

# CÉREQ ÉTUDES

26  
2019

## Enquête Génération 2010 2<sup>ème</sup> interrogation printemps 2015

Bilan méthodologique

*Coord.* Manon OLARIA

*Auteurs :* Christophe BARRET, Mady CISSE,  
Christophe DZIKOWSKI, Émilie GAUBERT,  
Zora MAZARI, Manon OLARIA, Florence RYK

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

# Enquête Génération 2010

## Bilan méthodologique

### 2<sup>ème</sup> interrogation – printemps 2015

**Coordinatrice :**  
Manon OLARIA

**Auteurs :**  
Christophe BARRET  
Mady CISSE  
Christophe DZIKOWSKI  
Émilie GAUBERT  
Zora MAZARI  
Manon OLARIA  
Florence RYK

*Membres ou ex-membres de l'Équipe ingénierie et gestion d'enquête (Eige) du Département entrées et évolutions dans la vie active (Deeva) du Centre d'études et de recherches sur les qualifications (Céreq).*

## Synthèse :

Ce document présente un panorama complet des travaux de production, de collecte et de mise à disposition des données de l'enquête 2015 auprès de la Génération 2010. Après une brève présentation du dispositif des enquêtes Génération et plus spécifiquement de l'enquête 2015, 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'avril à juillet 2015. Enfin, l'ensemble des traitements post collecte sont décrits (création des bases de données, apurement, codification, traitement de la non réponse, pondération...).

## Sommaire

<b>I. PRESENTATION DE L'ENQUETE 2015 AUPRES DE LA GENERATION 2010</b> .....	<b>5</b>
1. LE DISPOSITIF DES ENQUETES GENERATION .....	5
2. LE CHAMP DE L'ENQUETE .....	7
3. LES EXTENSIONS .....	8
4. DESCRIPTIF ET OBJECTIFS DE LA GENERATION 2010 A 5 ANS .....	9
<i>La Génération 2010 est une Génération « pleine ». Ainsi, elle a été interrogée une première fois en 2013.</i>	
<i>La seconde interrogation a eu lieu en 2015 et sera suivie d'une troisième en 2017.</i> .....	9
5. LE QUESTIONNAIRE .....	9
a) <i>Descriptif du questionnaire</i> .....	9
b) <i>Nouveau module sur l'acquisition de diplôme ou titres en post-initial</i> .....	11
c) <i>Intégration de questions issues de l'enquête Génération 1992 à 5 ans</i> .....	12
d) <i>Autres évolutions</i> .....	12
e) <i>Le calendrier professionnel</i> .....	13
6. LE CALENDRIER DE L'ENQUETE .....	14
<b>II. PREPARATION DE LA COLLECTE</b> .....	<b>15</b>
1. CONSTITUTION DE L'ECHANTILLON .....	15
a) <i>Retour sur la probabilité d'inclusion du tirage de la première interrogation</i> .....	15
b) <i>Calcul de la nouvelle probabilité d'inclusion</i> .....	15
c) <i>Tirage équilibré</i> .....	16
d) <i>Cibles et caractéristiques de l'échantillon</i> .....	16
2. DEVELOPPEMENT DU CATI (ENQUETE TELEPHONIQUE) .....	19
a) <i>Logiciel utilisé</i> .....	19
b) <i>Technique de développement du calendrier</i> .....	19
c) <i>Test du CATI et du calendrier associé</i> .....	19
3. MISE A JOUR DES COORDONNEES POSTALES .....	20
4. FIDELISATION .....	21
a) <i>Site internet du Céreq</i> .....	21
b) <i>Réseaux sociaux</i> .....	21
c) <i>Pré-fidélisation</i> .....	22
d) <i>Mailings de remerciement</i> .....	22
5. ACTUALISATION DES COORDONNEES TELEPHONIQUES .....	23
6. LETTRE-AVIS .....	26
<b>III. LA COLLECTE PAR TELEPHONE</b> .....	<b>28</b>
1. CALENDRIER ET ORGANISATION GENERALE DE LA COLLECTE .....	28
2. LE SUIVI DE LA COLLECTE EN CHIFFRES .....	28
3. LES REGLES DE RAPPEL .....	30
4. DES MESSAGES LAISSES SUR REPONDEUR .....	31
5. SUIVI TECHNIQUE ET PERSONNE « QUALITE » .....	32
6. LES RESIDENTS A L'ETRANGER A LA DATE DE L'ENQUETE .....	32
<b>IV. BILAN DE COLLECTE</b> .....	<b>33</b>
<b>V. LES TRAITEMENTS EN AVAL</b> .....	<b>34</b>
1. LES TABLES BRUTES LIVREES PAR LE PRESTATAIRE .....	34
2. LA CREATION DES TABLES.....	34
a) <i>Traitements et contrôles des fichiers de base transmis par le prestataire</i> .....	34
b) <i>Création de la base individus</i> .....	35

c) <i>Calendriers professionnel et habitat</i> .....	35
d) <i>Création des séquences d’emplois et de non-emploi</i> .....	36
e) <i>Les recopies</i> .....	38
3. LA CODIFICATION DES VARIABLES .....	38
a) <i>Contrats de travail</i> .....	38
b) <i>Variables à créer en amont de la codification du secteur d’activité</i> .....	39
c) <i>Secteur d’activité de l’établissement employeur</i> .....	41
d) <i>Profession des périodes d’emploi</i> .....	42
e) <i>Variables de localisation géographique</i> .....	43
f) <i>Codification des ouverts</i> .....	44
4. LE TRAITEMENT DES SALAIRES (PRIMES INCLUSES).....	44
5. ANONYMISATION DES DONNEES ET FINALISATION .....	48
6. FORMATS, LABELS ET DICTIONNAIRE DES VARIABLES .....	48
7. LES TABLES FINALES.....	49
<b>VI. LA PONDERATION FINALE .....</b>	<b>50</b>
<b>VII. ANNEXES.....</b>	<b>55</b>
ANNEXE 1 - TABLE DES ILLUSTRATIONS .....	55
ANNEXE 2 - LETTRE DE PRE-FIDELISATION .....	56
ANNEXE 3 - LETTRE AVIS DE CONTACT AVEC LES JEUNES .....	56
1. LETTRE ENVOYEE PAR COURRIER POSTAL .....	57
1.....	57
2. LETTRE ENVOYEE PAR COURRIER ELECTRONIQUE .....	57
3. PIECES JOINTES AUX COURRIERS .....	59
ANNEXE 4 - CAWI DE FIDELISATION .....	62
a) <i>Écran d’accueil</i> .....	62
b) <i>Authentification avec le mois de naissance et l’année de naissance</i> .....	62
c) <i>Affichage des coordonnées</i> .....	63
d) <i>Après validation des coordonnées, page de prise de rendez-vous</i> .....	63
e) <i>Après validation</i> .....	64
ANNEXE 5 - NOMENCLATURE NAF REV2 EN 88 DIVISIONS (A88) .....	65
ANNEXE 6 - DEFINITION DES VARIABLES ANNEXES UTILISEES DANS SICORE PCS .....	68
ANNEXE 7 - LISTE DES QUESTIONS OUVERTES.....	71

## I. Présentation de l'enquête 2015 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.

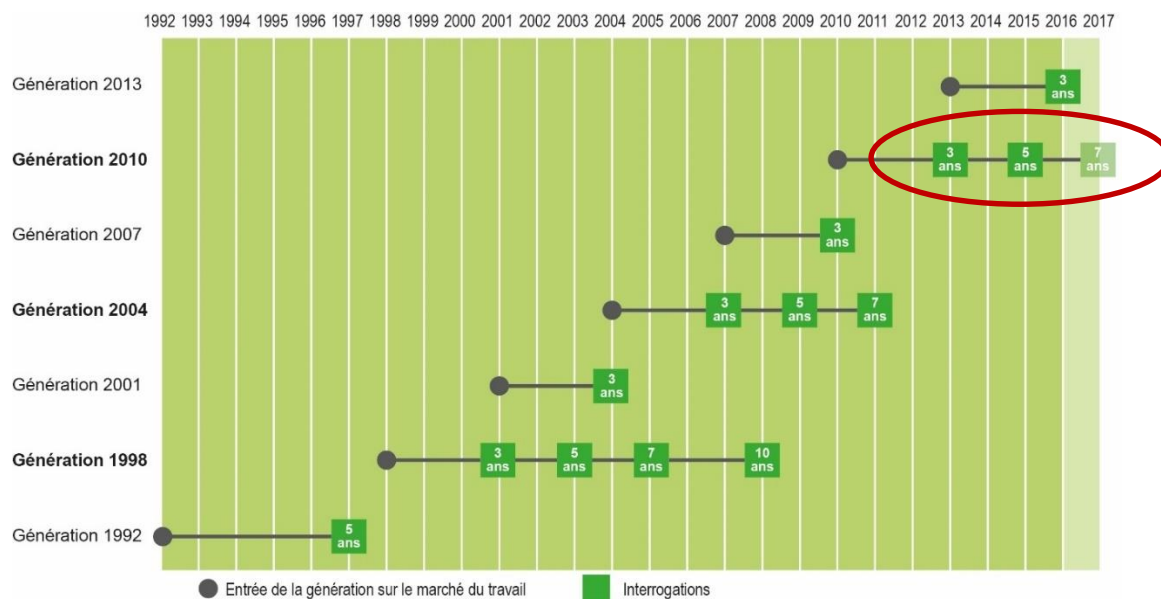
#### ***Encadré 1 - Le Centre d'études et de recherches sur les qualifications***

Créé en 1971, 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
<b>Génération 1992</b>		1997	26 000	-
<b>Génération 1998</b>	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
<b>Génération 2001</b>		2004	25 000	15 000
<b>Génération 2004</b>	Première interrogation	2007	65 000	37 000
	Deuxième interrogation	2009	20 000	2 700
	Troisième interrogation	2011	12 400	2 000
<b>Génération 2007</b>		2010	29 000	26 000
<b>Génération 2010</b>	Première interrogation	2013	38 500	13 500
	Deuxième interrogation	2015	13 710	1 500

## 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 initiale en France (Métropolitaine + DOM) durant l'année scolaire 2009-2010 ;
- avoir quitté le système éducatif entre octobre 2009 et octobre 2010<sup>1</sup>, 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) ;

---

<sup>1</sup> 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 2015 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 2015 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 mars 2015 (label n° 2015X714AU).

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. Les extensions**

Les acteurs intervenant dans le domaine de la formation sont multiples, avec leurs questionnements propres mais aussi avec un questionnement commun sur le devenir d'une population bien précise. À la demande de certains ministères, conseils régionaux ou observatoires de branches, des extensions sont régulièrement adossées aux enquêtes Génération. Elles permettent de situer les analyses ciblées sur des publics, des filières ou des territoires particuliers dans un contexte plus large, par rapprochement avec des indicateurs de référence nationaux.

Plusieurs types d'extensions existent, parfois combinées :

- L'extension d'échantillon, pour disposer d'un nombre de répondants suffisant pour permettre des analyses statistiquement pertinentes sur les catégories de sortants ciblées ;
- l'extension de champ, pour interroger également sur les catégories de sortants ciblées les personnes non retenues dans le champ Céreq (par exemple en levant la restriction de primo sortie) ;
- l'extension de questionnement, pour poser quelques questions complémentaires sur des catégories de sortants ciblées.

De manière générale, il y a très peu d'extensions sur les ré-interrogations des enquêtes Génération. Parmi les 12 extensions présentes lors de la première interrogation (cf Céreq Etudes n°25), seulement une a été conservée pour cette enquête : l'extension d'échantillon des sortants de thèses hors santé, financée par la DGESIP. Le nombre de questionnaires conventionnés était de 1 021. Nous en avons obtenu 1 463 exploitables, soit 43 % de plus que ce qui avait été demandé.

## 4. Descriptif et objectifs de la Génération 2010 à 5 ans

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

L'objectif de cette seconde interrogation est d'élargir la fenêtre d'observation au-delà des toutes premières années d'entrée dans la vie active, d'analyser les premières mobilités professionnelles des jeunes et de permettre le lien entre la première et la troisième interrogation en maintenant le contact avec les jeunes et en décrivant leur situation professionnelle entre trois et cinq ans.

Pour cette seconde interrogation, sont concernés uniquement les jeunes ayant répondu à la première interrogation en 2013.

## 5. Le questionnaire

### a) *Descriptif du questionnaire*

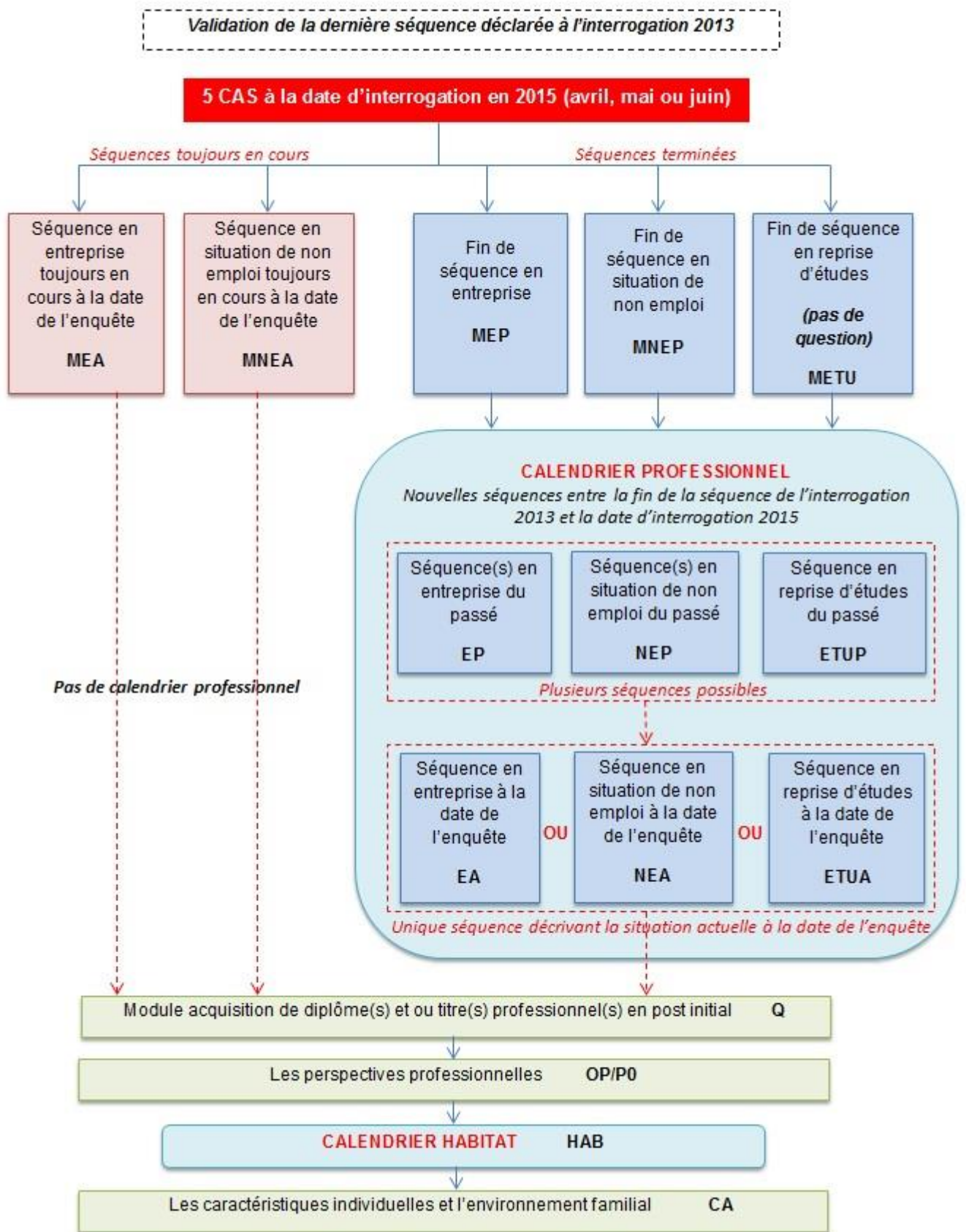
L'architecture du questionnaire de la première interrogation se composait de différentes thématiques : le parcours scolaire, le calendrier professionnel qui constitue le cœur du questionnaire et permet la description des séquences, de modules spécifiques liés à des extensions, les questions relatives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

Le questionnaire de la ré-interrogation s'appuie sur le même questionnaire mais avec une suppression du module sur le parcours scolaire et ceux relatifs aux extensions ou encore un allègement sur les caractéristiques sociodémographiques. La logique de conception de ce questionnaire a consisté à s'appuyer sur la structure de la précédente enquête à 5 ans (Génération 2004 à 5 ans, enquête 2009) en y intégrant les nouveaux questionnements de l'enquête Génération 2010 à 3 ans (en 2013) mais aussi quelques questions de la précédente enquête à 7 ans (Génération 2004 à 7 ans, enquête 2011). Des questions de Génération 1992 à 5 ans ont également été introduites, notamment dans le module « perspectives professionnelles », pour permettre des comparaisons sur une longue période.

