



HAL
open science

Rapprochement exploratoire de l'enquête Génération 2017 avec des données administratives issues du système d'information sur le suivi de l'étudiant (SISE)

Stéphane Jugnot

► To cite this version:

Stéphane Jugnot. Rapprochement exploratoire de l'enquête Génération 2017 avec des données administratives issues du système d'information sur le suivi de l'étudiant (SISE). 2024. halshs-04499960

HAL Id: halshs-04499960

<https://shs.hal.science/halshs-04499960>

Preprint submitted on 11 Mar 2024

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

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



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

Céreq WORKING PAPER

24
2024

Rapprochement
exploratoire
de l'enquête
Génération 2017
avec des données
administratives issues
du système d'information
sur le suivi de l'étudiant
(SISE)

DOCUMENT DE TRAVAIL

ENQUÊTE GÉNÉRATION 2017
MÉTHODOLOGIE D'ENQUÊTE

INSERTION
PROFESSIONNELLE

STATISTIQUE DE
L'ÉDUCATION

ENQUÊTE

COMPARAISON

Stéphane JUGNOT
Céreq > Direction scientifique

Sommaire

Synthèse	6
Les données du SIES mobilisées pour le rapprochement expérimental	7
Des choix qu'il faudra questionner en cas de renouvellement de l'exercice	8
Des résultats globalement satisfaisants, conditionnellement à la capacité à disposer de l'INE des répondants à l'enquête.....	9
En guise de conclusion	11
1. Présentation succincte des données mobilisées	13
1.1. L'enquête Génération 2017 (Céreq)	13
1.2. Les fichiers annuels SISE-Inscriptions (SIES).....	16
1.3. Les fichiers annuels BPBAC (SIES)	19
1.4. Les fichiers annuels SISE-Résultats (SIES)	20
1.5. Quelques statistiques descriptives sur les fichiers annuels des inscriptions	20
2. La réalisation de l'appariement expérimental	23
2.1. L'INE comme clef d'appariement	23
2.2. Un cadre de travail sécurisé strict.....	27
2.3. Première étape : la construction d'un fichier historique des parcours dans SISE-Inscriptions	28
2.3.1. Création des fichiers globaux annuels d'inscription	29
2.3.2. Création d'un fichier historique de parcours avec les données administratives	32
2.4. Deuxième étape : l'appariement avec l'échantillon de l'enquête Génération 2017	40
2.5. Troisième étape : l'appariement avec les fichiers de SISE-Résultats	48
3. Les parcours dans SISE des jeunes de la Génération 2017	50
3.1. Les parcours d'inscription	50
3.2. Les parcours de diplomation dans le supérieur	62
3.3. Le plus haut diplôme identifié	65

Annexes

Annexe 1. Fiche programme	67
Annexe 2. Extraits du questionnaire de la première interrogation de l'enquête Génération 2017	68
Annexe 3. Les informations de SISE pouvant intéresser le Céreq	77
SISE-Inscriptions.....	78
SISE-Diplomation.....	81
Annexe 4. Quelques statistiques descriptives de suivi des parcours dans les fichiers SISE- Inscriptions mobilisés	82
Annexe 5. Programmes SAS	94
Partie A. Création d'une table unique d'inscrits par année (INSCRITSaa, aa=2010 à 2017) et exploitations exploratoires.....	94
Partie B. Création d'une table de trajectoires 2010 à 2017 (table HISTORIQUE SISE) et exploitations exploratoires.....	103
Partie C. Rapprochement de la table HISTORIQUE SISE et de la table de l'échantillon de l'enquête Génération (table SISE_ECHANT) et exploitations exploratoires sur la capacité à faire ce rapprochement	115
Partie D. Rapprochement des répondant l'enquête et des informations issues de SISE-Inscriptions (appariement du fichier « Progedo » Génération 2017 et HISTORIQUE SISE).....	125
Partie E. Rapprochement des répondant l'enquête et des informations issues de SISE-Résultats	137

Liste des encadrés et tableaux

Encadrés

Encadré 1 • Le champ de l'enquête Génération 2017.....	13
Encadré 2 • Construire la base de sondage de l'enquête Génération 2017	24

Tableaux

Tableau 1 • Nature des informations collectées dans l'enquête Génération 2017 sur le parcours dans l'enseignement supérieur	15
Tableau 2 • Répartition des répondants ayant obtenu un diplôme de l'enseignement selon la partie du questionnaire ayant permis d'identifier leur plus haut diplôme.	15

