



HAL
open science

Enquête Génération 2010 3 ème interrogation automne 2017. Bilan méthodologique

Mady Cissé, Emilie Gaubert, Zora Mazari, Manon Olaria

► To cite this version:

Mady Cissé, Emilie Gaubert, Zora Mazari, Manon Olaria. Enquête Génération 2010 3 ème interrogation automne 2017. Bilan méthodologique. [Rapport Technique] Céreq Etudes n°27, Céreq. 2019, 76p. halshs-02889462

HAL Id: halshs-02889462

<https://shs.hal.science/halshs-02889462>

Submitted on 3 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CÉREQ ÉTUDES

27
2019

Enquête Génération 2010 3^{ème} interrogation automne 2017

Bilan méthodologique

Coord. Manon OLARIA

*Auteurs : Mady CISSE, Émilie GAUBERT,
Zora MAZARI, Manon OLARIA.*

*Équipe ingénierie et gestion d'enquête,
Département des Entrées et Évolutions dans la Vie Active (DEEVA),
Céreq.*

Enquête Génération 2010

Bilan méthodologique

3^{ème} interrogation – automne 2017

Coordinatrice :

Manon OLARIA

Auteurs :

Mady CISSE

Émilie GAUBERT

Zora MAZARI

Manon OLARIA

*Membres de l'Équipe ingénierie et gestion d'enquête
Département entrées et évolutions dans la vie active du Céreq.*

Synthèse :

Ce document présente un panorama complet des travaux de production, de collecte et de mise à disposition des données liés à l'enquête 2017 auprès de la Génération 2010. Après une brève présentation du dispositif des enquêtes Génération et plus spécifiquement de l'enquête 2017, ce document détaille toutes les étapes de réalisation de cette enquête dans un ordre chronologique.

Après une étape de préparation (constitution de l'échantillon, développement du CATI, mise à jour des coordonnées, pré-fidélisation, lettre avis...), la collecte par téléphone s'est déroulée d'octobre à décembre 2017. Enfin, l'ensemble des traitements post collecte sont décrits (création des bases de données, apurement, codification, traitement de la non réponse, pondération...).

Sommaire

I. PRESENTATION DE L'ENQUETE 2017 AUPRES DE LA GENERATION 2010	5
1. LE DISPOSITIF DES ENQUETES GENERATION	5
2. LE CHAMP DE L'ENQUETE	7
3. DESCRIPTIF ET OBJECTIFS DE LA GENERATION 2010 A 7 ANS	8
4. LE QUESTIONNAIRE	8
a) <i>Descriptif du questionnaire</i>	8
b) <i>Nouveau module sur la thématique de l'encadrement</i>	11
c) <i>Module sur la vie en couple et les enfants</i>	12
d) <i>Autres évolutions</i>	12
e) <i>Le calendrier professionnel</i>	13
5. LE CALENDRIER DE L'ENQUETE	14
II. PREPARATION DE LA COLLECTE	15
1. CONSTITUTION DE L'ECHANTILLON	15
2. DEVELOPPEMENT DU CATI (ENQUETE TELEPHONIQUE)	15
a) <i>Logiciel utilisé</i>	15
b) <i>Technique de développement du calendrier</i>	15
c) <i>Test du CATI et du calendrier associé</i>	15
3. MISE A JOUR DES COORDONNEES POSTALES	16
4. FIDELISATION	17
a) <i>Site internet du Céreq</i>	17
b) <i>Réseaux sociaux</i>	18
c) <i>Adresse de messagerie</i>	18
d) <i>Pré-fidélisation</i>	18
5. ACTUALISATION DES COORDONNEES TELEPHONIQUES	19
a) <i>Actualisation par l'enquêté lors de la pré-fidélisation</i>	19
b) <i>Actualisation par un prestataire</i>	19
c) <i>Ordonnance des numéros de téléphone</i>	21
6. LETTRE AVIS	22
III. LA COLLECTE PAR TELEPHONE	23
1. CALENDRIER ET ORGANISATION GENERALE DE LA COLLECTE	23
2. LE SUIVI DE LA COLLECTE EN CHIFFRES	24
3. LES REGLES DE RAPPEL	25
4. DES MESSAGES LAISSES SUR REPONDEUR	26
5. SUIVI TECHNIQUE ET PERSONNE « QUALITE »	26
IV. BILAN DE COLLECTE	27
V. LES TRAITEMENTS EN AVAL	28
1. LES TABLES BRUTES LIVREES PAR LE PRESTATAIRE	28
2. LA CREATION DES TABLES	28
a) <i>Traitements et contrôles des fichiers de base transmis par le prestataire</i>	28
b) <i>Création de la base individus</i>	29
c) <i>Calendriers professionnels et habitat</i>	29
d) <i>Création des nouvelles séquences d'emploi et de non-emploi</i>	30
e) <i>Les recopies</i>	32
3. LA CODIFICATION DES VARIABLES	33

a) Contrats de travail.....	33
b) Variables à créer en amont de la codification du secteur d'activité	34
c) Secteur d'activité de l'établissement employeur.....	35
d) Profession des périodes d'emploi.....	37
e) Variables de localisation géographique.....	39
f) Codification des ouverts.....	39
4. LE TRAITEMENT DES SALAIRES (PRIMES INCLUSES).....	40
5. ANONYMISATION DES DONNEES ET FINALISATION	44
6. FORMATS, LABELS ET DICTIONNAIRE DES VARIABLES.....	44
7. LES TABLES FINALES.....	45
VI. LA PONDERATION FINALE	46
1. ÉCHANTILLONNAGE.....	46
2. EXPLOITATION DE L'HISTORIQUE D'APPEL POUR OBTENIR DES PARADONNEES POUR LA MODELISATION DE LA NON-REPONSE....	46
3. MODELISATION DE LA NON-REPONSE.....	47
a) Modèle logistique de non-réponse.....	47
b) Groupes de réponse homogène	48
c) Pondération avant calage.....	49
4. CALAGE SUR LES MARGES DES ENQUETES A 3 ET 5 ANS.....	49
a) Calage de l'enquête à 5 ans sur l'enquête à 3 ans	51
b) Calage de l'enquête à 7 ans sur les précédentes.....	51
5. LA PONDERATION FINALE.....	52
6. DIFFICULTES RENCONTREES.....	53
a) Une modélisation en deux étapes : contact puis réponse.....	53
b) Des variables trop explicatives de la non-réponse	54
c) Calage sur la situation un mois donné plutôt qu'à la date d'enquête	54
VII. ANNEXES.....	56
ANNEXE 1 - TABLE DES ILLUSTRATIONS.....	56
ANNEXE 2 - PRESENTATION DE L'OUTIL RNVP.....	58
ANNEXE 3 - LETTRE DE PRE-FIDELISATION.....	60
a) Lettre envoyée par courrier postal.....	60
b) Lettre envoyée par mail	61
ANNEXE 4 - CAWI DE PRE-FIDELISATION	62
a) Écran d'accueil	62
b) Affichage des coordonnées	62
c) Premiers résultats présentés après validation	63
ANNEXE 5 - LETTRE AVIS DE CONTACT AVEC LES JEUNES	65
a) Lettre envoyée par courrier postal.....	65
b) Lettre envoyée par courrier électronique.....	66
c) Pièce jointe aux courriers	67
ANNEXE 6 - NOMENCLATURE NAF REV2 EN 88 DIVISIONS (A88).....	68
ANNEXE 7 - DEFINITION DES VARIABLES UTILISEES DANS SICORE PCS	70
ANNEXE 8 - LISTE DES QUESTIONS OUVERTES.....	73
ANNEXE 9 - VARIABLES EN ENTREE DU MODELE DE REGRESSION LOGISTIQUE POUR LA PONDERATION.....	74

I. Présentation de l'enquête 2017 auprès de la Génération 2010

1. Le dispositif des enquêtes Génération

Les enquêtes Génération s'intéressent à l'insertion et au cheminement des sortants du système éducatif lors de leurs premières années de vie active. Elles ont pour objectifs principaux de produire des indicateurs d'insertion (taux d'emploi, taux de chômage, taux d'emploi à durée indéterminée, etc.), selon les niveaux de formation, les filières, les spécialités, à destination des acteurs publics et sociaux. Elles proposent ainsi des informations qui contribuent à la compréhension des processus d'insertion et des différenciations des parcours en début de carrière.

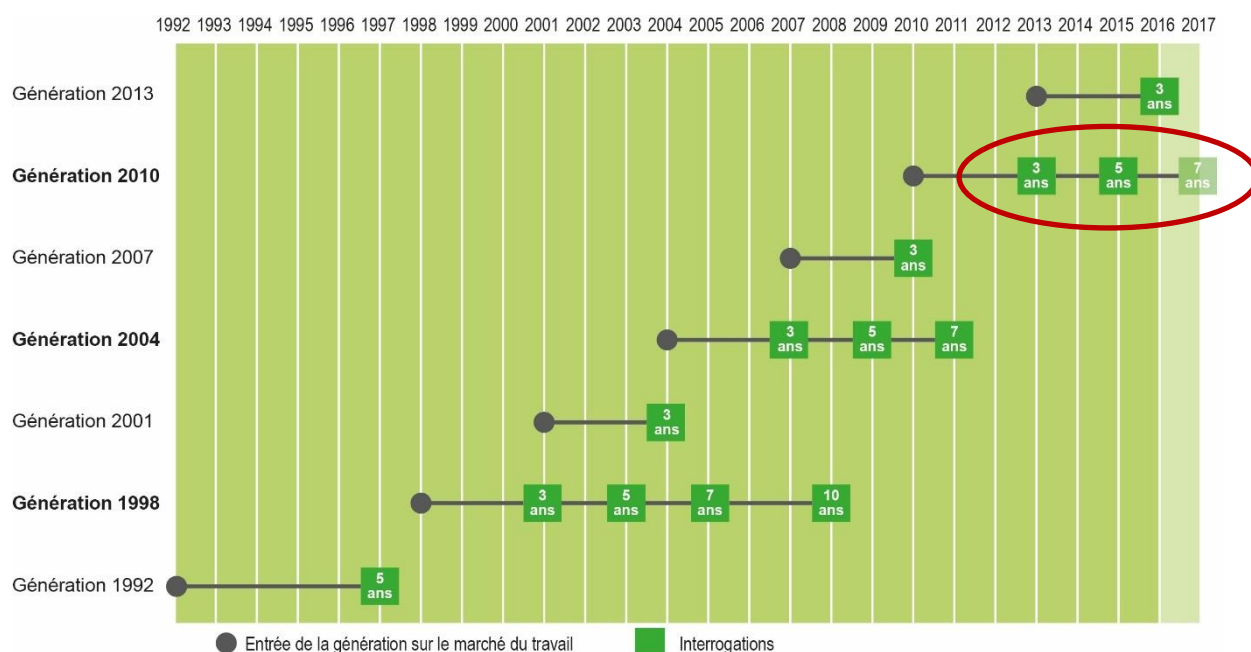
Encadré 1 - Le Centre d'études et de recherches sur les qualifications

Créé au début des années 1970, devenu établissement public en 1985, le Centre d'études et de recherches sur les qualifications (Céreq) est placé sous la tutelle des ministères en charge de l'Éducation nationale et du Travail. Il a pour missions de développer des études et des recherches, de collecter et d'exploiter des données originales dans le domaine de la relation formation-emploi, et de formuler des avis et propositions destinés à éclairer les choix en matière de politiques de formation.

La première enquête Génération a été effectuée en 1997 auprès de jeunes sortis du système éducatif en 1992 et issus de tous les niveaux de formation. Un dispositif régulier d'interrogations a été mis en place à partir de l'enquête effectuée en 2001 auprès de sortants du système éducatif en 1998 : les enquêtes s'enchaînent au rythme d'une génération nouvelle de sortants interrogée tous les trois ans et avec une alternance entre une enquête Génération pleine » et une « mini » enquête Génération.

L'enquête Génération « pleine » comprend plusieurs interrogations. La première interrogation, trois ans après la sortie du système éducatif, est principalement tournée vers la production d'indicateurs et des usages institutionnels. Les interrogations suivantes, à cinq et sept ans, sont plutôt centrées sur les usages analytiques (notamment sur la question des parcours, des mobilités sur moyen terme). Ces ré-interrogations permettent aussi d'approfondir certains constats issus de l'exploitation de la première interrogation (par exemple sur les insertions les plus problématiques).

La « mini » enquête Génération se limite à la première interrogation à trois ans, avec une taille d'échantillon plus réduite, et un questionnement allégé. Cette enquête a vocation à réactualiser les indicateurs d'insertion sur les trois premières années d'insertion selon une grille d'analyse moins fine que pour les Générations « pleines ».

Figure 1 - Calendrier des enquêtes Génération

Le cœur de l'enquête, pour la première interrogation comme pour les autres quand il y a lieu, est constitué par le calendrier professionnel qui permet de suivre mois par mois la situation des jeunes à l'issue de leur formation initiale, et de décrire les séquences successives d'emploi et de non-emploi.

Les enquêtes Génération permettent aussi de répondre à des demandes d'extensions nationales ou régionales portant sur les jeunes issus de certains niveaux ou spécialités de formations ou sur les jeunes ayant bénéficié de certaines mesures pour la formation. Les échantillons peuvent donc être élargis en conséquence.

Tableau 1 - Détail des enquêtes Génération et effectifs de répondants

		Année d'interrogation	Nombre de répondants	Dont liés aux extensions d'échantillon
Génération 1992		1997	26 000	-
Génération 1998	Première interrogation	2001	55 000	19 900
	Deuxième interrogation	2003	22 000	1 900
	Troisième interrogation	2005	16 000	1 000
	Quatrième interrogation	2008	11 000	900
Génération 2001		2004	25 000	15 000
Génération 2004	Première interrogation	2007	65 000	37 000
	Deuxième interrogation	2009	20 000	2 700
	Troisième interrogation	2011	12 400	2 000
Génération 2007		2010	29 000	26 000
Génération 2010	Première interrogation	2013	38 500	13 500
	Deuxième interrogation	2015	13 710	1 500
	Troisième interrogation	2017	8 900	-
Génération 2013		2016	22 700	13 800

2. Le champ de l'enquête

La Génération 2010 concerne les « primo sortants » de formation initiale en 2009-2010 (année scolaire). Cette restriction aux primo-sortants signifie que les sortants de formation qui avaient déjà interrompu leurs études pour une durée supérieure à un an avant l'année scolaire considérée sont hors champ. Tous les niveaux et domaines de formations sont concernés par l'enquête. De façon plus précise, les critères d'éligibilité pour être dans le champ retenu, nommé ensuite « champ Céreq », sont les suivants :

- avoir été inscrit dans un établissement de formation en France (Métropolitaine + DOM) durant l'année scolaire 2009-2010 ;
- avoir quitté le système éducatif entre octobre 2009 et octobre 2010¹, diplômé ou non ;
- ne pas avoir interrompu ses études durant une année ou plus avant l'année scolaire 2009-2010 (sauf pour raison de santé) ;
- ne pas avoir repris ses études pendant l'année qui a suivi l'entrée sur le marché du travail ;
- avoir 35 ans ou moins en 2010 ;
- être localisé en France (Métropolitaine + DOM) au moment de l'enquête (ce qui exclut donc notamment les personnes résidant à l'étranger à la date d'enquête).

Quelques points particuliers concernent l'application de ces critères :

- Bien qu'il s'agisse de contrats de travail au même titre que les contrats de professionnalisation ou que les contrats de qualification, les contrats d'apprentissage sont considérés comme relevant de la formation initiale. Une personne repérée comme sortant de formation en 2009-2010 qui poursuit par un contrat d'apprentissage en 2010-2011 est donc considérée en poursuite d'études, donc hors champ. Ce n'est pas le cas s'il s'agit d'un contrat de professionnalisation.
- Une personne sortie d'un établissement de formation en 2009-2010 qui poursuit des cours par correspondance ou des cours du soir en 2010-2011 est considérée comme en poursuite d'études, donc hors champ, si elle n'a pas d'emploi en parallèle. Si elle a un emploi en parallèle, sa situation d'emploi prime.
- Une personne sortie d'un établissement de formation en 2009-2010 mais poursuivant ses études en 2010-2011 comme élève fonctionnaire n'est pas considérée comme étant en poursuite d'études ; elle est considérée en emploi.

Les différences de champ qui existent entre les quatre cohortes enquêtées sont mineures :

- Les sortants de classes de 6ème et 5ème, inclus dans le champ de la Génération 1998 et de la Génération 2001 sont exclus du champ depuis la Génération 2004 ;
- les étudiants étrangers sortants de l'Université (repérables dans le Système d'Information sur le Suivi de l'Étudiant, SISE) étaient hors champ pour la Génération 1998 et la Génération 2001 ; ils sont inclus dans le champ depuis la Génération 2004 ;
- la période de référence pour la date de sortie était l'année civile pour la Génération 1998 et la Génération 2001 ; c'est l'année scolaire depuis la Génération 2004 (octobre n à octobre n+1) ;

¹ Pour certaines formations (école fonction publique, sport/animation, santé/social, et thèse) il y a une prolongation jusqu'à décembre 2010

- la réforme de mastérisation des concours d'enseignement, mise en place sous régime transitoire l'année 2010, et surtout le rattachement administratif des IUFM aux universités, n'ont pas permis d'intégrer dans cette enquête Génération les sortants de première année de cette formation.

Encadré 2 - Une enquête de la statistique publique

L'enquête 2017 auprès de la Génération 2010, extensions comprises, est une enquête de la statistique publique et relève à ce titre de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Celle-ci définit le cadre de procédures destinées à garantir l'intérêt, la pertinence et la qualité des enquêtes publiques, ainsi que la confidentialité des informations collectées.

L'opportunité de l'enquête 2017 auprès de la Génération 2010 a ainsi été discutée au Comité national de l'information statistique (CNIS), qui a émis un avis favorable en mars 2012 puis lui a attribué un label d'intérêt général et de qualité statistique en juillet 2017 (label n° 2017X717AU).

Les réponses apportées à l'enquête sont confidentielles. Les fichiers des réponses qui sont mis à disposition des chercheurs sont anonymes et ne comportent pas d'informations susceptibles de permettre une identification directe ou indirecte des personnes enquêtées. La réalisation de l'enquête et sa diffusion a fait l'objet d'une déclaration auprès de la Commission nationale informatique et libertés (CNIL).

3. Descriptif et objectifs de la Génération 2010 à 7 ans

La Génération 2010 est une Génération « pleine ». Ainsi, elle a été interrogée une première fois en 2013 et une seconde fois en 2015. La troisième interrogation a eu lieu en 2017.

L'objectif de cette troisième interrogation est, comme pour la seconde interrogation, d'élargir la fenêtre d'observation au-delà de l'entrée dans la vie active, d'analyser les mobilités professionnelles des jeunes et les débuts de carrière qui se dessinent après sept ans passés sur le marché du travail. De plus, l'enquête explore, parallèlement aux trajectoires professionnelles, les trajectoires familiales des jeunes (accès à l'autonomie résidentielle, construction de la famille, etc.).

Pour cette troisième interrogation, sont concernés uniquement les jeunes ayant répondu à la seconde interrogation en 2015.

4. Le questionnaire

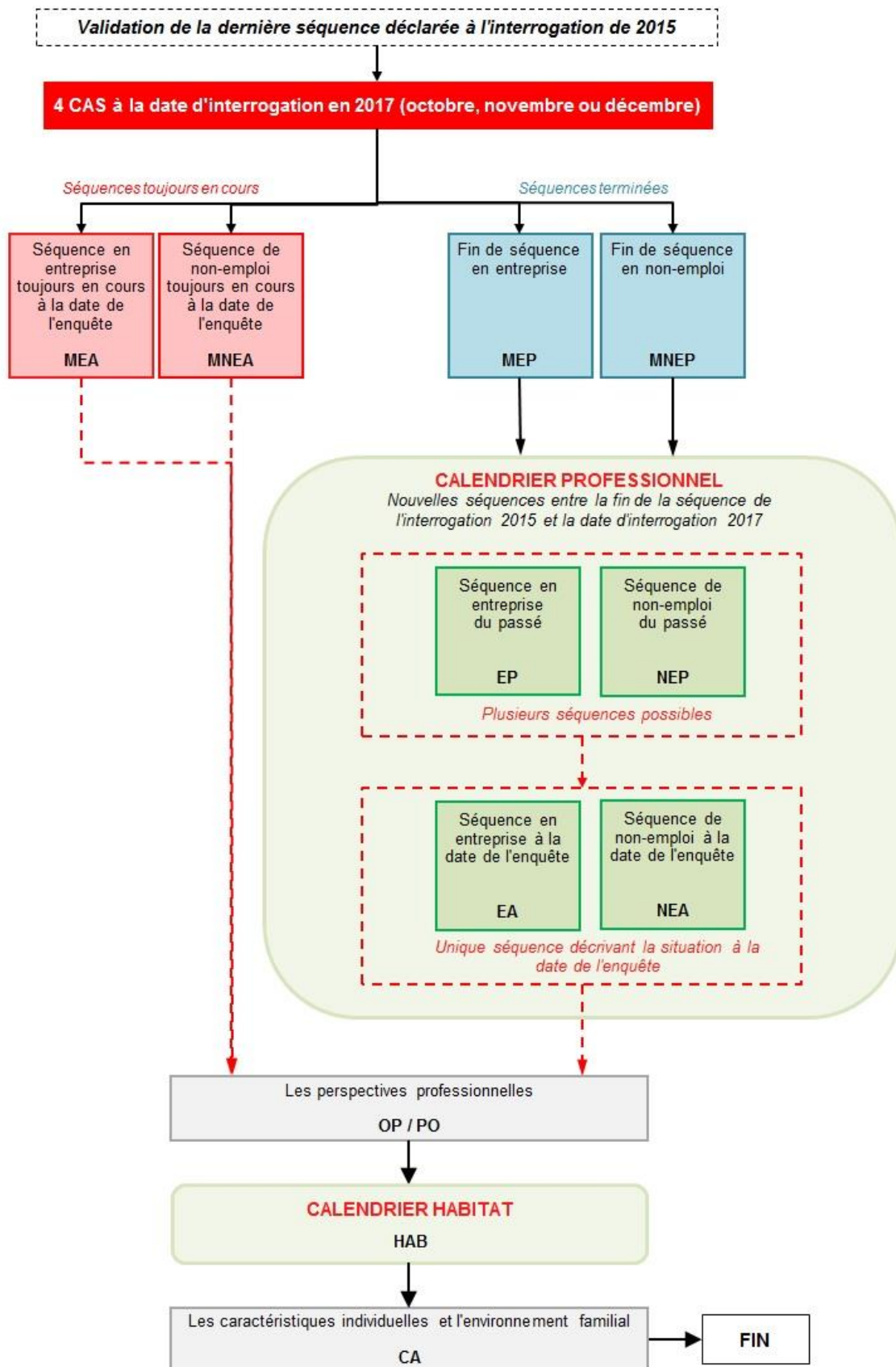
a) Descriptif du questionnaire

L'architecture du questionnaire de la première interrogation se composait de différentes thématiques : le parcours scolaire, le calendrier professionnel qui constitue le cœur du questionnaire et permet la description des séquences, de modules spécifiques liés à des extensions, les questions relatives aux perspectives professionnelles, le calendrier décrivant les modes de cohabitation, les informations sur les caractéristiques sociodémographiques.

Le questionnaire de la seconde interrogation s'appuyait sur le même questionnaire, sans le module sur le parcours scolaire et ceux relatifs aux extensions ou encore un allègement sur les caractéristiques sociodémographiques.

