives (contact me), the general philosophy being rather to separate the projects into different teams working on different subjects. The list of projects is on the CASD website, but without the name of the principal applicant yet.

Contacts and web links

The census at INSEE

Information on the population census on the INSEE website (files exist for all years since 2004, but it is a bit complicated to find the addresses. I suggest that go to the INSEE website and Search "Housing, individuals, activity, educational and occupational mobility, residential migration"

- <https://www.insee.fr/en/metadonnees/source/s1156> (documentation)
- <https://www.insee.fr/fr/information/2409289> file presentation (in French)
- <https://www.insee.fr/fr/statistiques/2409559> 2013 files access to files (in French)

Réseau Quetelet

- <http://www.reseau-quetelet.cnrs.fr/spip/?lang=en>

Committee on Statistical Confidentiality

- http://www.cnis.fr/cms/Accueil/activites/trois_comites/Comite_du_secret_statistique the home page (in French)
- http://www.cnis.fr/cms/Accueil/activites/trois_comites/Comite_du_secret_statistique/Les_procedures_dacces_aux_donnees_confidentielles the page describing the procedures, with a PDF document describing the procedure and the link to the form.
- comite-secret@cnis.fr the contact address

Secure Data Access Center (CASD)

- <https://casd.eu/en/home>
- <https://casd.eu/en/sources> the page that lists available sources (at CASD or elsewhere)
- <https://casd.eu/en/projects> the list of projects

Last but not least

This text is produced as part of the Big_Stat project, see http://big_stat.site.ined.fr/en/. It can be updated and corrected. Please send your opinions, additions and corrections to toulemon@ined.fr and giulia.ferrari@ined.fr.