

# Note de présentation du fichier anonyme de l'enquête Envie

Service des Enquêtes et Sondages - Ined

Avril 2026

La présente note a pour but de décrire succinctement le fichier de données anonymisées issues de l'Enquête sur la vie affective des jeunes adultes (Envie - 2023) ainsi que les méthodes d'anonymisation utilisées pour la production de ce fichier.

Ce fichier a été produit par le Service des Enquêtes et Sondages (SES) de l'Ined dans le cadre de l'école d'été « Quantification, genre et sexualité » organisée par l'Observatoire français des parcours de vie (Equipex+ LifeObs) et l'École Universitaire de Recherche Sciences sociales du genre et de la sexualité (EUR GSST). Cette formation a pour objectif de promouvoir l'utilisation de grandes enquêtes socio-démographiques auprès d'étudiant·e·s en Master en leur fournissant une base de données permettant de reproduire, aussi fidèlement que possible, les résultats d'une publication tirée du bulletin d'information scientifique de l'Ined : *Population & Sociétés*<sup>1</sup>. En concertation avec l'équipe scientifique de l'enquête Envie, la publication retenue est la suivante :

Marie Bergström, Florence Maillochon et Equipe ENVIE, 2024, "Couples, histoires d'un soir, « sexfriends » : diversité des relations intimes des moins de 30 ans", *Population et Sociétés* : 1-4. <https://doi.org/10.3917/popsoc.623.0001>

A la différence d'un Fichier de Production et de Recherche (FPR), ce fichier anonyme est téléchargeable directement depuis le catalogue de l'Ined et ce sans qu'aucun engagement ne soit pris par l'utilisateur·rice des données<sup>2</sup>. De ce fait, et pour assurer l'anonymat des répondant·e·s, les variables nécessaires à la reproduction des résultats des publications listées précédemment sont disponibles dans la base de données. A celles-ci sont adjointes d'autres variables ne présentant pas d'enjeux de réidentification particuliers et permettant de compléter les analyses. En plus de cela, plusieurs variables ont fait l'objet d'une procédure d'anonymisation. Le principe ainsi que quelques exemples de méthodes d'anonymisation pour les fichiers d'enquêtes sont décrits ci-après.

---

1. Les données ayant fait l'objet de plusieurs procédures d'anonymisation, les résultats tirés de ce fichier peuvent ne pas correspondre exactement à ceux présentés dans la publication en question (cf. partie 1).

2. La procédure permettant d'accéder aux FPR de l'Ined est détaillée ici : <https://archined.ined.fr/view/AYDXWW55Bnm4X3q6Cr5P>

# 1 Principe et méthodes pour l'anonymisation de fichiers d'enquêtes

L'anonymisation a pour but de limiter, autant que possible, le risque qu'un individu ou une organisation reconnaisse ou puisse apprendre quoi que ce soit de nouveau à propos d'une autre personne ou organisation à partir des données mises à disposition. Autrement dit, l'anonymisation a pour but de limiter, autant que possible, ce que nous appelons plus communément le risque de divulgation (Hundepool et al., 2012).

Les procédures d'anonymisation utilisées peuvent être de différentes natures. Tout d'abord, il existe des méthodes dites « non-perturbatrices » : les réponses des individus répondants sont préservées en l'état ou généralisées. Ces méthodes peuvent impliquer la suppression de variables (jugées trop sensibles par exemple) ou d'individus (présentant des caractéristiques trop rares), ou la combinaison de modalités (généralisation).

Ensuite, il existe des méthodes dites « perturbatrices » qui, quant à elles, induisent des modifications dans les données collectées : les données finales ne correspondent plus exactement aux réponses données par l'individu. Les méthodes de perturbation sont nombreuses : suppressions locales d'information, échanges de valeurs entre individus, ajout de bruit aux variables quantitatives, etc. Ces méthodes permettent d'éviter les suppressions de variables ou d'individus, mais ont un impact fort sur l'information contenue dans le fichier de données.

## 2 Conséquences sur les résultats des analyses

Le fichier anonyme de l'enquête Envie (2023) a été produit en combinant méthodes perturbatrices et non-perturbatrices. Ainsi, et même si divers procédés ont pu être mis en place pour limiter la perte d'information attribuable aux perturbations réalisées sur le fichier, les résultats tirés de ce fichier ne peuvent être utilisés à des fins scientifiques.

L'utilisation de ce fichier anonyme ne peut donc donner lieu à des publications au même titre que les Fichiers de Production et de Recherche (FPR) mis à disposition via Quetelet-Progedo ou que les fichiers détails diffusés via le Centre d'Accès Sécurisé aux Données (CASD). Les informations contenues dans les fichiers mis à disposition au CASD sont plus détaillées que celles mises à disposition via Quetelet-Progedo, qui sont elles-mêmes plus détaillées que celles mises à disposition dans le fichier anonyme.

Les Fichiers de Production et de Recherche (FPR) mis à disposition via Quetelet-Progedo peuvent être demandés via l'application de commande<sup>3</sup> et sont accessibles aux chercheur-es français-es et étranger-es, doctorant-es, post-doctorant-es, et étudiant-es à des fins de recherche, de production scientifique et dans certains cas d'enseignement. Pour plus de détails sur les conditions d'accès aux FPR, veuillez consulter la note "Accéder aux données FPR de l'Ined"<sup>4</sup>.