La construction du questionnaire de la troisième interrogation repose très fortement sur le questionnaire de la seconde interrogation. En particulier, la structure du calendrier professionnel ainsi que les modules de description de ses séquences n'ont été que marginalement modifiés pour aboutir au questionnement de l'enquête Génération 2010 à 7 ans sur ces points. L'objectif de cette ré-interrogation, tout comme celui de l'interrogation à 5 ans, est d'actualiser le calendrier professionnel de l'individu depuis la précédente interrogation. Plus précisément, si l'individu à l'interrogation de 2017 est dans la même situation que celle déclarée à la date d'interrogation de 2015 (dans la même entreprise, ou dans la même situation de non emploi qui a persisté entre 2015 et 2017), il ne passe pas par le calendrier professionnel. Dans l'autre cas, la date de fin de la séquence qui était en cours au moment de l'enquête de 2015 est enregistrée puis le calendrier professionnel est renseigné à partir de cette date pour reconstituer mensuellement son parcours jusqu'à la date d'interrogation de 2017.

Figure 2 : Schéma d'organisation du questionnaire de la 3^{ème} interrogation de la Génération 2010



L'enquête Génération 2010 à 5 ans intégrait un module sur l'acquisition de diplômes et titres professionnels en post initial. Ce focus sur la reprise d'études ayant été effectué deux ans auparavant, le choix a été fait de ne pas reconduire ce module. Il en va de même pour les courts modules de description des séquences de reprise d'études.

L'enquête Génération 2010 à 7 ans intègre des nouveautés par rapport à l'interrogation à 5 ans. Tout d'abord, au sein du module d'emploi, le module permettant de détailler l'activité des jeunes qui occupent des fonctions d'encadrement de l'enquête Génération 2010 à 5 ans a été intégralement révisé. Un nouveau module a ainsi été créé et est axé sur les cadres occupant des fonctions d'encadrement. D'autre part, au sein du module Caractéristiques individuelles et environnement familial, des questions sur la répartition des tâches domestiques dans le couple ainsi que sur les enfants, questions classiquement posées pour une troisième interrogation, ont été ajoutées. Ces modules font l'objet d'une présentation plus précise dans les parties ci-dessous.

b) Nouveau module sur la thématique de l'encadrement

Dans le prolongement d'études réalisées ou en cours de réalisation au Céreq sur l'évolution de l'accès des femmes aux postes de cadres et aux fonctions d'encadrement en France² et en Europe³, il est apparu intéressant et opportun de développer un module spécifique de questions sur l'encadrement autour de quatre thématiques :

- les conditions d'accès à ces fonctions ;
- l'emprise temporelle de ces postes ;
- le contenu et la répartition de l'activité dans ces fonctions ;
- le mode de rétribution associé à ces responsabilités d'encadrement.

Parmi les questions sous-jacentes, au regard de la problématique qui intéresse le Céreq, on peut évoquer les suivantes : les formes d'accès à ces fonctions diffèrent-elles entre hommes et femmes ? La nature des postes occupés diffère-t-elle selon le sexe (postes plus administratifs et fonctionnels chez les femmes et plus opérationnels chez les hommes... ?) Les femmes exercent-elles leur fonction dans des cadres temporels plus resserrés, plus contraints ?

L'horizon des 7 ans semble suffisamment éloigné de la fin des études pour tenter de cerner quelques constantes dans l'exercice de ces fonctions managériales. En outre, une expérience de 7 ans permet potentiellement de toucher une plus grande diversité de parcours, et de contextes de sortie de formation initiale, qu'en tout début de vie active et donc des modalités d'accès diversifiées (promotion interne versus recrutements externes...).

² Di Paola V., Dupray A., Epiphane D. et Moullet S. (2017). Accès des femmes et des hommes aux positions de cadres en début de vie active : une convergence en marche ? in *Femmes et Homme, l'égalité en question*, Collection Insee Références, pp. 31-47.

Epiphane D. (2016). Les femmes dans les filières et les métiers 'masculins' : des paroles et des actes, *Travail, genre et sociétés*, n° 36, pp. 161-166.

Coupié T. et Epiphane D. (2009). Allongement séculaire des scolarités féminines et évolution récente de la ségrégation dans les professions : quelles relations ? , *Économies et Sociétés*, Série « Socio-économie du travail », AB, n° 30, pp. 87-122.

³ Participation du Céreq au projet « Wo.Man » (Women in Management) financé par l'A.N.R (Agence Nationale de la Recherche) comparant l'accès des jeunes femmes aux fonctions d'encadrement en France, Suisse, Royaume Uni et Suède

c) *Module sur la vie en couple et les enfants*

Le module sur les caractéristiques individuelles et l'environnement familial a été enrichi par des questions portant sur la **répartition des tâches domestiques au sein du couple**, ainsi que des questions sur les **conséquences de la naissance d'un enfant** sur la vie professionnelle.

Ce questionnement prend sa source dans la continuité de celui du questionnaire de l'enquête Génération 1998 à 7 ans qui avait permis une étude approfondie sur les liens entre débuts de carrière professionnelle et début de « carrière » familiale⁴. Dans le souhait de comparer les générations (évolutions sur 12 ans), les questions de la troisième interrogation de l'enquête Génération 2010 sur ce thème sont identiques à celles de la troisième interrogation de l'enquête Génération 1998.

d) *Autres évolutions*

Outre l'intégration de deux nouveaux modules et la non-reconduction de modules de la précédente interrogation, le questionnaire de 2017 est essentiellement construit à partir de celui de 2015. Des ajustements ont été effectués pour certaines questions au sein du module Non-emploi, de la partie Opinion sur l'emploi (module d'emploi) et des Perspectives professionnelles.

Jugé assez long par les enquêtés lors de l'interrogation de 2010 à 5 ans, le module Non-emploi a été allégé de certaines questions. Les questions suivantes ont été supprimées des modules de Non-emploi du passé et à la date de l'enquête :

- utilité des entretiens avec un intermédiaire du marché du travail (Pôle Emploi, mission locale, PAIO)
- personne(s) à qui l'individu a parlé de sa recherche d'emploi
- nombre de formations suivies (information en clair)
- principal financeur de la (des) formation(s)

Par ailleurs, certaines questions ont été ajoutées, ne concernant que le module de Non-emploi à la date de l'enquête :

- aide financière de la part de la famille (OL7 / MOL7). Si oui, le montant (OL8/MOL8, OL8A/MOL8A, OL9/MOL9)
- plus petit salaire mensuel que l'individu serait prêt à accepter pour un travail à temps plein (OL10/MOL10, OL10A/MOL10A, OL11/MOL11)

Les questions suivantes, également posées à la date d'enquête uniquement, ont été introduites dans le but de mieux comprendre le comportement des individus qui sont ni en emploi, ni en formation, ainsi de mieux mesurer le concept du « halo » du chômage.

- pour les individus en situation de recherche d'emploi ou d'inactivité : souhait d'occuper un emploi le mois prochain (OL1 / MOL1). Si non, pour quelles raisons (OL2 / MOL2)
 - o Pour les individus en situation de recherche d'emploi ou d'inactivité et qui n'ont pas suivi de formation : leur a-t-on proposé de suivre une formation (OL3 / MOL3) ? Si oui, raison principale pour laquelle l'individu ne l'a pas suivie (OL4 / MOL4)
 - o Pour les individus qui n'ont pas suivi de formation ou ne sont pas en formation à la date d'interrogation : souhait de suivre une formation le mois prochain (OL5 / MOL5). Si non, raison principale (OL6 / MOL6)

⁴Couppié T. et Epiphane D. (2007). Vivre en couple et être parent : impacts sur les débuts de carrière, *Céreq-Bref* n°241.

La question EA79 était posée de la manière suivante pour la Génération 2010 à 5 ans : Diriez-vous que vous êtes : (Très bien payé / Plutôt bien payé / Normalement payé / Plutôt mal payé / Très mal payé). La modalité 3, modalité centrale, a été supprimée afin de rendre la question plus exploitable statistiquement.

Dans la partie Perspectives professionnelles, ont été ajoutées la question sur l'atout ou le handicap de la formation initiale dans son parcours professionnel (EA0) et celle portant sur les raisons du handicap de la formation jugée par l'individu.

L'ensemble des questions sur les projets de l'individu en matière d'avenir professionnel ne font pas partie de cette interrogation. Celles-ci sont :

- envisage de se mettre à son compte (OP7)
- poste le plus élevé que l'individu pense pouvoir atteindre en restant chez son employeur (P03A)
- prêt à changer d'employeur dans les cinq ans à venir (P03B)
- prêt à quitter la fonction publique (P03BF)
- intégrer la fonction publique dans les cinq ans à venir (P03BG)
- quitter la région dans les cinq ans à venir (P03C)
- changer de métier dans les cinq ans à venir (P03D)
- occuper une fonction d'encadrement dans les cinq ans à venir (P03H)
- qu'est-ce que pour vous travailler ? (P03I)

Comme pour les enquêtes Génération 2004 et 2007, une géolocalisation des adresses des enquêtés a été réalisée. Il s'agit des adresses de l'individu et de son établissement (s'il est en emploi) à chaque date d'enquête (2013, 2015, 2017) complétées par l'adresse disponible à la sortie de formation initiale (2010). Cette géolocalisation, réalisée en partenariat avec le CGET, permet de disposer d'une variable d'appartenance à un « quartier prioritaire politique de la ville » et ainsi d'étudier les effets de l'inscription socio-spatiale des individus sur leurs parcours. Pour la première fois, sur cette enquête, une normalisation des adresses à la volée a permis de faciliter la saisie des enquêteurs tout en améliorant la qualité de l'information collectée. Un descriptif de cet outil est disponible en ANNEXE 2 - Présentation de l'outil RNVP.

e) *Le calendrier professionnel*

Avant d'ouvrir le calendrier professionnel, l'enquêté doit valider, via une série de questions, la situation déclarée à la date de l'interrogation précédente, c'est-à-dire au printemps 2015. La validation porte sur la situation (en emploi, en recherche d'emploi, en reprise d'études, en formation, en inactivité) ; dans le cadre d'une situation en emploi, la validation porte également sur l'entreprise. En effet, une attention particulière est portée aux « raccords » nécessaires avec l'interrogation précédente de manière à éviter les ruptures de situations. Pour cela, des données sur la situation de l'enquêté à la date de l'interrogation précédente sont importées dans le CATI.

Plusieurs situations peuvent être rencontrées :

1. L'individu valide la situation à la date de l'enquête précédente et déclare être toujours dans la même situation : le calendrier ne s'ouvre pas et des questions lui sont posées afin de compléter et d'actualiser les informations données lors de la précédente interrogation.
2. L'individu valide la situation à la date de l'enquête précédente et déclare être en 2017 dans une autre situation : il lui est demandé à quelle date (mois et année) s'est terminée cette situation ainsi que quelques questions sur cette nouvelle situation. Suite à cela, le calendrier professionnel s'ouvre à la date de fin de la situation déclarée lors de la précédente interrogation.
3. L'individu valide partiellement la situation (cas de l'emploi). Il confirme être en emploi en 2013 mais pas dans l'entreprise citée. Le nom de cette dernière lui est alors demandé ainsi que des informations pour comprendre l'erreur :
 - s'il s'agit de la même entreprise (nom mal orthographié ou changement de nom) alors l'individu se retrouve en situation 1 ou 2.
 - s'il s'agit d'une autre entreprise alors il lui est demandé s'il a travaillé au préalable dans l'entreprise renseignée. Si oui : il lui est demandé la date à laquelle il a quitté l'entreprise et le calendrier s'ouvre à cette date. Si non : le calendrier s'ouvre à la date de l'enquête précédente.
4. L'individu ne valide pas la situation : le calendrier s'ouvre au mois suivant la date de la précédente interrogation.

Lorsque le calendrier s'ouvre, l'enquêté y situe les différentes séquences d'emploi, de recherches d'emplois, de reprises d'études, de formations ou autres intervenues après le printemps 2015, avec le même protocole que pour l'interrogation précédente.

5. Le calendrier de l'enquête

La réalisation de l'enquête s'est étendue sur deux mois et demi.

La chronologie des principales étapes de l'enquête Génération 2010 à 7 ans a été la suivante :

Tableau 2 - Calendrier de l'enquête

Étapes	Début	Fin
Conception du CATI	23 février	30 septembre
Test CATI 1	20 avril 2017	24 avril 2017
Test CATI 2	29 juin 2017	3 juillet 2017
Mise à jour des adresses postales	12 juin 2017	14 juin 2017
Actualisation des coordonnées téléphoniques	19 juin 2017	19 juillet 2017
Terrain d'enquête	2 octobre 2017	16 décembre 2017
Constitution des bases, apurements, codifications, pondération, rédaction, documentation	Janvier 2018	Juillet 2018
Bilan méthodologique	Juillet 2018	Septembre 2018

II. Préparation de la collecte

1. Constitution de l'échantillon

La population mise en production pour Génération 2010 à 7 ans est celle ayant répondu au questionnaire complet à 5 ans. Seuls 3 individus ont été supprimés (2 étant décédés et 1 refus CNIL détectés lors de la phase de fidélisation).

Ainsi, 13 707 individus ont été envoyés en enquête en 2017.

2. Développement du CATI (enquête téléphonique)

a) Logiciel utilisé

Le questionnaire CATI a été développé sous Interviewer Voxco ©. La version utilisée était une solution en client léger gérée par command center.

Ce questionnaire a pu être testé en ligne.

b) Technique de développement du calendrier

Pour cette enquête un calendrier interactif était disponible. Ce dernier était une application externe qui n'a pas été développée sous Voxco. Cette application a été réalisée en technologie Web par les équipes du prestataire de collecte et intégré en complément du questionnaire. Une question, transparente pour l'enquêté, permettait à l'enquêteur d'ouvrir dans une nouvelle fenêtre le calendrier. Une fois ce dernier validé et fermé il ne pouvait plus être modifié.

c) Test du CATI et du calendrier associé

Deux phases de test du questionnaire et de la méthodologie ont été réalisées.

Les jeunes ayant répondu aux enquêtes expérimentales de la seconde interrogation de l'enquête Génération 2010, dans le champ Céreq ont été sollicités pour réaliser les deux tests de cette enquête. Cet échantillon a été divisé aléatoirement en deux sous-échantillons de manière à avoir tous les niveaux de formation représentés pour chacun des tests.

Le nombre de numéros de téléphone par individu était de 2 au maximum alors qu'en enquête réelle il varie de 3 à 8 numéros disponibles.

Tableau 3 - Synthèse des tests

	Global		Test n°1		Test n°2	
Total enquêtes	400	12,8 %	223	12,4 %	177	13,3 %
Refus de répondre	27	0,9 %	10	0,6 %	17	1,3 %
Abandons en cours	21	0,7 %	18	1,0 %	3	0,2 %
Individus identifiés sur répondeur	433	13,8 %	112	6,3 %	320	24,0 %
Total individus contactés	2 127	66,9 %	960	53,4 %	1 167	87,5 %
Total appels	3 133	100,0 %	1 799	100,0 %	1 334	100,0%
DMC			16mn 58		13mn 46	
Durée du test	8 jours		4 jours		4 jours	
Dates du test			20 au 24 avril 2017		29 juin au 3 juillet 2017	

Ces tests ont permis :

- de peaufiner la structure du questionnaire (questions, items, consignes...) et du discours
- d'identifier les évolutions dans les consignes ou pour les programmes de formation enquêteurs
- de mesurer la DMC et son évolution, en vue de l'enquête
- de tester les calendriers interactifs d'activité et habitat
- de tester l'outil RNVP (mis en place lors de la réalisation du test n° 2)

3. Mise à jour des coordonnées postales

Restructuration, normalisation et validation postale (RNVP) :

Une mise à jour des adresses postales a été effectuée de manière à disposer d'adresses fiables pour l'envoi des courriers de fidélisation et lettre avis, mais aussi pour les recherches de numéros de téléphone basées sur l'adresse des individus, en amont de l'enquête. Afin de limiter le phénomène d'attrition, il est nécessaire d'actualiser les coordonnées que l'individu a déclarées deux ans auparavant en cas de déménagement, de changement de téléphone par exemple.

Les recherches menées par notre prestataire sur notre fichier ont donné les résultats suivants :

Tableau 4 - Résultats RNVP

	Effectif	%
Adresses valides	12 965	94,6 %
Adresses litigieuses	18	5,3 %
Adresses rejetées	724	0,1 %
Total	13 707	100 %

- Adresses valides : adresses corrigées et enrichies et réintégrables dans la base
- Adresses litigieuses : doute sur l'adresse
- Adresses en rejet : concernent des voies inconnues dans la localité, déménagés dont l'adresse n'est pas commercialisable (fichier CNIL), etc. ...

Recherche des déménagés

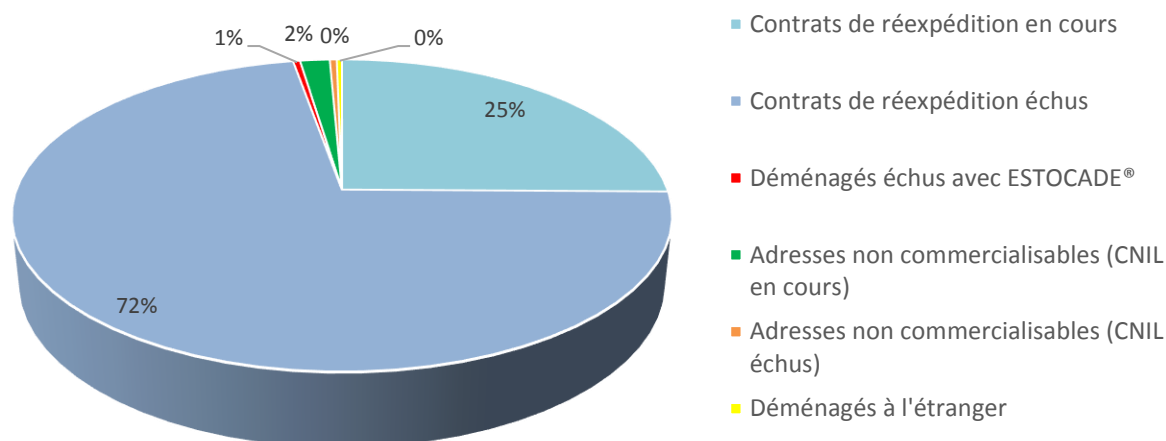
Sur les 13 707 adresses envoyées, 1 619 individus ont déménagé dont 1 574 sont exploitables (soit près de 11,5 % des adresses).

Tableau 5 - Résultats des recherches des déménagés

NOUVELLES ADRESSES	QUANTITE	% total déménagé	% total adr.
Contrats de réexpédition en cours	408	25,2 %	3,0 %
Contrats de réexpédition échus	1 166	72,0 %	8,5 %
TOTAL ADRESSES RACHETEES	1 574	97,2 %	11,5 %

REPERAGE DES DEMENAGES	QUANTITE	% total déménagés	% total adr.
Déménagés échus avec ESTOCADE®	7	0,4 %	0,1 %
Adresses non commercialisables (CNIL en cours)	28	1,7 %	0,2 %
Adresses non commercialisables (CNIL échus)	6	0,4 %	0,0 %
Déménagés à l'étranger	4	0,3 %	0,0 %
TOTAL ADRESSES INEXPLOITABLES	45	2,8 %	0,3 %

Figure 3 - Répartition des déménagés



Quelques définitions :

- **Contrats de réexpédition en cours** : nouvelles adresses de déménagés dont le contrat de réexpédition souscrit auprès de La Poste est en cours (pendant 6 mois renouvelable 1 fois)
- **Contrats de réexpédition échus** : nouvelles adresses de déménagés dont le contrat de réexpédition souscrit auprès de La Poste est échu
- **Adresses non commercialisables CNIL** : anciennes adresses des déménagés qui n'ont pas donné l'autorisation à La Poste de commercialiser leurs nouvelles adresses
- **Déménagés à l'étranger** : anciennes adresses de déménagés à l'étranger
- **Estocade** : déménagés dont le contrat de réexpédition est échu

4. Fidélisation

Dans l'optique de fidéliser la population de jeunes répondants à l'enquête et d'amortir les effets de l'attrition pour les réinterrogations, communiquer à travers les réseaux sociaux ou par l'intermédiaire de courrier électronique, permet de développer une proximité complémentaire à celle du courrier papier.

a) *Site internet du Céreq*

Pour permettre aux jeunes de s'assurer de l'origine et de l'intérêt de l'enquête, des informations ont été diffusées sur le site du Céreq (www.cereq.fr). Dès la page d'accueil, un encart était visible sur toute la durée du plateau d'enquête et une page spécifique Génération 2010 précisait les objectifs de l'enquête et l'importance de la participation de chacun.

b) Réseaux sociaux

Au-delà du site du Céreq, une page Facebook « Céreq Génération 2010 » (www.facebook.com/cereqG10) a été créée lors de la première interrogation pour pallier le manque d'interactions et d'échanges avec l'équipe en charge de l'enquête ; l'alimentation de ce compte a été réalisée par l'équipe.

La création de la page Facebook répond à plusieurs objectifs :

- créer un lien entre le Céreq et les jeunes afin de les fidéliser pour les prochaines interrogations
- recueillir les commentaires des jeunes sur l'enquête
- donner plus de visibilité au Céreq et à l'enquête Génération

Au-delà de l'information donnée sur le Céreq et ses missions, la page Facebook était principalement destinée à informer les jeunes répondants sur l'enquête Génération elle-même. La page contenait les objectifs de l'enquête, les modalités, le déroulement, la périodicité et les dates d'interrogation, la présentation de l'équipe en charge de l'enquête, les actualités, les événements en lien avec l'enquête, les informations post-enquête, des vidéos et les premiers résultats.

La page Facebook a par ailleurs permis d'ouvrir les portes du plateau d'enquête ce qui a facilité le premier contact avec les télénquêteurs.

Audience Facebook : 567 fans (condition : avoir un compte Facebook personnel)

Post le + populaire :

Génération 2010 - CEREQ
Publié par Zora Mazari [?] · 3 octobre 2017 · 🌐

#Generation2010 nous vous sollicitons une nouvelle fois pour une troisième enquête. Que devenez-vous?

CEREQ.FR
Sept ans après votre sortie de formation en 2010, que devenez-vous ? / actualites / accueil - Céreq - Centre d'études et de recherches sur les qualifications
Content Management System

647	10	28
Personnes touchées	Interactions	Clics

La page Facebook a rendu l'enquête plus accessible tout en conservant son caractère officiel. Cet outil de communication est adapté au public cible et permet de créer une interaction avec lui.

Il existe également un compte Twitter du Céreq (<https://twitter.com/PRESSECEREQ>), orienté presse, destiné plus particulièrement aux journalistes et au réseau institutionnel du Céreq.

c) Adresse de messagerie

L'adresse de messagerie de contact mise en place pour faciliter les échanges lors de la première interrogation (generation2010@cereq.fr) a été réutilisée.

d) Pré-fidélisation

Une étape de pré-fidélisation a eu lieu en décembre 2016. Un e-mail ou un courrier a été envoyé aux répondants de la seconde interrogation pour les informer du démarrage de la troisième interrogation à l'automne 2017. A ce courrier était jointe une note présentant quelques premiers résultats de

l'enquête Génération 2010 à 5 ans. Les deux modèles de lettre de pré-fidélisation figurent dans ANNEXE 3 - Lettre de pré-fidélisation».