Tableau 3 • Estimation du champ couvert par les données administratives mobilisées dans la présente étude (sur la base des inscrits pour l'année universitaire 2018-2019)	17
Tableau 4 • Nombre d'observations dans les fichiers annuels BPBAC, selon le type de « module élémentaire de formation »	19
Tableau 5 • Nombre d'observations dans SISE-Inscriptions, selon la composante et l'année	20
Tableau 6 • Proportion d'observations repérées comme étant en doubles comptes selon la composante et l'année de SISE-Inscriptions	21
Tableau 7 • Proportion d'observations "principales" selon la composante et l'année de SISE-Inscriptions ..	21
Tableau 8 • Proportion d'observations repérées comme des ré-immatriculations abusives selon la composante et l'année de SISE-Inscriptions	21
Tableau 9 • Proportion d'observations pour lesquelles l'année de première inscription dans le supérieur est disponible selon la composante et l'année de SISE-Inscriptions	22
Tableau 10 • Proportion d'observations pour lesquelles l'année du baccalauréat est renseignée selon la composante et l'année de SISE-Inscriptions	22
Tableau 11 • Qualité de l'INE disponible dans l'échantillon de l'enquête Génération, selon la composante de la base de sondage	25
Tableau 12a • Répartition des répondants de l'enquête Génération par plus haut diplôme atteint et groupes de qualité de l'INE des composantes de la base de sondage, en nombre	26
Tableau 12b • Répartition des répondants de l'enquête Génération par plus haut diplôme atteint et groupes de qualité de l'INE des composantes de la base de sondage, en %	26
Tableau 13 • Application de l'indicateur de qualité de l'INE sur les fichiers SISE et BPBAC mobilisés	27
Tableau 14 • Répartition des étudiants (INE), selon leur nombre d'inscriptions repérées en compilant toutes les composantes, par année	30
Tableau 15 • Nombre d'étudiants (INE) sans aucune inscription principale et nombre d'étudiants (INE) ayant au moins deux inscriptions principales, par année	30
Tableau 16 • Répartition des étudiants (INE) ayant au moins une inscription principale à l'université (composante UNIV), selon leur nombre total d'inscriptions principales.....	31
Tableau 17 • Répartition des étudiants (INE) ayant au moins une inscription principale dans une école (composantes INGE et MANA), selon leur nombre total d'inscriptions principales	31
Tableau 18 • Répartition des étudiants (INE) ayant au moins une inscription principale dans une E.N.S., selon leur nombre total d'inscriptions principales	32
Tableau 19 • Cohérence des informations d'état-civil (sexe, année et mois de naissance) en cas d'inscriptions multiples dans SISE	33
Tableau 20 • Cohérence de l'année du bac en cas d'inscriptions multiples dans SISE	34
Tableau 21 • Critères mobilisés pour la sélection de l'observation privilégiée pour le suivi des trajectoires .	36

Tableau 22 • Répartition des étudiants (INE) selon le nombre d'années de présence dans SISE-Inscriptions, par degré d'étude et composante d'inscription en 2016	38
Tableau 23 • Années d'absence dans les fichiers de SISE-Inscriptions 2010 à 2016 des étudiants repérés en 2016, par degré d'étude et composante d'inscription en 2016	39
Tableau 24 • Capacité à apparier l'échantillon de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, par composante de la base de sondage de Génération	41
Tableau 25 • Capacité à apparier la table des répondants de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, par composante de la base de sondage de Génération	42
Tableau 26 • Capacité à apparier la table des répondants de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le niveau de sortie (NIVSOR)	43
Tableau 27 • Capacité à apparier la table des répondants de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le plus haut diplôme atteint (PHDIP)	44
Tableau 28 • Données pondérées - Capacité à apparier la Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le niveau de sortie (NIVSOR)	45
Tableau 29 • Données pondérées - Capacité à apparier la Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le plus haut diplôme atteint (PHDIP)	46
Tableau 30 • Proportion de répondants retrouvés dans chacun des fichiers annuels SISE, selon le plus haut diplôme atteint	47
Tableau 31 • Nombre de répondants à l'enquête Génération 2017 pour lesquels un diplôme est récupéré dans les fichiers SISE-Résultats de l'année considérée	49
Tableau 32 • Codification synthétique des diplômes	49
Tableau 33 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un DOCTORAT	53
Tableau 34 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un DIPLÔME D'ÉCOLE (commerce ou ingénieur)	54
Tableau 35 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un DIPLÔME BAC+5 hors diplôme d'école.....	55
Tableau 36 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme une LICENCE PROFESSIONNELLE	57
Tableau 37 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un DIPLÔME BAC+3 (hors licence professionnelle).....	58
Tableau 38 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un DIPLÔME BAC+2 hors santé/social.....	59
Tableau 39 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un BACCALAURÉAT GÉNÉRAL.....	60
Tableau 40 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un BACCALAURÉAT TECHNOLOGIQUE	61