3. Le FPR de l'enquête Envie sera mis à disposition sur <https://data.progedo.fr/> à la fin de l'été 2026

4. <https://archined.ined.fr/view/AYDXWW55Bnm4X3q6Cr5P>

## 3 Les opérations sur le fichier de données

La première étape a consisté à identifier et sélectionner les variables nécessaires pour reproduire les résultats du *Population & Sociétés*. Ainsi, une dizaine de variables ont été sélectionnées initialement. A celles-ci ont été ajoutées des variables auxiliaires, liées aux thématiques abordées dans le bulletin. Une partie de ces variables a ensuite été traitée pour garantir l'anonymat, selon les modalités décrites ci-après<sup>5</sup>.

### 3.1 Corrections générales

Pour chaque variable, les répondants qui n'ont pas souhaité ou su répondre ont été recodés en non réponse. Pour toutes les variables catégorielles, des regroupements de modalités ont été réalisés pour que les effectifs soient systématiquement supérieurs à 20.

### 3.2 Brouillage du genre

La variable `genre_id` a été brouillée. La distribution générale reste cependant la même : on conserve le même nombre de femmes, d'hommes et de personnes non-binaires.

### 3.3 Brouillage des âges

Les âges des enquêté(e)s ont été brouillés. La structure par classe d'âge a cependant été respectée (il est donc possible de retrouver la variable `age_cat` à partir de la variable `age_rec`).

### 3.4 Brouillage des PCS

Les PCS des parents des enquêté(e)s ont été brouillés. La structure par PCS du ménage en 5 classes a cependant été respectée.

### 3.5 Construction de nouvelles variables

Concernant les variables indiquant l'apprentissage de choses sur la sexualité par les sœurs d'une part et les frères d'autre part, elles ont été regroupées pour ne garder que l'information de l'apprentissage de choses sur la sexualité par un ou plusieurs membres de la fratrie.

La variable du poids de l'enquêté (en kg) a été construite à partir de la distribution du poids déclaré dans le fichier brut.

### 3.6 Recodage des variables

Les variables de cadre de rencontre ont été simplifiées pour correspondre aux catégories utilisées dans le *Population & Sociétés*.

Les valeurs extrêmes des âges au premier baiser et au premier rapport sexuel consenti ont été regroupées dans des catégories uniques.

La variable `r1_act_rec` a été modifiée pour rester cohérente avec la variable `couple12M` malgré les regroupements effectués.

---

5. Pour plus d'informations sur le contenu du fichier anonyme, veuillez consulter le dictionnaire des codes.

### 3.7 Tableau récapitulatif

Le tableau ci-dessous donne la correspondance entre les variables d'origine du FPR et les variables du fichier anonyme.

TABLE 1 – Table de passage des variables du FPR aux variables anonymisées

Variables d'origine	Variables anonymisées
genre_id	genre_id_rec
b_age_rec	b_age_rec
infsoe, infrer	infrat
m_pcs	m_pcs_rec
p_pcs	p_pcs_rec
poids	poids_rec
r_age_c	r_age_c_rec
r_cadre	r_cadre_rec
r1_act	r1_act_rec
r1_cadre	r1_cadre_rec
r2_cadre	r2_cadre_rec
r3_cadre	r3_cadre_rec
age	age_rec

## Bibliographie - pour aller plus loin

- [1] Josep DOMINGO-FERRER et Vicenç TORRA. « Disclosure risk assessment in statistical data protection ». en. In : *Journal of Computational and Applied Mathematics* 164-165 (mars 2004), p. 285-293. ISSN : 03770427. DOI : 10.1016/S0377-0427(03)00643-5. URL : <https://linkinghub.elsevier.com/retrieve/pii/S0377042703006435>.
- [2] *Guide du secret statistique*. Rapp. tech. Institut national de la statistique et des études économiques, p. 2023. URL : [https://www.insee.fr/fr/statistiques/fichier/1300624/guide\\_secret\\_janv\\_2023.pdf](https://www.insee.fr/fr/statistiques/fichier/1300624/guide_secret_janv_2023.pdf).
- [3] Anco HUNDEPOOL et al. *Statistical disclosure control*. anglais. ISSN : 1942-9088. Chichester, West Sussex, United Kingdom, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord : John Wiley & Sons Inc., 2012. ISBN : 978-1-119-97815-2.
- [4] *L'anonymisation de données personnelles*. Mai 2020. URL : <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>.
- [5] *Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte. Guide pour la recherche*. Rapp. tech. Institut des sciences humaines et sociales du Centre National de la Recherche Scientifique, 2021. URL : [https://www.inshs.cnrs.fr/sites/institut\\_inshs/files/pdf/Guide\\_rgpd\\_2021.pdf](https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/Guide_rgpd_2021.pdf).
- [6] Michał PIETRZAK. « Statistical Disclosure Control Methods for Microdata from the Labour Force Survey ». In : *Acta Universitatis Lodzianensis. Folia Oeconomica* 3.348 (juin 2020), p. 7-24. ISSN : 2353-7663, 0208-6018. DOI : 10.18778/0208-6018.348.01. URL : <https://czasopisma.uni.lodz.pl/foe/article/view/3992>.