5. Actualisation des coordonnées téléphoniques

Pour chaque ré-interrogation, les coordonnées téléphoniques disponibles sont celles validées par l'enquêté lors de la vague précédente, conservées d'une enquête à l'autre. Néanmoins, deux ans s'écoulant entre deux interrogations, il est nécessaire d'actualiser ces coordonnées en cas, par exemple, de déménagement, de changement de téléphone, etc. L'objectif est de limiter le phénomène d'attrition en conservant une taille d'échantillon suffisante le moins biaisé possible. L'un des moyens pour limiter l'attrition est l'enrichissement de coordonnées téléphoniques.

a) Actualisation par l'enquêté lors de la pré-fidélisation

L'étape de pré-fidélisation a permis aux individus de mettre à jour leurs coordonnées via un lien mentionné dans le mail ou le courrier qui renvoyait vers un formulaire internet permettant de modifier ou confirmer les informations sur l'adresse et les numéros de téléphones. Ce formulaire en ligne fait intervenir une phase d'authentification en demandant le mois et l'année de naissance. Pour garantir la sécurité des données, les mots de passe des tables sont cryptés.

Afin d'augmenter les chances d'enrichissement des coordonnées, les courriers et mails ont été envoyés en doublons lorsque cela était possible. Ainsi, sur les 13 710 individus de la base, 89,8 % ont reçu à la fois un mail et un courrier de pré-fidélisation, 9,7 % seulement un courrier et 0,5 % seulement un mail.

1 925 (soit 14 %) individus ont complété le questionnaire de pré-fidélisation. Parmi eux, 980 ont validé leurs coordonnées et 935 les ont modifiées.

Le CAWI mis en place est disponible en ANNEXE 4 - CAWI de pré-fidélisation.

b) Actualisation par un prestataire

Cependant, cela ne suffisant pas, une autre actualisation des coordonnées téléphoniques a été réalisée en juin 2017 via un prestataire. Le mode de recherche est le même que celui de l'enrichissement des données des deux précédentes interrogations.

Un protocole de recherche par étape a été retenu.

- **Phase A** : dans une première phase, l'objectif était de retrouver le jeune lui-même ou à défaut ses parents dans la base France Télécom, en recherchant les individus pour lesquels les noms, prénoms et adresses aboutissaient à un seul écho, en acceptant un relâchement léger des critères de recherche : légers écarts d'orthographe sur le nom, le prénom ou l'adresse, voire en l'absence de l'adresse ou du prénom ;
- **Phase D** : l'objectif était de retrouver le jeune lui-même à partir de cette base partenaire ;
- **Phase B** : le critère géographique est assoupli pour rechercher des échos sur les critères « nom et prénom et commune », « nom et prénom et département », « nom et prénom et région », « nom et prénom et Ile de France », « nom et prénom et reste de la France », toujours en autorisant des légers écarts d'orthographe ;
- **Phase C** : tout recours au prénom est supprimé pour rechercher des échos sur les critères « nom et commune » et « nom et département de l'adresse ».

L'ordonnancement de ces différentes étapes s'est fait comme suit :

Phase A :

1. NOM PRENOM COMMUNE France Télécom si présence d'un à 5 échos

Sinon Phase D :

1. NOM PRENOM COMMUNE bases partenaires si présence d'un à 5 échos

Sinon (phase B et C) :

2. NOM COMMUNE France Télécom et NOM PRENOM DEPARTEMENT France télécom si NOM COMMUNE + NOM PRENOM DEPARTEMENT compris entre 1 et 5 échos (PHASE B)
3. NOM COMMUNE France Télécom si présence d'un à 5 échos
4. NOM PRENOM DEPARTEMENT France Télécom si présence d'un à 5 échos
5. NOM COMMUNE France télécom et NOM PRENOM DEPARTEMENT bases partenaires si NOM COMMUNE + NOM PRENOM DEPARTEMENT compris entre 1 et 5 échos
6. NOM COMMUNE base partenaire si présence d'un à 5 échos
7. NOM PRENOM DEPARTEMENT base partenaire si présence d'un à 5 échos
8. NOM DEPARTEMENT France Télécom si présence d'un à 5 échos
9. NOM DEPARTEMENT base partenaire si présence d'un à 5 échos
10. NOM PRENOM REGION France Télécom si présence d'un à 5 échos
11. NOM PRENOM REGION base partenaire si présence d'un à 5 échos
12. NOM PRENOM IDF France Télécom si présence d'un écho
13. NOM PRENOM RESTE FRANCE base partenaire si présence d'un écho

Le bilan de ces recherches est le suivant :

- Phase A :

Sur les 13 707 individus envoyés en recherche téléphonique, 1 241 ont été retournés en écho des recherches phase A (9,1%), c'est-à-dire que pour ces individus un écho unique est retrouvé suite à une recherche sur le nom (éventuellement le prénom) et l'adresse de l'individu.

Tableau 6 - Bilan de la phase A (téléphone fixe)

CLASSE	SIGNIFICATION	Nbr tel enrichis	%
U11	Nom, prénom et adresse exacts	252	1,8
U12	Nom et prénom exacts, adresse approchée, une seule réponse	26	0,2
U13	Nom et prénom identiques, adresse absente, une seule réponse	14	0,1
U21	Nom ou prénom approché, adresse exacte, une seule réponse	23	0,2
O111	Nom exact, adresse exacte, prénom différent, une seule réponse	615	4,5
O11	Nom exact, adresse exacte, prénom différent ou approché, plusieurs réponses	66	0,5
O101	Nom exact, prénom exact, adresse différente, une seule réponse	87	0,6
O121	Nom identique, adresse approchée, une seule réponse	67	0,5
O12	Nom exact, adresse approchée	21	0,2
O131	Nom identique, adresse absente, une seule réponse	58	0,4
O13	Nom identique, adresse absente, plusieurs réponses	2	0,0
O211	Nom approché, adresse exacte, une seule réponse	10	0,1

Total	1 241	9,1
--------------	--------------	------------

- Phase D :

Sur les 13 707 individus envoyés en recherche téléphonique, 3 567 ont été retournés en écho des recherches phase D (26 %), c'est-à-dire que pour ces individus un écho unique est retrouvé suite à une recherche sur le nom, (éventuellement le prénom) et l'adresse de l'individu.

Tableau 7 - Bilan de la phase D (téléphone portable)

CLASSE	SIGNIFICATION	Nbr tel enrichis	%
U11	Nom, prénom et adresse exacts	2 072	15,1
U12	Nom et prénom exacts, adresse approchée, une seule réponse	91	0,7
U13	Nom et prénom identiques, adresse absente, une seule réponse	94	0,7
U21	Nom ou prénom approché, adresse exacte, une seule réponse	67	0,5
U22	Nom ou prénom approché, adresse approchée, une seule réponse	7	0,1
O111	Nom exact, adresse exacte, prénom différent, une seule réponse	650	4,7
O11	Nom exact, adresse exacte, prénom différent ou approché, plusieurs réponses	116	0,8
O101	Nom exact, prénom exact, adresse différente, une seule réponse	309	2,3
O121	Nom identique, adresse approchée, une seule réponse	40	0,3
O12	Nom exact, adresse approchée	18	0,1
O131	Nom identique, adresse absente, une seule réponse	78	0,6
O13	Nom identique, adresse absente, plusieurs réponses	3	0,0
O211	Nom approché, adresse exacte, une seule réponse	22	0,2
Total		3 567	26,0

(Recherches excluant les individus retrouvés en phase A)

- Phase B et C :

Tableau 8 - Bilan phase B et C – base France télécom

Phase	Description	Nb total	%
B	Rapprochement, nom, prénom, commune	74	0,3
B	Rapprochement, nom, prénom, département	104	0,4
B	Rapprochement, nom, prénom, France hors IDF	164	0,6
B	Rapprochement, nom, prénom, IDF	45	0,2
B	Rapprochement, nom, prénom, région	14	0,1
C	Rapprochement, nom, commune	9 798	38,2
C	Rapprochement, nom, dept	15 427	60,2
Total		25 626	100,0

c) Ordonnance des numéros de téléphone

Plusieurs numéros sont disponibles pour un même individu. Certains numéros sont plus fiables que d'autres, l'ordre d'appel a donc son importance.

Le choix du numéro selon le type de recherches a été choisi de la manière suivante :

1. Numéro issu de la pré-fidélisation
2. Numéro ayant permis de réaliser l'enquête à 5 ans
3. Numéro phase D

4. Numéro phase A
5. Numéro issu de la recherche élargie

6. Lettre avis

La lettre avis a en premier lieu un rôle informatif. Elle cherche ensuite à sensibiliser le futur enquêté pour faciliter son adhésion lors du contact téléphonique. Elle l'incite à préparer le calendrier d'activité des deux années écoulées, élément clé du questionnaire nécessitant un effort de mémoire. Elle propose également un numéro vert, une adresse courriel et un site internet pour que l'individu puisse indiquer ses nouvelles coordonnées et/ou puisse prendre rendez-vous. De plus, le caractère non obligatoire est rappelé de même que les références aux lois de 1951 sur l'obligation, la coordination et le secret en matière de statistiques et de 1978 relative à l'informatique, aux fichiers et aux libertés.

Pour cette enquête, il a été décidé d'envoyer à la fois une lettre avis mails aux individus disposant d'un mail et une lettre avis papier à ceux disposant d'une adresse postale fiable.

Les deux modèles de lettres avis figurent dans «ANNEXE 5 - Lettre avis de contact avec les jeunes».

Finalement, 99,4 % des individus ont reçu à minima une lettre-avis.

Tableau 9 - Répartition du nombre de lettres avis envoyées selon le mode

	Effectif
Lettre papier	1 233
Lettre mail	651
Lettre mail et papier	11 737
Total	13 621

III. La collecte par téléphone

La collecte de l'enquête s'est déroulée par téléphone d'octobre à décembre 2017. 13 707 individus ont été mis en production.

L'enquête est assistée par ordinateur (CATI) : l'enquêteur interroge les individus par téléphone tout en suivant sur un écran d'ordinateur un script préétabli qui affiche les questions qu'il doit poser, les modalités et les éventuelles consignes. Les réponses sont saisies directement sur informatique. Le CATI est capable de gérer des filtres et d'orienter l'individu vers des questions différentes en fonction des réponses précédentes.

Afin de spécialiser les télé-enquêteurs sur des cibles particulières, la collecte a été répartie entre trois terrains d'enquête par niveau de formation.

1. Calendrier et organisation générale de la collecte

Tableau 10 - Calendrier et organisation générale avec le prestataire

	Délais
Réunion de démarrage	28 février 2017
Transmission du questionnaire	28 février 2017
Programmation et tests du questionnaire CATI	28 février 2017 – 30 septembre 2017
Transmission au Céreq d'un lien test	28 février 2017 – 30 septembre 2017
Premier pilotage CATI	20 – 24 avril 2017
Deuxième pilotage CATI	29 juin 2017 – 3 juillet 2017
Envoi lettre avis papier	25 septembre 2017
Ouverture du numéro vert	26 septembre 2017
Ouverture du CAWI de fidélisation	26 septembre 2017
Formation des enquêteurs :	
Superviseurs	Semaine du 25 septembre
Agents	2-3 octobre 2017
Envoi lettre avis mail en 3 lots	2-4 octobre 2017
Réalisation des enquêtes	2 octobre 2017 – 16 décembre 2017
Envoi des mails de relance	21 novembre 2017
Envoi SMS – 1^{er} lot (terra 2 et 3)	27 novembre
Envoi SMS – 2^{ème} lot (terra 1)	29 novembre
Remise fichier final CATI	20 décembre 2017

2. Le suivi de la collecte en chiffres

L'échantillon se composait de 13 707 individus. Parmi eux, 8 898, soit 65 %, ont répondu à l'enquête.

Figure 4 - Nombre d'enquêtes réalisées selon la période

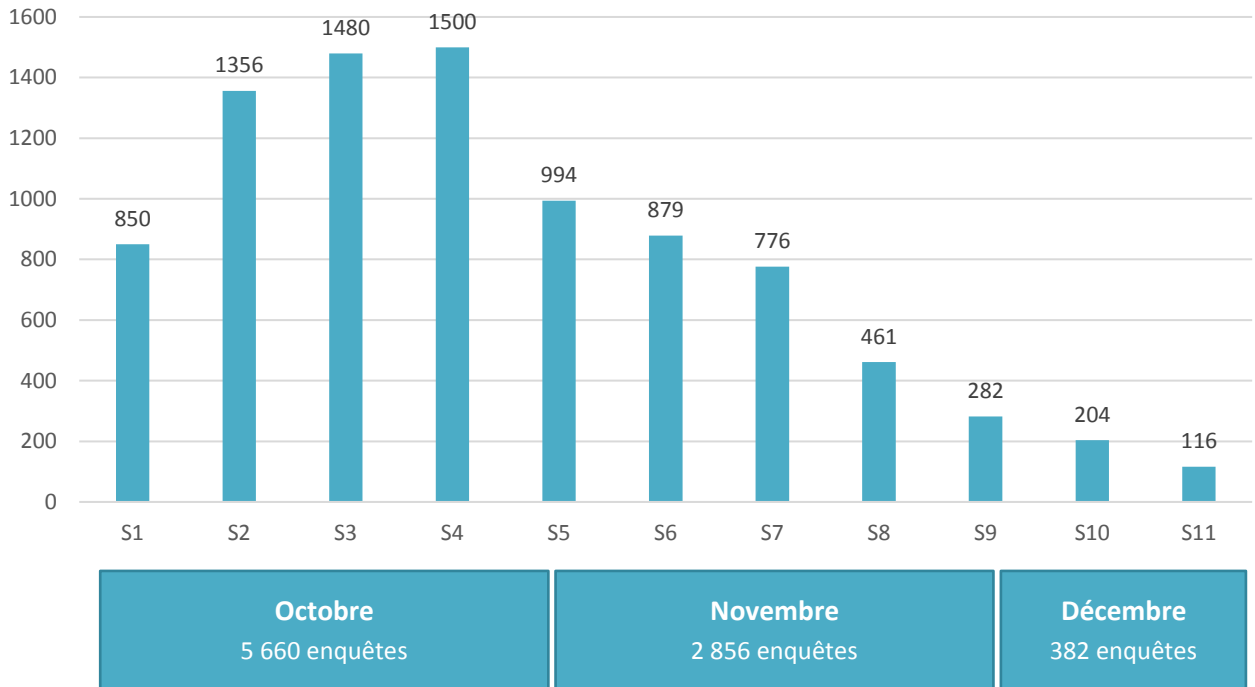
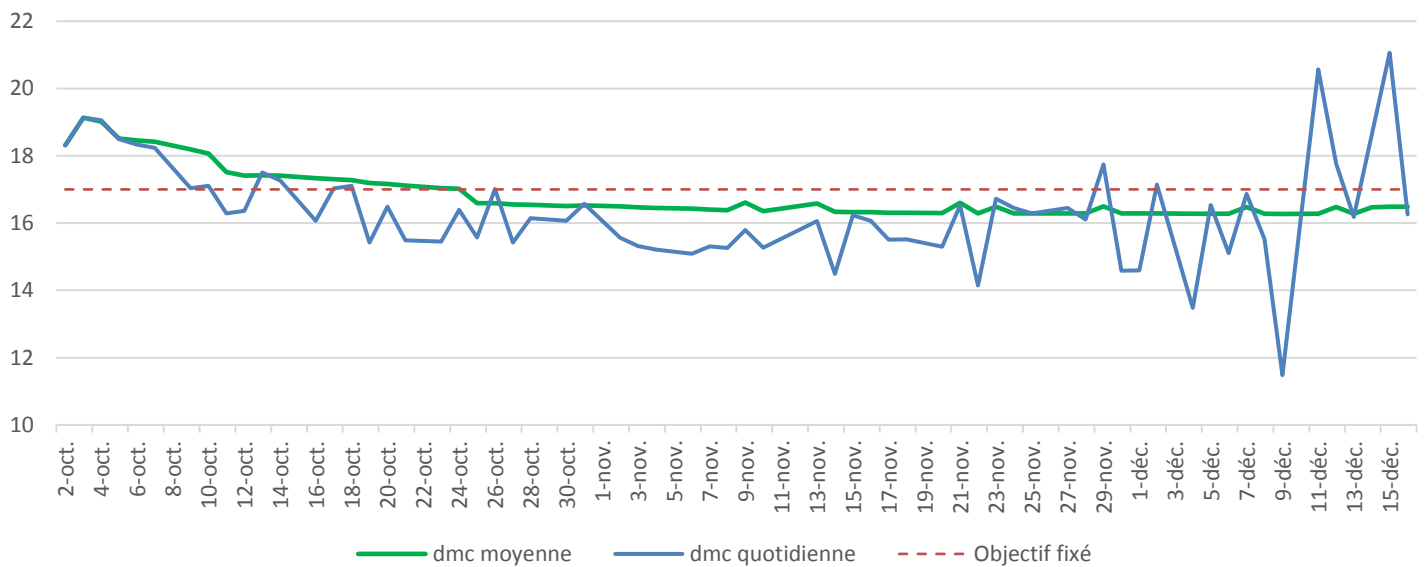


Figure 5 - Durée moyenne de communication



3. Les règles de rappel

En début d'enquête, nous avons programmé les règles de rappel suivantes :

Tableau 11 - Règles de rappel initiales

Règle de réitération		Priorité
Rappel	Date et heure choisies par l'enquêteur dans le calendrier	1
Fiche vierge	Présentation automatique après les rappels et avant les autres catégories	2
Ne répond pas Occupés	<ul style="list-style-type: none"> Rappel automatique à « J+1h » Si la seconde tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi	3
Répondeur	<ul style="list-style-type: none"> Rappel automatique à « J +15mn » Avec un maximum de 2 tentatives sur la même journée Présentation obligatoire sur un samedi	3

Ces règles de rappel ont été révisées en fin de production de la manière suivante, les priorités restant inchangées :

Tableau 12 - Règles de rappel modifiées pour la fin de l'enquête

	22 novembre 2017	15 décembre 2017 16h
Rappel	Inchangé	
Fiche vierge	Non concerné	
Ne répond pas	<ul style="list-style-type: none"> Rappel automatique à « J+1h » Si la seconde tentative est infructueuse rappel automatique à « J+3h » Si la troisième tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi	<ul style="list-style-type: none"> Rappel automatique à « J+45min » Si la seconde et troisième tentative est infructueuse rappel automatique à « J+45min » Si la quatrième tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi
Occupés	<ul style="list-style-type: none"> Rappel automatique à « J+15min » Si la seconde tentative est infructueuse rappel automatique à « J+3h » Si la troisième tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi	<ul style="list-style-type: none"> Rappel automatique à « J+15min » Si la seconde et troisième tentative est infructueuse rappel automatique à « J+45min » Si la quatrième tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi
Répondeur	<ul style="list-style-type: none"> Rappel automatique à « J+45mn » Si la seconde tentative est infructueuse rappel automatique à « J+3h » Si la troisième tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi	<ul style="list-style-type: none"> Rappel automatique à « J+45mn » Si la seconde et troisième tentative est infructueuse rappel automatique à « J+45min » Si la quatrième tentative est infructueuse rappel automatique à « J+1j » avec décalage de 1h Présentation obligatoire sur un samedi

4. Des messages laissés sur répondeur

En milieu de terrain (début novembre), les télé-enquêteurs ont commencé à laisser des messages sur répondeur afin d'inciter les individus à rappeler sur le numéro vert ou bien à décrocher lors du prochain appel. Les messages sur répondeurs ont été laissés au bout de six tentatives d'appels infructueuses.

5. Suivi technique et personne « qualité »

En raison de la distance entre le plateau d'enquête et le Céreq, une personne a été recrutée pour suivre le plateau téléphonique en continu. Cela a notamment permis au Céreq d'avoir des informations en temps réel. Cette personne « qualité » a réalisé des écoutes quotidiennes, a assisté les superviseurs et assuré la formation des enquêteurs lorsque les agents du Céreq n'étaient pas présents. Elle a fait part des difficultés rencontrées sur place et permis au Céreq de réagir efficacement face aux problèmes humains, techniques ou d'organisation du terrain. Un rapport de fin de mission incluant des données de collectes et des retours a été réalisé par cette personne « qualité ».

IV. Bilan de collecte

Sur les 13 707 individus répondant à la vague précédente mis en production, 8 898 ont répondu à l'enquête Génération 2010 à 7 ans.

Tableau 13 - Résultats d'appels des numéros mis en production

	Global		Terrain 1		Terrain 2		Terrain 3	
	Effectif	%	Effectif	%	Effectif	%	Effectif	%
Enquêtes Réalisées	8 898	63,15 %	6 012	71,44 %	2 470	53,19 %	416	40,39 %
Hors Cible (Mauvais Individus)	1 444	10,25 %	616	7,32 %	639	13,76 %	189	18,35 %
Hors cible (homonyme parfait)	109	0,77 %	55	0,65 %	44	0,95 %	10	0,97 %
Rappels pour finir l'enquête	4	0,03 %	2	0,02 %	2	0,04 %	0	0,00 %
A déjà répondu	27	0,19 %	9	0,11 %	14	0,30 %	4	0,39 %
Abandons en Cours	49	0,35 %	18	0,21 %	20	0,43 %	11	1,07 %
Refus de Répondre	207	1,47 %	87	1,03 %	96	2,07 %	24	2,33 %
Va rappeler sur le N° Gratuit (n° vert)	71	0,50 %	48	0,57 %	21	0,45 %	2	0,19 %
Résident à l'étranger	40	0,28 %	25	0,30 %	14	0,30 %	1	0,10 %
Injoignables Permanents	194	1,38 %	108	1,28 %	71	1,53 %	15	1,46 %
Refus d'un tiers	545	3,87 %	223	2,65 %	262	5,64 %	60	5,83 %
Décédés	21	0,15 %	7	0,08 %	9	0,19 %	5	0,49 %
Demandent à être supprimés des fichiers d'enquête	0	0,00 %	0	0,00 %	0	0,00 %	0	0,00 %
Faux Numéro (le numéro ne connaît pas la cible)	0	0,00 %	0	0,00 %	0	0,00 %	0	0,00 %
Ne plus jamais appeler	437	3,10 %	211	2,51 %	178	3,83 %	48	4,66 %
Refus CNIL définitifs	70	0,50 %	32	0,38 %	26	0,56 %	12	1,17 %
Abandons après Max Appels	1 621	11,50 %	859	10,21 %	597	12,86 %	165	16,02 %
Bug question ou calendrier	2	0,01 %	0	0,00 %	2	0,04 %	0	0,00 %
Ne connaît pas ses coordonnées	351	2,49 %	104	1,24 %	179	3,85 %	68	6,60 %
TOTAL DES CONTACTS UTILES	14 090	100,00 %	8 416	100,00 %	4 644	100,00 %	1 030	100,00 %
Faux Numéro Disque France Télécom	962		405		379		178	
NRP	316		119		165		32	
À Rappeler	32		14		13		5	
Occupé	4		1		2		1	
Répondeur	515		285		179		51	
N'accepte pas les N° masqués	19		14		5		0	
Message laissé sur le répondeur	9		1		4		4	
Individu identifié sur nom prénom sur répondeur	5		1		4		0	
TOTAL DES FICHES TRAITEES	15 952		9 256		5 395		1 301	

V. Les traitements en aval

Cette partie présente l'ensemble des traitements réalisés pour construire les fichiers finaux de diffusion à partir des fichiers bruts, à l'exception du calcul de la pondération finale présentée dans la partie VI, qui aborde le traitement de la non réponse globale et le calage sur marge.