Tableau 41 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un BACCALAURÉAT PROFESSIONNEL	62
Tableau 42 • Nombre de répondants à l'enquête Génération 2017 ayant des diplômes repérés* dans les fichiers SISE-Résultats, selon leur plus haut diplôme atteint dans l'enquête	63
Tableau 43 • Proportion des répondants à l'enquête Génération 2017 repérés dans SISE-Résultats comme ayant eu certains diplômes, selon leur plus haut diplôme atteint à l'enquête	65
Tableau 44 • Répartition des répondants à l'enquête Génération 2017 selon leur plus haut diplôme dans l'enquête et leur plus haut diplôme repéré dans SISE-Résultats	66
Tableau 45 • Répartition de la Génération 2017 selon leur plus haut diplôme dans l'enquête et leur plus haut diplôme repéré dans SISE-Résultats (données pondérées)	66

Tableaux de l'annexe 4

Tableau A4.1. • Où les inscrits de 2016 sont-ils inscrits en 2015, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?	82
Tableau A4.2. • Où les inscrits de 2016 sont-ils inscrits en 2014, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?	84
Tableau A4.3. • Où les inscrits de 2016 sont-ils inscrits en 2013, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?	86
Tableau A4.4. • Où les inscrits de 2016 sont-ils inscrits en 2012, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?	88
Tableau A4.5. • Où les inscrits de 2016 sont-ils inscrits en 2011, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?	90
Tableau A4.6. • Où les inscrits de 2016 sont-ils inscrits en 2010, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?	92

Rapprochement exploratoire de l'enquête Génération 2017 avec des données administratives issues du système d'information sur le suivi de l'étudiant (SISE)

Stéphane Jugnot*

Synthèse

Depuis le début des années 1990, le dispositif d'enquêtes « Génération » permet au Centre d'études et de recherches sur les qualifications (Céreq) de documenter régulièrement les premières années de vie active des jeunes après leur sortie de formation initiale. Ces enquêtes abordent le parcours scolaire et les diplômes obtenus. Elles recueillent d'autres informations susceptibles de jouer sur le lien entre formation et emploi. Un calendrier mensuel d'activité collecte des informations sur la situation professionnelle mois par mois depuis la fin des études pour étudier les conditions d'accès à l'emploi. Tous les jeunes sortis du système éducatif, des jeunes arrêtant leurs études au niveau du collège sans aucun diplôme, jusqu'aux docteurs, sont interrogés de façon homogène. Comme tous les enquêtés sont sortis de formation initiale au cours ou à l'issue d'une même année scolaire, il est facile de comparer les conditions d'accès au marché du travail en fonction du niveau et de la nature de la formation initiale suivie : niveau de diplôme, spécialités, passage par une filière professionnelle, sortie d'apprentissage, etc. Les enquêtes Génération sont la seule source labellisée « statistique publique » qui le permettent.

Ces dernières années, le contexte juridique et institutionnel a fortement évolué offrant de nouvelles opportunités pour mobiliser davantage les données d'origine administrative et réduire les coûts de collecte des enquêtes¹. Le Céreq a donc initié une série de rapprochements expérimentaux pour voir dans quelle mesure ces opportunités pouvaient permettre de faire évoluer les enquêtes Génération. Ce travail a déjà été réalisé avec les données de Pôle emploi² et avec les données relatives au parcours dans l'enseignement secondaire produites par la direction de l'évaluation, de la prospective et de la performance (DEPP)³.

Le présent travail poursuit ces travaux avec les données d'origine administrative sur le parcours dans l'enseignement supérieur qui sont produites par la sous-direction des systèmes d'information et d'études statistiques (SIES), le service statistique du ministère chargé de l'enseignement supérieur et de la recherche. À cet effet, l'enquête Génération 2017, collectée en 2020⁴ auprès d'un échantillon de jeunes sortis de formation initiale au cours ou à l'issue de l'année 2016-2017, est rapprochée des données annuelles du SIES relatives aux années 2006 à 2016.

* Céreq > Projets appariements (N° ORCID : 0000-0001-6819-7012).

