



Conseil national
de l'information statistique

COMITÉ DU LABEL
DE LA STATISTIQUE PUBLIQUE

Montrouge, le 16 décembre 2022
N°2022_31544_DG75-L002

AVIS DE CONFORMITÉ

Enquête pour la collecte de la base de sondage de l'enquête Génération 2021

Service producteur : Centre d'études et de recherches sur les qualifications (Céreq)

Opportunité : avis favorable émis le 16 mai 2022 par la Commission « Emploi, qualification et revenus du travail »

Réunion du Comité du label du 19 octobre 2022 (commission « Ménages »)

Commission	Ménages
Type d'avis	Avis de conformité
Label d'intérêt général et de qualité statistique	Oui
Caractère obligatoire	Oui
Période de validité	2023
Publication JO	Oui
Périodicité	Ponctuelle

Descriptif de l'opération

Le dispositif Génération a été mis en place à la fin des années 1990. Les deux ministères de tutelle du Céreq, c'est-à-dire le ministère en charge de l'Éducation et le ministère en charge de l'Emploi, sont à l'origine de la demande.

Pour chaque enquête Génération, une collecte de fichiers d'élèves a été réalisée pour construire la base de sondage : pour partie auprès de divers ministères et organismes centralisateurs de fichiers d'élèves, pour partie directement auprès des établissements de formation. Jusqu'à la Génération 2017, l'enquête auprès des établissements de formation n'avait pas fait l'objet d'une demande de labellisation par le Comité du label. En vue de l'enquête auprès de la Génération 2021, le Céreq a décidé de faire de cette collecte une enquête de la statistique publique à part entière.

Le Comité du label lui-même a encouragé la labellisation de cette enquête auprès des établissements de formation face aux difficultés dont le Céreq a fait part au sujet de cette phase préliminaire à l'enquête Génération 2017. L'entrée en vigueur du RGPD a rendu les établissements de formations plus réticents à transmettre des données nominatives. L'objectif de la labellisation de cette collecte,

avec caractère obligatoire, est de fournir aux établissements de formation une base légale pour fournir les données nominatives demandées par le Céreq.

Cette collecte est une étape indispensable pour la réalisation de l'enquête Génération 2021 du Céreq. Il n'existe pas de base nationale d'élèves nominative couvrant l'ensemble des sortants de formations initiales dispensées en France (métropole et DOM) pour une année donnée. Le Céreq doit donc constituer cette base de sondage en amont de chaque nouvelle enquête Génération, à partir de différentes sources et en opérant divers traitements.

Deux principales opérations de collecte ont lieu pour récupérer des listes nominatives d'élèves ou d'étudiants inscrits dans un établissement de formation initiale français en 2020-2021 et 2021-2022. Dans un premier temps, une collecte de données nominatives est réalisée en récupérant des bases nationales centralisées auprès de divers ministères et organismes. Dans un second temps, une collecte complémentaire auprès de tous les autres établissements de formation est réalisée par le Céreq : le Céreq demande l'opportunité pour cette partie de la collecte.

La liste des établissements à enquêter est établie à partir d'une sélection dans la BCE (Base centrale des établissements – ministère en charge de l'Éducation), en sélectionnant les établissements délivrant de la formation initiale ne figurant pas dans les bases centralisées.

La collecte d'inscrits en 2020-2021 et 2021-2022 permet de faire des comparaisons de fichiers afin de repérer les poursuites d'études : les inscrits en 2020-2021 et poursuivants en 2021-2022 ne seront pas conservés dans la base de sondage de l'enquête Génération 2021.

L'enquête s'adresse aux établissements de formation délivrant de la formation initiale certifiante, situés en France. Les établissements de formation pour lesquels le Céreq collecte des fichiers d'inscrits directement auprès des ministères centralisateurs ne sont pas concernés par cette enquête. Elle a pour champ géographique les établissements de formation situés en France métropolitaine ou dans un département ou région d'outre-mer.