1. Les tables brutes livrées par le prestataire

À l'issue de la collecte, le prestataire a livré les fichiers bruts au Céreq. Différents traitements ont été réalisés sur ces fichiers pour aboutir aux fichiers finaux : apurement des données, récupération des données de la vague précédente et raccordements, décomposition de la table du calendrier professionnel en deux tables séparant les périodes d'emploi et les périodes de non emploi, codage du secteur d'activité et de la catégorie socioprofessionnelle, redressement des salaires, anonymisation des données...

Plusieurs formats de fichiers reçus : sous Excel, Access et texte. En parallèle, des compléments de fichiers reçus sous format sas : libellés des questions, des formats (avec définitions des modalités).

2. La création des tables

La création des bases pour une ré-interrogation est substantiellement plus longue que pour une interrogation à trois ans. La principale difficulté réside en effet dans le « raccordement » des bases avec la précédente interrogation en 2015.

La première étape consiste à alimenter la base de l'enquête 2017 par les réponses obtenues par ces mêmes enquêtés lors des deux précédentes interrogations.

Remarque : Les identifiants des individus sont les mêmes dans les fichiers des différentes vagues, les informations spécifiques non présentes dans les fichiers à 7 ans peuvent être obtenues par appariement de tables.

a) **Traitements et contrôles des fichiers de base transmis par le prestataire**

Une fois l'enquête terminée, le prestataire chargé de l'enquête envoie au Céreq des fichiers au format DBF (fichier Access). Une fois convertis au format SAS, des contrôles de validation sont effectués. En théorie, un calendrier d'activité mal renseigné ne pouvait être validé et l'enquête ne pouvait alors pas se poursuivre.

En pratique, il est arrivé dans certains cas, que des calendriers soient mal définis. Certains calendriers d'activité sont donc non exploitables, à cause de certaines périodes qui se chevauchent, ou qui ne sont pas renseignées ou encore parce que des périodes d'emploi mentionnées ne sont pas associées à une description de l'emploi alors que cela aurait dû être le cas.

Des contrôles de validation sont donc effectués sur ces tables. Il s'agit de réaliser des vérifications sur la cohérence des raccords de séquences avec l'enquête précédente, de corriger des problèmes de calendrier (d'activité et habitat), notamment en repérant les problèmes de chevauchement de séquences (d'emplois et de non emploi) et éventuellement de pointer des problèmes d'écriture de filtres du questionnaire développé sous-système CATI.

Les observations jugées non exploitables sur le noyau du questionnaire ont été supprimées.

16 individus (soit 0,2 % des individus) sont alors supprimés pour les raisons suivantes :

- 4 suite à une ouverture injustifiée du calendrier
- 4 suite à l'absence d'ouverture du calendrier
- 4 suite à l'absence d'ouverture des modules permettant le raccord avec l'enquête précédente
- 1 suite à l'ouverture simultanée des 3 modules permettant le raccord avec l'enquête précédente
- 3 suite à l'absence d'ouverture des modules suite au calendrier

Au final, le nombre total des enquêtes exploitables est de 8 882.

b) Création de la base individus

L'objectif est d'avoir in fine toutes les informations individuelles à 3 ans, 5 ans et 7 ans. Pour cela, on récupère d'une part l'ensemble des variables présentes à 3 ans et 5 ans (car on souhaite les conserver à 7 ans dans la table individu) et d'autre part les variables présentes à 7 ans qu'on suffixe « _17 » pour certaines de façon à dater l'information collectée et distinguer une même question présente à la fois en 2013 (suffixée « _13 »), en 2015 (suffixée « _15 ») et en 2017.

En plus des variables présentes à 3 ans, ont été intégrées dans la base individus les variables relatives au nouveau module « Obtention de diplôme(s) ou titre(s) professionnel(s) ». Ce module a pour but de savoir si l'enquêté a obtenu un diplôme depuis la fin de sa formation initiale, ou s'il a tenté d'en obtenir un. Il spécifie aussi le titre de ce diplôme ainsi que les circonstances de la formation suivie pour parvenir à l'obtention, ou non, de ce diplôme.

c) Calendriers professionnels et habitat

– Calendrier professionnel

La poursuite de séquences d'emploi/non emploi et l'ajout de nouvelles séquences d'emploi/non emploi nous obligent à renuméroter les séquences (*nseq*) car les numéros de séquences fournis par le prestataire dans le fichier brut ne tient pas compte des interrogations précédentes. Il faut également recalculer les dates de début et de fin des séquences raccordées. Ce travail suppose de travailler sur une base regroupant les séquences d'emploi et de non emploi.

Cela implique de mettre à jour certaines variables telles que la variable précisant le type de séquence (exemple : une séquence d'emploi à la date de l'enquête en 2015 qui se termine entre les deux interrogations doit être transformée en séquence d'emploi du passé) mais également les variables sur les temps d'accès à un premier emploi, à un EDI, ainsi que les variables nombre de mois passés en emploi etc...

– Calendrier habitat

La poursuite de séquence et l'ajout de nouvelles nous amènent à réaliser un travail similaire à celui mené sur le calendrier professionnel.

d) *Création des nouvelles séquences d'emploi et de non-emploi*

À cette étape, il faut pouvoir raccorder les informations de 2015 avec les nouvelles informations de 2017. Deux cas sont à gérer :

- pour les individus qui ont changé d'emploi ou de situation de non emploi entre 2015 et 2017 : le raccord s'établit avec la validation de la dernière séquence à 5 ans. Il faut recalculer le numéro de séquence pour respecter l'enchaînement des séquences d'emploi et de non-emploi et réinitialiser les variables de début et de fin de séquence.
- pour les individus toujours dans la même séquence d'emploi ou de non emploi (autrement dit dont la situation n'a pas changé entre 2015 et 2017) : le raccord se fait par une transformation des noms des nouvelles variables correspondant à la séquence déjà présente et une réinitialisation des variables de début et de fin de séquence.

En ce qui concerne les informations à la date de l'enquête en 2015, l'idée, pour ne pas les écraser, est de créer des variables intermédiaires en conservant le nom de la variable contenant l'information tout en lui rajoutant une extension. Pour cette enquête trois extensions sont possibles :

- "_13" pour une information date d'enquête à 3 ans (interrogation 2013)
- "_15" pour une information date d'enquête à 5 ans (interrogation 2015)
- "_17" pour une information date d'enquête à 7 ans (interrogation 2017)

Enfin, on regroupe dans une base globale les séquences issues de l'enquête à 3 ans, celles à 5 ans, celles « raccordées » et les nouvelles de l'enquête à 7 ans.

Quelques définitions :

MEP : Ces séquences concernent les emplois qui étaient en cours en 2015 (lors de l'interrogation à 5 ans) et qui se sont finis entre l'interrogation 2015 (à 5 ans) et l'interrogation 2017 (à 7 ans) ("même emploi dans la séquence du passé").

MEA : Ces séquences concernent les emplois en cours en 2015 et qui sont toujours en cours en 2017 ("même emploi dans la séquence actuelle").

MNEP : Ces séquences concernent les périodes de non emploi qui étaient en cours en 2015 (lors de l'interrogation à 5 ans) et qui se sont finies entre l'interrogation 2015 (à 5 ans) et l'interrogation 2017 (à 7 ans)

MNEA : Ces séquences concernent les périodes de non emploi en cours en 2015 et qui sont toujours en cours en 2017

META : Ces séquences concernent les reprises d'études en cours en 2015 et qui sont toujours en cours en 2017

EP : Ces séquences concernent les emplois qui ont été déclarés entre les deux interrogations 2015 et 2017 ("nouvelles séquences du passé").

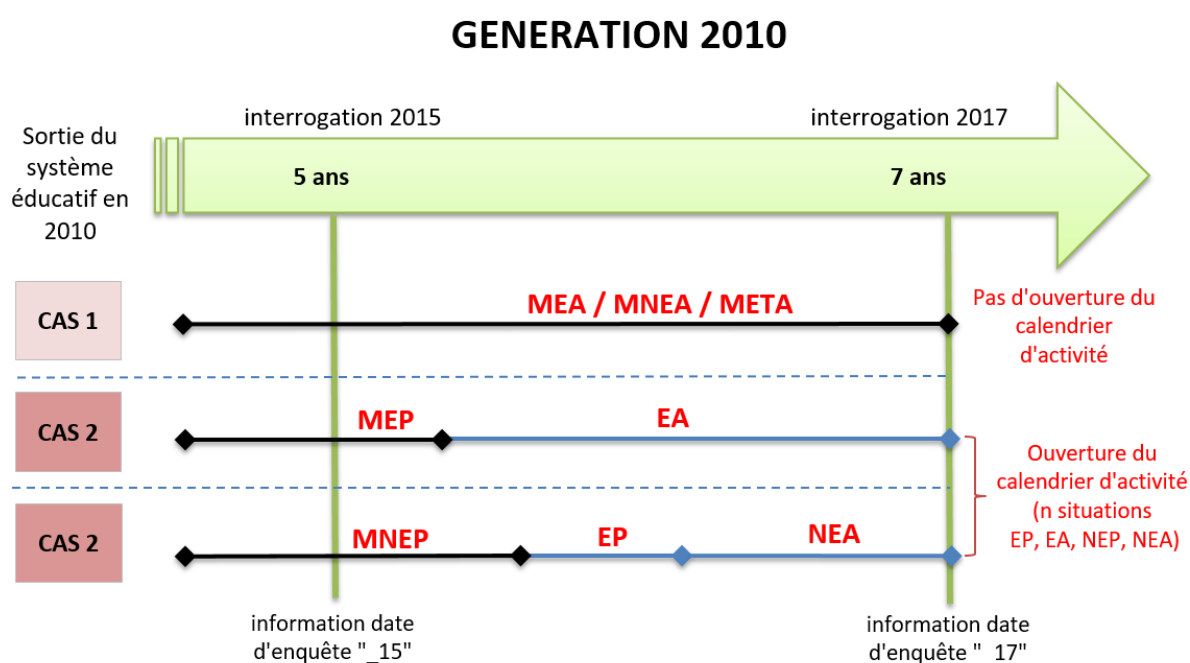
EA : Ces séquences concernent les emplois qui débutent après la précédente interrogation (2015) et qui sont toujours en cours en 2017 ("nouvelles séquences actuelles").

NEA : Ces séquences concernent les périodes de non emploi qui débutent après la précédente interrogation (2015) et qui sont toujours en cours en 2017 ("nouvelles séquences actuelles")

NB : Pour chaque séquence d'emploi décrite, il n'y a qu'une seule ouverture de module possible : MEP, ou MEA ou EP ou EA.

Pour une meilleure compréhension de la définition des séquences, ci-après un schéma récapitulatif :

Figure 6 - Schéma récapitulatif expliquant le raccord des séquences entre la seconde et la troisième interrogation



Il existe une multitude de parcours professionnels, deux seront décrits. Pour chaque individu, deux situations de départ sont possibles :

- Soit la séquence renseignée à la date d'enquête précédente (c'est-à-dire en 2015) continue à la date d'enquête actuelle (c'est-à-dire en 2017) alors aucune autre séquence n'est décrite. La personne n'a donc pas changé de situation depuis l'interrogation à 5 ans.
- Soit la séquence renseignée à la date d'enquête précédente (c'est-à-dire en 2015) s'est terminée entre les deux interrogations (c'est-à-dire entre 2015 et 2017) alors la description du parcours professionnel se poursuit via le calendrier d'activité.

CAS 1 : l'individu était dans une situation au moment de l'interrogation à 5 ans et l'est toujours à la date d'enquête à 7 ans alors on est en MEA (toujours dans la même entreprise) ou en MNEA (toujours en non emploi) ou en META (toujours en reprise d'étude). On ne décrit aucune autre séquence.

CAS 2 : l'individu était en emploi au moment de l'interrogation à 5 ans et a quitté cet emploi entre l'interrogation à 5 et 7 ans (MEP). On décrit ensuite une nouvelle séquence d'emploi qui commence et se termine entre 5 et 7 ans (EP). Enfin, la dernière séquence d'emploi, commencée entre 5 et 7 ans, continue à la date d'enquête à 7 ans on est alors en EA.

La variable TABSEQ, qui permet de décrire la temporalité de la séquence, est recalculée. Créée dans les bases à 5 ans, elle regroupait 3 types de séquences :

- 1 = séquence terminée à la première interrogation
- 2 = séquence en cours à la première interrogation
- 3 = séquence débutée après la première interrogation

À présent elle regroupe 9 types de séquence :

- 1 = Séquence commencée et terminée à la 1ère interrogation (entre 2010 et 2013)
- 2 = Séquence commencée à la 1ère interrogation et terminée à la 2ème interrogation (entre 2010 et 2015)
- 3 = Séquence commencée à la 1ère interrogation et terminée à la 3ème interrogation (entre 2010 et 2017)
- 4 = Séquence commencée à la 1ère interrogation et toujours en cours à la 3ème interrogation (entre 2010 et 2017)
- 5 = Séquence commencée et terminée à la 2ème interrogation (entre 2013 et 2015)
- 6 = Séquence commencée à la 2ème interrogation et terminée à la 3ème interrogation (entre 2013 et 2017)
- 7 = Séquence commencée à la 2ème interrogation et toujours en cours à la 3ème interrogation (entre 2013 et 2017)
- 8 = Séquence commencée et terminée à la 3ème interrogation (entre 2015 et 2017)
- 9 = Séquence commencée et toujours en cours à la 3ème interrogation (entre 2015 et 2017)

e) *Les recopies*

Les recopies ne concernent que les séquences d'emploi :

- emplois courts inférieurs à 6 mois du passé (CAL=01) ou actuels (CAL=03) pour les salariés et intérimaires (INTER=1 ou INTER=2).
- emplois longs inférieurs ou égal à 12 mois du passé (CAL=02) ou actuels (CAL=04) pour les salariés et intérimaires dans une seule entreprise (INTER=2) **hors** intérimaires plusieurs missions dans plusieurs entreprises différentes (INTER=1).

On distingue plusieurs situations :

- pour les emplois longs du passé ou actuels (CAL=02 ou CAL=04), si l'individu change de situation, c'est à dire un changement de (et/ou) :
 - o Contrat de travail
 - o Profession
 - o Catégorie socio professionnelle

- Temps de travail
- Salaire, primes et 13ème mois

Alors on pose les questions de début et de fin de séquence. Il n'y a donc pas de recopie. En revanche, s'il indique que l'une de ses situations n'a pas changé, le début est recopié sur la fin.

- pour les emplois courts du passé (CAL=01), seules les questions de début de séquence sont posées systématiquement. Les informations saisies dans les variables de début de séquence qui concernent l'embauche sont recopiées dans les variables de fin de séquence qui concernent le départ de l'entreprise.
- pour les emplois courts actuels (CAL=03), seules les questions de fin de séquence sont posées systématiquement. Les informations saisies dans les variables de fin de séquence qui concernent le départ de l'entreprise sont recopiées dans les variables de début de séquence qui concernent l'embauche. C'est le seul cas où les informations de fin sont copiées sur le début de séquence.

3. La codification des variables

a) *Contrats de travail*

Pour les périodes d'emploi, plusieurs questions abordent la nature du contrat de travail en début de séquence et en fin de séquence, permettant par exemple de repérer les personnes qui accèdent à une entreprise via un emploi court (emploi aidé, mission d'intérim) puis y restent après passage en contrat à durée indéterminée.

Deux variables de synthèse sont donc construites pour définir le contrat de travail en début de séquence (STAT_EMB) et le contrat de travail en fin de séquence (STAT_FIN). Le tableau ci-dessous présente les différentes modalités existantes et les questions mobilisées pour les constituer.

Tableau 14 : La codification du contrat de travail

Modalité*	Nature du contrat	Questions mobilisées
01	Indépendant (à son compte)	EP1
02	Aide familiale	EP1
03	Fonctionnaire	EP20 /EP26
04	CDI	EP20 /EP26
05	CDD sans distinction privé public	EP20 /EP26
06	Apprentissage	EP21 /EP27
07	Intérim	INTER, EP20 /EP26
08	Contrat de professionnalisation	EP21 /EP27
09	CAV	Recodage de O_EP21/O_EP27
10	CAE, CEC	Recodage de O_EP21/O_EP27
11	Autre contrat aidé (dont CIRMA, CIE)	Recodage de O_EP21/O_EP27
12	Saisonnier	EP21 /EP27
13	Stagiaire	Recodage de O_EP21/O_EP27
14	Pas de contrat	Recodage de O_EP21/O_EP27
16	Autres	Recodage de O_EP21/O_EP27
20	ATER	Recodage de O_EP21/O_EP27
21	Post-doctorat + contrats recherche	Recodage de O_EP21/O_EP27
22	Vacataire	Recodage de O_EP21/O_EP27
23	Contrat étranger	Recodage de O_EP21/O_EP27
24	Bourse	Recodage de O_EP21/O_EP27
25	CNE	Recodage de O_EP21/O_EP27
26	Intermittents du spectacle	Recodage de O_EP21/O_EP27
27	CUI - CAE	Ep21 /ep27
28	CUI - CIE	Ep21 /ep27
29	Contrat d'avenir	Ep21 /ep27
30	CDD secteur public	Ep21 /ep27
31	CDD secteur privé	Ep21 /ep27
32	Allocataire de recherche	Recodage de O_EP21/O_EP27
33	Autre contrat aidé (CDDI)	Recodage de O_EP21/O_EP27
34	Contrat engagement éducatif	Recodage de O_EP21/O_EP27
35	Service civique	Recodage de O_EP21/O_EP27
36	Volontariat	Recodage de O_EP21/O_EP27

*Ce découpage est le plus large existant. Pour les différentes publications, une agrégation de données a été utilisée.

b) Variables à créer en amont de la codification du secteur d'activité

- PUB = distinction secteur public ou secteur privé

La variable PUB détermine si une entreprise appartient au secteur public ou au secteur privé.

Pour faciliter la codification, on utilise un menu qui contient des noms d'entreprise connus et dont on connaît leur statut. Dans le cas où l'entreprise n'est pas renseignée dans le menu, nous récupérons l'information déclarée par l'individu. Les modalités sont :

- 1 = État
- 2 = Collectivités, HLM, Hôpitaux, ...
- 3 = Sécurité sociale
- 4 = Entreprises publiques nationalisées
- 5 = Privé
- X ou manquant = Indéterminé

Remarque : aucune mise à jour n'a été effectuée de la base des noms d'entreprise qui permet de coder la variable PUB et ce depuis la création des bases génération.

PUB est une variable intermédiaire qui permet de calculer la variable NATENTR (nature de l'entreprise) ainsi que la PCS, elle ne sera pas disponible dans les bases finales.

- NATENTR = Statut de l'établissement employeur, 10 modalités

La variable NATENTR est construite à partir de la variable PUB. Les modalités de cette dernière sont détaillées afin d'en définir des plus précises pour NATENTR :

- 11 = Éducation nationale <=> "1" ou "2" modalité PUB
- 12 = Armée <=> "1" modalité PUB
- 13 = Autre État <=> "1" ou "2" modalité PUB
- 21 = Collectivités territoriales <=> "2" modalité PUB
- 22 = Hôpitaux <=> "2" modalité PUB
- 23 = Divers secteur public <=> "2" modalité PUB
- 30 = Sécurité sociale <=> "3" modalité PUB
- 40 = Entreprises publiques nationalisées <=> "4" modalité PUB
- 50 = Secteur Privé <=> "5" modalité PUB
- 60 = Indéterminé <=> X ou manquant modalité PUB

- STATUT = Statut dans l'emploi

Les deux variables créées STAT_EMB et STAT_FIN permettent de voir les changements de contrat de travail durant la séquence d'emploi.

La variable STAT_EMB est construite à partir de la situation à l'embauche décrite dans chacune des séquences d'emploi du calendrier professionnel (emploi issu de la période des 2 ans après la dernière interrogation).

La variable STAT_FIN est le statut de fin d'emploi des périodes d'emploi décrites dans le calendrier, si elles se terminent avant la date d'interrogation, ou du statut au moment de l'interrogation, pour la séquence d'emploi en cours.

Si l'individu interrogé n'a pas changé d'emploi ni de contrat de travail depuis la précédente interrogation seule STAT_FIN sera calculée.

- POSPRO = Classification professionnelle

Idem que pour le statut dans l'emploi, deux variables sont créées pour la position professionnelle déclarée POSPRO_EMB et POSPRO_FIN à l'embauche et en fin d'emploi.

- FONCT = Fonction principale

Deux variables créées également pour la fonction professionnelle FONCT_EMB et FONCT_FIN.

c) *Secteur d'activité de l'établissement employeur*

Dans l'enquête Génération 2010, le secteur d'activité des établissements est codé en NAF rev2 (nomenclature des activités économiques en vigueur en France depuis le 1er janvier 2008). La NAF a la même structure que la nomenclature d'activités de la Communauté européenne (NACE rév.2) mais

elle comporte un niveau supplémentaire, spécifique à la France, celui des sous-classes. L'environnement Sicore activité a été mobilisé pour cette codification. La NAF révision 2 comporte cinq niveaux comprenant respectivement : 21, 88, 272, 615 et 732 postes. Avec les informations recueillies dans l'enquête, le choix s'est porté sur la nomenclature de niveau 2 (88 divisions). Cette nomenclature est disponible en ANNEXE 6 -

La procédure mise en place pour Génération 2010 à 7 ans :

Dans les enquêtes générations, l'activité économique de l'établissement est définie à partir des questions EP2A à EP10. L'enchaînement des questions permet de définir l'activité *principale* et de recueillir l'information sur deux niveaux :

- un menu déroulant optimisé avec des noms d'entreprise/administration pour coder facilement l'activité en amont (sans question subsidiaire)
- à défaut, une combinaison de questions permettant d'identifier l'activité en aval
 - o Soit par l'intermédiaire de questions fermées
 - o Soit de questions fermées assorties d'une question ouverte contenant clairement la nature de l'activité

Par conséquent, avec cet ensemble d'information hétérogène, deux méthodes de codage ont été utilisées séquentiellement :

- Imputation du code NAF à partir du menu entreprise (26 % des périodes d'emploi) :

Pour réduire le temps de passation du questionnaire et faciliter le travail *a posteriori* de codification, un menu déroulant a été construit au moment de la préparation de la Génération 2004 à 3 ans. Ce menu est une table contenant des noms de grandes entreprises du CAC 40 ainsi que des noms génériques d'entreprises/administrations dont l'activité principale est facilement identifiable. Ainsi, lors de l'interrogation, si le nom de l'entreprise/administration cité existe dans ce menu, l'activité est codée selon la nomenclature NAF associée sans question supplémentaire.

L'intérêt de cette méthode est d'éviter de poser des questions de description de l'entreprise (et donc du temps de questionnement) ainsi que des questions dont la réponse est évidente dès lors que le nom de l'entreprise ou de l'administration est disponible. En revanche, ce menu a plusieurs limites : certains grands groupes ont de nombreux établissements avec des activités principales différentes, certains sigles sont ambigus et il est très incomplet, en particulier pour les PME.

- Imputation par codage automatique

Cas 1 : Traitement de l'information résultant d'un questionnement fermé (14 % des périodes d'emploi) :

Le questionnement permet de préciser la nature de l'activité et de trouver le code NAF le plus approchant. Il s'agit de déterminer l'activité : est-elle liée à la fabrication ou au commerce de produits, alimentaires ou non alimentaires ? Ou plutôt en lien avec de la vente de services ? Pour des entreprises ou des particuliers ? À partir de ces informations, un codage automatique a été réalisé. Si l'enquêté arrivait à se positionner dans cet enchaînement de questions (sans jamais répondre « autre »), il ne renseignait pas l'activité en clair de son établissement.