À l'issue de cette première phase de conception, le questionnaire a été réparti par binôme entre tous les chargés d'études et chercheurs du Deeva pour contrôles et révisions mineures (ajouts/suppressions de questions et/ou de modalités).

En parallèle de cette construction, un groupe de travail interne au Deeva a mené une réflexion sur l'acquisition de diplômes en cours de vie active. Ce groupe de travail, qui s'est réuni 4 fois entre juin et octobre 2014, a abouti sur une rénovation majeure du module sur l'obtention (ou la tentative d'obtention) de diplômes et de titres professionnels postérieurs à la formation initiale.

Figure 2 - Schéma d'organisation du questionnaire de la 2<sup>ème</sup> interrogation de la Génération 2010



## b) Nouveau module sur l'acquisition de diplôme ou titres en post-initial

Cette enquête intègre un module rénové sur l'obtention (ou la tentative d'obtention) de diplômes ou titres professionnels en post-initial (*i.e.* postérieure à la première sortie du système éducatif en 2009-2010).

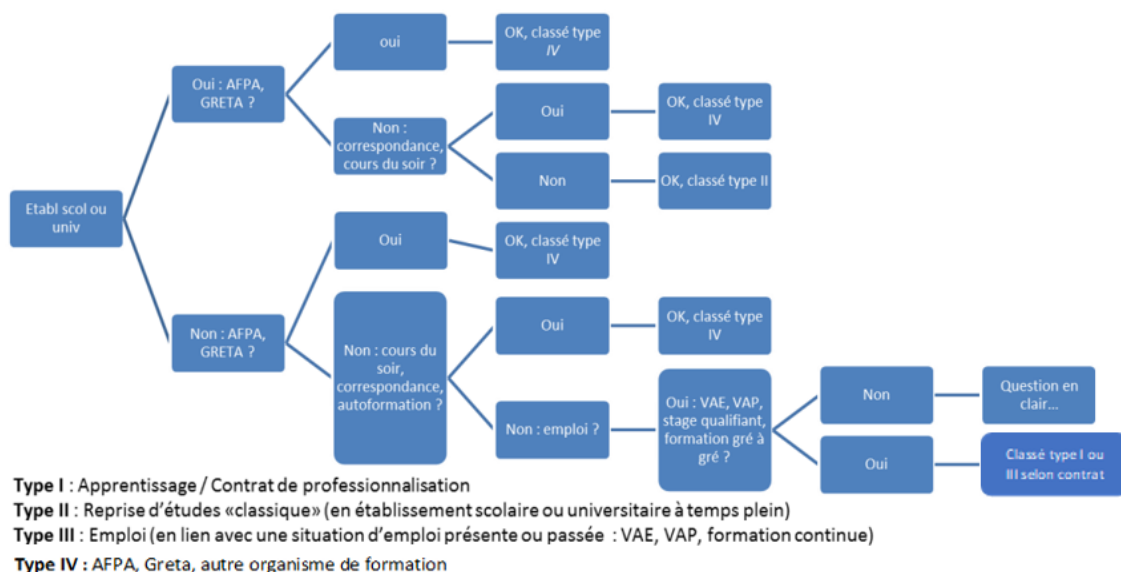
Ce module doit permettre d'appréhender la problématique de la deuxième chance, la porosité de plus en plus importante entre formation et emploi en début de vie active, d'alimenter les acteurs publics de la lutte contre le décrochage scolaire (pour les bas niveaux ou dans le supérieur), d'analyser plus finement les déterminants de ces retours en formation ainsi que mesurer l'efficacité de l'obtention de diplôme en post-initial. Cela répond également à une demande du Comité de concertation des enquêtes Génération exprimée en 2009 et un intérêt du Conseil scientifique du Céreq.

Le premier module de ce type a été développé lors de l'enquête 2011 auprès de la Génération 2004 (interrogation à 7 ans). La première exploitation de ce module a fait l'objet d'une publication « Quand les bacheliers reprennent le chemin des diplômes », Mora, Bref n° 325, 2014. L'exploitation de ce module s'est révélée difficile. Elle a, d'une part, nécessité de lourds traitements manuels pour discerner les « vrais » diplômes et, d'autre part, soulevé des problèmes de concordance avec les questions posées lors du calendrier professionnel. Le module rénové tente de corriger ces principales limites. Il se décompose en deux parties très proches :

- l'obtention de diplômes ou titres professionnels
- la tentative d'obtention de diplômes ou titres professionnels

Pour chaque partie, le questionnement commence par la détermination du diplôme ou du titre professionnel le plus élevé obtenu (ou visé) et sa date d'obtention, le cas échéant, la date de démarrage et la durée de la formation, le type de formation suivi, pour terminer sur des questions d'opinion.

**Figure 3 - Schéma du mode d'obtention du diplôme ou titre**



### c) *Intégration de questions issues de l'enquête Génération 1992 à 5 ans*

Le dispositif des enquêtes Génération fête ses 20 ans en 2017. La Génération 1992 a en effet été interrogée une seule fois en 1997, soit 5 ans après la sortie de formation initiale. L'enquête 2015 auprès de la Génération 2010 est donc une opportunité pour produire des analyses avec un recul temporel de 18 ans.

Des intentions d'études et recherche exploitant cette comparaison sur longue période se sont déjà manifestées au sein du Deeva. Le séminaire de direction du Céreq de septembre 2014 a exprimé le souhait de réaliser un évènement autour des 20 ans des enquêtes Génération (thématique principale de la biennale du Céreq organisée en décembre 2017).

Les principaux ajouts concernent notamment l'opinion sur l'emploi ou sur les perspectives professionnelles. Les questions et modalités ont été volontairement formulées de la même manière qu'en 1997, pour permettre la meilleure comparaison possible avec l'enquête 2015.

### d) *Autres évolutions*

Au-delà du nouveau module sur l'acquisition de diplôme en post-initial et l'intégration de questions de Génération 1992 à 5 ans, d'autres évolutions du questionnaire sont à mentionner.

- Modification du recueil de l'information sur l'activité de l'établissement employeur : auparavant l'activité se codait en nomenclature NES (les questions s'y afféraient) ; pour l'enquête Génération 2010 à 3 ans, la formulation des questions n'a pas été modifiée mais la codification de l'activité a été réalisée *a posteriori* dans la nomenclature NAF rev2. Pour cette interrogation (Génération 2010 à 5 ans) les questions sont modifiées pour une codification directe dans cette nouvelle nomenclature NAF rev2.
- le module de questions sur les périodes de non-emploi est réintégré. La longueur du questionnaire a incité à réduire le questionnement de ce module par rapport à la précédente interrogation (Génération 2010 à 3 ans).
- un module « encadrant » a été ajouté dans le questionnement sur l'emploi. Cette proposition se justifie par le démarrage d'un projet de recherche (en collaboration avec l'Apec) sur les caractéristiques de l'emploi cadre en début de carrière (sous l'angle comparatif hommes/femmes).

Ce module comporte six questions qui n'ont pas été intégrées dans le test.

- la question OP6B sur les raisons d'une inquiétude sur l'avenir professionnel a été ajoutée dans le module « perspectives professionnelles » en complément de la question OP6 (sentiment sur l'avenir professionnel). La question OP6 est très rudimentaire et ne permet pas d'avancer sur les causes ou les raisons subjectives de l'inquiétude des enquêtés par rapport à leur avenir professionnel, elle est donc complétée par la OP6B pour tenter d'apporter des éclairages sur ces raisons.
- les questions EA0 et EA0B sur la correspondance entre emploi et formation, ajoutées dans la partie « opinion sur l'emploi », permettent d'interroger le jeune sur son sentiment d'adéquation ou de déclassement entre la formation initiale suivie et l'emploi occupé.

### 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 2013. 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 2015 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 recherche d'emploi, de reprise d'étude, de formation ou autres, intervenues après le printemps 2013, avec le même protocole que pour l'interrogation précédente.

## 6. Le calendrier de l'enquête

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

Une expérimentation multimode (internet/téléphone) a également été réalisée à partir de mai 2015 <sup>2</sup>.

La chronologie des principales étapes de l'enquête 2015 auprès de la Génération 2010 a été la suivante :

**Tableau 2 - Calendrier de l'enquête**

Étapes	Début	Fin
Mise à jour des adresses postales	Octobre 2014	Novembre 2014
Conception du CATI	26 janvier 2015	13 mars 2015
Élaboration des conventions	Février 2015	11 mars 2015
Échantillonnage	Février 2015	Mars 2015
Actualisation des coordonnées téléphoniques	Février 2015	Février 2015
Test CATI 1	23 mars 2015	25 mars 2015
Test CATI 2	2 avril 2015	3 avril 2015
Terrain d'enquête	13 avril 2015	11 juillet 2015
Constitution des bases, apurements, codifications, pondération, rédaction, documentation	Septembre 2015	Juillet 2016
Livraison des bases aux extensions	Septembre 2016	Septembre 2016
Bilan méthodologique	Décembre 2015	Avril 2017

<sup>2</sup> L'ensemble des expérimentations menées a fait l'objet d'une publication dans les actes des Journées de méthodologie statistique de l'Insee de 2018 : Barret C., Cissé M., Gaubert E., Mazari Z., Olaria M. 2018. Efficacité d'un protocole multimode (téléphone et internet). *Journées de méthodologie statistique de l'Insee*, Paris, 12-14 juin 2018.



## II. Préparation de la collecte

### 1. Constitution de l'échantillon

L'échantillon de la Génération 2010 à 5 ans est proche de celui des individus ayant répondu au questionnaire complet à 3 ans. Ont été exclus de la ré-interrogation des individus provenant d'extensions non reconduites (santé / social, sport) et les jeunes ayant exprimé leur refus d'être recontactés à l'issue de la première interrogation. Un échantillonnage a donc été effectué pour ne plus prendre en compte les sur échantillonnages qui avaient été introduits pour satisfaire les objectifs des partenaires d'extensions. En revanche, les sur échantillonnages qui avaient été introduits concernant les populations qui répondent le moins, et ainsi limiter les conséquences de l'attrition sur ces populations, ont été conservés.

Finalement, sur les 33 547 individus ayant répondu au questionnaire complet en 2013, 23 482 ont été envoyés en enquête en 2015.

#### a) *Retour sur la probabilité d'inclusion du tirage de la première interrogation*

La probabilité d'inclusion du tirage de la deuxième interrogation est déterminée à partir de la probabilité d'inclusion construite pour l'échantillon de la première interrogation.

La probabilité d'inclusion pour le tirage de l'échantillon de la première interrogation était de la forme suivante :  $P(i) = P1(i) * C(i)$

$P1(i)$  correspond à la probabilité de tirage permettant d'obtenir une taille d'échantillon simulé de 25 000 observations. Cette probabilité d'inclusion est calculée en prenant en compte les probabilités de réponse anticipée, les taux de couverture et les strates de formations.

En fonction des caractéristiques des échantillons simulés en utilisant la probabilité de tirage  $P1(i)$ , un coefficient multiplicatif  $C(i)$  a été intégré pour atteindre les objectifs des extensions.

#### b) *Calcul de la nouvelle probabilité d'inclusion*

Deux coefficients interviennent dans le calcul de la probabilité d'inclusion pour tirer l'échantillon de la deuxième vague.

Pour chaque individu répondant de la première interrogation, la nouvelle probabilité d'inclusion, pour obtenir l'échantillon de la deuxième interrogation, fera intervenir un premier coefficient  $1/C(i)$ . Ce coefficient est obtenu en faisant le ratio entre la probabilité d'inclusion  $P1(i)$  qui ne prend pas en compte les surreprésentations liées aux extensions et la probabilité d'inclusion initiale  $P(i)$ .

Si l'observation  $i$  n'appartient à aucune extension,  $1/C(i)$  est égale à 1. Si l'observation  $i$  appartient à une ou plusieurs extensions, on a  $1/C(i) < 1$ . En appliquant à toutes les observations cette probabilité d'inclusion dans le cadre d'un tirage équilibré, nous obtiendrions l'échantillon de 20 790 individus correspondant aux observations financées par le Céreq.

Un léger sur-échantillonnage a été conservé pour certaines sous-populations pour s'assurer un effectif plus important et pour répondre aux besoins des chargés d'études sur les questions à enjeux. Le coefficient multiplicatif pour conserver un léger sur-échantillonnage est défini par :

$C2k(i) = 1 + st\_Ek(i)$ , avec  $k$  l'indice de la sous-population considérée, et

$$st\_Ek(i) = \frac{\text{Nb d'individus souhaités dans la catégorie } Ek}{\text{Nb d'individus dans } Ek, \text{ sans sur échantillonnage}} - 1$$

Si l'individu  $i$  appartient à la catégorie  $Ek$



On a  $C2k(i) \geq 1$  pour toutes les observations.

Au final, la probabilité d'inclusion pour le tirage de l'échantillon de la deuxième vague est :  
 $P(i) = C2k(i)/C(i)$

### c) Tirage équilibré

Le tirage est équilibré d'une part sur la nouvelle probabilité d'inclusion, pour contrôler la taille de l'échantillon. Il est aussi équilibré sur les variables contenant, pour chaque niveau de formation agrégé, les poids finaux de la première interrogation afin de contrôler la structure de l'échantillon par rapport au niveau de formation.

Les tirages équilibrés ont été effectués en utilisant la macro Cube développée par l'Insee.

Les équations d'équilibrage sur la probabilité d'inclusion assurent que

$$\sum_{i \in S} \frac{\pi_i}{\pi_i} = \sum_{i \in U} \pi_i$$

Donc que l'échantillon est de taille fixe.