Les informations demandées aux établissements de formation portent sur les élèves inscrits au cours des deux années scolaires 2020-2021 et 2021-2022. Elles doivent permettre d'identifier les individus et de caractériser la formation suivie au cours de l'année scolaire concernée. Plus précisément, les informations demandées sont les suivantes : nom et prénom(s), INE, adresse, numéro(s) de téléphone, mail, date et lieu de naissance, sexe, classe et spécialité de formation, obtention ou non du diplôme préparé.

Le Céreq sera le seul utilisateur des données collectées. Un plan de gestion des données décrit l'ensemble des opérations réalisées ainsi que les durées de conservation des données. Environ 3 500 établissements sont sollicités.

~~~

### ***Le Comité du label de la statistique publique émet les recommandations ou observations suivantes :***

#### **Préambule**

- Le projet de collecte d'information soumis à l'examen du Comité du label est une phase préliminaire de l'enquête Génération 2021. L'objectif est de contribuer à la constitution d'une base de données nationale des jeunes sortant du système éducatif afin de pouvoir y sélectionner l'échantillon qui sera ultérieurement interrogé dans le cadre de Génération 2021.
- Deux principales opérations ont lieu pour compiler les listes nominatives constituant cette base de données :
  - un recueil de bases nationales centralisées auprès de divers ministères et organismes ;
  - une collecte complémentaire, par voie d'enquête, auprès de tous les établissements de formation non présents dans les bases centralisées, ou pour lesquels les informations de

contact disponibles dans ces bases sont insuffisantes (universités, écoles d'ingénieurs, écoles de commerce...) :

- deux fichiers d'inscrits sont collectés auprès de chaque établissement, celui des inscrits durant l'année scolaire 2020-2021 et celui des inscrits durant l'année scolaire 2021-2022 ;
  - les fichiers d'un même établissement sont d'abord comparés pour identifier les sortants potentiels, les fichiers des différents établissements sont ensuite comparés à l'aide de procédures statistiques d'appariement.
- La portée du présent avis couvre exclusivement le recueil des bases de la collecte complémentaire. L'inscription de cette dernière dans le cadre des enquêtes de la statistique publique vise à sécuriser le processus de production et à garantir auprès des institutions l'application des règles du secret statistique dans une perspective d'augmentation des taux de retour et, *in fine*, l'amélioration de la qualité de l'enquête Génération. L'inscription dans le cadre de la loi de 1951 sur l'obligation, la coordination et le secret en matière de statistiques circonscrit cette collecte de données à des finalités exclusivement statistiques.

### Remarques générales

- Compte tenu de la spécificité du dossier, le Comité rappelle au service qu'il lui revient de s'assurer du respect de la réglementation « *informatique et libertés* » et de tenir compte des observations énoncées ci-après :
  - ne faisant pas partie du service statistique public, le Céreq doit se conformer au régime de droit commun s'agissant de la protection des données. Si des données sensibles devaient être collectées, le Céreq devrait mettre en œuvre la procédure de saisine préalable de la Cnil ou rechercher un autre fondement parmi ceux énumérés à l'article 9.2 du RGPD ;
  - la base légale du traitement au sens de l'article 6 du RGPD n'a pas été indiquée dans le dossier et doit être précisée ;
  - la nécessité d'une extension de la collecte d'informations aux responsables de l'individu, afin d'accroître les chances de contact, doit être évaluée en conformité avec le principe de minimisation disposé par l'article 5 du RGPD ;
  - les mentions informatives des différents courriers ou plaquettes informatives à l'attention des directeurs d'établissements ou des correspondants techniques devront être complétées ou précisées afin de comporter l'ensemble des éléments énumérés par les articles 13 et 14 du RGPD ;
  - le cahier des charges de l'appel d'offres présenté au Comité ne comporte pas de clauses relatives à la protection des données, en dehors d'une annexe intitulée « préconisations en vue de l'application du RGPD », alors que la relation avec un sous-traitant doit être régie par un contrat contenant les clauses énumérées à l'article 28-3 du RGPD ;
  - compte tenu du nombre important des personnes concernées par le traitement et de la collecte de l'identifiant national élève (INE), une analyse d'impact relative à la protection des données (AIPD) pourrait devoir être effectuée et, le cas échéant, transmise pour avis à la Cnil préalablement à la mise en œuvre du traitement ;