L'outil INSEE de recherche dans la nomenclature NAF s'est révélé précieux. Il est disponible sur le site : http://recherche-naf.insee.fr/SIRENET_Template/Accueil/template_page_accueil.html.

Si les questions fermées permettaient de conclure sur le code en NAF rev2, celui-ci était affecté. En revanche si cela ne permettait pas de conclure, le nom de l'entreprise/administration et sa localisation étaient alors mobilisés pour trouver l'activité de l'établissement.

Cas 2 : Traitement de l'information résultant d'un questionnaire fermé + information en clair de la nature de l'activité (47 % des périodes d'emploi) :

Dans le questionnaire fermé, l'enquêté a la possibilité de préciser en clair la nature de l'activité dans une modalité « autre » lorsqu'il n'est pas parvenu à se positionner dans les modalités proposées.

La généralisation de la codification automatique avec l'outil SICORE dans l'enquête Génération 2010 a facilité la transition pour aboutir à une codification de qualité selon la nomenclature NAF.

Encadré 3 - SICORE Environnement Activité (millésime 2012)

Sicore permet de coder l'activité en NAF rév. 2. Cet environnement de codage est communément appelé Sicape (i.e. Sicore APE). Le codage n'utilise pas de variables annexes ; il s'appuie uniquement sur le libellé d'activité déclaré. Cet environnement Sicore a la particularité de retourner plusieurs codes APE, avec les probabilités associées.

Pour l'enquête Génération 2010, des règles de décisions ont été définies afin de repérer le meilleur écho.

- 1^{ère} règle : si Sicore fournit 1 SEUL écho dont la probabilité est supérieure à 0,40, l'écho en 5 positions est choisi.
- 2^{ème} règle : si Sicore fournit des échos multiples :
 - sélection des 2 premières positions des différents échos
 - calcul de la somme des probabilités des échos identiques sur 2 positions successives (premier écho inclus).

Dans ce deuxième cas, si la somme des probabilités est supérieure à 0,40, l'écho sur 2 positions est choisi. Le taux de codage final est de 87,4 %.

– Codification manuelle (13 % des périodes d'emploi)

Pour les périodes d'emploi dont le secteur d'activité n'a pas pu être codé par les deux méthodes précédentes, un codage manuel a été réalisé par une société extérieure à partir du nom de l'entreprise/administration, des informations en clair et des questions fermées (EP2A à EP10).

d) Profession des périodes d'emploi

Dans l'enquête Génération, pour chaque séquence longue en entreprise, l'enquêté décrit sa profession à l'embauche et à la fin de la séquence (si celle-ci a évolué).

La profession a été codée selon la nomenclature des professions et catégories socioprofessionnelles (PCS) dans sa version actuelle datant de 2003.

Cette nomenclature a été créée par l'Insee pour disposer d'une classification pertinente pour l'analyse des comportements socio-économiques des personnes. Elle classe la population selon une synthèse de la profession (ou de l'ancienne profession), de la position hiérarchique et du statut (salarié ou non).

La PCS comporte trois niveaux d'agrégation emboîtés : 8 groupes socioprofessionnels, 24 ou 42 catégories socioprofessionnelles et 486 professions.

La construction de la PCS s'appuie d'abord sur le libellé de l'emploi demandé à l'individu, complété par d'autres informations relatives à la position professionnelle de la personne et aux caractéristiques de l'entreprise dans laquelle elle travaille (appartenance au secteur public, taille de l'entreprise, secteur d'activité). Des corrections orthographiques et une normalisation ont été effectuées sur une partie des libellés, la plupart étant cependant issus d'un menu déroulant très détaillé proposé par l'application de collecte.

À partir de ces informations, le logiciel SICORE, permet de coder automatiquement la PCS.

Encadré 4 - SICORE Environnement PCS 2013

Le logiciel Sicore permet d'effectuer un codage automatique des professions à partir d'un questionnement en face à face ou bien à partir d'un fichier en entrée (fonctionnement en « Batch »). C'est le fonctionnement qui a été utilisé pour l'enquête Génération 2010.

Sicore en Batch nécessite en entrée un fichier en format texte qui comporte les variables suivantes (cf ANNEXE 7 - Définition des variables utilisées dans Sicore PCS)

- Identifiant
- Libellé de la profession
- STATUT : statut dans l'emploi (3 modalités)
- PUB : statut de l'établissement employeur (5 modalités)
- CPF : classification professionnelle ou qualification (10 modalités)
- FN : fonction principale (9 modalités)
- NBS : nombre de salariés employés (4 modalités)
- NAF4 : activité principale de l'établissement sur 4 caractères
- NAF2 : activité principale de l'établissement sur 2 caractères
- T : taille de l'entreprise (4 modalités)
- OPA : orientation des productions agricoles (8 modalités)
- DEP : département
- SAU : surface agricole utilisée (88 modalités)
- S : sexe
- SP : statut précaire
- STRE : emploi actuellement (Oui ou Non)

Il procède en 3 étapes successives :

- Étape 1 : il analyse le libellé de la profession
- Étape 2 : il entre dans une table de décision et fait éventuellement appel à certaines variables annexes (cf. la liste ci-dessus)
- Étape 3 : il donne en sortie au mieux un code PCS, ainsi qu'un indice de confiance du code proposé.

Dans le cadre de l'enquête Génération 2010, l'utilisation de la NAF2 a été privilégiée (la NAF4 n'étant que très partiellement présente) et les variables OPA, DEP, SAU et STRE n'ont pas été utilisées car informations non disponibles dans l'enquête. Toutes les autres variables sont utilisées comme variables annexes.

Ce système a permis d'atteindre un taux de codification de 65 %.

Les professions qui n'ont pas pu être codées par cette procédure (35 % des séquences d'emploi) ont été transmises à une société externe qui a procédé à une codification manuelle.

e) *Variables de localisation géographique*

Pour l'ensemble des répondants, on dispose de la localisation géographique de la commune de résidence à la date de l'enquête en 2017. Pour chaque séquence en entreprise, la commune du lieu de travail est également renseignée.

Les communes sont repérées à partir de leur code officiel géographique ce qui permet de mobiliser les différents zonages administratifs (région département) et d'études développés par l'Insee (unités urbaines, aires urbaines, zone d'emploi).

Pour le zonage en aires urbaines (ZAU), les informations retenues sont la taille de l'aire urbaine et le type d'espace auquel la commune appartient (pôles urbain, couronne périurbaine, commune multipolarisée, communes isolées).

Chaque variable contenant une localisation géographique a été déclinée sous plusieurs variables, du format « nom de variable + nom du type code géographique ».

Voici un exemple de résultat pour la variable « CA0A » qui est la résidence à la date de l'enquête :

- CA0ADEP : Département de résidence de la commune à la date de l'enquête
- CA0AREG : Code de la région de résidence à la date de l'enquête
- CA0AREGIONF : Libellé de la région de résidence à la date de l'enquête
- CA0ACATAEU : Catégorie d'espace du ZAU 2010 pour la commune de résidence à la date de l'enquête
- CA0ATAU : Tranche d'aire urbaine 2010 pour la commune de résidence à la date de l'enquête
- CA0ASTATUTUU : Statut de communes selon la définition des unités urbaines 2010 pour la commune de résidence à la date de l'enquête (ville centre, banlieue, isolée, rurale)
- CA0ATYPEUU : Type de communes selon la définition des unités urbaines 2010 pour la commune de résidence à la date de l'enquête (rural/urbain)
- CA0AZE : Code de la zone d'emploi de la commune de résidence à la date de l'enquête
- CA0ALIBZE : Libellé de la zone d'emploi de la commune de résidence à la date de l'enquête

f) *Codification des ouverts*

Certaines variables présentent une modalité « Autre », qui offre la possibilité d'une réponse en clair. Dans la mesure du possible, les réponses en clair données par les individus ont été reclassées.

19 variables ont fait l'objet de recodifications. Toute réponse ouverte donnée par un individu a été analysée. Lorsqu'un reclassement dans une modalité existante était possible, il a été effectué. Lorsque des modalités non proposées dans le questionnaire étaient citées de manière récurrente dans les ouverts, ces nouvelles modalités ont été créées. Les réponses n'ayant pas pu être reclassées selon les précédents critères sont restées dans la modalité « Autre ».

Les variables concernées sont signalées dans le dictionnaire des variables. Une indication « modalité ajoutée » est notée face aux nouvelles modalités créées.

La liste des variables ouvertes est disponible en ANNEXE 8 - Liste des questions ouvertes.

4. Le traitement des salaires (primes incluses)

Au moment de l'interrogation, il est demandé systématiquement aux individus, pour chaque séquence d'activité déclarée, le salaire net perçu à l'embauche et le salaire de fin de séquence ou actuel (si augmentation du salaire dans la séquence). Uniquement les observations des salariés sont concernées par cette phase d'imputation. Les séquences d'activité correspondant aux situations d'indépendants ou du statut d'aide familial ne sont pas prises en compte puisqu'il s'agit de revenu. Pour ces derniers, les rémunérations fournies dans les bases sont REVDEB et REVFIN (information brute). Cette question sur les salaires est dite "sensible", on donne donc la possibilité aux enquêtés de répondre soit en clair, soit par tranches de salaire. Le cas échéant, ils ont aussi la possibilité de ne pas répondre. Les structures du type de déclaration des salaires sont les suivantes pour les débuts et les fins des séquences d'emploi.

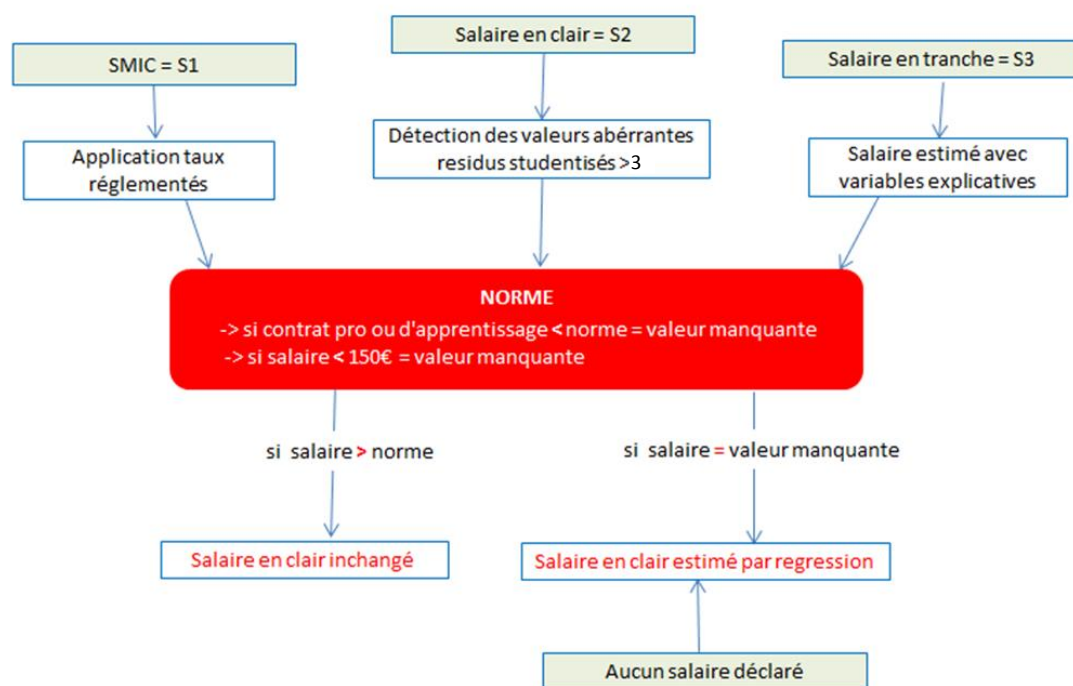
Tableau 15 - Structures du type de déclaration pour les débuts de séquences entre 2015 et 2017 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations)

Type de déclaration	Nombre de séquences	Pourcentage
Clair	4 148	83,7
Smic	335	6,8
Tranche	356	7,2
Manquant	119	2,3

Tableau 16 - Structures du type de déclaration pour les fins de séquences entre 2015 et 2017 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations)

Type de déclaration	Nombre de séquences	Pourcentage
Clair	9 937	80,6
Smic	405	3,3
Tranche	706	5,7
Manquant	1 283	10,4

Le traitement des salaires se fait en fonction du mode de réponse. Le déroulement des étapes est présenté dans le schéma suivant :

Figure 7 - Gestion des salaires

Tous les salaires sont rapportés à une « norme » construite à partir de la valeur du SMIC d’une période donnée (deux revalorisations du SMIC sur les deux nouvelles années observées) mais aussi en fonction de la quotité de travail. Pour les contrats de professionnalisation et les contrats apprentissage, on intègre l’âge et l’avancement dans la formation.

Procédure de calcul

Pour l’ensemble des modèles estimés, une sélection « stepwise » du meilleur modèle a été effectuée à partir de critères d’informations. Deux modèles distincts sur les salaires manquants (temps plein/temps partiel) ont été spécifiés. Pour les salaires déclarés en tranche, un modèle différent a été utilisé pour réaliser l’imputation.

Le modèle est réalisé sur l’ensemble des salaires déclarés, c’est-à-dire toutes interrogations confondues. Pour prendre en compte l’inflation et la conjoncture, la variable « mois de début de séquence », pour les salaires de début, ou la variable « mois de fin de séquence » pour les salaires de fin de séquences est ajoutée au modèle.

Étape 1 : Détermination des salaires aberrants

Une première détermination des salaires aberrants a été faite à partir d’un modèle de régression basé sur les variables sexe, âge, niveau de sortie, plus haut diplôme, spécialité, type de contrat, catégorie sociale, région, taille de l’entreprise, activité de l’entreprise, ancienneté, mois. À l’issue de ce modèle, les observations dont les résidus studentisés sont supérieurs à 3 ont été considérés comme aberrants, et les salaires correspondants ont été mis à valeur manquante. Cette opération concerne environ 100 séquences que ce soit pour les salaires de début ou de fin.

Étape 2 : Imputation des salaires déclarés à partir des tranches

Lorsque le salaire a été déclaré via les tranches, un modèle de régression mobilisant les variables suivantes a été spécifié : sexe, âge, niveau de sortie, le plus haut diplôme, le type de contrat, la catégorie sociale, l'activité de l'entreprise, la région, l'ancienneté, « la classe de la tranche de salaire », le mois. Ce modèle a été estimé sur les observations ayant le salaire déclaré en clair. La variable à expliquer de ce modèle est le salaire en clair, les tranches de salaires ont été recalculées à partir des salaires en clair.

Étape 3 : Comparaison des salaires déclarés aux normes en vigueur.

Les salaires déclarés sont comparés à une norme « basse » en fonction de la quotité de travail et du statut et calculée à partir du taux horaire du SMIC net et des taux réglementés (pour les contrats spécifiques tels que le contrat de professionnalisation et le contrat d'apprentissage). Si les salaires sont inférieurs à cette norme ils sont mis à valeur manquante, dans le cas contraire ils sont conservés.

Le SMIC a été réévalué deux fois entre avril 2015 et décembre 2017 (période d'interrogation)

Tableau 17 – Revalorisation du SMIC entre avril 2015 et décembre 2017

Date de revalorisation	Montant net du SMIC
Janvier 2016	1 141,61
Janvier 2017	1 149,07

*Source INSEE

Le contrat de professionnalisation (nouveau contrat de qualification)

Ce contrat s'adresse aux jeunes de 16 à 25 ans, aux demandeurs d'emplois de 26 ans et plus et aux bénéficiaires de certaines allocations ou contrats. Les bénéficiaires de 16 à 25 ans révolus sont rémunérés en pourcentage du SMIC selon leur âge et leur niveau de formation. Les salariés âgés de 26 ans et plus perçoivent une rémunération qui ne peut être ni inférieur au SMIC, ni à 85% du salaire minimum conventionnel.

Le calcul du salaire prend en compte l'âge et le niveau de formation pour chaque individu. La variable NISOR (niveau de sortie sur 15 positions) est utilisée comme référence pour le niveau de formation.

Tableau 18 – Normes de rémunérations des contrats de professionnalisation

	Inférieur au BAC	BAC et plus
Moins de 21 ans	55 %	65 %
21 ans et plus	70 %	80 %

*pourcentage du SMIC

Ce salaire ne peut être inférieur à 55 % du smic pour les bénéficiaires âgés de moins de 21 ans et à 70 % du smic pour les bénéficiaires de 21 ans et plus. Ces rémunérations ne peuvent être inférieures,

respectivement à 65 % et 80 % du smic, dès lors que le bénéficiaire est titulaire d'une qualification au moins égale à celle d'un baccalauréat professionnel ou d'un titre ou diplôme à finalité professionnelle de même niveau.

Le contrat d'apprentissage

Ce contrat s'adresse aux jeunes de 16 à 25 ans mais des dérogations à ces limites d'âges sont possibles. L'apprenti perçoit un salaire déterminé en pourcentage du smic et dont le montant varie en fonction de l'âge du bénéficiaire et de sa progression dans le ou les cycles de formation faisant l'objet de l'apprentissage.

Tableau 19 - Normes de rémunérations des contrats d'apprentissage

Année d'exécution du contrat	Moins de 18 ans	De 18 ans à moins de 21 ans	21 ans et plus
1^{ère} année	25 %	41 %	53 %
2^{ème} année	37 %	49 %	61 %
3^{ème} année	53 %	65 %	78 %

**Pourcentage du smic*

Au titre de sa progression dans un cycle de formation, l'apprenti bénéficie d'une rémunération variant en fonction de l'année d'exécution du contrat. Ainsi, le salaire minimum perçu par l'apprenti correspond au pourcentage du SMIC (ou, dans certains cas, du salaire minimum conventionnel), allant de 25 % à 78 % déterminé en fonction de son âge et de sa progression dans le cycle.

Pour déterminer l'année d'exécution du contrat, on suppose que la période d'emploi sur laquelle il déclare être apprenti correspond à son lieu d'apprentissage. La variable durée est utilisée par défaut comme paramètre pour déterminer l'année d'exécution du contrat.

Un test de cohérence est réalisé comme pour les autres contrats pour vérifier si le salaire déclaré est proche du salaire calculé en fonction des conditions de rémunération dans le cadre d'un contrat d'apprentissage ou d'un contrat de professionnalisation.

Étape 4 : Imputation des salaires manquants

Les salaires inférieurs à 150 € et les salaires manquants sont imputés par un modèle faisant intervenir les variables : sexe, âge, niveau de sortie, indicatrice de qualification, plus haut diplôme, spécialité, type de contrat, catégorie sociale, région, taille de l'entreprise, activité de l'entreprise, ancienneté. Les modèles de régression expliquent environ la moitié de la variance totale. Des modèles différents ont été utilisés pour imputer les observations à temps plein ou les observations à temps partiel.

Étape 5 : Le traitement des primes

Comme pour les salaires, des questions sur les primes sont posées à chaque interrogation aux individus qui déclarent un salaire hors primes. Les enquêtés ont la possibilité de déclarer des primes en clair, en tranches ou de refuser de déclarer leur montant.

Remarque : dans les bases mises à disposition, les informations sur les primes ne sont pas fournies car intégrées directement dans le calcul du salaire d'embauche ou de fin.

Il existe deux types de primes :

- la prime dite de « 13^{ème} mois » : si l'individu déclare toucher un 13^{ème} mois, une prime mensuelle est calculée en fonction du salaire déclaré corrigé. On ajoute cette prime au salaire.
- les « autres primes » : des primes mensuelles sont collectées en clair ou en tranches. Le même traitement que pour le salaire est effectué ; pour chaque individu dont la prime est manquante, une prime est déterminée par imputation sur modèle de régression.

L'objectif étant de déterminer un salaire « primes incluses » en clair pour toutes les séquences d'emplois (hormis les jobs de vacances et les emplois de statut aide familiale).

5. Anonymisation des données et finalisation

Afin d'assurer la confidentialité des réponses des personnes enquêtées et conformément au principe du secret statistique garanti par la loi de n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, les fichiers de résultats ont été rendus anonymes.

L'anonymisation consiste d'abord à supprimer les informations directement nominatives (nom et prénom des personnes), ainsi que les coordonnées téléphoniques, postales et mails, nécessaires à la réalisation de la collecte de l'enquête.

Elle consiste aussi à s'assurer qu'une identification indirecte par recoupement de différentes informations disponibles est impossible. Les localisations géographiques trop fines (la commune) et les noms des entreprises sont supprimés ; les questions ouvertes sont supprimées, seule est conservée la codification de ces variables.

6. Formats, labels et dictionnaire des variables

Les labels des questions de l'enquête ont été récupérés pour la majeure partie des tables SAS de l'enquête à trois ans mais aussi du questionnaire.

Un fichier SAS a été créé contenant tous les formats des variables de la base dont l'ordre de création est basé sur l'enchaînement des questions du questionnaire d'origine de Génération 2010 à 7 ans. Pour la grande majorité, les formats ont été créés en tenant compte des modifications des modalités de certaines variables du questionnaire, en particulier des questions ouvertes (traitées en amont avec leurs nouvelles modalités) mais aussi de certains modules du questionnaire ajoutés.

En parallèle, le dictionnaire des variables a été créé à partir du dictionnaire de l'interrogation de 2015 et complété avec les nouvelles variables ou modifié pour les variables de 2015 qu'il a fallu suffixer. Le dictionnaire contient un index papier et un index créé automatiquement par ordre alphabétique avec les numéros de pages en face des noms de variables, ce qui permet de rediriger directement l'utilisateur sur l'information via un lien hypertexte.

7. Les tables finales

À l'issue des différents traitements, les fichiers de résultats de l'enquête Génération 2010 à 7 ans se répartissent en trois tables. Ces tables concernent les personnes du champ Céreq ayant répondu à la vague précédente et au questionnaire actuel.

Tableau 20 : Caractéristique des tables créées

Nom	L'observation dans la table est :	Contenu
Table « Individus »	Le répondant (8 882 observations)	Ensemble des réponses individuelles de l'enquête, y compris le calendrier du mode de cohabitation. Informations issues de la précédente vague. Pondérations. <i>Environ 1000 variables</i>
Table « Seqentr »	La séquence d'emploi (26 056 observations)	Ensemble des périodes d'emploi présentes dans le calendrier professionnel et de la vague précédente. Une clé (IDENT,NSEQ) permet d'associer les périodes au répondant concerné de la table « Individus ». <i>Environ 300 variables</i>
Table « Nonempl »	La séquence de non emploi (16 487 observations)	Ensemble des périodes de non emploi présentes dans le calendrier professionnel et de la vague précédente. Une clé (IDENT,NSEQ) permet d'associer les périodes au répondant concerné de la table « Individus ». <i>Environ 250 variables</i>

Les bases comparables

Pour usage interne, le Céreq a réalisé une base nationale à champ comparable pour les enquêtes à 7 ans des Générations 1998, 2004, 2010.

Les principaux traitements réalisés sont les suivants :

- Définition d'une nomenclature de diplôme agrégée commune aux trois générations
- Recalcul des pondérations sur les générations précédentes (lié à l'affinement des marges de calage)
- Suppression des Dom pour la Génération 2010
- Suppression des sortants d'IUFM pour les Générations 2004 et 1998.

Il est important de noter que la comparaison d'une même variable nécessite de revenir aux formulations et aux modalités de la question dans chacune des enquêtes.