Les équations d'équilibrage sur le poids  $P_k$ , où  $k$  est le niveau de formation, et  $P_k(i)$  est égale à la pondération finale de la première interrogation si l'individu  $i$  appartient au niveau de formation  $k$ , ou est égale à zéro sinon, assurent que

$$\sum_{i \in S} \frac{P_i}{\pi_i} = \sum_{i \in U} P_i$$

Donc, pour un niveau de formation donné  $k$ , les effectifs pondérés de l'échantillon sont égaux à ceux de la population répondante de la première interrogation. En ajoutant ces contraintes d'équilibrage, on s'assure que l'échantillon aura la même structure que la population répondante à la première interrogation.

### d) Cibles et caractéristiques de l'échantillon

Le tableau suivant présente les effectifs attendus dans l'échantillon de la deuxième interrogation. La première colonne correspond aux effectifs répondant à la première interrogation dans le champ Céreq. La deuxième colonne donne les chiffres dans le cas où le sur-échantillonnage n'est pas conservé. Enfin, la troisième colonne indique les effectifs souhaités dans l'échantillon de la seconde vague. Sur certains niveaux de diplôme, un faible sur-échantillonnage est conservé. Les populations toujours sur-échantillonnées sont les non diplômés et les sortants de niveau bac. La DGESIP a souhaité poursuivre l'extension sur les docteurs hors santé.

**Tableau 3 - Effectifs attendus dans l'échantillon de la deuxième interrogation par plus haut diplôme**

Plus haut diplôme obtenu	Effectif des répondants à la première interrogation	Effectif sans sur échantillonnage	Effectif souhaité dans l'échantillon
<b>NON DIPLÔMÉ</b>	3 888	2 808	3 251
<b>CAP-BEP-MC INDUSTRIEL</b>	2 649	2 031	2 031
<b>CAP-BEP-MC TERTIAIRE</b>	2 620	1 906	1 906
<b>BAC PRO-BT-BP INDUSTRIEL</b>	1 804	1 090	1 246
<b>BAC PRO-BT-BP TERTIAIRE</b>	2 428	1 520	1 728
<b>BAC TECHNO INDUSTRIEL</b>	349	175	201
<b>BAC TECHNO TERTIAIRE</b>	1 646	1 147	1 277
<b>BAC GÉNÉRAL</b>	1 991	1 213	1 447

BTS-DUT INDUSTRIEL	1 018	740	740
BTS-DUT TERTIAIRE	1 495	1 308	1 308
AUTRE BAC+2 INDUSTRIEL	66	41	41
AUTRE BAC+2 TERTIAIRE	311	208	208
BAC+2/3 SANTÉ SOCIAL	2 933	891	891
LICENCE PRO LSH GESTION DROIT	311	238	238
LICENCE PRO MATHS SCIENCE TECHNIQUE	785	457	457
L3 LSH GESTION DROIT	1 141	606	606
L3 MATHS SCIENCE TECHNIQUE	486	160	160
AUTRE BAC+3 LSH GESTION DROIT	171	148	148
AUTRE BAC+3 MATHS SCIENCE TECHNIQUE	40	19	19
BAC+4 LSH GESTION DROIT	385	186	186
BAC+4 MATHS SCIENCE TECHNIQUE	291	123	123
M2 LSH GESTION DROIT	1 805	1326	1 326
M2 MATHS SCIENCE TECHNIQUE	916	479	479
AUTRE BAC+5 LSH GESTION DROIT	199	157	157
AUTRE BAC+5 MATHS SCIENCE TECHNIQUE	134	108	108
BAC+5 ÉCOLE DE COMMERCE	384	369	369
INGÉNIEUR	1 037	672	672
DOCTORAT SANTÉ	281	236	236
DOCTORAT HORS SANTÉ LSH GESTION DROIT	607	129	608
DOCTORAT HORS SANTÉ MATHS SCIENCE TECHNIQUE	1 376	296	1 377
ENSEMBLE	33 547	20 787	23 544

Les probabilités d'inclusion ont été adaptées pour atteindre ces objectifs, comme présenté dans la partie II. 1. b) Calcul de la nouvelle probabilité d'inclusion.

Les tableaux suivants présentent les caractéristiques de l'échantillon, en termes d'effectifs par niveau de formation et de structure de cet échantillon par rapport au niveau de formation avec pondération.

**Tableau 4 - Effectifs dans l'échantillon de la 2ème interrogation par plus haut diplôme**

Plus haut diplôme obtenu	Effectifs dans l'échantillon de la 2 <sup>ème</sup> interrogation
NON DIPLÔMÉ	3 231
CAP-BEP-MC INDUSTRIEL	2 040
CAP-BEP-MC TERTIAIRE	1 896
BAC PRO-BT-BP INDUSTRIEL	1 230
BAC PRO-BT-BP TERTIAIRE	1 719
BAC TECHNO INDUSTRIEL	199
BAC TECHNO TERTIAIRE	1 277
BAC GENERAL	1 439
BTS-DUT INDUSTRIEL	745
BTS-DUT TERTIAIRE	1 298
AUTRE BAC+2 INDUSTRIEL	39
AUTRE BAC+2 TERTIAIRE	209
BAC+2/3 SANTÉ SOCIAL	892
LICENCE PRO LSH GESTION DROIT	241
LICENCE PRO MATHS SCIENCE TECHNIQUE	452
L3 LSH GESTION DROIT	600
L3 MATHS SCIENCE TECHNIQUE	171

AUTRE BAC+3 LSH GESTION DROIT	150
AUTRE BAC+3 MATHS SCIENCE TECHNIQUE	15
BAC+4 LSH GESTION DROIT	185
BAC+4 MATHS SCIENCE TECHNIQUE	128
M2 LSH GESTION DROIT	1 323
M2 MATHS SCIENCE TECHNIQUE	480
AUTRE BAC+5 LSH GESTION DROIT	157
AUTRE BAC+5 MATHS SCIENCE TECHNIQUE	110
BAC+5 ÉCOLE DE COMMERCE	369
INGÉNIEUR	671
DOCTORAT SANTÉ	238
DOCTORAT HORS SANTÉ LSH GESTION DROIT	607
DOCTORAT HORS SANTÉ MATHS SCIENCE TECHNIQUE	1 376
ENSEMBLE	23 487

Le faible écart entre les effectifs obtenus et les cibles recherchées tient au fait que pour certaines observations il a fallu ramener à 1 les probabilités d'inclusion obtenues. En effet, le calcul des coefficients multiplicatifs, tel que précisé dans la partie précédente, ne tient pas compte du fait que certaines probabilités de tirage puissent être supérieures à 1. Certains coefficients multiplicatifs C2 ont donc été ajustés pour mieux approcher les cibles. Une seconde justification aux écarts observés est que le tirage a donné lieu à une phase d'atterrissage pour laquelle il n'y a plus d'équilibrage.

Le tableau suivant présente les structures, selon le plus haut diplôme, de l'échantillon et de la population répondante à la première interrogation. Les structures sont obtenues avec les pondérations.

**Tableau 5 - Structures, selon le plus haut diplôme, de l'échantillon et de la population répondante**

Plus haut diplôme obtenu	Échantillon de la deuxième interrogation	Population répondante première interrogation
NON DIPLÔMÉ	16,79	16,79
CAP-BEP-MC INDUSTRIEL	7,50	7,47
CAP-BEP-MC TERTIAIRE	6,95	6,99
BAC PRO-BT-BP INDUSTRIEL	5,55	5,55
BAC PRO-BT-BP TERTIAIRE	7,90	7,90
BAC TECHNO INDUSTRIEL	0,90	0,90
BAC TECHNO TERTIAIRE	5,29	5,29
BAC GÉNÉRAL	9,18	9,18
BTS-DUT INDUSTRIEL	3,81	3,78
BTS-DUT TERTIAIRE	5,87	5,90
AUTRE BAC+2 INDUSTRIEL	0,20	0,21
AUTRE BAC+2 TERTIAIRE	1,11	1,11
BAC+2/3 SANTÉ SOCIAL	4,01	4,02
LICENCE PRO LSH GESTION DROIT	1,09	1,06
LICENCE PRO MATHS SCIENCE TECHNIQUE	1,91	1,93
L3 LSH GESTION DROIT	2,58	2,62
L3 MATHS SCIENCE TECHNIQUE	0,72	0,68
AUTRE BAC+3 LSH GESTION DROIT	0,54	0,53
AUTRE BAC+3 MATHS SCIENCE TECHNIQUE	0,05	0,06
BAC+4 LSH GESTION DROIT	0,75	0,77
BAC+4 MATHS SCIENCE TECHNIQUE	0,53	0,52
M2 LSH GESTION DROIT	6,37	6,39
M2 MATHS SCIENCE TECHNIQUE	2,16	2,14

AUTRE BAC+5 LSH GESTION DROIT	0,72	0,73
AUTRE BAC+5 MATHS SCIENCE TECHNIQUE	0,55	0,55
BAC+5 ÉCOLE DE COMMERCE	1,47	1,47
INGÉNIEUR	2,94	2,94
DOCTORAT SANTÉ	1,16	1,16
DOCTORAT HORS SANTÉ LSH GESTION DROIT	0,43	0,43
DOCTORAT HORS SANTÉ MATHS SCIENCE TECHNIQUE	0,96	0,96

## 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.

**Tableau 6 - Synthèse des tests**

	Global		Test n° 1		Test n° 2	
Total Enquêtes	224	18,2 %	106	15,4 %	118	21,8 %
Refus de répondre	9	0,7 %	3	0,4 %	6	1,1 %
Refus d'un tiers	8	0,6 %	5	0,7 %	3	0,6 %
Abandons en cours	6	0,5 %	2	0,3 %	4	0,7 %
Hors cible	10	0,8 %	8	1,2 %	2	0,4 %
Résident à l'étranger	4	0,3 %	2	0,3 %	2	0,4 %
Total appels	1 232	100,0 %	690	100,0 %	542	100,0 %
DMC			17mn 51		16mn 00	
Durée du test	5 jours		3 jours		2 jours	
Dates du test			23 au 25 mars 2015		1 <sup>er</sup> et 2 avril 2015	

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 (durée moyenne de communication) et son évolution, en vue de l'enquête ;
- de tester les calendriers interactifs d'activité et habitat.

### 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 déclaré deux ans auparavant en cas de déménagement, de changement de téléphone par exemple.

Les recherches menées par notre prestataire sur notre fichier initial (composé de l'ensemble des répondants de Génération 2010 à 3 ans, enquête principale et expérimentale) ont donné les résultats suivants :

**Tableau 7 - Résultats RNVP**

	Effectif	%
<b>Adresses valides</b>	32 605	83 %
<b>Adresses litigieuses</b>	5 445	14 %
<b>Adresses rejetées</b>	1 113	3 %
<b>Total</b>	39 163	100 %

- Adresses valides : adresses corrigées et enrichies par le prestataire 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. ...

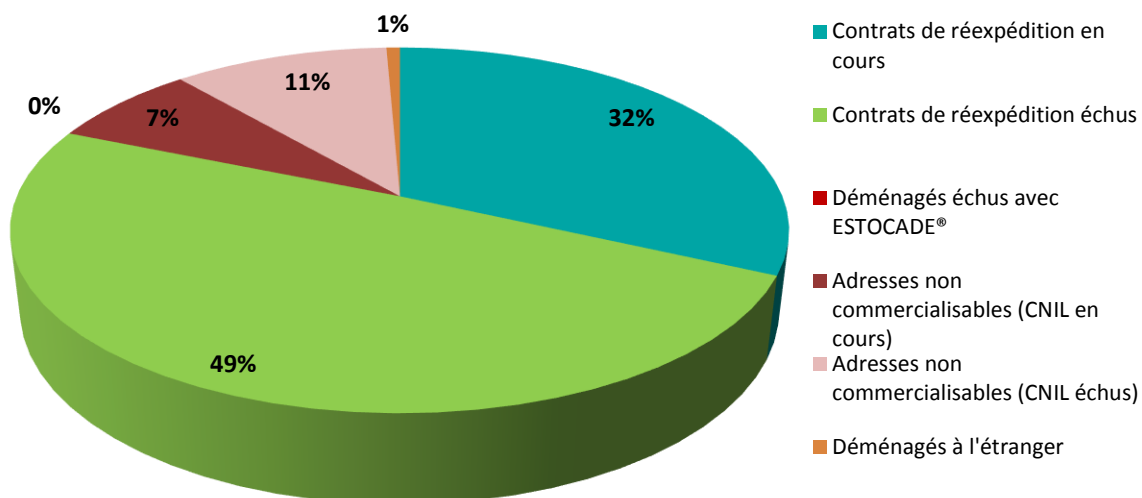
Recherche des déménagés :

Sur les 39 163 adresses envoyées, 3 884 individus ont déménagé dont 3 155 sont exploitables (soit près de 8 % des adresses).

**Tableau 8 - Résultats des recherches des déménagés**

NOUVELLES ADRESSES	QUANTITE	% total déménagé	% total adr.
<b>Contrats de réexpédition en cours</b>	1 236	31,82 %	3,16 %
<b>Contrats de réexpédition échus</b>	1 919	49,41 %	4,90 %
<b>TOTAL ADRESSES RACHETÉES</b>	3 155	81,23 %	8,06 %

REPERAGE DES DEMENAGES	QUANTITE	% total déménagé	% total adr.
<b>Déménagés échus avec ESTOCADE®</b>	0	0,00 %	0,00 %
<b>Adresses non commercialisables (CNIL en cours)</b>	288	7,42 %	0,74 %
<b>Adresses non commercialisables (CNIL échus)</b>	415	10,68 %	1,06 %
<b>Déménagés à l'étranger</b>	26	0,67 %	0,07 %
<b>TOTAL ADRESSES INEXPLOITABLES</b>	729	18,77 %	1,86 %

**Graphique 1 - 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 échoué
- **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 échoué

## 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](http://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](http://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 afin de les faire évoluer ;
- 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.

Audience Facebook :

- 383 fans (condition : avoir un compte Facebook personnel)
- Plus de 2 000 visiteurs / par semaine (ayant interagi ou non avec la page)
- Profil des fans/visiteurs : majoritairement des femmes de 25-34 ans

Le sentiment des jeunes interrogés :

Des commentaires positifs et de nombreux encouragements. Les participants ont trouvé que la démarche était utile, que leur participation était valorisante. Certains ont signalé que le questionnaire était long mais pour la bonne cause. Et d'autres ont indiqué que le calendrier professionnel leur a permis de faire une rétrospection de leur parcours.

Ce qui leur a paru négatif relève de la longueur de l'entretien, de la répétition de certaines questions et de l'inadéquation avec certaines professions. Enfin, plusieurs d'entre eux regrettaient de n'avoir la possibilité de s'exprimer plus.

La page Facebook a rendu l'enquête plus accessible tout en conservant son caractère officiel. Cet outil de communication est adapté au public cible, avec en plus du côté interactif, beaucoup de commentaires utiles pour améliorer la qualité de la collecte des enquêtes futures (notamment concernant l'expérimentation internet).

