



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

DONNÉES D'ENQUÊTES À L'INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES (INED) : DE LA PRODUCTION À L'ARCHIVAGE

Journée d'études du 26 mai 2021

« Documents et données : l'université à l'ère de la transition numérique »

Arianna CAPORALI, ingénieure de recherche et responsable du DataLab

Lindsay FUSFELD, archiviste et déléguée à la protection des données (DPD)

Julie BARON, chargée de la documentation et la mise à disposition des enquêtes

L'Institut national d'études démographiques (Ined)

- Un institut de recherche public spécialisé dans les études sur la population, notamment au moyen d'enquêtes
- Environ 250 personnes; 11 unités de recherche et 7 services d'appui à la recherche
- Un Service des enquêtes et sondages (SES) pour la collecte de données
- Un DataLab pour la mise à disposition des données en collaboration avec Quetelet PROGEDO Diffusion
- Une archiviste/déléguée à la protection des données (DPD)

« Aussi ouvert que possible... »

- Loi n°2016-1321 du 7 octobre 2016 pour une République numérique
- Plan national pour la science ouverte, juillet 2018 : « structurer et ouvrir les données de la recherche »
- Réflexions en cours à l'Ined pour une politique d'ouverture des données (obligation de mise à disposition)
- Obligations pour certains projets (ANR, par ex.)

« ...aussi fermé que nécessaire »

- Informations non librement communicables
 - Protection de la vie privée, données de santé ...
 - Secret en matière de statistiques lorsque sont en cause des données collectées au moyen de questionnaires ayant trait aux faits et comportements d'ordre privé (art. L213-2 du Code du patrimoine)
- Règlement général sur la protection des données (RGPD)
 - Données à caractère personnel, dont données « sensibles »
 - Limitation des finalités, mais réutilisation à des fins de recherche = finalité ultérieure compatible (art. 5.1.b du RGPD)
 - Les données anonymes/anonymisées ne sont pas/plus des données à caractère personnel

Fichiers de données d'une enquête (1)

Type de fichier	Contenu	Accès	Conservation
Fichier des contacts	Coordonnées des personnes enquêtées	Accès restreint à quelques personnes de l'équipe du projet	Destruction dès que possible
Fichier brut	Toutes les informations collectées, sauf données nominatives	Accès restreint à quelques personnes de l'équipe du projet	Conservation sécurisée et versement aux Archives Nationales
Fichier de production intermédiaire (FPI)	Fichier brut après traitements statistiques (ex., apurements, pondérations, variables calculées)	Accès restreint à l'équipe du projet	Conservation sécurisée et versement aux Archives Nationales
Fichier d'étude	FPI après premiers traitements statistiques visés à limiter les risques de ré-identification des enquêtés	Accès restreint à l'équipe du projet et aux chercheurs associés aux publications	Conservation sécurisée et versement aux Archives Nationales

Fichiers de données d'une enquête (2)

Type de fichier	Contenu	Accès	Conservation
Fichier de production et de recherche (FPR)	Fichier d'étude après tous les traitements nécessaires pour limiter les risques de ré-identification des enquêtés	Accessible à la communauté scientifique via Quetelet PROGEDO Diffusion	Conservation sécurisée (Ined et Huma-Num) et versement aux Archives Nationales
Fichier des données contextuelles	Données agrégées (en libre accès ou pas) qui peuvent venir éclairer les analyses des données de l'enquête	Accès restreint à l'équipe du projet si le respect du secret statistique le nécessite (ex. niveau géographique fin)	Conservation sécurisée et versement aux Archives Nationales
Fichier d'étude (ou FPR) enrichi des données contextuelles	Fichier mettant en relation les données de l'enquête avec les données contextuelles	Accès restreint à l'équipe du projet si nécessaire (ex. niveau géographique fin)	Conservation sécurisée et versement aux Archives Nationales

Mise à disposition des FPR

- Accès gratuit pour une utilisation dans une finalité de recherche; toute utilisation commerciale est exclue
- Accès possibles uniquement aux étudiants et chercheurs ayant une adresse e-mail institutionnelle
- Procédure d'accès mise en place suite à la Loi pour une République Numérique et à l'avis du Comité du Secret Statistique (CSS) du 14 décembre 2018



Pour accéder aux FPR il faut être rattaché à une institution habilité par le CSS. La liste des institutions habilitées est régulièrement mise à jour

Mise à disposition des FPR

- 1^{ère} demande d'accès exclusivement :
 - L'Ined fait signer au demandeur un engagement de confidentialité qu'il transmet au secrétariat du CSS pour habilitation
 - Le CSS transmet une dérogation pour signature aux Archives de France
 - En cas de première demande, délai à prévoir de plusieurs semaines
- Pour chaque demande d'accès ultérieur, signature d'un formulaire d'engagement à destination de l'Ined uniquement, sans passer par le CSS
- Demande d'accès d'étudiants: pour accélérer le processus, l'enseignant peut anticiper leur demande d'habilitation en fournissant la liste des intéressés (même avant le début du cours)

Articulation avec la politique d'archivage (1)

- Actualisation du tableau de gestion des archives avec intégration des différents fichiers de données d'enquête suite aux travaux internes sur les données confidentielles
- Conservation des données en conformité avec la réglementation
 - Principe de conservation limitée, en particulier en « base active » (archives courantes)
 - Art. 5.1.e du RGPD permet une conservation plus longue à des fins de recherche scientifique (archives intermédiaires)
 - Obligation de verser certaines archives aux Archives nationales (archives définitives)

Articulation avec la politique d'archivage (2)

- Archivage intermédiaire assuré par le service des enquêtes
 - « Data and documentation acquisition procedure » : procédure en cours relative au « versement » au SES par l'équipe de l'enquête
 - DUA longue afin de répondre aux besoins de la recherche
 - Mesures de sécurité en place, conformément à l'art. 89 du RGPD
- Versement définitif des données en lien avec l'archiviste de l'Ined
 - Versement des versions successives des données collectées, métadonnées
 - Ne sont pas versées : les données directement identifiantes (coordonnées, données nominatives)

- Équilibre entre **volonté d'ouverture** des données d'enquêtes et la **nécessaire protection** des données
- Une meilleure articulation à trouver entre la **conservation limitée** des données (RGPD), les besoins de la recherche sur le long terme, et **l'archivage historique**