¹ Jugnot Stéphane, « Les opportunités créées par la facilitation et le développement des utilisations des données administratives à des fins de recherche et de production statistique », Céreq, Working paper, n° 9, septembre 2021, 66p.

² Jugnot Stéphane, « Analyse du rapprochement expérimental entre le fichier historique des demandeurs d'emploi et l'enquête Génération 2004 », Céreq, Working paper, n° 11, février 2022, 90p.

³ Jugnot Stéphane, « Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants 2005 à 2016 de la DEPP : des résultats prometteurs », Céreq, Working paper, n° 19, avril 2023, 144p.

⁴ La crise sanitaire de la Covid 19 a conduit à repousser la collecte du printemps 2020 à la fin de l'année 2020. Une minorité de répondants n'a été jointe que début 2021.

Les données du SIES mobilisées pour le rapprochement expérimental

Comme les autres données d'origine administrative mobilisées dans les précédents travaux expérimentaux, les données du SIES ne sont pas des données administratives brutes, mais des données ayant déjà fait l'objet d'un certain nombre de traitements et de contrôles dans le but de permettre une utilisation statistique. Contrairement aux données de la DEPP sur l'enseignement secondaire, elles ne résultent pas d'un système d'information centralisé couvrant l'ensemble de l'enseignement supérieur. Pour les réunir, le SIES doit procéder chaque année à une collecte d'informations auprès des établissements concernés pour alimenter son système d'Information sur le suivi de l'étudiant (SISE). Ces collectes portent, d'une part, sur les étudiants inscrits (« SISE-Inscriptions ») et, d'autre part, sur les résultats aux diplômes (« SISE-Résultats »).

Ces remontées de données individuelles ne couvrent pas l'ensemble des établissements de l'enseignement supérieur, mais le SIES étend progressivement le champ couvert. Les universités constituent la première composante à avoir été collectée, à partir de 1995. Les établissements publics et privés habilités à délivrer un titre d'ingénieur ont suivi en 2001 puis les écoles normales supérieures en 2004, les écoles de management en 2006, les écoles vétérinaires (depuis 2009), etc. Au total, sur la base des éléments présentés en mars 2020 par le SIES devant le Conseil national de l'information statistique (tableau 3), les composantes mobilisées par la présente étude couvriraient un peu moins de 90 % des effectifs inscrits dans l'enseignement supérieur et près des deux tiers des établissements. Les établissements dépendants du ministère de la Culture ont été intégrés dans le dispositif en 2017, après notre période d'intérêt. À cette date, près de 7 % des effectifs restaient non couverts, notamment en raison de l'absence des établissements de la sphère médicosociale relevant du ministère chargé des affaires sociales. Au-delà, il restera toujours une petite proportion d'effectifs non couverts : certains établissements des composantes couvertes ne remontent pas leurs effectifs (le taux de réponse des établissements oscillerait entre 90 % et 100 % selon les composantes). Par ailleurs, chaque année, de nouveaux établissements privés sont créés. Ils n'intègrent pas immédiatement le dispositif de remontées de données individuelles.

Le SIES ne produit pas de fichier global annuel ni pour les inscriptions ni pour les résultats, mais une collection de fichiers construits par composante : les universités, les écoles d'ingénieurs, les écoles de management, etc. **Reconstruire des parcours à partir des données administratives nécessite donc d'apparier les différents fichiers du SIES.** Pour ce travail, nous avons d'abord produit un fichier historicisé de SISE-Inscriptions, en construisant d'abord des fichiers globaux annuels puis en appariant ces fichiers globaux annuels entre eux. Nous avons ensuite procédé de même pour produire un fichier historicisé de SISE-Résultats.

Ces rapprochements sont réalisés en utilisant l'INE comme variable d'identification de l'individu. Cet identifiant national de l'élève et de l'étudiant est attribué et géré grâce à un répertoire national d'identification géré par la DEPP. Créé en 1995, cet identifiant joue, dans la sphère de l'éducation, le même rôle que le « numéro de sécurité sociale » dans la sphère de l'emploi, la santé et la Sécurité sociale. Il permet de fiabiliser l'identification des personnes dans les outils de gestion et facilite ensuite la réutilisation à des fins statistiques des données administratives. D'emblée, l'un des objectifs de l'INE était ainsi de faciliter le suivi des parcours de formation initiale. Pour différentes raisons, cet objectif n'a pas été pleinement atteint, un même élève pouvant se voir attribuer plusieurs INE successifs, notamment à l'occasion de changements d'académie dans l'enseignement secondaire ou lors d'inscriptions dans certains établissements de l'enseignement supérieur⁵. **De nouveaux outils ont été mis en œuvre par la DEPP et le SIES pour que les établissements utilisent plus systématiquement un INE certifié. Leur déploiement a commencé en 2017. Les données mobilisées dans la présente étude ne bénéficient donc pas de l'amélioration attendue de la qualité de l'INE disponible dans les fichiers de gestion des établissements. Théoriquement, le suivi des parcours à partir des données administratives devrait être de meilleure qualité à l'avenir que celui constaté dans la présente étude.**