### **c) Pré-fidélisation**

Une étape de pré-fidélisation a eu lieu en décembre 2014. Un e-mail ou un courrier a été envoyé aux répondants de la première interrogation pour les informer du démarrage de la seconde interrogation au printemps 2015. Les deux modèles de lettre de pré-fidélisation figurent dans ANNEXE 2 - Lettre de pré-fidélisation.

Cette étape permet également au répondant de confirmer ou modifier ses coordonnées (postales et téléphoniques) avant le début de la campagne. Cette opération s'effectue via un formulaire web dont le lien est mentionné dans le mail.

Les répondants de la première interrogation qui ont reçu le courrier papier ont aussi eu la possibilité de modifier leur coordonnées en renvoyant un coupon détachable par enveloppe T.

### **d) Mailings de remerciement**

Pour fidéliser les jeunes ayant répondu à la seconde interrogation en vue de l'interrogation en 2017, des messages électroniques ont été envoyés à chacun des participants. Pour cela, l'adresse mail avait été collectée à la fin du questionnaire et l'adresse de messagerie de contact mise en place pour faciliter les échanges lors de la première interrogation ([generation2010@cereq.fr](mailto:generation2010@cereq.fr)) a été réutilisée.

Créer le lien passe par la connaissance de l'enquête mais aussi de celui qui la produit : le Céreq. Chaque mail contenait un lien vers le site internet du Céreq mais aussi l'adresse de la page Facebook. Le mailing a permis notamment de vérifier au fil de l'eau la validité des mails recueillis.

C'est pourquoi le mailing de remerciement suivant a été envoyé :

Bonjour <pre> <nom>,</pre>

Nous tenons à vous remercier chaleureusement pour votre participation à l'enquête Génération 2010 à laquelle vous avez répondu par téléphone au cours du mois de <mois d'interrogation> 2015.

Si vous souhaitez vous exprimer sur l'enquête, utilisez sa page Facebook « Céreq Génération 2010 : [www.facebook.com/cereqG10](http://www.facebook.com/cereqG10)

L'enquête 2015 est désormais terminée, vous avez été plus de 14 000 à répondre au questionnaire. Nous vous informerons de la sortie des études de premiers résultats prévues début 2016. Vous pouvez retrouver dès à présent des informations sur les enquêtes Génération en cliquant [ici](#).

L'équipe gestion d'enquêtes du Céreq

Le taux d'échec était aux alentours de 9 %. Les échecs étaient pour la plupart des mails qui ont été spammés mais aussi des mails erronés, des boîtes de réception pleine etc.

## 5. Actualisation des coordonnées téléphoniques

Pour chaque ré-interrogation, nous disposons des coordonnées téléphoniques validées par l'enquêté lors de la vague précédente. Néanmoins, deux ans s'écoulant entre deux interrogations, il est nécessaire d'actualiser ces coordonnées afin de pouvoir recontacter l'individu souhaité en cas, par exemple, de déménagement, de changement de téléphone, etc.

L'enrichissement via la pré-fidélisation de décembre 2014 ne suffisant pas, une autre actualisation des coordonnées téléphoniques a été réalisée début février 2015 via un prestataire. Le mode de recherche est le même que celui de l'enrichissement des données de l'enquête Génération 2010 à 3 ans.

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 28 937 individus envoyés en recherches téléphoniques, 10 464 échos ont été retournés (36 %), c'est-à-dire que pour ces individus, on retrouve un écho unique suite à une recherche sur le nom, (éventuellement le prénom) et l'adresse de l'individu.

**Tableau 9 - Bilan de la phase A (téléphone fixe)**

CLASSE	SIGNIFICATION	Nombre de téléphone enrichis	%
<b>U11</b>	Nom, prénom et adresse exacts	2 319	8,01
<b>U12</b>	Nom et prénom exact, adresse approchée, une seule réponse	116	0,40
<b>U13</b>	Nom et prénom identiques, adresse absente, une seule réponse	49	0,17
<b>U21</b>	Nom ou prénom approché, adresse exacte, une seule réponse	146	0,50
<b>U22</b>	Nom ou prénom approché, adresse approchée, une seule réponse	7	0,02
<b>O111</b>	Nom exact, adresse exacte, prénom différent, une seule réponse	4 973	17,19
<b>O11</b>	Nom exact, adresse exacte, prénom différent ou approché, plusieurs réponses	1 729	5,98
<b>O101</b>	Nom exact, prénom exact, adresse différente, une seule réponse	328	1,13
<b>O121</b>	Nom identique, adresse approchée, une seule réponse	440	1,52
<b>O12</b>	Nom exact, adresse approchée	174	0,60
<b>O131</b>	Nom identique, adresse absente, une seule réponse	106	0,37
<b>O13</b>	Nom identique, adresse absente, plusieurs réponses	8	0,03

<b>O211</b>	Nom approché, adresse exacte, une seule réponse	69	0,24
	Total	10 464	36,2

- Phase D :

Sur les 28 937 individus envoyés en recherches téléphoniques, 11 517 échos ont été retournés (40 %), c'est-à-dire que pour ces individus, on retrouve un écho unique suite à une recherche sur le nom, (éventuellement le prénom) et l'adresse de l'individu.

**Tableau 10 - Bilan de la phase D (téléphone portable)**

CLASSE	SIGNIFICATION	Nombre téléphone enrichis	de %
<b>U11</b>	Nom, prénom et adresse exacts	5 579	19,28
<b>U12</b>	Nom et prénom exact, adresse approchée, une seule réponse	263	0,91
<b>U13</b>	Nom et prénom identiques, adresse absente, une seule réponse	112	0,39
<b>U21</b>	Nom ou prénom approché, adresse exacte, une seule réponse	206	0,71
<b>U22</b>	Nom ou prénom approché, adresse approchée, une seule réponse	6	0,02
<b>O111</b>	Nom exact, adresse exacte, prénom différent, une seule réponse	3 314	11,45
<b>O11</b>	Nom exact, adresse exacte, prénom différent ou approché, plusieurs réponses	757	2,62
<b>O101</b>	Nom exact, prénom exact, adresse différente, une seule réponse	812	2,81
<b>O121</b>	Nom identique, adresse approchée, une seule réponse	215	0,74
<b>O12</b>	Nom exact, adresse approchée	95	0,33
<b>O131</b>	Nom identique, adresse absente, une seule réponse	93	0,32
<b>O13</b>	Nom identique, adresse absente, plusieurs réponses	2	0,01
<b>O211</b>	Nom approché, adresse exacte, une seule réponse	63	0,22
	Total	11 517	39,80

(Recherches excluant les individus retrouvés en phase A)

- Phases B et C :

**Tableau 11 - Bilan phase B et C**

PHASE	DESCRIPTION	NB TOTAL	%
B	Rapprochement, nom, prénom, commune	3	0,01
B	Rapprochement, nom, prénom, dept	168	0,41
B	Rapprochement, nom, prénom, France hors IDF	403	0,99
B	Rapprochement, nom, prénom, IDF	97	0,24
B	Rapprochement, nom, prénom, région	23	0,06
C	Rapprochement, nom, commune	16 037	39,48
C	Rapprochement, nom, dept	23 886	58,81
<b>Total</b>		<b>40 617</b>	<b>100,0</b>

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 préfidélisation
2. Numéro ayant permis de réaliser l'enquête à 3 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 afin de 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 mail 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. À la fin de l'enquête 2013 auprès de la Génération 2010, une question permettait de récupérer l'adresse mail des individus qui était d'accord pour nous la communiquer. Pour une question de coûts et de rapidité, le mail a été privilégié pour eux pour l'envoi des lettres avis par rapport à l'envoi de courrier papier. L'impact de cette procédure (sur les taux de réponse notamment) n'a pas été mesuré. Les deux modèles de lettre avis figurent dans ANNEXE 3 - Lettre avis de contact avec les jeunes.

Pour tous les individus disposant d'un mail, la lettre avis a été envoyée par mail. Pour les individus restants, et dont les adresses paraissaient suffisamment fiables, une lettre-avis a été envoyée par voie postale.

**Tableau 12 - Répartition du nombre de lettres avis envoyées selon le mode**

	Effectif
<b>Lettre papier</b>	8 391
<b>Lettre mail</b>	13 487
<b>Lettre mail et papier</b>	1 604
<b>Total</b>	<b>23 482</b>

Le CAWI mis en place est disponible en ANNEXE 2 - Lettre de pré-fidélisation de fidélisation. L'actualisation des coordonnées pouvait aussi se réaliser par téléphone en appelant le numéro vert spécialement dédié à l'enquête.

Au final, 261 individus se sont connectés au CAWI dont 65 pour valider leurs coordonnées et 196 pour les modifier.

### III. La collecte par téléphone

La collecte de l'enquête s'est déroulée par téléphone d'avril à juillet 2015 par le prestataire H2A basé à Bagneux. 23 937 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 par le télé-enquêteur. Le CATI est programmé pour gérer des filtres et orienter le questionnement 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 spécialisés par niveau de formation.

#### 1. Calendrier et organisation générale de la collecte

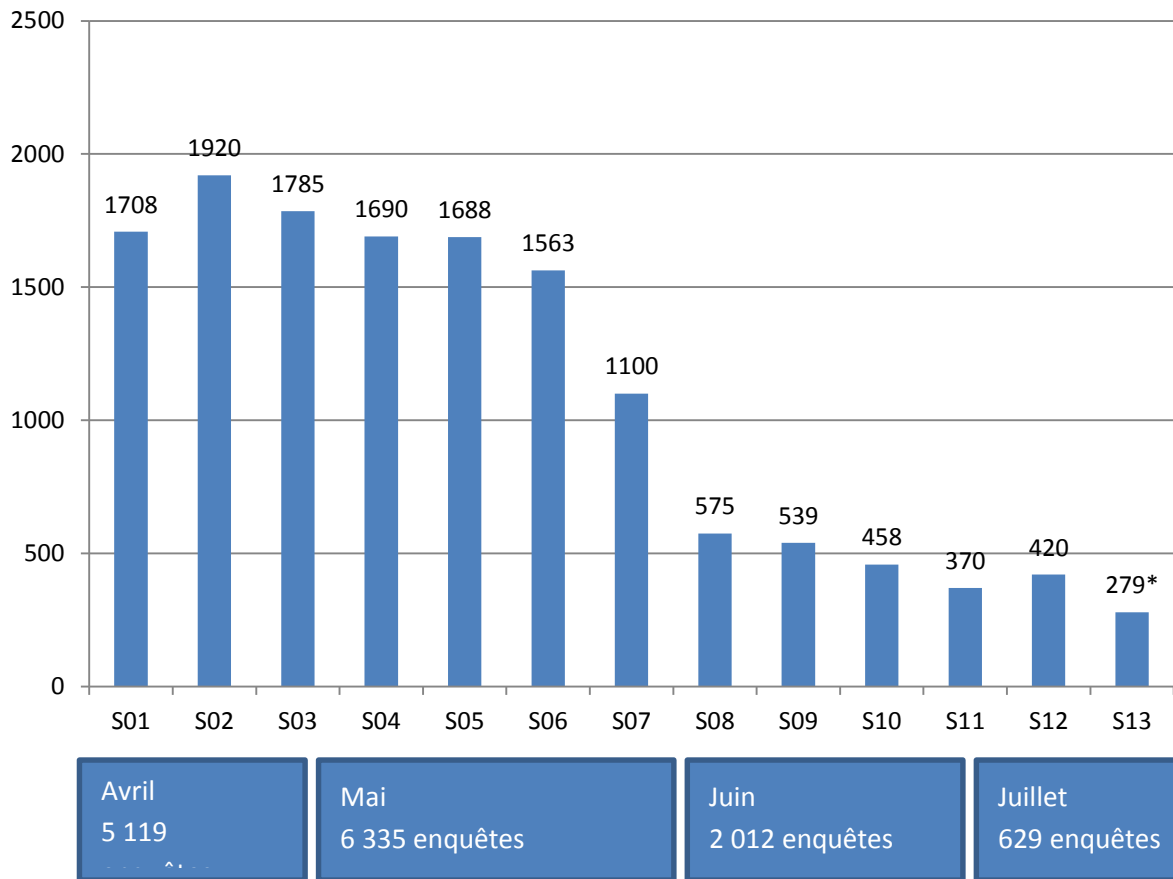
**Tableau 13 - Calendrier et organisation générale avec le prestataire**

	Délais
Réunion de démarrage Transmission du questionnaire	23 janvier 2015
Programmation et tests du questionnaire CATI Transmission au Céreq d'un lien test	26 janvier 2015 – 13 mars 2015
Transmission lien CAWI d'enrichissement des coordonnées	27 février 2015
Réunion programmation	2 mars 2015
Transmission fichier pilote 1	16 mars 2015
Premier pilotage CATI – 106 questionnaires complétés – 17 min 50s en moyenne	23 mars 2015 – 25 mars 2015
Deuxième pilotage CATI auprès de 118 individus – 16 min 00s en moyenne	2-3 avril 2015
Ouverture du numéro vert	8 avril 2015
Formations enquêteurs : Superviseurs Agents	2-3 avril 2015 7-10 avril 2015
Réalisation des enquêtes	13 avril 2015 – 11 juillet 2015
Remise fichier final CATI	29 juillet 2015

#### 2. Le suivi de la collecte en chiffres

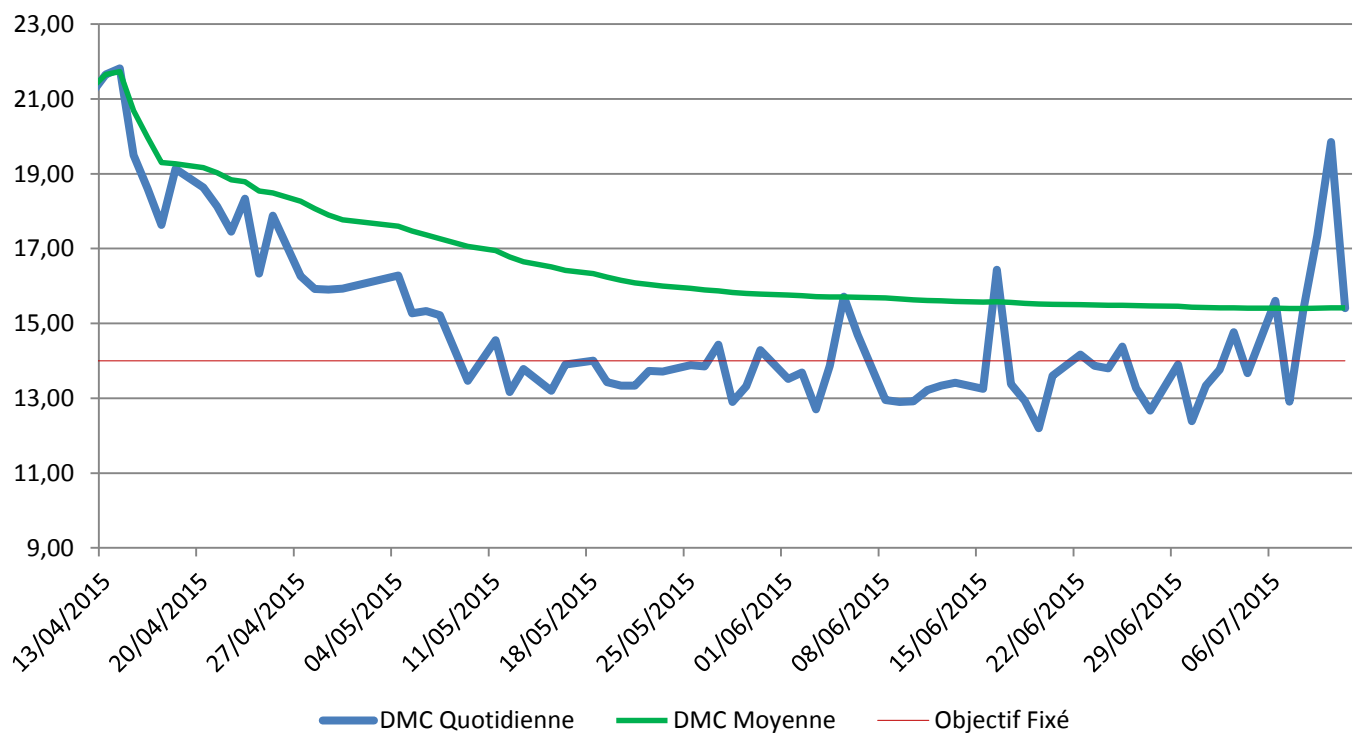
Notre échantillon se composait de 23 482 individus. Parmi eux, 14 095, soit 60 %, ont répondu à l'enquête.