VI. La pondération finale

Le principe général de la repondération de l'enquête Génération 2010 à 7 ans est le suivant : il s'agit de multiplier la pondération finale de l'enquête à 7 ans par un coefficient lié à la non-réponse pour obtenir le poids que chaque individu représente dans l'enquête. Ensuite, ces poids sont calés sur des indicateurs des enquêtes à 3 ans et 5 ans de la génération 2010 afin de garantir une cohérence entre les estimations à des paramètres de l'interrogation à 3 ans et 5 ans dans la population de répondants à 7 ans (et de corriger en partie certains défauts de l'étape de correction de non-réponse).

1. Échantillonnage

Pour la ré-interrogation à 7 ans de la Génération 2010, il n'y a pas eu de ré-échantillonnage des répondants à l'interrogation à 5 ans. Cependant, les proches de deux individus ont indiqué que ces personnes étaient décédées. De plus, une personne a exprimé ne plus vouloir être réinterrogée par mail suite à cette opération de préfidélisation. Au final, sur les 13 710 répondants à l'interrogation à 5 ans, la totalité est envoyée en production de l'enquête 2010 à 7 ans.

2. Exploitation de l'historique d'appel pour obtenir des parodonnées pour la modélisation de la non-réponse

À partir des fichiers de bilan du protocole de l'enquête G10-7 ans, envoyés par le prestataire, des variables ont été créées en vue du traitement de la non-réponse.

Dans un premier temps, pour chaque individu envoyé en production, ont été créées les variables indiquant s'il a :

- Répondu à l'enquête
- Arrêté en cours d'enquête
- Été contacté directement, ou a eu un tiers contacté
- Été Identifié sur son répondeur
- Été identifié hors champ (décédé, injoignable permanent ou résident à l'étranger) s'il a été contacté. De fait ce ne sont que des décès

Des variables concernant les différentes lettres-avis, et relances ont été créées :

- Lettre-avis papier envoyée
- Lettre-avis mail envoyée
- Lettre-avis mail reçue
- Relance mail envoyée
- Relance SMS envoyée

À partir de l'historique d'appel, on a pu obtenir les variables concernant :

- Le nombre d'appels reçus
- Le nombre de jours où ces appels ont été reçus
- Le nombre de numéros de téléphone appelés

- Le fait d'avoir ou non pris un rendez-vous téléphonique
- Le fait d'avoir reçu un message sur son répondeur par un enquêteur

Les individus avaient possibilité, suite aux courriers avis, de valider ou de mettre à jour leurs coordonnées sur un CAWI de fidélisation, ainsi que d'y prendre un rendez-vous téléphonique. Le prestataire a envoyé un fichier permettant de savoir si :

- L'individu s'est connecté au CAWIFID
- L'individu a modifié ou validé ses coordonnées

Cependant, il est impossible de savoir s'il a pris un rendez-vous, ou ses préférences de jour d'appel, car nous n'avons pas reçu ces informations.

Dans les lettres-avis, il était indiqué que les individus pouvaient appeler sur un n° gratuit pour avoir plus d'informations sur l'enquête. Les fichiers permettent de reconstituer :

- L'appel sur un n° gratuit
- Le fait d'être tombé ou non sur un répondeur

Enfin, en fin de plateau, les individus ayant exprimé un refus de participer à l'enquête, ceux ayant déclaré qu'ils allaient rappeler sur le n° gratuit mais ne l'ont pas fait, ainsi que les tiers ayant refusé de transmettre les coordonnées de l'enquêté ont été rappelés dans le but d'obtenir une enquête. Pour ces personnes, on a pu créer les variables concernant :

- Le type de refus
- Le fait d'avoir été rappelé ou non
- Le nombre de tentatives d'appel après le refus.

3. Modélisation de la non-réponse

La modélisation de la non-réponse se fait en deux étapes : la probabilité de réponse de chaque individu de l'échantillon est estimée par modélisation logistique, puis des groupes homogènes de réponse ont été effectués à partir de ces probabilités de réponse estimées par modélisation pour estimer finalement empiriquement la probabilité de réponse des individus de chacun de ces groupes.

a) *Modèle logistique de non-réponse*

Dans un premier temps, la probabilité de réponse de chaque individu a été estimée par régression logistique. Le modèle final résulte d'une sélection STEPWISE des variables présentes en Annexe 9. Celle-ci retient les variables suivantes :

Tableau 21 - Variables sélectionnées dans le modèle logistique

Étape	Effet	DDL	Pr>chi-2	Libellé de la variable
1	PHD	7	<.0001	Plus haut diplôme obtenu
2	I_relance*relancemail	1	<.0001	Avoir reçu une relance par mail pour ceux n'ayant pas répondu avant la date de relance
3	CA32	2	<.0001	Disposé pour un entretien qualitatif
4	Refus_all*rappel_all	1	<.0001	Être rappelé après refus parmi ceux ayant refusé
5	Appelgratuit	1	<.0001	Avoir appelé le numéro gratuit
6	Appelngra*repondeurn	1	<.0001	Être tombé sur le répondeur du numéro gratuit
7	Ca8	5	<.0001	Nationalité de la mère
8	Papier	1	<.0001	Avoir reçu une lettre avis-papier
9	Sitde_15	4	<.0001	Situation professionnelle à la date d'enquête en 2015
10	Appelngra*appeldirec	1	<.0001	Avoir eu une réponse à l'appel du numéro gratuit
11	CA0_13	2	0,0005	Avoir changé de commune de résidence entre 2013 et 2015
12	Habd_15	2	0,0009	Situation d'habitat à la date d'enquête en 2015
13	Connexion_cawifid	1	0,0007	S'être connecté au Cawi de fidélisation
14	Connexion_cawifid	2	<.0001	Avoir modifié/validé ou non ses coordonnées après connexion au Cawi de fidélisation
15	Op6_15	2	0,0029	Optimisme avenir professionnel (Question de G2010 à 5 ans)
16	Op2_15	2	0,0006	Opinion sur le parcours et l'avenir professionnel (Interrogation de 2013)
17	Ca20_15	2	0,0053	Enfants (en 2015)
18	Age10	1	0,0391	Age de l'individu en 2010 (à la sortie du système éducatif)
19	Sitde_13	4	0,0477	Situation professionnelle à la date d'enquête en 2013

La distribution des probabilités estimées pour les individus répondants est la suivante :

Tableau 22 - Probabilités de réponse des répondants estimées par le modèle logistique

N	Moy	Ecart-type	Min	1 ^{er} ctl	5 ^{ème} ctl	10 ^{ème} ctl	25 ^{ème} ctl	50 ^{ème} ctl	75 ^{ème} ctl	90 ^{ème} ctl	95 ^{ème} ctl	99 ^{ème} ctl	Max
8919	0,697	0,134	0,018	0,293	0,447	0,515	0,623	0,721	0,797	0,849	0,865	0,934	0,963

Les probabilités minimums sont de 0.018, ce qui conduit à une multiplication des poids par 55 ce qui est beaucoup trop. On a alors recours à des groupes de réponse homogène.

b) *Groupes de réponse homogène*

On construit les groupes de réponse en séparant l'ensemble des individus envoyés en production en 10 groupes selon les déciles des probabilités estimées par le modèle de régression logistique. La probabilité de répondre est estimée empiriquement à l'aide de ces groupes de réponse homogène par la proportion de répondants dans chaque groupe :

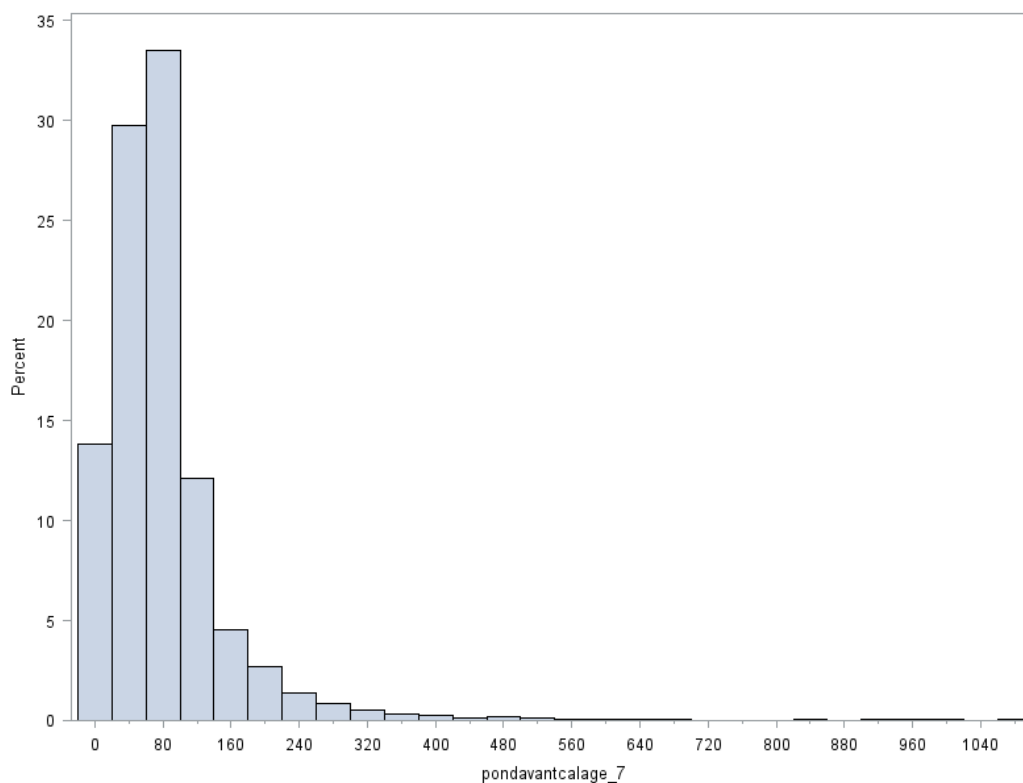
Tableau 23 - Groupes de Réponse homogène

Groupe	Nombre de répondants	Probabilité de réponse estimée
1	285	0,208
2	653	0,476
3	814	0,594
4	889	0,649
5	961	0,701
6	1019	0,743
7	1013	0,739
8	1070	0,780
9	1117	0,815
10	1098	0,801

Les probabilités minimums sont maintenant de 0.21, ce qui correspond à une multiplication du poids initial par 4.8, ce qui est plus raisonnable.

c) *Pondération avant calage*

Les poids avant calage sont obtenus pour chaque individu en multipliant son poids à l'interrogation à 5 ans par l'inverse de la probabilité de réponse estimée par groupes de réponse homogène. Leur distribution est la suivante :

Figure 8 - Distribution des poids avant calage

4. Calage sur les marges des enquêtes à 3 et 5 ans

Pour assurer que les estimations de grands indicateurs de l'enquête à partir de variables des enquêtes antérieures soient les mêmes à l'interrogation à 7 ans, un calage sur les marges de ces indicateurs a été effectué. Les indicateurs retenus sont les suivants :

- Région d'habitat en 2013 (regroupée en nouvelles régions) : *CA0_REG13*
- Région d'habitat en 2015 (regroupée en nouvelles régions) : *CA0_REG15*
- Situation professionnelle en Avril 2013 : *MOIS42*
- Situation professionnelle en Mars 2015 : *MOIS65*
- Mode de cohabitation à la date d'enquête en 2013 : *HABDE_13*
- Mode de cohabitation à la date d'enquête en 2015 : *HABDE_15*
- Lieu de naissance (France ou étranger) : *LIEUNAIS*
- Sexe : *Q1*
- Age en 2010 : *AGE10*
- Sexe*Plus Haut Diplôme (*Q1*PHD*) regroupés de cette manière car certaines cases manquaient d'individus :

Tableau 24 - Regroupements effectués pour le calage de PHD*Q1

PHD (Plus haut diplôme obtenu)	Q1 (Sexe de l'enquête)	
	Homme	Femme
Fréquence		
NON DIPLÔMÉ	399	257
CAP-BEP-MC INDUSTRIEL	470	87
CAP-BEP-MC TERTIAIRE	142	345
BAC PRO-BT-BP INDUSTRIEL	379	66
BAC PRO-BT-BP TERTIAIRE	185	387
BAC TECHNO INDUSTRIEL	54	14
BAC TECHNO TERTIAIRE	160	245
BAC GENERAL	231	263
BTS-DUT INDUSTRIEL	295	49
BTS-DUT TERTIAIRE	196	319
AUTRE BAC+2 INDUSTRIEL	11	8
AUTRE BAC+2 TERTIAIRE	51	39
BAC+2/3 SANTÉ SOCIAL	73	419
LICENCE PRO LSH GESTION DROIT	49	66
LICENCE PRO MATHS SCIENCE TECHNIQUE	147	74
L3 LSH GESTION DROIT	77	187
L3 MATHS SCIENCE TECHNIQUE	53	35
AUTRE BAC+3 LSH GESTION DROIT	31	40
AUTRE BAC+3 MATHS SCIENCE TECHNIQUE	5	3
BAC+4 LSH GESTION DROIT	37	49
BAC+4 MATHS SCIENCE TECHNIQUE	27	33
M2 LSH GESTION DROIT	242	385
M2 MATHS SCIENCE TECHNIQUE	159	99
AUTRE BAC+5 LSH GESTION DROIT	34	39
AUTRE BAC+5 MATHS SCIENCE TECHNIQUE	26	28
BAC+5 ÉCOLE DE COMMERCE	89	82
INGENIEUR	280	87
DOCTORAT SANTÉ	38	102
DOCTORAT HORS SANTÉ LSH GESTION DROIT	147	215
DOCTORAT HORS SANTÉ MATHS SCIENCE TECHNIQUE	494	316

a) Calage de l'enquête à 5 ans sur l'enquête à 3 ans

Puisque l'enquête à 5 ans n'avait pas été calée sur un ensemble aussi complet d'indicateurs de G10-3 (uniquement PHD*Sexe), il s'agit dans un premier temps de caler l'enquête de Génération 2010-5 ans sur l'ensemble des indicateurs de G10-3 ans, afin de pouvoir estimer des marges sur les indicateurs à 3 ans et à 5 ans issus de la même population.

On cale alors les individus répondants de G10-5ans sur les marges de G10-3 sur les variables *MOIS42*, *HABDE_13*, *CAOREG_13*, *Q1*PHD*, *Q1*, *AGE10* et *LIEUNAIS*. Les rapports de poids après utilisation de la méthode linéaire de la macro CALMAR (Insee) sont proches de 1, signe que les estimations à 5 ans de ces paramètres étaient proches des marges à 3 ans :

Tableau 25 - Distribution des rapports de poids du calage de G105

Niveau	Quantile
Max	1,257
99%	1,139
95%	1,098
90%	1,037
Q3	1,014
Médiane	0,995
Q1	0,976
10%	0,955
5%	0,944
1%	0,908
Min	0,853

À partir de cette pondération à 5 ans, on peut estimer toutes les marges souhaitées pour le calage des répondants à 7 ans.

b) Calage de l'enquête à 7 ans sur les précédentes

Le calage est réalisé à l'aide de la méthode linéaire de la macro CALMAR. La distribution des rapports de poids est la suivante :

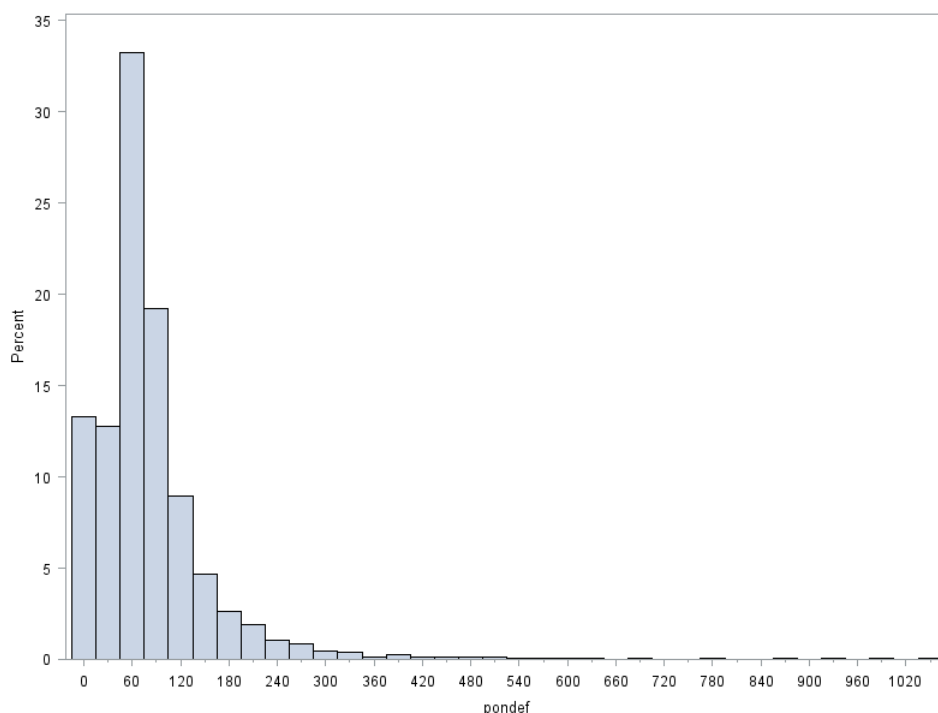
Tableau 26 - Distribution des rapports de poids du calage de G107

Niveau	Quantile
Max	1,488
99%	1,283
95%	1,177
90%	1,129
Q3	1,065
Médiane	1,003
Q1	0,947
10%	0,894
5%	0,855
1%	0,752
Min	0,481

5. La pondération finale

La distribution finale de la pondération de l'enquête Génération 2010 à 7 ans est la suivante :

Figure 9 - Distribution des poids finaux



On remarque quelques poids qui se détachent de la distribution. Ce sont des individus non-diplômés que la difficulté à contacter à chacune des interrogations conduit à repondérer fortement à chaque fois. Les distributions de poids par niveau de diplôme sont cependant moins étalées :

Tableau 27 - Distribution des poids finaux par niveau de diplôme

Plus haut diplôme obtenu	N	Moyenne	Écart-type	Min	P10	Q1	Médiane	Q3	P90	Max
NON DIPLOME	648	183,39	131,09	32,75	77,75	104,11	147,41	212,44	322,93	1057,94
CAP-BEP-MC	1033	99,05	63,71	14,91	47,50	59,78	82,40	114,86	173,76	691,14
BACCALAUREAT	1979	103,11	59,63	7,73	48,00	64,16	87,28	126,27	177,43	492,62
BTS-DUT, AUTRE BAC+2	966	80,57	38,29	9,68	44,04	56,07	73,31	94,65	126,40	330,48
BAC+2/3 SANTE SOCIAL	492	57,82	27,09	6,41	27,97	42,23	55,12	70,68	90,58	224,01
BAC+3/4 HORS SANTE SOCIAL	912	63,37	34,86	7,08	30,42	43,91	57,25	75,24	98,31	429,53
M2, GRANDES ECOLES, AUTRES BAC+5	1545	65,12	35,82	5,63	34,87	45,36	57,75	76,56	100,38	541,45
DOCTORAT	1307	13,80	22,87	2,42	3,92	4,80	6,45	10,16	40,79	496,26
TOUS	8882	79,71	69,43	2,42	8,76	43,92	66,17	97,85	148,32	1057,94

6. Difficultés rencontrées

Les principales difficultés rencontrées lors de la repondération de l'enquête Génération 2010 à 7 ans sont majoritairement liées au fait que c'est la troisième interrogation de cette Génération, et que les individus répondants ont déjà subi deux amplifications de poids liées à la non-réponse aux vagues précédentes. De ce fait, et sans surprise, les individus ayant le moins de chance de répondre à l'interrogation à 7 ans sont également ceux les moins susceptibles d'avoir répondu aux enquêtes à 3 ans et 5 ans, qui ont donc subi une forte modification de leur poids lors de ces enquêtes.

Pour éviter une trop forte dispersion des poids à la suite de cette troisième repondération, plusieurs choix dans la modélisation de la non-réponse ont été effectués, de sorte à ce que les coefficients de correction de non-réponse de la troisième interrogation ne soient pas trop élevés. Ainsi, les coefficients de multiplication des poids liés à la non-réponse sont distribués de cette manière :

Tableau 28 - Distribution des coefficients de non-réponse antérieurs de la Génération

Variable	Moyenne	Écart-type	Min	5 ^{ème} ctl	25 ^{ème} ctl	50 ^{ème} ctl	75 ^{ème} ctl	95 ^{ème} ctl	Max
Poids de tirage à 3 ans	5,85	4,23	1,00	1,00	2,13	5,30	8,21	14,08	20,95
Coefficient de NR à 3 ans	3,59	1,94	1,59	1,93	2,44	3,05	4,17	6,78	29,00
Pondef G103	24,48	16,84	1,56	3,06	10,73	23,23	33,80	55,26	134,36
Coefficient de NR à 5 ans	1,68	0,31	1,00	1,32	1,43	1,60	1,87	2,25	4,46
Pondef G105	49,23	30,46	2,07	4,32	32,17	45,92	63,48	104,22	403,29
Coefficient de NR à 3 et 5 ans	6,12	3,67	2,02	2,80	3,83	5,17	7,10	12,83	44,07

a) Une modélisation en deux étapes : contact puis réponse

Dans un premier temps, il a été tenté de recourir à une modélisation de la non-réponse en deux étapes : une modélisation de la probabilité de contact, puis pour les individus contactés une modélisation de la probabilité de répondre à l'aide des parodonnées disponibles uniquement pour les individus contactés (nombre d'appels avant contact par exemple). Cependant, une telle modélisation conduit à des coefficients multiplicatifs liés au contact et à la réponse trop importants, donc des poids avant calage maximums trop élevés, distribués de cette manière :

Tableau 29 - Distribution des poids de la modélisation en 2 étapes

Variable	Moyenne	Écart-type	Min	5 ^{ème} ctl	25 ^{ème} ctl	50 ^{ème} ctl	75 ^{ème} ctl	95 ^{ème} ctl	Max
Coefficient de contact	1,04	0,10	1,00	1,01	1,01	1,01	1,03	1,15	3,67
Coefficient de réponse	1,39	0,95	1,00	1,00	1,00	1,00	1,07	4,44	5,05
Coefficient de Non-réponse total	1,49	1,20	1,00	1,01	1,01	1,03	1,08	4,81	13,25
Poids avant calage	76,24	89,55	2,12	4,57	34,82	52,03	83,22	238,19	1770,33

b) *Des variables trop explicatives de la non-réponse*

Certaines variables utilisées dans les modèles de non-réponse ont été supprimées car elles expliquaient trop bien la non-réponse et certaines probabilités de réponse estimées par la modélisation étaient trop faibles, même après groupes de réponse homogène. En effet, les individus répondant avaient des probabilités de réponse trop supérieures à ceux qui ne répondaient pas :

Tableau 30 - Probabilités de réponse du modèle "complet" selon la réponse effective

Rep=0

N	Moyenne	Écart-type	Min	1 ^{er} ctl	5 ^{ème} ctl	10 ^{ème} ctl	25 ^{ème} ctl	50 ^{ème} ctl	75 ^{ème} ctl	90 ^{ème} ctl	95 ^{ème} ctl	99 ^{ème} ctl	Max
4788	0,29	0,25	0,00	0,01	0,02	0,03	0,07	0,23	0,44	0,70	0,78	0,88	0,98

Rep=1

N	Moyenne	Écart-type	Min	1 ^{er} ctl	5 ^{ème} ctl	10 ^{ème} ctl	25 ^{ème} ctl	50 ^{ème} ctl	75 ^{ème} ctl	90 ^{ème} ctl	95 ^{ème} ctl	99 ^{ème} ctl	Max
8919	0,84	0,23	0,00	0,03	0,21	0,41	0,89	0,93	0,95	0,97	0,97	0,98	0,99

Parmi les variables en question se trouvent celle concernant le fait d'avoir ou non reçu un sms de relance. En effet, cette relance a été faite tardivement (les 27/11 et 29/11), à tous ceux qui n'avaient pas répondu. Une grande partie des réponses ayant été faites avant cet envoi, il reste alors parmi les individus relancés beaucoup de non répondants.