  - le comité demande à être informé des dispositions que le service jugera utile de mettre en place pour répondre à ses obligations.
- Le Comité souligne les enjeux en termes de sécurisation de la base de données. Il note que le Céreq a revu son protocole afin de davantage sécuriser la transmission des données et prend acte des diverses dispositions prises pour assurer la sécurité des données personnelles

collectées. Il prend note en outre de l'engagement du Céreq à ne conserver la base de sondage que le temps nécessaire à l'échantillonnage de l'enquête Génération 2021.

- Le Comité encourage le service à poursuivre sa collaboration avec l'ensemble des parties prenantes afin de bénéficier de toute amélioration des systèmes d'information permettant de minimiser la charge de collecte auprès des établissements.

## Méthodologie

- Le comité invite le service à se rapprocher de l'Insee et du CASD afin d'étudier les conditions d'utilisation du Code Recherche Non Signifiant (CRNS) ainsi que les avantages que ce dernier pourrait procurer aux différentes opérations d'appariements nécessaires à la constitution de la base de sondage. Une première description du CRNS est disponible dans l'article « *Appariements sécurisés de données personnelles : un exemple pour la recherche* » (Yacine El Bouhairi, Kamel Gadouche, Rémy Marquier, JMS 2022).
- Le Comité note positivement les échanges du service avec l'équipe en charge du dispositif *Inserjeunes* dans une optique de capitalisation, de mutualisation et de partage d'expérience sur les dispositifs d'appariements de données personnelles.
- Le Comité demande au service de documenter les éventuelles conséquences de l'intégration dans le champ de la base de sondages, et ultérieurement dans celui de l'enquête, des étudiants étrangers en France ou français étudiant à l'étranger.
- Afin d'identifier les conséquences des innovations successives apportées aux modalités de constitution de la base de sondages, le Comité demande au service de présenter l'évolution du taux de sur-couverture dans une perspective historique. Cette présentation pourra être intégrée lors du prochain dossier qui sera soumis à l'examen du Comité.
- Le Comité prend acte que les demandes d'extension formulées par les partenaires du service se traduiront par un suivi de collecte adapté garantissant la qualité des données sur les différents sous-champs concernés.

## Lettres-avis

- Le Comité suggère au service d'insérer dans les lettres-avis un lien vers les résultats de l'enquête précédente afin d'attester de l'utilité et de la pertinence du recueil d'information.
- Le Comité informe le service que les conclusions de travaux en cours au sein de l'Insee, et s'appuyant sur la littérature de référence pour une langue administrative de qualité et la simplification des documents administratifs, préconisent de démarrer les lettres-avis en s'adressant directement à l'enquêté, avec des informations concrètes et qui le concernent lui-même plutôt que de présenter, dès la phrase d'accroche, l'enquête et sa finalité. Le Comité invite le service à évaluer si ces recommandations s'appliquent aux courriers à destination des directeurs d'établissements ou des correspondants techniques.
- Le Comité suggère d'intégrer dans les lettres-avis la formule standard de la mention du Cnis et du label d'intérêt général et de qualité statistique : « Vu l'avis favorable du Conseil national de l'information statistique, cette enquête est reconnue d'intérêt général et de qualité statistique, en application de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Elle a obtenu le visa n°XXXX du Ministre de l'Économie, des Finances et de la Souveraineté industrielle et numérique, valable pour l'année 2023 – Arrêté en date du XX/XX/XXXX. Cette enquête est obligatoire. En cas de défaut de réponse, les établissements enquêtés peuvent être l'objet de l'amende prévue à l'article 131-13 du Code pénal. » Le numéro mentionné est le numéro de visa obtenu sur la proposition du Comité du label et non le numéro de l'avis d'opportunité.

**Le Comité du label de la statistique publique émet un avis de conformité à l'enquête pour la collecte de la base de sondage de l'enquête Génération 2021 et, par délégation du Président du Cnis, lui attribue le label d'intérêt général et de qualité statistique avec proposition d'octroi de l'obligation.**

***Cet avis est valide pour l'année de collecte 2023.***

La Présidente du Comité du label de la  
statistique publique

Signé : Dominique BONNANS