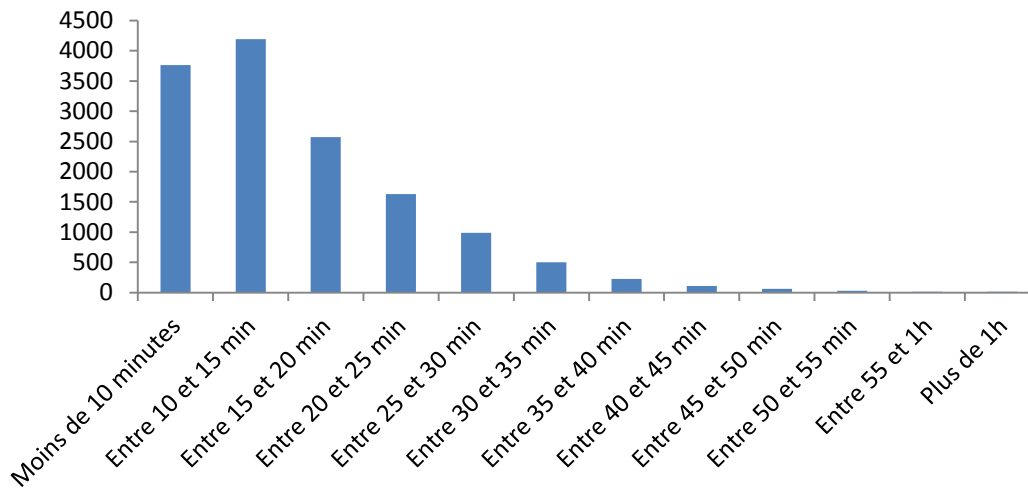
**Graphique 2 - Nombre d'enquêtes réalisées selon la période**



\*Dans S13 ont été également mises les 13 enquêtes récupérées via le numéro vert après le 11 juillet

**Graphique 3 - Durée moyenne de communication**



**Graphique 4 - Durée de passation du questionnaire**

### 3. Les règles de rappel

**Tableau 14 - Les règles de rappel**

Statut adresse	Passage au tel+1	Compteur + 1 sur le tel en cours	Enquêteur	Délai de rappel	Règle
<b>Ne répond pas</b>	OUI	OUI	Libre	80 minutes	Appel sur le 1 <sup>er</sup> et le 2 <sup>ème</sup> numéro en alternance avec jusqu'à 20 appels sur chaque numéro (sauf si classement hors cible sur l'un des numéros, dans ce cas continuer les 20 appels sur le numéro restant puis passage au TEL+1)
<b>Occupé</b>	OUI	OUI	Libre	10 minutes	
<b>Rendez-vous prioritaire</b>					Rappel en priorité des individus ayant pris rendez-vous à partir du CAWI de pré-fidélisation Honorer tous les rendez-vous
<b>Rendez-vous général</b>					
<b>Rendez-vous questionnaire</b>					
<b>Faux numéro (entreprise, fax, dialer...)</b>	OUI	NON			
<b>Change le numéro de téléphone</b>		NON	Même enquêteur que l'appel précédent	Immédiat	Appel sur le numéro enrichi
<b>Répondeur</b>	NON	OUI	Libre	110 minutes	
<b>Hors cible</b>					
<b>Individu non Francophone/physique ou mental</b>	NON / OUI	NON			S'il s'agit de la bonne personne, il s'agit bien d'un hors cible mais il ne faut pas passer au TEL+1.

Injoignable durée d'étude (avant fin juin 2015)	NON	NON			
Il/elle habite à l'étranger	NON	NON			Il faut les renvoyer sur CAWI (géré)
Il/elle est décédé(e)	NON	NON			
Faux numéro nominatif	OUI	NON			Passage au TEL+1
Homonyme parfait	OUI	NON			Passage au TEL+1
Ne connaît pas les coordonnées	OUI	NON			
Autres Hors cibles (contact)	OUI	NON			
<b>Refus</b>					
Raccroche au nez après 2ème tentative	NON	OUI	Même enquêteur que l'appel précédent		Rappel fixé à un autre jour (défini par le système) et une heure différente
de la personne concernée - ne plus jamais appeler	NON	NON			
d'un tiers de passer la personne concernée	OUI	NON			
d'un tiers de donner les coordonnées	OUI	NON			
Refus définitif CNIL	NON	NON			Il ne faut plus du tout rappeler cet individu.
Autre refus	OUI	NON			
Abandon questionnaire					Appliquer les règles précédentes selon le cas.

#### 4. Des messages laissés sur répondeur

En milieu de terrain (mi-mai), 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 à partir du dixième appel.

##### ***Encadré 3 - Message laissé sur les répondeurs des interlocuteurs***

Bonjour Madame/Monsieur XXX du Centre d'études et de recherches sur les qualifications. Nous poursuivons une étude à laquelle vous avez répondu en 2013 et je souhaiterais vous poser quelques questions sur la suite de votre parcours professionnel. Nous avons des difficultés à vous joindre, donc nous aimerions que vous nous contactiez (indiquer SI POSSIBLE les horaires d'ouverture) au numéro vert suivant 0 800 001 939 afin de fixer un rendez-vous pour terminer notre étude. Ce numéro vert est gratuit depuis un téléphone fixe. Dans l'attente de votre appel, merci et très bonne fin de journée.



## 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 collecte et des retours a été réalisé par cette personne « qualité ».

## 6. Les résidents à l'étranger à la date de l'enquête

Une expérimentation de l'enquête Génération en version internet (CAWI) a été testée. Comme pour la première interrogation et dans le cadre de l'enquête principale téléphonique de 2015, les jeunes résidant à l'étranger à la date de l'enquête ne sont pas dans le champ de l'enquête principale et donc non contactés par téléphone. Par ailleurs, pour les besoins de l'enquête expérimentale menée en parallèle, leurs adresses mail ont été récupérées par l'intermédiaire d'un tiers quand cela était possible. Ils ont ainsi eu la possibilité de répondre au questionnaire par internet. Les 193 questionnaires recueillis ne sont pas disponibles dans les bases car les résidents étrangers ne sont pas intégrés dans l'enquête principale afin d'éviter des problèmes d'ordre méthodologique, notamment pour l'agrégation des données. Cependant, ils ont été basculés sur l'enquête expérimentale. Cette expérimentation fera l'objet d'une expertise afin de déterminer la possibilité d'une future intégration de ces jeunes dans le champ des enquêtes.

## IV. Bilan de collecte

Sur les 23 482 individus répondants à la vague précédente mise en production, 14 095 ont répondu à l'enquête Génération 2010 à 5 ans.

**Tableau 15 - Résultats d'appels des numéros mis en production**

	GLOBAL		Terrain 1		Terrain 2		Terrain 3		Recherche Individus	
Enquêtes Réalisées	14 095	60,8 %	7 217	72,1 %	5 323	56,7 %	1 515	50,8 %	40	4,9 %
Hors Cible (Mauvais Individus)	1 603	6,9 %	326	3,3 %	615	6,6 %	254	8,5 %	408	50,3 %
Hors cible (homonyme parfait)	270	1,2 %	14	0,1 %	34	0,4 %	8	0,3 %	214	26,4 %
Rappels pour finir l'enquête	2	0,0 %	1	0,0 %	0	0,0 %	1	0,0 %	0	0,0 %
A déjà répondu	61	0,3 %	16	0,2 %	30	0,3 %	9	0,3 %	6	0,7 %
Abandons en Cours	60	0,3 %	10	0,1 %	38	0,4 %	11	0,4 %	1	0,1 %
Refus de Répondre	627	2,7 %	250	2,5 %	274	2,9 %	97	3,3 %	6	0,7 %
Va rappeler sur le N° Vert	38	0,2 %	10	0,1 %	18	0,2 %	6	0,2 %	4	0,5 %
Résident à l'étranger	194	0,8 %	129	1,3 %	51	0,5 %	13	0,4 %	1	0,1 %
Injoignables Permanents	265	1,1 %	74	0,7 %	111	1,2 %	53	1,8 %	27	3,3 %
Refus d'un tiers	647	2,8 %	149	1,5 %	375	4,0 %	82	2,8 %	41	5,1 %
Décédés	24	0,1 %	2	0,0 %	9	0,1 %	3	0,1 %	10	1,2 %
Demandent à être supprimés des fichiers d'enquête	10	0,0 %	3	0,0 %	5	0,1 %	2	0,1 %	0	0,0 %
Faux Numéro (le numéro ne connaît pas la cible)	3	0,0 %	0	0,0 %	3	0,0 %	0	0,0 %	0	0,0 %
Ne plus jamais appeler	681	2,9 %	286	2,9 %	331	3,5 %	61	2,0 %	3	0,4 %
Refus Cnil définitif	198	0,9 %	55	0,5 %	102	1,1 %	35	1,2 %	6	0,7 %
Abandon après Max Appels	3 929	16,9 %	1 352	13,5 %	1 834	19,5 %	743	24,9 %	0	0,0 %
Ne connaît pas ses coordonnées	482	2,1 %	118	1,2 %	233	2,5 %	87	2,9 %	44	5,4 %
<b>TOTAL DES CONTACTS UTILES</b>	<b>23 189</b>	<b>100,0 %</b>	<b>10 012</b>	<b>100,0 %</b>	<b>9 386</b>	<b>100,0 %</b>	<b>2 980</b>	<b>100,0 %</b>	<b>811</b>	<b>100,0 %</b>
Faux Numéro Disque France Télécom	1 779		408		680		313		378	
NRP	565		62		74		51		378	
A Rappeler	25		1		2		10		12	
Occupé	527		47		79		52		349	
Répondeur	928		60		124		74		670	
N'accepte pas les N° masqués	4		0		3		0		1	
Individu identifié sur nom prénom sur répondeur	461		260		160		23		18	
<b>TOTAL DES FICHES TRAITÉES</b>	<b>27 478</b>		<b>10 850</b>		<b>10 508</b>		<b>3 503</b>		<b>2 617</b>	
Fichier livré (Individus)	23 482		10 454		9 797		3 231			
% Enquêtes / Individus livrés	60,0 %		69,0 %		54,3 %		46,9 %			
Durée moyenne du questionnaire	15 min 25		13 min 44		17 min 13		17 min 06		18 min 01	

## 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 – La pondération finale, 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...

Les fichiers ont été reçus sous divers formats : Excel, Access et texte. En parallèle, des fichiers complémentaires ont été reçus sous format sas : libellés des questions, 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 première interrogation. La principale difficulté réside en effet dans le « raccordement » des bases avec la précédente interrogation en 2013.

En effet, la première étape consiste à alimenter la base de l'enquête 2015 par les réponses obtenues par ces mêmes enquêtés lors de l'enquête 2013 concernant les caractéristiques individuelles, le parcours scolaire ainsi que les calendriers professionnels et d'habitat.

Remarque 1 : Les identifiants des individus sont les mêmes dans les fichiers des différentes vagues, les informations spécifiques non présentes dans les fichiers à 5 ans peuvent être obtenues par appariement de tables.

Remarque 2 : Les observations entrant dans le champ Céreq et celles relevant d'extensions spécifiques n'ont pas été séparées pour les phases d'apurement, de codification, de création de variables.

#### 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). Ceux-ci sont découpés par "terrain" :

- le terrain 1 regroupe les sortants du supérieur (7 217 enquêtes réalisées) ;
- le terrain 2 regroupe les sortants du secondaire (5 323 enquêtes réalisées) ;
- le terrain 3 regroupe les non diplômés (1 515 enquêtes réalisées) ;
- le terrain 4 correspond au terrain de pré-qualification (40 enquêtes réalisées).

Au total, 14 095 enquêtes ont été réalisées.

Les quatre terrains sont ensuite regroupés et la constitution des différents fichiers se fait selon l'objet étudié :

- une table d'enquête par individu ;

- une table sur le calendrier professionnel par individu et séquence d'activité ;
- une table sur le calendrier habitat par individu et séquence d'habitat.

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 se sont donc révélés non exploitables, à cause de certaines périodes se chevauchant non renseignées, ou encore parce que des périodes d'emploi mentionnées ne sont pas associées à une description de celui-ci 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 (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 de questionnaire développé sous le système CATI.

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

**Tableau 16 - Motifs de suppressions des individus**

Motif de suppression	Effectif
<b>Invalidation de la situation déclarée en 2013</b>	162
<b>Problème de calendrier (chevauchement, pas de situation actuelle,...)</b>	76
<b>Invalidation de l'entreprise déclarée en 2013</b>	74
<b>Pas d'ouverture de module en lien avec le calendrier (données manquantes)</b>	46
<b>Pas d'ouverture du calendrier (individu dans une situation différente de 2013)</b>	17
<b>Ouverture du calendrier (individu toujours dans la même situation qu'en 2013)</b>	9
<b>Doublon</b>	1
<b>Total</b>	385

Au final, 385 individus (soit 2,7 %) ont été supprimés. Le nombre total des enquêtes exploitables est donc de 13 710.

#### **b) Création de la base individus**

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

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) Calendrier professionnel et habitat**

##### **– Calendrier professionnel**

La poursuite de séquences d'emploi/non-emploi et l'ajout de nouvelles séquences imposent une renumérotation des séquences (*nseq*) car les numéros de séquences fournis par le prestataire dans le fichier brut ne tiennent pas compte de l'interrogation précédente. 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 2013 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 emplois etc...

– Calendrier habitat

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

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

À cette étape, il faut pouvoir raccorder les informations de 2013 avec les nouvelles informations de 2015. Deux cas sont à gérer :

- pour les individus qui ont changé d'emploi ou de situation de non-emploi entre 2013 et 2015 : le raccord s'établit avec la validation de la dernière séquence à 3 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 2013 et 2015) : le raccord se fait par une transformation des noms des nouvelles variables correspondantes 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 2013, 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, 2 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).

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

Quelques définitions :

**MEP** : Ces séquences concernent les emplois qui étaient en cours en 2013 (lors de l'interrogation à 3 ans) et qui se sont finis entre l'interrogation 2013 (à 3 ans) et l'interrogation 2015 (à 5 ans) ("même emploi dans la séquence du passé").