De plus, les variables concernant le nombre d'appels sont trop efficaces en cas de modèle sans distinction contact/réponse. En effet les individus non contactés sont souvent des individus appelés une unique fois sur un numéro qui n'existe plus. Une autre partie des non-répondants sont ceux qui ont été appelés le plus grand nombre de fois.

Bien que possiblement existante, la corrélation entre ces variables et les variables d'intérêt de l'enquête ne semble pas forcément importante, donc apportent peu en termes de précision à l'enquête. De plus, une partie de leur apport sur la précision est captée par les autres parodonnées du modèle avec lesquelles elles sont corrélées. Ces variables ont donc été retirées des modèles de réponse.

c) *Calage sur la situation un mois donné plutôt qu'à la date d'enquête*

Dans une première version, le calage de l'enquête Génération 2010 à 7 ans sur les situations professionnelles en fin de parcours des enquêtes 2013 et 2015 était réalisé sur les variables de situation à la date d'enquête en 2013 et 2015, plutôt qu'à un mois donné (Avril 2013 et Mars 2015).

La situation professionnelle des individus dans le temps étant une variable clé des enquêtes Génération, il est important d'estimer de la même manière au moins un indicateur de situation professionnelle sur les enquêtes à 3 et 5 ans, ainsi que sur les enquêtes à 5 et 7 ans.

Cependant, la situation à la date d'enquête est une variable bâtarde, qui synthétise des situations professionnelles à des dates différentes sur une période de quatre mois, au sein de laquelle la conjoncture du marché du travail, ainsi que les situations des individus ont évolué.

De plus, il apparaît que les situations professionnelles des répondants ne sont pas constantes sur la durée de l'enquête : les individus répondant à la fin de la période d'interrogation des enquêtes 2013 et 2015 ont plus tendance à être en emploi que ceux répondant durant les premières semaines. Il

existe alors un risque de déformation artificielle non souhaitée en calant sur les marges de la situation professionnelle à la date d'enquête. En effet, augmenter par calage le nombre d'individus en emploi, par exemple, peut conduire à augmenter également, le nombre d'individus répondant en fin de période d'enquête. Cette augmentation injustifiée peut créer des biais dans les analyses si la date d'interrogation est corrélée aux variables d'intérêt.

Dans le but d'avoir une homogénéité au sein des répondants dans la situation professionnelle sur laquelle le calage est réalisé, il a été décidé d'utiliser comme marges de calage les situations du mois précédent le premier mois d'enquête des interrogations 2013 et 2015, qui sont les derniers mois où tous les individus ont renseigné leur situation professionnelle.

VII. Annexes

ANNEXE 1 - Table des illustrations

Les tableaux

Tableau 1 - Détail des enquêtes Génération et effectifs de répondants	6
Tableau 2 - Calendrier de l'enquête	14
Tableau 3 - Synthèse des tests	15
Tableau 4 - Résultats RNVP	16
Tableau 5 - Résultats des recherches des déménagés	16
Tableau 6 - Bilan de la phase A (téléphone fixe)	20
Tableau 7 - Bilan de la phase D (téléphone portable)	21
Tableau 8 - Bilan phase B et C – base France télécom	21
Tableau 9 - Répartition du nombre de lettres avis envoyées selon le mode	22
Tableau 10 - Calendrier et organisation générale avec le prestataire	23
Tableau 11 - Règles de rappel initiales	25
Tableau 12 - Règles de rappel modifiées pour la fin de l'enquête	25
Tableau 13 - Résultats d'appels des numéros mis en production	27
Tableau 14 - La codification du contrat de travail	34
Tableau 15 - Structures du type de déclaration pour les débuts de séquences entre 2015 et 2017 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations)	40
Tableau 16 - Structures du type de déclaration pour les fins de séquences entre 2015 et 2017(nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations)	40
Tableau 17 - Revalorisation du SMIC entre avril 2015 et décembre 2017	42
Tableau 18 - Normes de rémunérations des contrats de professionnalisation	42
Tableau 19 - Normes de rémunérations des contrats d'apprentissage	43
Tableau 20 - Caractéristique des tables créées	45
Tableau 21 - Variables sélectionnées dans le modèle logistique	48
Tableau 22 - Probabilités de réponse des répondants estimées par le modèle logistique	48
Tableau 23 - Groupes de Réponse homogène	49
Tableau 24 - Regroupements effectués pour le calage de PHD*Q1	50
Tableau 25 - Distribution des rapports de poids du calage de G105	51
Tableau 26 - Distribution des rapports de poids du calage de G107	51
Tableau 27 - Distribution des poids finaux par niveau de diplôme	52
Tableau 28 - Distribution des coefficients de non-réponse antérieurs de la Génération	53
Tableau 29 - Distribution des poids de la modélisation en 2 étapes	53
Tableau 30 - Probabilités de réponse du modèle "complet" selon la réponse effective	54

Les figures

Figure 1 - Calendrier des enquêtes Génération	6
Figure 2 : Schéma d'organisation du questionnaire de la 3 ^{ème} interrogation de la Génération 2010 ..	10
Figure 3 - Répartition des déménagés	17
Figure 4 - Nombre d'enquêtes réalisées selon la période	24
Figure 5 - Durée moyenne de communication	24
Figure 6 - Schéma récapitulatif expliquant le raccord des séquences entre la seconde et la troisième interrogation	31
Figure 7 - Gestion des salaires	41
Figure 8 - Distribution des poids avant calage	49
La distribution finale de la pondération de l'enquête Génération 2010 à 7 ans est la suivante : Figure 9 - Distribution des poids finaux	52

Les encadrés

Encadré 1 - Le Centre d'études et de recherches sur les qualifications.....	5
Encadré 2 - Une enquête de la statistique publique	8
Encadré 3 - SICORE Environnement Activité (millésime 2012)	37
Encadré 4 - SICORE Environnement PCS 2013	38

ANNEXE 2 - Présentation de l'outil RNVP

Le traitement Restructuration, Normalisation et Validation Postal (RNVP) à la volée consiste à normaliser les données de type adresse collectées en temps réel. Cela permet d'optimiser la qualité des données et d'obtenir, ainsi, une base de données postales propre et fiable.

Le traitement de l'adresse s'effectue en trois étapes :

- la première est la restructuration de l'adresse qui consiste à l'analyse des éléments de l'adresse afin de détecter le contenu et le positionner aux endroits voulus pour la présentation normée d'une adresse postale
- la seconde est la normalisation de l'adresse postale qui permet d'une part de retirer l'accentuation et les caractères parasites ; et d'autre part de réaliser un redressement des abréviations selon la norme Afnor NF Z 10-011 pour la France, en restant dans une longueur de champs de 38 caractères maximum.
- la troisième est la validation postale qui permet de contrôler et corriger l'adresse à partir des référentiels postaux (Hexaposte et Hexavia pour la France par exemple).

Concernant le remplissage, il se réalise en trois étapes :

- Premièrement, comme illustré ci-dessous il s'agit de saisir dans la zone prévue à cet effet, la rue de l'adresse telle qu'énoncée par l'individu. Il apparait la liste de l'ensemble des rues existant en France.

Quelle est votre adresse actuelle ?

Saisir l'adresse (CP y compris) et choisir dans la liste pour obtenir l'adresse correcte et complète.		52 avenue aristide br	X
Numéro		52 AVENUE ARISTIDE BRIAND	
Adresse*		02100 ST QUENTIN	
Complément d'adresse*		52 AVENUE ARISTIDE BRIAND	
Code postal*		10600 LA CHAPELLE ST LUC	
Commune*		52 AVENUE ARISTIDE BRIAND	
		13800 ISTRES	
		52 AVENUE ARISTIDE BRIAND	
		14800 TOUQUES	
		52 AVENUE ARISTIDE BRIAND	
		15000 AURILLAC	
		52 AVENUE ARISTIDE BRIAND	
		17400 ST JEAN D ANGELY	
		52 AVENUE ARISTIDE BRIAND	
		24100 BERGERAC	
		52 AVENUE ARISTIDE BRIAND	

- Ensuite il suffit de sélectionner l'adresse et la ville correspondantes dans la liste. Si la liste est trop longue, il suffit de l'affiner avec le code du département.

Quelle est votre adresse actuelle ?

Saisir l'adresse (CP y compris) et choisir dans la liste pour obtenir l'adresse correcte et complète.		52 avenue aristide briand 922	X
Numéro		52 AVENUE ARISTIDE BRIAND	
Adresse*		92220 BAGNEUX	
Complément d'adresse*		52 RUE ARISTIDE BRIAND	
Code postal*		92290 CHATENAY MALABRY	
Commune*		AIX EN PROVENCE	

- Une fois l'adresse et la ville sélectionnées, le formulaire se remplit automatiquement, y compris le code Insee.

Quelle est votre adresse actuelle ?

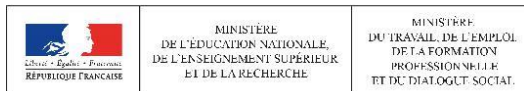
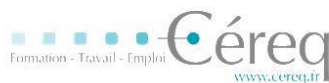
Saisir l'adresse (CP y compris) et choisir dans la liste pour obtenir l'adresse correcte et complète.	52 avenue aristide briand 922
Numéro	52
Adresse*	AVENUE ARISTIDE BRIAND
Complément d'adresse*	
Code postal*	92220
Commune*	BAGNEUX
INSEE	92007

Un menu déroulant peut également s'ouvrir si des compléments d'adresses existent (résidence, lieu-dit...) dans lequel il faudra sélectionner, le cas échéant, celui qui correspond.

Dans le cas où l'adresse n'est pas référencée, il est possible de saisir manuellement l'adresse dans chaque case.

ANNEXE 3 - Lettre de pré-fidélisation

a) Lettre envoyée par courrier postal



<Nom> <Prenom>
<Adr1>
<Adr2>
<Adr3>
<CP> <Commune>

Marseille, le 5 décembre 2016

Objet : Confirmation des coordonnées pour la troisième interrogation de l'enquête « Génération 2010 »

<Prenom> <Nom>

Vous avez participé à la première et la deuxième interrogation, respectivement en 2013 et 2015, de l'enquête « Génération 2010 » menée par le Céreq sur l'insertion professionnelle des jeunes.

Cette enquête se poursuit à l'automne 2017. Il s'agira de la troisième et dernière interrogation concernant la Génération 2010. Afin d'avoir les meilleures chances de vous contacter, nous vous proposons dès à présent **de confirmer ou de modifier l'intégralité de vos coordonnées**. Une fois vos coordonnées validées ou actualisées vous accéderez aux premiers résultats de l'enquête 2015 concernant votre génération.



Nous vous remercions chaleureusement pour votre participation depuis le début de cette étude.

En vous remerciant par avance,

Alberto LOPEZ
Directeur du Céreq

Les coordonnées dont on dispose pour vous contacter sont les suivantes :

Téléphone : <tel_fin>

E-mail : <email>

Pour les actualiser ou les confirmer vous pouvez accéder à votre espace en vous connectant à l'adresse suivante :

<lien>

Vos identifiants sont : Identifiant : <IDENT>
Mot de passe : <MDP>

Vous pouvez également nous contacter par email à l'adresse generation2010@cereq.fr en mentionnant votre identifiant : <IDENT>

Retrouvez-nous sur  [facebook.com/cereqG10](https://www.facebook.com/cereqG10)

Les destinataires des données sont le Céreq. Conformément à la loi « informatique et libertés » du 6 janvier 1978 modifiée en 2004, vous bénéficiez d'un droit d'accès et de rectification aux informations qui vous concernent, que vous pouvez exercer en vous adressant par courrier à Céreq, 10, place de la Joliette - BP 21321 - Marseille Cedex 02, ou par mail à l'adresse cil-generation@cereq.fr. Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant.

b) *Lettre envoyée par mail*



Objet : Enquête « Génération 2010 » - Actualisation de vos coordonnées

Bonjour <Prenom> <Nom>,

Vous avez participé à la première et la deuxième interrogation, en 2013 et en 2015, de l'enquête « Génération 2010 » menée par le Céreq sur l'insertion professionnelle des jeunes.

Cette enquête se poursuit à l'automne 2017. Il s'agira de la troisième et dernière interrogation concernant la Génération 2010. Afin d'avoir les meilleures chances de vous contacter, nous vous proposons dès à présent **de confirmer ou de modifier l'intégralité de vos coordonnées**. Une fois vos coordonnées validées ou actualisées vous accéderez aux premiers résultats de l'enquête 2015 concernant votre génération.

Nous vous remercions chaleureusement pour votre participation depuis le début de cette étude.

En vous remerciant par avance,

Le Céreq

Les coordonnées dont on dispose pour vous contacter sont les suivantes :

Téléphone : <tel_fin>

E-mail : <email>

Pour les actualiser ou les confirmer vous pouvez accéder à votre espace en vous connectant à l'adresse suivant :

<https://generation20107ans.cereq.fr>

Vos identifiants sont : Identifiant : <IDENT>
Mot de passe : <MDP>

Vous pouvez également nous contacter par email à l'adresse generation2010@cereq.fr en mentionnant votre identifiant : <IDENT>

Retrouvez-nous sur  facebook.com/cereqG10

Les destinataires des données sont le Céreq. Conformément à la loi « informatique et libertés » du 6 janvier 1978 modifiée en 2004, vous bénéficiez d'un droit d'accès et de rectification aux informations qui vous concernent, que vous pouvez exercer en vous adressant par courrier à Céreq, 10, place de la Joliette - BP 21321 - Marseille Cedex 02, ou par mail à l'adresse cil-generation@cereq.fr. Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant.

ANNEXE 4 - CAWI de pré-fidélisation

a) Écran d'accueil

The screenshot shows the login interface for the 'Enquête 2017 auprès de la Génération 2010'. At the top, there are logos for Céreq, the 'Génération 2010' survey, the INSEE (Statistique Publique), and the French Republic. Below the logos, the text reads: 'Bonjour, Bienvenue sur la page du Céreq dédiée à la Génération 2010, Pour vous identifier, merci de saisir l'identifiant et le mot de passe que nous vous avons fournis dans le mail de contact ou par courrier postal.' There are two input fields: 'Identifiant' and 'Mot de passe'. A 'Se connecter' button is located at the bottom.

b) Affichage des coordonnées

The screenshot shows the form for entering contact information. It features the same header as the login page. The text says: 'Bonjour, Vous avez participé à la première et la deuxième interrogation, en 2013 et en 2015, de l'enquête « Génération 2010 » menée par le Céreq sur l'insertion professionnelle des jeunes. Nous allons vous recontacter à l'automne 2017 pour vous interroger sur la suite de votre parcours professionnel. Afin d'avoir les meilleures chances de vous contacter, nous vous proposons dès maintenant de confirmer ou de modifier les coordonnées suivantes.' The form includes fields for: 'Nom Usuel', 'Prénom', 'Adresse 1', 'Adresse 2', 'Adresse 3', 'Code postal', 'Commune', 'Numéro de téléphone principal', 'Numéro de téléphone secondaire', and 'Adresse mail principale'. Below these fields is a section for validation: '* Validation des coordonnées :'. It contains three radio buttons: 'Je souhaite valider ces coordonnées', 'Je souhaite modifier ces coordonnées (Coordonnées en France)', and 'Je souhaite modifier ces coordonnées (Coordonnées à l'étranger)'. A red asterisk indicates that the response is mandatory. At the bottom, there is a 'Page suivante' button and a small disclaimer at the very bottom of the page.

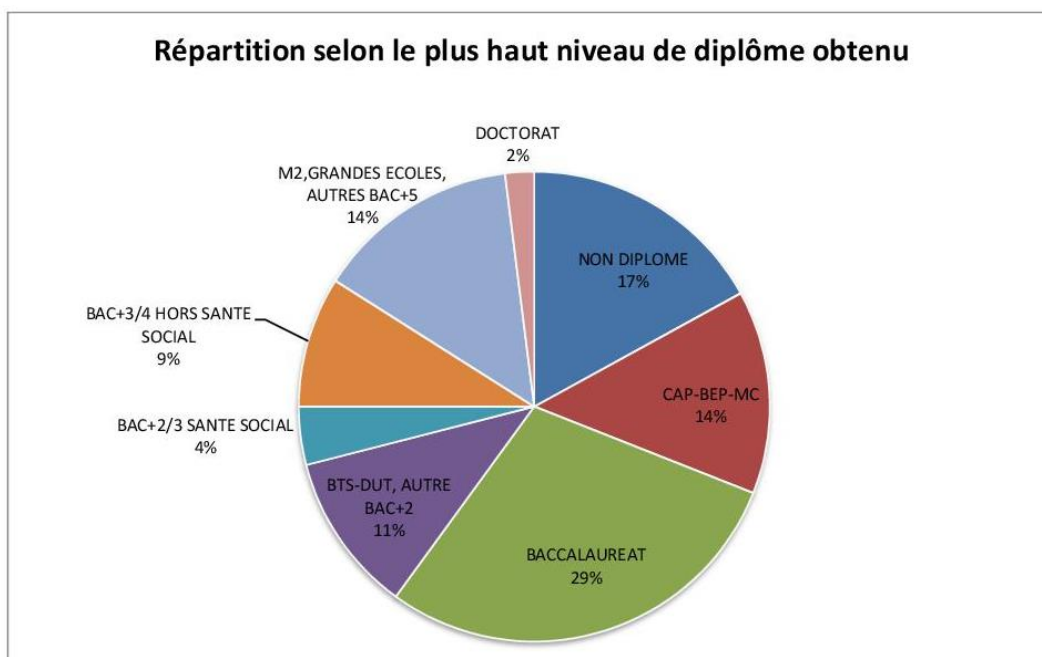
c) Premiers résultats présentés après validation

L'enquête 2015 auprès de la Génération 2010

La Génération 2010 est composée de l'ensemble des jeunes ayant terminé leur scolarité (pour la première fois) en 2009/2010. L'enquête 2015 conduite auprès de 14 000 jeunes de cette Génération 2010 permet d'étudier, de manière représentative, les cinq premières années de vie active des jeunes de cette Génération.

Le niveau de formation de la Génération 2010

Sur les 708 000 jeunes de la Génération 2010 (sortis du système éducatif en 2009/2010), près de 60% possédaient un niveau de diplôme égal ou inférieur au baccalauréat et 40 % possédaient un diplôme de l'enseignement supérieur.



Champ : Ensemble de la Génération 2010

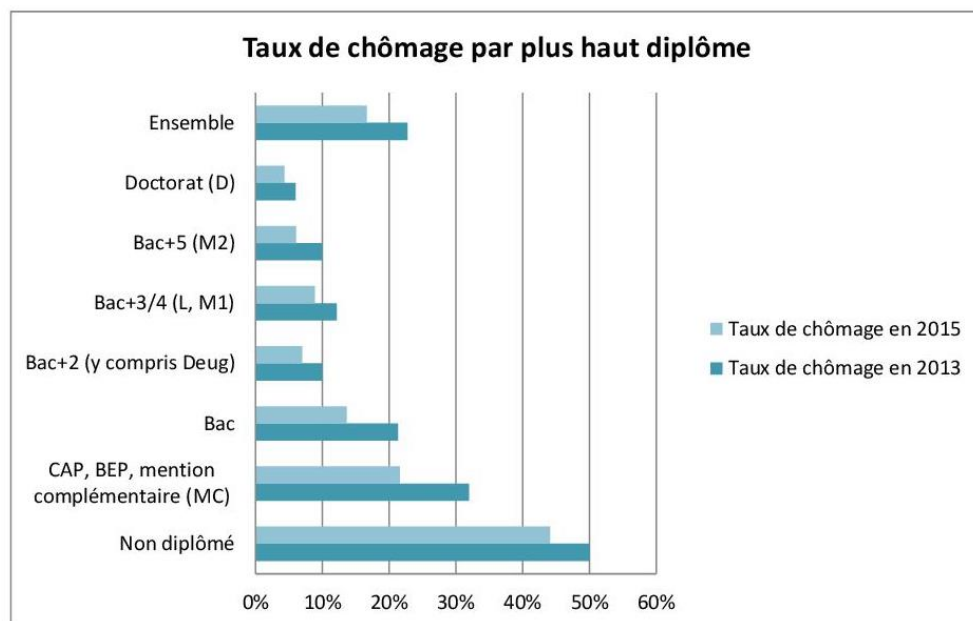
Source : Céreq, Enquête 2015 auprès de la "Génération 2010"

L'entrée dans la vie active et les parcours d'insertion

L'enquête Génération permet d'étudier les parcours d'insertion professionnelle grâce à un calendrier d'activités qui retrace mois par mois la situation de chaque jeune, entre emploi, recherche d'emploi, reprise d'études, formation et inactivité.

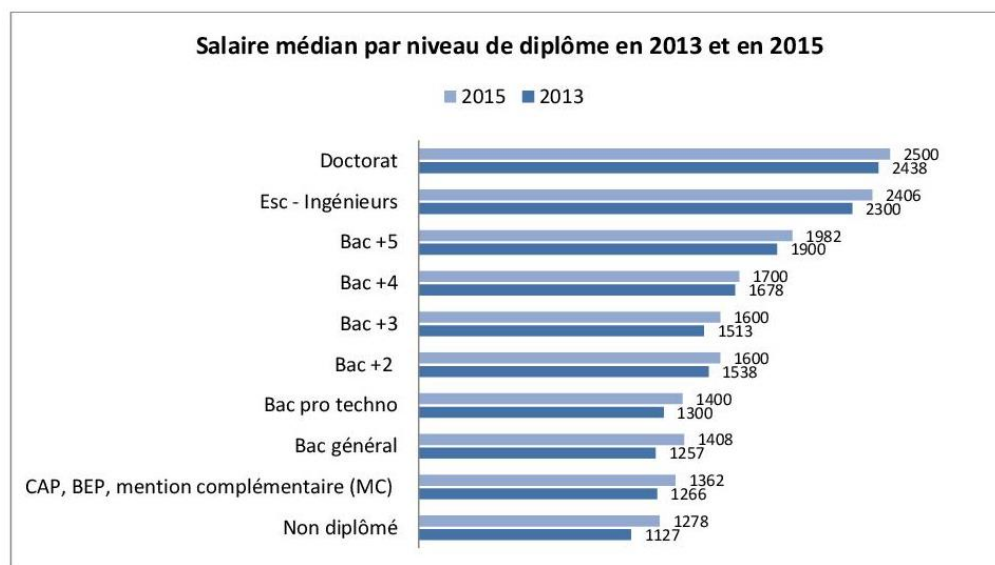
Ainsi près de trois jeunes sur quatre de la Génération 2010 ont accédé à leur premier emploi moins de six mois après leur sortie du système éducatif. Entre 2010 et 2015, les jeunes ont passé en moyenne 73% de leur temps en situation d'emploi, 17% de leur temps en recherche d'emploi et 10% de leur temps en formation, reprise d'études ou autres situations d'inactivité.