⁵ Plus précisément, l'organisation mise en place devait permettre d'assurer la continuité de l'INE, mais des facilités de gestion pouvaient conduire ponctuellement à ne pas tenir compte de cet impératif. Pour en savoir plus, voir la partie 2.1 de (Jugnot Stéphane, « Les opportunités créées par la facilitation et le développement des utilisations des données administratives à des fins de recherche et de production statistique », Céreq, Working paper, n° 9, septembre 2021, 66 p).

Les données mobilisées sont des données individuelles au sens du règlement général sur la protection des données. Les exploitations ont été réalisées dans le cadre technique sécurisé strict déployé par le Centre d'accès sécurisé aux données (CASD)⁶. Si son infrastructure a globalement donné satisfaction, les conditions de mise en œuvre du respect des règles du secret statistique nous semblent perfectibles. Les critères d'effectifs minimums « par case » (5 individus) ne devraient pas être appliqués indifféremment que les tabulations portent sur les données exhaustives ou qu'elles portent sur un faible échantillon tiré aléatoirement parmi l'ensemble des inscrits, tel que les répondants à l'enquête Génération.

Des choix qu'il faudra questionner en cas de renouvellement de l'exercice

Les rapprochements des fichiers SISE pour construire les fichiers historicisés et leur rapprochement avec les observations de l'enquête Génération supposent de faire des choix, qui ont une part de subjectivité. Ils peuvent être questionnés et devront l'être en cas de renouvellement de l'exercice. En particulier :

- Le SIES identifie certaines observations comme des **doubles comptes** . Nous avons fait le choix d'exclure ces observations qui représentent presque 10 % des inscriptions de la composante « Universités » en 2016.
- Un étudiant peut avoir **plusieurs inscriptions** . Il est alors présent plusieurs fois dans les fichiers SISE d'une même année. En compilant toutes les composantes d'une année donnée et en excluant les observations en doubles comptes, entre 6 % et 8 % des étudiants (identifiés par leur INE) ont deux inscriptions ou plus sur la période étudiée. Quand l'étudiant dispose de plusieurs inscriptions dans le même établissement, le SIES qualifie l'une d'inscription « principale » et les autres de « secondaires », sur la base de différents critères. En 2016, par exemple, 4,9 % des inscriptions de la composante « Universités » sont qualifiées de secondaires. Cette proportion est de 8,2 % pour la composante « ENS ». Elle est inférieure à 1 % pour les autres composantes. Au total, dans notre base de travail, le nombre d'étudiants ne disposant que d'inscriptions secondaires est marginal (au plus, 0,3 % sur la période étudiée). Les cas d'inscriptions principales multiples sont plus significatifs même s'ils restent peu nombreux (environ 4 % des étudiants).

De ce fait, pour observer des parcours d'inscription ou des passages par des types de formation donnés, il faut choisir les inscriptions à prendre en compte. Pour observer les passages par certains types de formation, nous avons fait le choix de ne conserver au plus que deux inscriptions principales et deux inscriptions secondaires. Pour les rares étudiants ayant trois inscriptions de même nature (principales ou secondaires), nous avons retenu les deux premières apparaissant dans le fichier par souci de simplicité, mais le choix aurait pu être plus raisonné. Pour définir l'inscription mise en avant dans le suivi des trajectoires parmi les, au plus, quatre inscriptions retenues, nous avons priorisé les inscriptions principales sur les inscriptions secondaires puis défini une hiérarchie des formations. Là encore, ces choix mériteront d'être rediscutés en cas de renouvellement de l'exercice.

- Pour les observations associées à un même INE, une même année et d'une année sur l'autre, quelques informations d'état civil sont disponibles, ainsi que l'année du baccalauréat. L'examen très partiel réalisé montre l'existence de cas de **divergences internes à SISE entre les informations disponibles pour un même INE** . Pour des raisons de temps et de difficulté à définir des critères de vraisemblance, nous avons fait le choix de ne pas tenir compte de ces divergences pour réaliser l'appariement, d'autant que ces cas sont peu nombreux lorsqu'on se limite aux répondants de l'enquête Génération.