**MEA** : Ces séquences concernent les emplois en cours en 2013 et qui sont toujours en cours en 2015 ("même emploi dans la séquence actuelle").

**MNEP** : Ces séquences concernent les périodes de non emploi qui étaient en cours en 2013 (lors de l'interrogation à 3 ans) et qui se sont finies entre l'interrogation 2013 (à 3 ans) et l'interrogation 2015 (à 5 ans)

**MNEA** : Ces séquences concernent les périodes de non emploi en cours en 2013 et qui sont toujours en cours en 2015

**META** : Ces séquences concernent les reprises d'études en cours en 2013 et qui sont toujours en cours en 2015

**EP** : Ces séquences concernent les emplois qui ont été déclarés entre les deux interrogations 2013 et 2015 ("nouvelles séquences du passé")

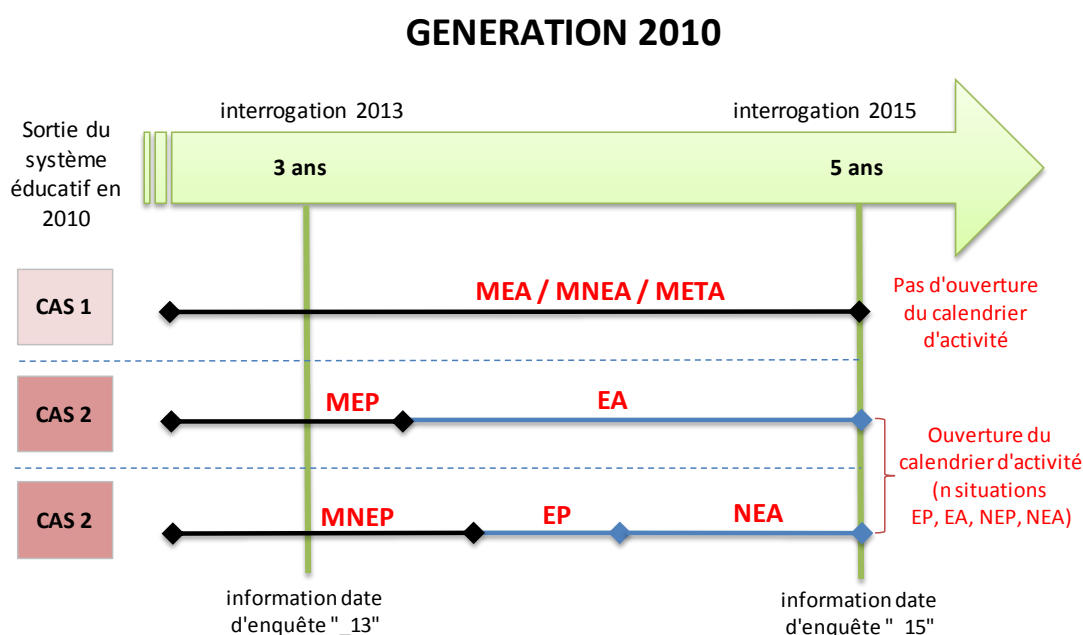
**EA** : Ces séquences concernent les emplois qui débutent après la précédente interrogation (2013) et qui sont toujours en cours en 2015 ("nouvelles séquences actuelles")

**NEA** : Ces séquences concernent les périodes de non emploi qui débutent après la précédente interrogation (2013) et qui sont toujours en cours en 2015 ("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 4 - Schéma récapitulatif expliquant le raccord des séquences entre la première et la seconde 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 2013) continue à la date d'enquête actuelle (c'est-à-dire en 2015) alors aucune autre séquence n'est décrite. La personne n'a donc pas changé de situation depuis l'interrogation à 3 ans.
- Soit la séquence renseignée à la date d'enquête précédente (c'est-à-dire en 2013) s'est terminée entre les deux interrogations (c'est-à-dire entre 2013 et 2015) 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 3 ans et l'est toujours à la date d'enquête à 5 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). Aucune autre séquence n'est décrite.

**CAS 2** : l'individu était en emploi au moment de l'interrogation à 3 ans et a quitté cet emploi entre l'interrogation à 3 et 5 ans (MEP). On décrit ensuite une nouvelle séquence d'emploi qui commence et se termine entre 3 et 5 ans (EP). Enfin, la dernière séquence d'emploi, commencée entre 3 et 5 ans, continue à la date d'enquête à 5 ans. La situation est alors EA.

On crée également la variable TABSEQ qui permet de décrire la temporalité de la séquence. Elle regroupe trois types de séquences :

- 1 = séquence terminée à la première interrogation
- 2 = séquence en cours à la première interrogation
- 3 = séquence débutée lors de la seconde interrogation

#### e) *Les recopies*

Les recopies ne concernent que les séquences d'emploi :

- emplois courts d'une durée inférieure à 6 mois du passé (CAL=01) ou actuels (CAL=03) pour les salariés et intérimaires (INTER=1 ou INTER=2).
- emplois longs d'une durée inférieure ou égale 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
  - o Temps de travail
  - o Salaire, primes et 13<sup>ème</sup> mois

Alors on pose les questions de début et de fin de séquence. Il n'y a donc pas de copie. 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 17 - 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 le calendrier professionnel (emplois issus 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.

La variable STAT\_FIN est le statut de fin d'emploi de la dernière période d'emploi décrite dans le calendrier, si elle se termine avant la date d'interrogation.

- 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 rév.2 (nomenclature des activités économiques en vigueur en France depuis le 1<sup>er</sup> 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 rév.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 5 - Nomenclature NAF rev2 en 88 divisions (A88).

La procédure mise en place pour Génération 2010 à 5 ans :

Dans les enquêtes Génération, 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 complémentaire) ;
- 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, trois méthodes de codage ont été utilisées séquentiellement :

- Imputation du code NAF à partir du menu entreprise (28 % 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
  - o Cas 1 : Traitement de l'information résultant d'un questionnement fermé (15 % 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](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 (40 % 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 4 - 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 à 5 ans, des règles de décisions ont été définies afin de repérer le meilleur écho.

- 1<sup>ère</sup> règle : si Sicore fournit 1 SEUL écho dont la probabilité est supérieure à 0,40, l'écho en 5 positions est choisi.
- 2<sup>ème</sup> 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 82 %.

- Codification manuelle (18 % 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, une codification manuelle a été réalisée à partir de la raison sociale déclarée, de l'information décrite en clair de l'activité et des réponses aux 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 5 - SICORE environnement PCS 2013**

Le logiciel Sicore permet d'effectuer un codage automatique des professions à partir d'un questionnaire 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 6 - Définition des variables annexes 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 et les variables OPA, DEP, SAU et STRE n'ont pas été utilisées car les informations ne sont pas disponibles dans l'enquête. Toutes les autres variables sont utilisées comme variable annexe.

Ce système a permis d'atteindre un taux de codification de 77 %.

Les professions qui n'ont pas pu être codées par cette procédure (23 % 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 2015. 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 (unité urbaine, aire urbaine, 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 (pôle urbain, couronne périurbaine, commune multi polarisée, commune isolée).

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, isolé, rural)
- 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 sont reclassées.

28 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 7 - 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 rémunérations 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. Sont prises en compte uniquement les séquences déclarées entre 2013 et 2015 c'est-à-dire les nouvelles

séquences ou les séquences qui se chevauchent entre les deux interrogations. Les structures du type de déclaration des salaires sont les suivantes pour les débuts et les fins des séquences d'emploi.

**Tableau 18 - Structures du type de déclaration pour les débuts de séquences entre 2013 et 2015 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations)**

Type de déclaration	Pourcentage
Clair	33,20
Smic	4,81
Tranche	4,03
Manquant*	57,95

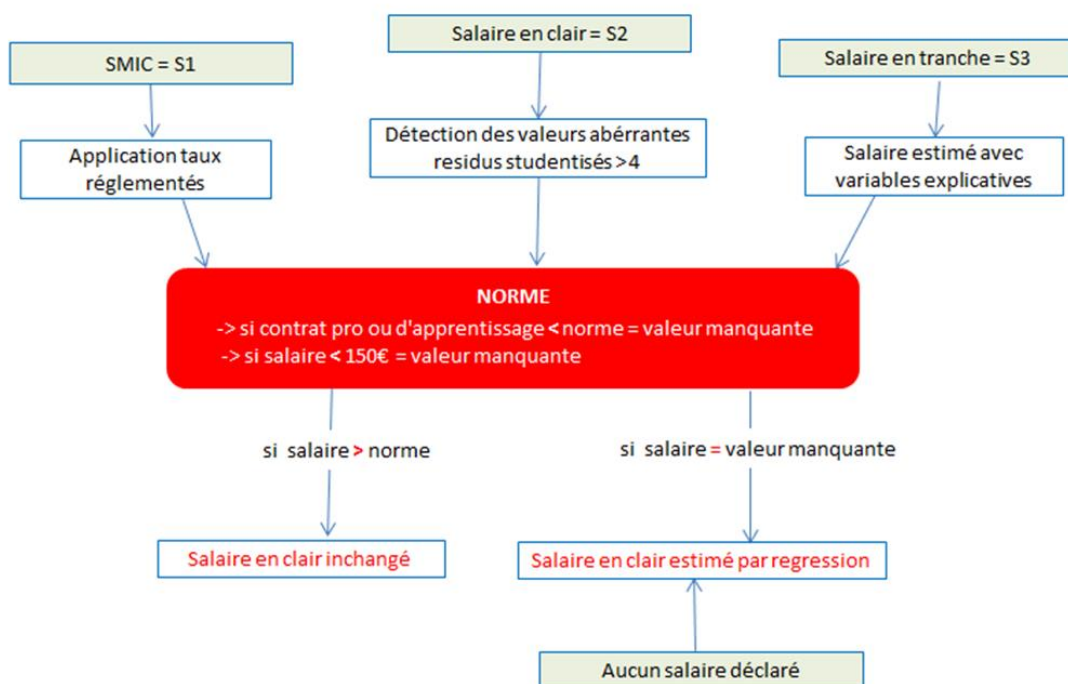
\* Les séquences d'emplois à cheval entre les deux interrogations ont un salaire d'embauche manquant dans l'interrogation de 2015 (question non posée). Ce dernier sera récupéré une fois l'imputation des salaires faites pour les autres séquences avec données manquantes, afin qu'ils n'interfèrent pas dans le modèle de régression

**Tableau 19 - Structures du type de déclaration pour les fins de séquences entre 2013 et 2015 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations)**

Type de déclaration	Pourcentage
Clair	61,55
Smic	8,57
Tranche	4,39
Manquant	25,49

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

**Figure 5 - 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 d'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.

**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. À l'issue de ce modèle, les observations dont les résidus studentisés sont supérieurs à 4 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 ». 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 calculés 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 juin 2013 et juillet 2015 (période d'interrogation)

**Tableau 20 - Revalorisation du SMIC entre avril 2013 et juillet 2015\***

Date de revalorisation	Montant net du SMIC
Janvier 2014	1 128,70
Janvier 2015	1 135,99

\*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 inférieure 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 21 - 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 22 - 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 <sup>ère</sup> année	25 %	41 %	53 %
2 <sup>ème</sup> année	37 %	49 %	61 %
3 <sup>ème</sup> 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 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 nous déclarer le 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<sup>ème</sup> mois » : si l'individu déclare toucher un 13<sup>ème</sup> 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 5 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 est rédigé à partir du dictionnaire de l'enquête de 2013 et complété avec les nouvelles variables ou modifié pour les variables de 2013 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 5 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 23 - Caractéristiques des tables créées**

Nom	L'observation dans la table est :	Contenu
Table « Individus »	Le répondant (13 710 observations)	Ensemble des réponses individuelles de l'enquêté, y compris le calendrier du mode de cohabitation. Informations issues de la précédente vague. Pondérations. Environ 800 variables
Table « Seqentr »	La séquence d'emploi (22 794 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 ». Environ 200 variables
Table « Nonempl »	La séquence de non emploi (32 641 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 » Environ 170 variables

### Les bases comparables

Pour usage interne, le Céreq a réalisé une base nationale à champ comparable pour les enquêtes à 5 ans des Générations 1992, 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 2007.

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

Contrairement à une première interrogation où la qualité des numéros de téléphone de la base de sondage, la difficulté à contacter certains types de public justifient le recours à une décomposition du mécanisme de la non réponse selon les trois étapes que sont le contact, l'acceptation de participer à l'enquête, et enfin le fait d'y répondre, la correction de la non réponse de cette ré-interrogation est ici plus simple.

Il est en effet observé sur plusieurs enquêtes que le taux de réponse est de l'ordre de 60 % ce qui est plus « confortable » que les taux de réponse d'une première interrogation. Ce qui s'explique en partie par le fait que les coordonnées ont été validées ou enrichies avant le terrain d'enquête en quasi-totalité. De ce fait le choix a été fait de recourir à une correction de la non réponse plus simple basée sur une modélisation de la probabilité de répondre en une étape, suivi d'un calage sur les marges issues du croisement SEXE\*PHD déterminé sur les données de Génération 2010 à 3 ans.

### *Bilan de la collecte*

Sur les 23 487 individus envoyés en production, 13 710 sont répondants et ont des questionnaires utilisables (un questionnaire inutilisable a été considéré comme une non-réponse totale).

### *Modélisation de la non-réponse*

La pondération initiale utilisée dans le modèle (PONDCORR dans le modèle ci-dessous) résulte du rapport entre la pondération de 2010 3 ans sur la probabilité d'échantillonnage pour le terrain de la seconde interrogation (coeffextb dans les bases).

#### **Encadré 6 - Script du modèle utilisé**

```
proc logistic data=production5ans outest=outest descending missing;
class q1 typotraj sitde phd habde AGE10 ca0areg zus ;
model repondant =q1 typotraj sitde phd habde AGE10 ca0areg zus / /*INFLUENCE*/ lackfit selection=stepwise
link=logit ;
output out=outmod2 predprobs = l;
weight pondcorr;
run;
```

### *Le modèle et ses caractéristiques*

L'option stepwise est utilisée pour que le modèle minimise les critères d'information AIC. Cela garantit que le modèle ne conserve pas trop de variables et n'aboutisse pas à une spécification inutilement complexe voire dommageable pour ce qui est de la précision des estimations.

L'option Lackfit utilisée pour réaliser le test d'adéquation de Hosmer-Lemeshow (pour vérifier l'adéquation du modèle aux données) fournit une p-valeur de 3,5 %. L'hypothèse nulle de bonne adéquation est conservée si on prend un niveau de 1 % (le test du khi-deux sous-jacent tend systématiquement à rejeter H0 lorsque les effectifs augmentent). En comparant les effectifs observés et prédits l'adéquation semble correcte.

**Encadré 7 - Sortie SAS du test d'adéquation de Hosmer-Lemeshow**

Partition for the Hosmer and Lemeshow Test					
Group	Total	repondant = 1		repondant = 0	
		Observed	Expected	Observed	Expected
1	2350	1001	952.06	1349	1397.94
2	2349	1103	1091.40	1246	1257.60
3	2349	1172	1167.14	1177	1181.86
4	2349	1230	1233.68	1119	1115.32
5	2349	1286	1301.10	1063	1047.90
6	2349	1381	1379.99	968	969.01
7	2349	1479	1475.47	870	873.53
8	2349	1634	1573.11	715	775.89
9	2351	1689	1655.16	662	695.84
10	2343	1735	1765.63	608	577.37

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
16.8318	8	0.0349

Avec les quantiles ci-dessous caractérisant la dispersion des probabilités de répondre estimées, on peut noter qu'il n'y a pas de valeurs infinitésimales. Par conséquent, il n'y a pas eu de recours aux classes de repondération.