La situation professionnelle en 2015



Champ : Ensemble de la Génération 2010

Source : Céreq, Enquête 2015 auprès de la "Génération 2010"



Champ : Jeunes en emploi à la date d'enquête (2013 ou 2015)

Source : Céreq, Enquête 2015 auprès de la "Génération 2010"

En 2015, les trois quart des jeunes occupent un emploi. Par rapport à 2013, le taux de chômage a reculé et les salaires ont progressé, et ce quel que soit le niveau de diplôme. Cela s'explique essentiellement par la poursuite du processus d'insertion dans l'emploi des jeunes de la Génération 2010.

Pour autant, la situation des non diplômés demeure préoccupante puisque leur taux de chômage cinq ans après la sortie avoisine les 40%.

ANNEXE 5 - Lettre avis de contact avec les jeunes

a) Lettre envoyée par courrier postal



<Nom> <Prenom>
<Adr1>
<Adr2>
<Adr3>
<CP> <Commune>



Marseille, le 31 août 2017

Objet : Troisième interrogation de l'enquête « Génération 2010 » sur l'insertion professionnelle des sortants du système éducatif en 2010

Bonjour <Prénom> <Nom>,

Nous vous remercions pour vos deux participations à l'enquête « Génération 2010 » en 2013 et 2015.

Le Centre d'études et de recherches sur les qualifications (Céreq) poursuit son enquête statistique sur l'avenir professionnel des jeunes sortis du système éducatif en 2010, à l'automne 2017. L'objectif est de mieux comprendre les parcours professionnels de votre Génération afin d'alimenter le débat public sur le lien formation-emploi. Les principaux enseignements seront diffusés auprès des acteurs institutionnels, des médias, des jeunes et leur famille.

Dans ce but, nous allons vous recontacter par téléphone dans les prochaines semaines pour un nouvel entretien destiné à décrire votre parcours professionnel depuis l'interrogation précédente. Cet entretien durera une quinzaine de minutes environ.

Vous trouverez ci-joint un calendrier qui vous aidera à reconstituer votre parcours professionnel depuis le printemps 2015. Nous vous invitons à le remplir et à le conserver afin de l'utiliser lors de notre entretien téléphonique.

Conformément à la loi, vos réponses resteront confidentielles et serviront uniquement à la réalisation de statistiques qui alimenteront des études sur l'insertion professionnelle.

En vous remerciant par avance de votre participation,

Le Directeur du Centre d'études et de recherches sur les qualifications

Albert LOPEZ

Le numéro de téléphone dont nous disposons pour vous joindre est le suivant : **<TEL1 >**

Pour actualiser ou confirmer votre numéro de téléphone, et pour faciliter l'entretien téléphonique, vous pouvez accéder à votre espace en flashant le QR-code ci-après ou en vous connectant au lien suivant :

<https://generation2010ans.cereq.fr/>

Vos identifiants sont : Identifiant : **<IDENT>**
Mot de passe : **<MDP>**



Vous pouvez également nous contacter au numéro vert **0 800 574 020** (appel gratuit, en semaine entre 12h et 20h45, le samedi entre 10h et 16h) en mentionnant votre identifiant : **<IDENT>**

Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue d'intérêt général et de qualité statistique sans avoir de caractère obligatoire.

Visa n°2017X717AU du Ministre de l'économie et des finances, valable pour l'année 2017.

En application de la loi n°51-711 du 7 juin 1951, les réponses à ce questionnaire sont protégées par le secret statistique et destinées au Céreq.

La loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Ce droit peut être exercé auprès du Céreq, 10 place de la Joliette, 13 567 Marseille Cedex 02, ou bien par un mail adressé au correspondant informatique et libertés du Céreq à l'adresse suivante : gil-generation@cereq.fr en mentionnant dans l'objet « Génération 2010 » suivi de votre identifiant apparaissant sur la première page de ce courrier.

b) *Lettre envoyée par courrier électronique*



De : Le Céreq (Centre d'Études et de Recherches sur les Qualifications)

Objet : Enquête nationale sur l'insertion professionnelle des jeunes « Génération 2010 »

Bonjour <Prénom> <Nom>,

Nous vous remercions pour vos deux participations à l'enquête « Génération 2010 » en 2013 et 2015.

Le Centre d'études et de recherches sur les qualifications (Céreq) poursuit son enquête statistique sur l'avenir professionnel des jeunes sortis du système éducatif en 2010, à l'automne 2017. L'objectif est de mieux comprendre les parcours professionnels de votre Génération afin d'alimenter le débat public sur le lien formation-emploi. Les principaux enseignements seront diffusés auprès des acteurs institutionnels, des médias, des jeunes et leur famille.

Dans ce but, nous allons vous recontacter par téléphone dans les prochaines semaines pour un nouvel entretien destiné à décrire votre parcours professionnel depuis l'interrogation précédente. Cet entretien durera une quinzaine de minutes environ.

Vous trouverez ci-joint un calendrier qui vous aidera à reconstituer votre parcours professionnel depuis le printemps 2015. Nous vous invitons à le remplir et à le conserver afin de l'utiliser lors de notre entretien téléphonique.

Conformément à la loi, vos réponses resteront confidentielles et serviront uniquement à la réalisation de statistiques qui alimenteront des études sur l'insertion professionnelle.

En vous remerciant par avance de votre participation,

Le Céreq

*Le numéro de téléphone dont nous disposons pour vous joindre est le suivant : <TEL1 >
Pour actualiser ou confirmer votre numéro de téléphone, et pour faciliter l'entretien téléphonique,
vous pouvez accéder à votre espace
en cliquant [ici](#)*

*Vos identifiants sont : Identifiant : <IDENT>
Mot de passe : <MDP>*

*Vous pouvez également nous contacter au numéro vert 0 800 574 020 (appel gratuit, en semaine entre 12h
et 20h45, le samedi entre 10h et 16h) en mentionnant votre identifiant : <IDENT>*

Des informations complémentaires sur notre enquête sont disponibles [ici](#).

Retrouvez-nous aussi sur notre page Facebook spéciale Génération 2010



Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue d'intérêt général et de qualité statistique sans avoir de caractère obligatoire.
Visa n°2017X717AU du Ministre de l'économie et des finances, valable pour l'année 2017.
En application de la loi n°51-711 du 7 juin 1951, les réponses à ce questionnaire sont protégées par le secret statistique et destinées au Céreq.
La loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Ce droit peut être exercé auprès du Céreq, 10 place de la Joliette, 13 567 Marseille Cedex 02, ou bien par un mail adressé au correspondant informatique et libertés du Céreq à l'adresse suivante : cil-generation@cereq.fr en mentionnant dans l'objet « Génération 2010 » suivi de votre identifiant.

Si vous ne souhaitez pas recevoir d'emails de notre part, [cliquez ici](#).

c) Pièce jointe aux courriers

DOCUMENT A CONSERVER – NE PAS ENVOYER

Votre parcours depuis 2015

► Nous vous incitons à remplir ce calendrier afin de faciliter l'entretien téléphonique. L'enquêteur disposera sur écran d'un calendrier vierge qu'il remplira avec vous oralement. Pour vous aider souvenez-vous de votre situation lors des événements marquants suivants :

		2015							2016							2017																								
		Avril	Mai	Juin	Juill.	Août	Sept	Oct	Nov	Déc.	Janv.	Fév.	Mars	Avril	Mai	Juin	Juill.	Août	Sept	Oct	Nov	Déc.	Janv.	Fév.	Mars	Avril	Mai	Juin	Juill.	Août	Sept	Oct	Nov	Déc.						
Périodes d'emploi	En Intérim																																							
	En Entreprise (ou administration)																																							
Périodes de non-emploi	En recherche d'emploi																																							
	En reprise d'études - A temps plein dans un établissement scolaire																																							
	En formation - Autre qu'une reprise d'études																																							
	Autres situations																																							

Coupe du monde 2015 de rugby en Angleterre
 Elections régionales en France
 Championnat d'Europe de football 2016
 Donald Trump est élu 45e président des États-Unis
 Elections présidentielles en France
 Elections législatives en France

► **Exemple de parcours :**

Dans cet exemple, lors de son interrogation en avril 2015, un jeune a renseigné travailler dans l'entreprise « Entreprise 1 » à Paris 10^{ème}. Il y est resté jusqu'en décembre 2015. Il a ensuite recherché un emploi de janvier 2016 à mars 2016. Après 3 mois passés en formation, il travaille depuis juillet 2016 dans l'entreprise « Entreprise 2 » à Paris 2^{ème}.

		2015							2016							2017																								
		Avril	Mai	Juin	Juill.	Août	Sept	Oct	Nov	Déc.	Janv.	Fév.	Mars	Avril	Mai	Juin	Juill.	Août	Sept	Oct	Nov	Déc.	Janv.	Fév.	Mars	Avril	Mai	Juin	Juill.	Août	Sept	Oct	Nov	Déc.						
Périodes d'emploi	En Intérim																																							
	En Entreprise (ou administration)																																							
Périodes de non-emploi	En recherche d'emploi																																							
	En reprise d'études - A temps plein dans un établissement scolaire																																							
	En formation - Autre qu'une reprise d'études																																							
	Autres situations																																							

← Salaire à relever sur sa dernière fiche de paie

← Salaire à relever sur sa première et dernière fiche de paie

ANNEXE 6 - Nomenclature NAF rev2 en 88 divisions (A88)

Section	Libellé des sections	Code Division	Intitulé
A	AGRICULTURE, SYLVICULTURE ET PÊCHE	01	Culture et production animale, chasse et services annexes
		02	Sylviculture et exploitation forestière
		03	Pêche et aquaculture
B	INDUSTRIES EXTRACTIVES	05	Extraction de houille et de lignite
		06	Extraction d'hydrocarbures
		07	Extraction de minerais métalliques
		08	Autres industries extractives
		09	Services de soutien aux industries extractives
C	INDUSTRIE MANUFACTURIÈRE	10	Industries alimentaires
		11	Fabrication de boissons
		12	Fabrication de produits à base de tabac
		13	Fabrication de textiles
		14	Industrie de l'habillement
		15	Industrie du cuir et de la chaussure
		16	Travail du bois et fabrication d'articles en bois et en liège, à l'exception des meubles ; fabrication d'articles en vannerie et sparterie
		17	Industrie du papier et du carton
		18	Imprimerie et reproduction d'enregistrements
		19	Cokéfaction et raffinage
		20	Industrie chimique
		21	Industrie pharmaceutique
		22	Fabrication de produits en caoutchouc et en plastique
		23	Fabrication d'autres produits minéraux non métalliques
		24	Métallurgie
		25	Fabrication de produits métalliques, à l'exception des machines et des équipements
		26	Fabrication de produits informatiques, électroniques et optiques
		27	Fabrication d'équipements électriques
		28	Fabrication de machines et équipements n.c.a.
		29	Industrie automobile
30	Fabrication d'autres matériels de transport		
31	Fabrication de meubles		
32	Autres industries manufacturières		
33	Réparation et installation de machines et d'équipements		
D	PRODUCTION ET DISTRIBUTION D'ÉLECTRICITÉ, DE GAZ, DE VAPEUR ET D'AIR CONDITIONNÉ	35	Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
E	PRODUCTION ET DISTRIBUTION D'EAU ; ASSAINISSEMENT, GESTION DES DÉCHETS ET DÉPOLLUTION	36	Captage, traitement et distribution d'eau
		37	Collecte et traitement des eaux usées
		38	Collecte, traitement et élimination des déchets ; récupération
		39	Dépollution et autres services de gestion des déchets
F	CONSTRUCTION	41	Construction de bâtiments
		42	Génie civil
		43	Travaux de construction spécialisés
G	COMMERCE ; RÉPARATION D'AUTOMOBILES ET DE MOTOCYCLES	45	Commerce et réparation d'automobiles et de motocycles
		46	Commerce de gros, à l'exception des automobiles et des motocycles
		47	Commerce de détail, à l'exception des automobiles et des motocycles
H	TRANSPORTS ET ENTREPOSAGE	49	Transports terrestres et transport par conduites
		50	Transports par eau
		51	Transports aériens
		52	Entreposage et services auxiliaires des transports
		53	Activités de poste et de courrier
I	HÉBERGEMENT ET RESTAURATION	55	Hébergement
		56	Restauration

J	INFORMATION ET COMMUNICATION	58	Édition
		59	Production de films cinématographiques, de vidéo et de programmes de télévision ; enregistrement sonore et édition musicale
		60	Programmation et diffusion
		61	Télécommunications
		62	Programmation, conseil et autres activités informatiques
		63	Services d'information
K	ACTIVITÉS FINANCIÈRES ET D'ASSURANCE	64	Activités des services financiers, hors assurance et caisses de retraite
		65	Assurance
		66	Activités auxiliaires de services financiers et d'assurance
L	ACTIVITÉS IMMOBILIÈRES	68	Activités immobilières
M	ACTIVITÉS SPÉCIALISÉES, SCIENTIFIQUES ET TECHNIQUES	69	Activités juridiques et comptables
		70	Activités des sièges sociaux ; conseil de gestion
		71	Activités d'architecture et d'ingénierie ; activités de contrôle et analyses techniques
		72	Recherche-développement scientifique
		73	Publicité et études de marché
		74	Autres activités spécialisées, scientifiques et techniques
N	ACTIVITÉS DE SERVICES ADMINISTRATIFS ET DE SOUTIEN	75	Activités vétérinaires
		77	Activités de location et location-bail
		78	Activités liées à l'emploi
		79	Activités des agences de voyage, voyagistes, services de réservation et activités connexes
		80	Enquêtes et sécurité
		81	Services relatifs aux bâtiments et aménagement paysager
O	ADMINISTRATION PUBLIQUE	82	Activités administratives et autres activités de soutien aux entreprises
		84	Administration publique et défense ; sécurité sociale obligatoire
P	ENSEIGNEMENT	85	Enseignement
Q	SANTÉ HUMAINE ET ACTION SOCIALE	86	Activités pour la santé humaine
		87	Hébergement médico-social et social
		88	Action sociale sans hébergement
R	ARTS, SPECTACLES ET ACTIVITÉS RÉCRÉATIVES	90	Activités créatives, artistiques et de spectacle
		91	Bibliothèques, archives, musées et autres activités culturelles
		92	Organisation de jeux de hasard et d'argent
		93	Activités sportives, récréatives et de loisirs
S	AUTRES ACTIVITÉS DE SERVICES	94	Activités des organisations associatives
		95	Réparation d'ordinateurs et de biens personnels et domestiques
		96	Autres services personnels
T	ACTIVITÉS DES MÉNAGES EN TANT QU'EMPLOYEURS ; ACTIVITÉS INDIFFÉRENCIÉES DES MÉNAGES EN TANT QUE PRODUCTEURS DE BIENS ET SERVICES POUR USAGE PROPRE	97	Activités des ménages en tant qu'employeurs de personnel domestique
		98	Activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre
U	ACTIVITÉS EXTRA-TERRITORIALES	99	Activités des organisations et organismes extraterritoriaux

ANNEXE 7 - Définition des variables utilisées dans Sicore PCS

STATUT : statut dans l'emploi (clivage salarié / indépendant).

Les modalités de STATUT pour les règles Sicore sont les suivantes :

- 1. indépendant
- 2. salariés de sa propre entreprise, gérant mandataire, PDG
- 3. salarié
- * (ou blanc) manquant

PUB : statut de l'établissement employeur (distinction public / privé). Cette variable ne concerne que les salariés.

Les modalités de PUB pour les règles Sicore sont les suivantes :

- 1. État
- 2. collectivités territoriales, HLM, hôpitaux
- 3. sécurité sociale
- 4. entreprises publiques nationalisées
- 5. privé
- * (ou blanc) manquant

SP : emploi précaire (distinction apprenti / autre). Cette variable ne concerne que les salariés.

Les modalités de SP pour les règles Sicore sont les suivantes :

- 1. apprenti
- * (ou blanc) manquant

CPF : classification professionnelle ou Qualification

Cette variable ne concerne a priori que les salariés (cf. PCS 2003 page 43). Les modalités de CPF pour les règles Sicore sont les suivantes :

!! Modifiées pour la PCS 2003 !!

- 1. manœuvre ou ouvrier spécialisé
- 2. ouvrier qualifié ou ouvrier hautement qualifié ou technicien d'atelier
- 3. agent de maîtrise
- 4. directeur général ou adjoint direct au directeur
- 5. technicien, dessinateur, VRP
- 6. instituteur, assistante sociale, infirmière et autres personnels de catégorie B de la fonction publique
- 7. ingénieur ou cadre
- 8. professeur et personnel de catégorie A de la fonction publique
- 9. employés de bureau, de commerce, agents de service, aides-soignantes, gardiennes d'enfants, personnels de catégorie C ou D de la fonction publique
- 0. autres
- * (ou blanc) manquant

FN : fonction principale

Les modalités de FN pour les règles Sicore sont les suivantes :

!! Modifiées pour la PCS 2003 !!

- 1. production, fabrication, chantiers
- 2. installation, réparation, maintenance
- 3. nettoyage, gardiennage, entretien ménager
- 4. manutention, magasinage, logistique
- 5. secrétariat, saisie, accueil
- 6. gestion, comptabilité
- 7. commerce, vente, technico-commercial
- 8. études, recherche et développement, méthodes
- 0. autres
- * (ou blanc) manquant

NBS : nombre de salariés employés

Cette variable ne concerne a priori que les chefs d'entreprise ou personnes installées à leur compte.

Les modalités de NBS pour les règles Sicore sont les suivantes :

- 1. aucun salarié
- 2. 1 ou 2 salarié(s)
- 3. 3 à 9 salariés
- 4. 10 salariés ou plus
- * (ou blanc) manquant

NAF2 et **NAF** : activité principale de l'établissement

Les modalités de NAF2 et NAF pour les règles Sicore sont les suivantes :

- Toutes les modalités possibles de la nomenclature d'activité
- ** ou ***** (ou blanc) manquant

La variable NAF2 comprend les deux premières positions du code NAF (niveau "division" de la nomenclature). La variable NAF comprend les codes NAF à quatre positions.

Si l'on dispose de la NAF à quatre positions, créer aussi la variable annexe NAF2. Certains codages ne nécessitant que la NAF2.

S : sexe

Les modalités de S pour les règles Sicore sont les suivantes :

- 1. masculin
- 2. féminin
- * (ou blanc) manquant

T : taille de l'entreprise

Les modalités de T pour les règles Sicore sont les suivantes :

- 0 de 0 à 9 salariés
- P de 10 à 49 salariés

- M de 50 à 499 salariés
- G 500 salariés et plus
- * (ou blanc) manquant

OPA : orientation des productions agricoles

Les modalités de OPA pour les règles Sicore sont les suivantes :

- 1. polyculture (culture des terres labourables)
- 2. maraîchage ou horticulture
- 3. vigne ou arbres fruitiers
- 4. élevage d'herbivores (bovins, ovins...)
- 5. élevage de granivores (volailles, porcs, ...)
- 6. polyculture-élevage
- 7. élevage d'herbivores et de granivores
- 8. autre
- * (ou blanc) manquant

DEP : département

Les modalités de DEP pour les règles Sicore sont les suivantes :

- Tous les numéros de départements métropolitains (pour la Corse, on peut distinguer 2A et 2B ou ne pas distinguer en mettant 20)
- 97 pour les DOM (pas de distinction)
- ** (ou blanc) manquant

Le département peut être utile pour la codification de certaines professions d'agriculteurs.

ANNEXE 8 - Liste des questions ouvertes

Nom de la variable "ouvert"	Libellé de la variable	Module dans le questionnaire
O_EP21	Emploi en CDD : Sur quel contrat de travail avez-vous été embauché ?	Séquences en entreprise
O_EP27	Emploi en CDD : Quel était votre contrat de travail juste avant de quitter cette entreprise ?	
O_EP83	Vous avez démissionné, principalement parce que vous souhaitez	
O_NEP5C	Pas de réponse à la dernière offre d'emploi proposée : Pourquoi ?	Séquences de non-emploi
O_NEP5CT	Pas de réponse à la dernière offre d'emploi proposée : Pourquoi ?	
O_NEP5D	Personne non disponible : Pour quelle raison ?	
O_NEP5DT	Personne non disponible : Pour quelle raison ?	
O_OL2	Personne n'étant pas prête à occuper un emploi à partir du mois prochain : Pour quelle raison ?	
O_OL6	Personne n'étant pas prête à suivre une formation à partir du mois prochain : Pour quelle raison ?	
O_DIS31	Victime de discrimination : raison ?	Discrimination
O_DIS41	Victime de discrimination : auteur ?	
O_DIS51	Victime de discrimination : forme ?	
O_EA0B	Emploi pas en rapport avec formation initiale : Pour quelle(s) raison(s) ?	Opinions sur l'emploi
O_OP6B	Pour quelle raison êtes-vous inquiet pour votre avenir professionnel ?	Les perspectives professionnelles
O_QXX3	Conjoint n'ayant pas pris le congé paternité pour le premier enfant : Pour quelle raison ?	Caractéristiques individuelles - Enfants
O_QXX3B	Conjoint n'ayant pas pris le congé paternité pour le second enfant : Pour quelle raison ?	
O_QYY3	Personne n'ayant pas pris le congé paternité pour le premier enfant : Pour quelle raison ?	
O_QYY3B	Personne n'ayant pas pris le congé paternité pour le second enfant : Pour quelle raison ?	
O_HI23B	Personne ayant le sentiment que son travail prend trop de place sur son temps « hors travail » : Est-ce quelque chose :	

ANNEXE 9 - Variables en entrée du modèle de régression logistique pour la pondération

ZUS	Appartenance à une ZUS (Zone Urbaine Sensible)
SITDE_13	Situation de l'enquêté à la date de l'enquête (en 2013)
SITDE_15	Situation de l'enquêté à la date de l'enquête (en 2015)
HABDE_15	Situation d'habitat à la date d'enquête (en 2015)
AGE10	Age de l'individu en 2010 (à la sortie du système éducatif)
CA0_13	Habite toujours dans la commune de fin d'études en 2013
CA0_15	Habite toujours dans la commune à la date de l'interrogation de 2013 en 2015
CA32	Disposé pour un entretien qualitatif
CA31	Disposé à répondre à un questionnaire internet
CA20_13	Enfants (en 2013)
CA20_15	Enfants (en 2015)
CA7	Nationalité du père
CA8	Nationalité de la mère
OP6_13	Optimisme avenir professionnel (Question de Géné2010 à 3 ans)
OP6_15	Optimisme avenir professionnel (Question de Géné2010 à 5 ans)
OP2_13	Opinion sur le parcours et l'avenir professionnel (interrogation de 2013)
OP2_15	Opinion sur le parcours et l'avenir professionnel (interrogation de 2015)
PHD	Plus haut diplôme obtenu
Q1	Sexe
LA	Type de lettre-avis reçu
mail	Lettre-avis mail
papier	Lettre-avis papier
i_relance*relancemai	Avoir reçu une relance par mail pour ceux n'ayant pas répondu avant la date de relance
connexion_cawifid	S'être connecté au CAWI de fidélisation
appelgratuit	Avoir appelé le numéro gratuit
refus_all*rappel_all	Etre rappelé après refus parmi ceux ayant refusé

Céreq

*Établissement public national sous la tutelle
du ministère chargé de l'Éducation
et du ministère chargé de l'Emploi.*

DEPUIS 1971


Mieux connaître les liens formation - emploi - travail.
Un collectif scientifique au service de l'action publique.

• 12 centres associés sur le territoire et de nombreuses coopérations internationales

 **+ d'infos**
et tous les travaux

À explorer
www.cereq.fr



 **+ de 600 publications**
Accessibles librement