**Encadré 8 - Sortie SAS de l'estimation des quantiles**

Quantile	Estimate
100% Max	0.999991
99%	0.799521
95%	0.754257
90%	0.730645
75% Q3	0.687644
50% Median	0.597623
25% Q1	0.514588
10%	0.459140
5%	0.430863
1%	0.369440
0% Min	0.142566

L'aire sous la courbe de ROC est de 62,4 %.

### *Les effets des facteurs :*

PHD : Globalement la probabilité de répondre augmente avec le niveau de diplôme

AGE : La probabilité de répondre augmente aussi avec l'âge (âge et diplôme sont corrélés positivement.)

SITDE : Sitde la référence est : sitde=5 reprise d'étude

Sitde=1 en emploi ++

Sitde=2 recherche d'emploi -

Sitde=3 autre situation - -

Sitde=4 formation –

SEXE : Référence = Femmes

Q1=hommes +

SITUATION D'HABITAT A LA DATE D'ENQUETE :

Référence : vit seul

Vit chez ses parents +

Vit en couple +

TRAJECTOIRE D'INSERTION :

Référence p01. Accès immédiat et durable à l'emploi

typotraj p02. Accès rapide et durable à l'emploi +

typotraj p03. Accès progressif à l'emploi après chômage - (limite)

typotraj p04. Accès progressif à l'emploi après inactivité +

typotraj p05. Sortie d'emploi vers le chômage -

typotraj p06. Sortie temporaire du marché du travail vers l'inactivité –

typotraj p07. Chômage durable ou récurrent+

typotraj p08. Inactivité durable +

typotraj p09. Périodes importantes ou récurrentes de RE et formations en cours de parcours +

### *Calage sur marges*

Le calage sur marges est obtenu par le croisement SEXE et PHD. Les 30 modalités de la grille de plus haut diplôme utilisées et le croisement avec le sexe conduit à 60 marges de calage. (Une matrice ligne avec 60 colonnes). En pratique, l'Insee utilise entre 50 et une centaine de marges de calage. Le calage a été fait avec la macro Calmar. La distance linéaire (M1) est suffisante du fait que l'analyse des rapports de poids ne montre pas de difficulté particulière. En effet, les rapports des poids calés sur les poids initiaux sont compris entre 0.84 et 1.65. Il n'y a pas de corrections extrêmes des poids durant cette étape de calage.

**Encadré 9 - Sortie SAS de l'estimation des quantiles pour le calage sur marges**

Quantile	Estimate
100% Max	1.644953
99%	1.104389
95%	1.059172
90%	1.036459
75% Q3	1.013272
50% Median	1.001349
25% Q1	0.981531
10%	0.969213
5%	0.968575
1%	0.861264
0% Min	0.849603

**Le contrôle par rapport aux variables d'intérêt.**

Des comparaisons entre les structures des variables obtenues sur les données de génération 2010 à 3 ans et Génération 2010 à 5 ans ont été faites, y compris sur des variables n'intervenant pas dans la correction de la non réponse. Il s'agit de vérifier que l'on retrouve bien les mêmes structures entre les individus pondérés de 2010 à 3 ans et ceux pondérés de 2010 à 5 ans par rapport aux mêmes questions posées lors de la première interrogation.

**Tableau 24 - Comparaisons entre les structures des variables obtenues sur les données de Génération 2010 à 3 ans et Génération 2010 à 5 ans**

Variable	Modalités	Génération 2010 à 3 ans		Génération 2010 à 5 ans	
		Effectif	Part en %	Effectif	Part en %
<b>SITDE</b>	1	478003.49	67.5199	478223.33	67.5301
	2	140709.17	19.8757	140980.37	19.9079
	3	28624.74	4.0434	28722.49	4.0559
	4	17503.17	2.4724	17561.07	2.4798
	5	43103.65	6.0886	42675.68	6.0263
<b>Permis de conduire</b>	1	548182.81	77.4367	554065.06	78.2398
	2	159728.37	22.5633	154097.89	21.7602
<b>Séjours à l'étranger</b>	1	71348.60	10.0788	71188.34	10.0525
	2	23017.68	3.2515	24001.50	3.3893
	3	613544.90	86.6698	612973.11	86.5582
<b>Sexe</b>	1	360410.40	50.9094	360411	50.9094
	2	347533.83	49.0906	347535	49.0906
<b>Plus haut diplôme obtenu</b>	01	118839.22	16.7865	118840	16.7866
	02I	52850.87	7.4654	52851	7.4654
	02T	49463.97	6.9870	49464	6.9870
	03I	39291.81	5.5501	39292	5.5501
	03T	55919.43	7.8988	55919	7.8988
	04I	6403.50	0.9045	6404	0.9046
	04T	37452.77	5.2904	37452	5.2902
	05	64996.59	9.1810	64997	9.1811
	06I	26749.21	3.7784	26749	3.7784
	06T	41742.73	5.8963	41743	5.8964
	07I	1475.50	0.2084	1476	0.2085
	07T	7862.90	1.1107	7863	1.1107
08	28446.88	4.0182	28446	4.0181	

	09L	7524.52	1.0629	7525	1.0629
	09M	13661.19	1.9297	13662	1.9298
	10L	18527.40	2.6171	18527	2.6170
	10M	4831.30	0.6824	4831	0.6824
	11L	3725.44	0.5262	3726	0.5263
	11M	432.76	0.0611	433	0.0612
	12L	5440.31	0.7685	5440	0.7684
	12M	3646.50	0.5151	3646	0.5150
	13L	45253.10	6.3922	45253	6.3922
	13M	15127.80	2.1369	15128	2.1369
	14L	5140.71	0.7261	5141	0.7262
	14M	3865.38	0.5460	3865	0.5459
	15	10385.35	1.4670	10386	1.4671
	16	20845.70	2.9445	20846	2.9446
	17	8180.12	1.1555	8180	1.1555
	18L	3076.64	0.4346	3077	0.4346
	18M	6784.62	0.9584	6784	0.9583

Sur les variables Sexe et plus haut diplôme, la correspondance est très forte voire complète puisque le calage a été fait sur ces variables.

## 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 - Effectifs attendus dans l'échantillon de la deuxième interrogation par plus haut diplôme .....	16
Tableau 4 - Effectifs dans l'échantillon de la 2ème interrogation par plus haut diplôme .....	17
Tableau 5 - Structures, selon le plus haut diplôme, de l'échantillon et de la population répondante .....	18
Tableau 6 - Synthèse des tests .....	19
Tableau 7 - Résultats RNVP .....	20
Tableau 8 - Résultats des recherches des déménagés .....	20
Tableau 9 - Bilan de la phase A (téléphone fixe) .....	24
Tableau 10 - Bilan de la phase D (téléphone portable) .....	25
Tableau 11 - Bilan phase B et C .....	26
Tableau 12 - Répartition du nombre de lettres avis envoyées selon le mode .....	26
Tableau 13 - Calendrier et organisation générale avec le prestataire .....	28
Tableau 14 - Les règles de rappel .....	30
Tableau 15 - Résultats d'appels des numéros mis en production .....	33
Tableau 16 - Motifs de suppressions des individus .....	35
Tableau 17 - La codification du contrat de travail .....	39
Tableau 18 - Structures du type de déclaration pour les débuts de séquences entre 2013 et 2015 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations) .....	45
Tableau 19 - Structures du type de déclaration pour les fins de séquences entre 2013 et 2015 (nouvelles séquences ou séquences qui se chevauchent entre les deux interrogations) .....	45
Tableau 20 - Revalorisation du SMIC entre avril 2013 et juillet 2015* .....	46
Tableau 21 - Normes de rémunérations des contrats de professionnalisation .....	47
Tableau 22 - Normes de rémunérations des contrats d'apprentissage .....	47
Tableau 23 - Caractéristiques des tables créées .....	49
Tableau 24 - Comparaisons entre les structures des variables obtenues sur les données de Génération 2010 à 3 ans et Génération 2010 à 5 ans .....	53

#### Les figures

Figure 1 - Calendrier des enquêtes Génération .....	6
Figure 2 - Schéma d'organisation du questionnaire de la 2 <sup>ème</sup> interrogation de la Génération 2010 .....	10
Figure 3 - Schéma du mode d'obtention du diplôme ou titre .....	11
Figure 4 - Schéma récapitulatif expliquant le raccord des séquences entre la première et la seconde interrogation .....	37
Figure 5 - Gestion des salaires .....	45

#### 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 - Message laissé sur les répondeurs des interlocuteurs .....	31
Encadré 4 - SICORE environnement Activité (millésime 2012) .....	42
Encadré 5 - SICORE environnement PCS 2013 .....	43
Encadré 6 - Script du modèle utilisé .....	50
Encadré 7 - Sortie SAS du test d'adéquation de Hosmer-Lemeshow .....	51
Encadré 8 - Sortie SAS de l'estimation des quantiles .....	51
Encadré 9 - Sortie SAS de l'estimation des quantiles pour le calage sur marges .....	53

#### Les graphiques

Graphique 1 - Répartition des déménagés .....	21
Graphique 2 - Nombre d'enquêtes réalisées selon la période .....	29
Graphique 3 - Durée moyenne de communication .....	29
Graphique 4 - Durée de passation du questionnaire .....	30





## 1. Lettre envoyée par courrier postal

1.



Centre d'études et de recherches  
sur les qualifications

Céreq DFFVA G2010  
10, place de la Joliette  
BP 21321  
13567 Marseille cedex 02  
[www.cereq.fr](http://www.cereq.fr)



Bonjour <prénom> <nom> ,

Le Céreq poursuit son enquête statistique sur l'avenir professionnel des jeunes sortis du système éducatif en 2010. 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.

Nous vous remercions pour votre première participation à cette enquête en 2013. 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 sera de plus courte durée, une quinzaine de minutes environ.

Vous trouverez ci-joint un calendrier qui vous aidera à reconstituer votre parcours professionnel depuis le printemps 2013. Nous vous invitons à le remplir et à le conserver afin de l'utiliser lors de notre entretien téléphonique.

Vos réponses serviront uniquement à produire des statistiques : elles resteront strictement confidentielles.

Nous vous remercions très vivement de votre participation à cette nouvelle enquête.

Le directeur du Céreq  
Alberto LOPEZ



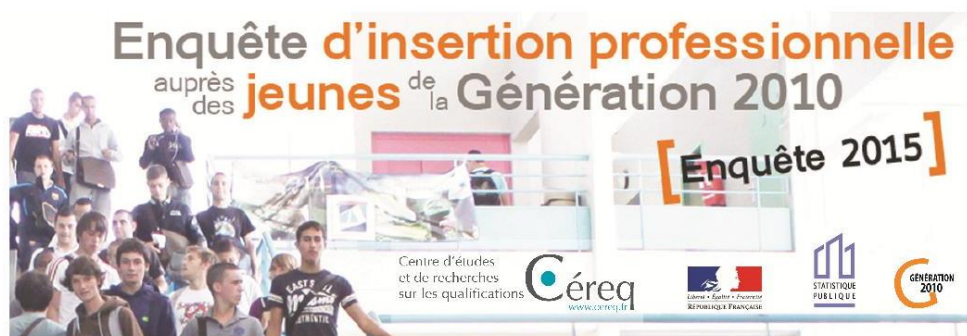
**<Si votre numéro de téléphone n'est plus le : TEL1>  
ou <Nous ne disposons pas de vos coordonnées téléphoniques>**

Merci de nous communiquer par internet votre numéro de téléphone en vous connectant à l'adresse suivante : <https://generation2010.cereq.fr> en utilisant l'identifiant suivant <IDENT> suivi du mot de passe <MDP>.

Vous pouvez également nous contacter au **numéro vert 0 800 00 19 39** (appel gratuit depuis un téléphone fixe, en semaine entre 9h et 20h30, le samedi entre 10h et 18h) en mentionnant la référence : <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°2015X714AU du Ministre de l'Education nationale, du Ministre du travail, de l'emploi, de la formation professionnelle et du dialogue social, du Ministre des finances et des comptes publics et du Ministre de l'économie, de l'industrie et du numérique, valable pour l'année 2015. 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 à la production de statistiques par le 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. Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous 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](mailto:cil-generation@cereq.fr) en mentionnant dans l'objet « Génération 2010 » suivi de votre identifiant apparaissant sur la première page de ce courrier.

## 2. 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 <pre> <nom> ,

Le Céreq poursuit son enquête statistique sur l'avenir professionnel des jeunes sortis du système éducatif en 2010. **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.

Nous vous remercions pour votre première participation à cette enquête en 2013. **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 sera de plus courte durée, une quinzaine de minutes environ.

Vous trouverez en pièce jointe un calendrier qui vous aidera à reconstituer votre parcours professionnel depuis le printemps 2013. Nous vous invitons à le remplir puis à l'imprimer, afin de l'utiliser lors de notre entretien téléphonique.

Vos réponses serviront uniquement à produire des statistiques : elles resteront strictement confidentielles.

Nous vous remercions vivement pour votre participation à cette nouvelle enquête.

Le directeur du Céreq  
Alberto Lopez



**<Si votre numéro de téléphone n'est plus le : TEL 1 >  
Ou <Nous ne disposons pas de vos coordonnées téléphoniques>**

Vous pouvez nous communiquer par internet votre numéro de téléphone en cliquant [ici](#).

Vous pouvez aussi nous contacter au **numéro vert 0 800 00 19 39**  
(appel gratuit depuis un téléphone fixe, en semaine entre 9h et 20h30, le samedi entre 10h et 18h)  
en mentionnant la référence : **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°2015X714AU du Ministre de l'Éducation nationale, du Ministre du travail, de l'emploi, de la formation professionnelle et du dialogue social, du Ministre des finances et des comptes publics et du Ministre de l'économie, de l'industrie et du numérique, valable pour l'année 2015. 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 à la production de statistiques par le 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. Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous 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](mailto:cil-generation@cereq.fr) en mentionnant dans l'objet « Génération 2010 » suivi de votre identifiant mentionné dans le cadre.

Si vous ne souhaitez pas recevoir d'emails de notre part, **cliquez ici**.



### 3. Pièces jointes aux courriers

#### ► LES ENQUÊTES GÉNÉRATIONS

Réalisées par le Céreq, celles-ci observent régulièrement l'entrée dans la vie active de l'ensemble d'une génération ayant quitté le système scolaire une même année. Emploi, formation, inactivité, chômage, reprise d'études... Elles apportent un éclairage sur la diversité des parcours et les effets du contexte économique et social sur l'insertion professionnelle des jeunes.

À ce titre, vous avez été interrogé(e) en 2013. Les résultats de cette enquête ont donné lieu à différentes publications, dont l'ouvrage *Quand l'école est finie* que vous pouvez consulter sur le site :

[www.cereq.fr](http://www.cereq.fr) Rubrique publications / ouvrages.



Vous pouvez également vous abonner à notre page Facebook, dédiée à votre Génération, la Génération 2010, afin de suivre l'évolution de l'enquête et de ces résultats. [facebook.com/cereqG10](https://facebook.com/cereqG10)

#### ► POURQUOI VOUS INTERROGE-T-ON À NOUVEAU ?

Cette enquête cherche à compléter votre parcours depuis la dernière interrogation au printemps 2013, soit deux ans de plus sur le marché du travail. Le Céreq souhaite observer l'insertion professionnelle des jeunes successivement sur trois, cinq et sept années de vie active. Les résultats permettront notamment d'alimenter les réflexions sur les politiques publiques d'emploi et de formation.

#### ...PAR TÉLÉPHONE

Un téléenquêteur spécifiquement formé à cette enquête vous interrogera par téléphone. Afin que l'entretien se déroule de manière fluide et confortable, et pour assurer une bonne compréhension de vos réponses, l'utilisation d'un téléphone fixe est préférable.

#### ► LA CONFIDENTIALITÉ DES DONNÉES

Vos réponses au questionnaire sont couvertes par le secret statistique et les données collectées ne donneront lieu qu'à des exploitations statistiques. Vos coordonnées ne seront utilisées que par le Céreq pour le dispositif « Génération 2010 ».

#### ► QU'EST-CE QUE LE CÉREQ ? QUEL EST SON RÔLE ?

Le Céreq (Centre d'études et de recherches sur les qualifications) est un établissement public sous tutelle du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche et du ministère du Travail, de l'Emploi, de la Formation professionnelle et du Dialogue social. Expert de la formation et de l'emploi, il produit des statistiques, des études, et collabore à de nombreux travaux aux côtés des acteurs de l'éducation et de l'emploi tant au niveau national qu'international afin d'éclairer les politiques dans ces domaines.

**Pour plus d'informations et pour retrouver des résultats des études précédentes, n'hésitez pas à consulter le site du Céreq :**

[www.cereq.fr](http://www.cereq.fr)



## Conseils pour préparer l'entretien téléphonique

- ▶ **Nous vous suggérons afin de faciliter l'entretien, de remplir le calendrier professionnel de 2013 à 2015** (voir au dos) :
  - en vous aidant de vos CV, contrats de travail et bulletins de salaire,
  - en faisant appel aux événements qui ont marqué votre vie afin de vous repérer dans le temps.

*Quelques repères chronologiques (quelle était votre situation professionnelle à ces dates ?):*

5 décembre 2013 mort de Nelson Mandela

7 au 23 février 2014 jeux olympiques d'hiver de Sotchi

12 juin au 13 juillet 2014 coupe du monde de football au Brésil

4 août 2014 centenaire de la deuxième guerre mondiale

### QUELQUES RÈGLES

- ▶ Les diverses situations sont exclusives, c'est-à-dire que **pour une période donnée, vous ne pouvez décrire qu'une seule situation.**
- ▶ Si vous avez vécu plusieurs situations différentes pendant une même période, **privilégiez toujours le travail sur les autres situations.**
- ▶ Si, pendant un mois donné, vous avez travaillé trois semaines et le reste du mois vous étiez en recherche d'emploi, alors considérez le mois entier comme une période de travail.

### QUELQUES PRÉCISIONS SUR LES SITUATIONS

- ▶ **« En Intérim »** : vous avez effectué plusieurs missions d'intérim dans des entreprises différentes durant une période donnée. Si vous étiez intérimaire dans une seule et même entreprise, classez-vous dans la situation **« En entreprise »**.
- ▶ **« En reprise d'études »** : uniquement si vous avez repris vos études à temps plein dans un établissement scolaire ou universitaire.
- ▶ **« En formation »** : cette rubrique ne concerne que les périodes de formation hors contrat de travail.

## DOCUMENT À CONSERVER – NE PAS RENVOYER

Vous trouverez au dos de cette feuille quelques conseils pour compléter le calendrier ci-dessous. Il vous aidera lors de l'entretien téléphonique. L'enquêteur disposera sur écran d'un calendrier vierge qu'il remplira avec vous oralement.

	2013												2014												2015					
	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin			
En intérim																														
En entreprise																														
En recherche d'emploi																														
En reprise d'études à temps plein dans un établissement scolaire																														
En formation autre qu'une reprise d'études																														
Autres situations																														

On vous demandera également de nous indiquer votre salaire au premier mois et au dernier mois de vos/votre contrat(s) de travail.

### Exemple de parcours pour un jeune réinterrogé en 2013 :

Un jeune a continué son emploi dans l'entreprise PLEXO. Il a recherché un emploi de septembre 2013 à janvier 2014. Après 3 mois passés en formation, il travaille depuis mai 2014 dans l'entreprise SOMATEC.

	2013												2014												2015					
	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin			
En intérim																														
En entreprise																														
En recherche d'emploi																														
En reprise d'études à 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 4 - CAWI de fidélisation

### a) Écran d'accueil

 MINISTÈRE DU TRAVAIL, DE L'EMPLOI, DE LA FORMATION PROFESSIONNELLE ET DU DIALOGUE SOCIAL	 * CENTRE D'ETUDES ET DE RECHERCHES SUR LES QUALIFICATIONS	 MINISTÈRE DE L'ÉDUCATION NATIONALE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
---	---	--

**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 ou le courrier de contact.

**ATTENTION :** lors de la saisie de votre login et votre mot de passe, veillez à bien respecter la casse (majuscules/minuscules)

Si vous rencontrez un problème technique pour vous connecter vous pouvez nous contacter à l'adresse suivante : [generation2010@cereq.fr](mailto:generation2010@cereq.fr).

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 [cil@cereq.fr](mailto:cil@cereq.fr). Vous pouvez également, pour

### b) Authentification avec le mois de naissance et l'année de naissance

 MINISTÈRE DU TRAVAIL, DE L'EMPLOI, DE LA FORMATION PROFESSIONNELLE ET DU DIALOGUE SOCIAL	 * CENTRE D'ETUDES ET DE RECHERCHES SUR LES QUALIFICATIONS	 MINISTÈRE DE L'ÉDUCATION NATIONALE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
---	---	---

**Merci,**

Pour des raisons de sécurité, veuillez entrer votre mois et année de naissance :

Mois (format MM) :

Année (format AA) :

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 [cil@cereq.fr](mailto:cil@cereq.fr). Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant.

© 2014 Céreq

c) *Affichage des coordonnées*

Déconnexion

**Bonjour,**

Vous avez répondu à notre enquête sur l'insertion professionnelle des jeunes au printemps 2013.  
Nous allons vous recontacter au printemps 2015 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 :

Nom :	PAPOT
Nom marital :	
Prénom :	BASTIEN
Complément d'identification :	
Complément de voie :	
N°, type et libellé de voie :	12 RUE GASTON BOCQUILLON
Complément de distribution :	
Ville :	JUVISY SUR ORGE
Code postal :	91260
Pays (si étranger) :	
Courriel :	
Numéro de fixe :	
Numéro de mobile :	0621360578
Autre numéro (si étranger) :	

Valider mes coordonnées

Modifier mes coordonnées

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 [oil@cereq.fr](mailto:oil@cereq.fr). Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant.

© 2014 Céreq

d) *Après validation des coordonnées, page de prise de rendez-vous*

Plan de dialogue :

Merci d'avoir mis à jour vos coordonnées !

Merci d'indiquer le meilleur moment pour vous joindre par téléphone dans les quinze prochains jours :

Le : (Réponses multiples, tranches de 2h)

- Lundi → Tranche horaire :  (menu déroulant de 9h -20h30)
- Mardi → Tranche horaire :  (menu déroulant de 9h -20h30)
- Mercredi → Tranche horaire :  (menu déroulant de 9h -20h30)
- Jeudi → Tranche horaire :  heures (menu déroulant de 9h -20h30)
- Vendredi → Tranche horaire :  (menu déroulant de 9h -20h30)



- Samedi → Tranche horaire : menu déroulant (menu déroulant de 9h -20h30)
- Pas de préférence
- N'est pas disponible dans les 15 prochains jours

e) *Après validation*



MINISTÈRE  
DE L'ÉDUCATION NATIONALE,  
DE L'ENSEIGNEMENT SUPÉRIEUR,  
DE LA FORMATION  
PROFESSIONNELLE  
ET DE LA RECHERCHE



**Céreq** • CENTRE D'ÉTUDES ET DE RECHERCHES SUR LES QUALIFICATIONS



MINISTÈRE  
DE L'ÉDUCATION NATIONALE,  
DE L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE

[Déconnexion](#)

**Merci d'avoir mis à jour vos coordonnées !**

Nous vous recontacterons au printemps 2015.

Si vous souhaitez consulter le site du Céreq cliquez sur le lien suivant : <http://www.cereq.fr/>

Si vous souhaitez consulter les premiers résultats de l'enquête, vous pouvez les télécharger ici :



L'ouvrage complet sur votre génération est disponible ici :



Si vous souhaitez réagir, une page facebook est dédiée à notre enquête :



Si vous souhaitez nous contacter pour toute information ou question, merci d'envoyer un message à l'adresse suivante [generation2010@cereq.fr](mailto:generation2010@cereq.fr).

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 [cl@cereq.fr](mailto:cl@cereq.fr). Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant.

© 2014 Céreq

**ANNEXE 5 - 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		36	Captage, traitement et distribution d'eau

	PRODUCTION ET DISTRIBUTION D'EAU ; ASSAINISSEMENT, GESTION DES DÉCHETS ET DÉPOLLUTION	37 38 39	Collecte et traitement des eaux usées Collecte, traitement et élimination des déchets ; récupération Dépollution et autres services de gestion des déchets
F	CONSTRUCTION	41 42 43	Construction de bâtiments Génie civil Travaux de construction spécialisés
G	COMMERCE ; RÉPARATION D'AUTOMOBILES ET DE MOTOCYCLES	45 46 47	Commerce et réparation d'automobiles et de motocycles Commerce de gros, à l'exception des automobiles et des motocycles Commerce de détail, à l'exception des automobiles et des motocycles
H	TRANSPORTS ET ENTREPOSAGE	49 50 51 52 53	Transports terrestres et transport par conduites Transports par eau Transports aériens Entreposage et services auxiliaires des transports Activités de poste et de courrier
I	HÉBERGEMENT ET RESTAURATION	55 56	Hébergement Restauration
J	INFORMATION ET COMMUNICATION	58 59 60 61 62 63	Édition Production de films cinématographiques, de vidéo et de programmes de télévision ; enregistrement sonore et édition musicale Programmation et diffusion Télécommunications Programmation, conseil et autres activités informatiques Services d'information
K	ACTIVITÉS FINANCIÈRES ET D'ASSURANCE	64 65 66	Activités des services financiers, hors assurance et caisses de retraite Assurance Activités auxiliaires de services financiers et d'assurance
L	ACTIVITÉS IMMOBILIÈRES	68	Activités immobilières
M	ACTIVITÉS SPÉCIALISÉES, SCIENTIFIQUES ET TECHNIQUES	69 70 71 72 73 74 75	Activités juridiques et comptables Activités des sièges sociaux ; conseil de gestion Activités d'architecture et d'ingénierie ; activités de contrôle et analyses techniques Recherche-développement scientifique Publicité et études de marché Autres activités spécialisées, scientifiques et techniques Activités vétérinaires
N	ACTIVITÉS DE SERVICES ADMINISTRATIFS ET DE SOUTIEN	77 78	Activités de location et location-bail 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
		82	Activités administratives et autres activités de soutien aux entreprises
O	ADMINISTRATION PUBLIQUE	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 6 - Définition des variables annexes 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 7 - 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 ?	Séquences en entreprise
O_EP83	Vous avez démissionné, principalement parce que vous souhaitez	Séquences en entreprise
O_NEP5C	Pas de réponse à la dernière offre d'emploi proposée : Pourquoi ?	LES SÉQUENCES DE NON-EMPLOI
O_NEP5D	Personne non disponible : Pour quelle raison ?	LES SÉQUENCES DE NON-EMPLOI
O_DIS31	Victime de discrimination : raison ?	Discrimination
O_DIS41	Victime de discrimination : auteur ?	Discrimination
O_DIS51	Victime de discrimination : forme ?	Discrimination
O_EA0B	Emploi pas en rapport avec formation initiale : Pour quelle(s) raison(s) ?	Opinions sur l'emploi
O_ETA2	Aujourd'hui, vous préparez ?	Séquence de reprise d'études
O_Q42	Équivalent au bac : Quel est ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q42B	Bac général : Dans quelle série ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q43	Inférieur au bac : Quel est ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q44	Bac +2 : Quel est ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q45	Bac +3/4 : Quel est ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q46	Bac +5 : Quel est ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q49	En quelle année avez-vous obtenu ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q52	En quelle année avez-vous commencé cette formation pour obtenir ce diplôme ?	Acquisition de diplôme(s) ou titre(s) professionnel(s) post-initiaux
O_Q74	Équivalent au bac : Quel est ce diplôme ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q74B	Bac général : Dans quelle série ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q74C	Bac Technologique : Dans quelle série ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q75	Inférieur au bac : Quel est ce diplôme ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q76	Bac +2 : Quel est ce diplôme ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q77	Bac +3/4 : Quel est ce diplôme ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q78	Bac +5 : Quel est ce diplôme ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q81	En quelle année avez-vous commencé cette formation pour préparer ce diplôme ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_Q104	Quelle est la raison principale pour laquelle vous n'avez pas suivi toute la formation ?	Tentative d'obtention de diplôme(s) ou titre(s) professionnel(s)
O_OP6B	Pour quelle raison êtes-vous inquiet pour votre avenir professionnel ?	Les perspectives professionnelles

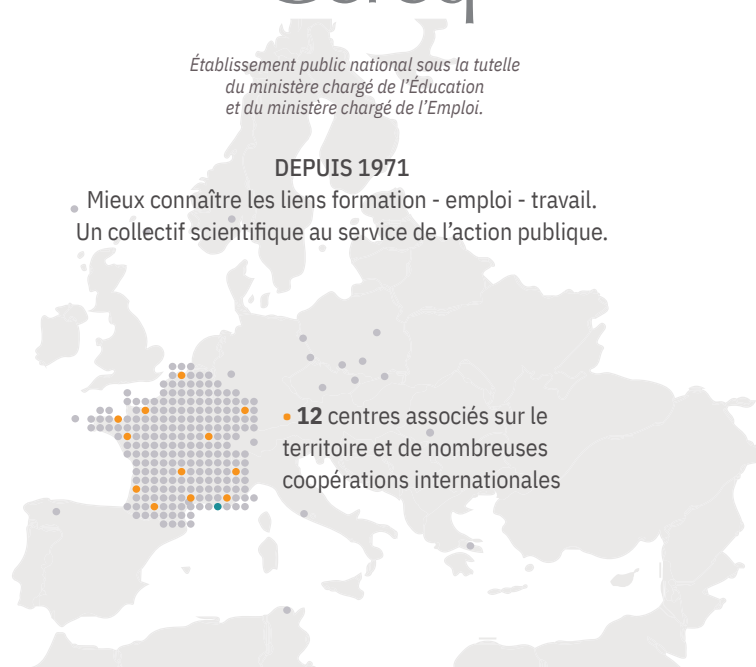


# Céreq

*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](http://www.cereq.fr)



 **+ de 600 publications**  
Accessibles librement