⁶ <https://www.casd.eu/>

- Lors de la constitution de la base de sondage de l'enquête Génération, **plusieurs INE peuvent avoir été récupérés pour un même répondant de l'enquête Génération**. L'essentiel des cas concerne les sortants du secondaire. Au total, seuls 44 individus de l'échantillon du Céreq sont associés à deux observations (INE) du fichier historicisé construit à partir de SISE-Inscriptions. Le plus souvent, les observations relèvent d'années d'inscription différentes. Il n'y a qu'un seul cas de présence une même année avec deux INE différents. Par simplicité, nous avons fait le choix pour ces personnes de prendre en compte les informations associées aux deux INE.
- Globalement, quand la comparaison est possible, il y a le plus souvent une **cohérence externe entre SISE et Génération pour les quelques informations caractérisant l'étudiant qui sont disponibles** (sexe, mois et année de naissance, année du baccalauréat) mais quelques cas de divergence existent aussi. Nous avons fait le choix de ne pas tenir compte de ces divergences pour effectuer l'appariement.
- **Il est nécessaire de construire des variables synthétiques pour exploiter les informations de parcours et de diplomation**, de façon à disposer d'une variable agrégée sur les formations d'une part, et sur les diplômes, d'autre part. Pour identifier les formations dans lesquelles les étudiants sont inscrits, la variable synthétique que nous avons construite se limite à combiner l'identification de la composante de SISE-Inscriptions (université, écoles d'ingénieurs, enquête 26b, etc.) et le degré d'études en nombre d'années post-baccalauréat⁷. Pour identifier les diplômes obtenus dans SISE-Résultats, la variable synthétique a été construite à grosse maille à partir de la nomenclature très détaillée disponible dans les bases. En cas de renouvellement de l'exercice, il faudra évidemment questionner ces choix qui, pour se faire de façon totalement rigoureuse, nécessitent de connaître parfaitement les nomenclatures détaillées de base et leurs évolutions dans le temps.
- **Un même étudiant repéré par son INE peut avoir plusieurs diplômes la même année**. Par exemple, 6 805 répondants à l'enquête Génération sont identifiés comme ayant réussi un diplôme du supérieur en 2016 d'après le fichier historicisé de SISE-Résultats, 467 ont réussi deux diplômes la même année et 34 trois diplômes. Nous avons fait le choix d'utiliser toutes les informations disponibles pour définir les diplômes obtenus et le plus haut diplôme obtenu dans le fichier historicisé de SISE-Résultats.

Des résultats globalement satisfaisants, conditionnellement à la capacité à disposer de l'INE des répondants à l'enquête

Sans détailler ici tous les résultats présentés dans les parties 2.5, 2.6 et 3 et dans l'annexe 4, on notera notamment les résultats ci-dessous.

Le Céreq a des difficultés à récupérer un INE valide sur une partie de son champ. C'est une limite majeure pour l'appariement de l'enquête avec les données administratives du SIES. Son impact quantitatif est plus important que les défauts de couverture des données administratives pour suivre les parcours dans l'enseignement supérieur en France.

Plus précisément, si l'on regarde chacune des composantes de la base de sondage de l'enquête Génération :

- tous les jeunes de la composante « Universités » sont retrouvés dans le fichier historicisé de SISE ;
- pour les composantes « Autres établissements » et « DGAFP », la difficulté principale est de disposer de l'INE dans la base de sondage, car, quand l'INE est disponible, les personnes sont la plupart du temps retrouvées dans le fichier historicisé de SISE ;
- pour les personnes relevant des composantes « Culture » et « DREES », les deux difficultés se combinent, celle à disposer de l'INE dans la base de sondage et un défaut de couverture dans SISE ;
- pour les personnes relevant de la composante « CIFRE », l'INE n'est pas récupéré dans la base de sondage. Il n'est donc pas possible de les retrouver dans SISE ;

⁷ Variable DEGETU de SISE-Inscriptions.

- pour les personnes relevant des composantes « SIFA » et « SYSCA », l'INE est disponible dans la base de sondage, mais peu sont retrouvés dans SISE. Ce résultat est logique puisque ces composantes couvrent essentiellement des sortants du secondaire.

En déplaçant le point de vue pour traduire ces constats selon le plus haut diplôme atteint dans l'enquête, à partir du niveau bac+3, plus de 90 % des jeunes sont retrouvés dans le fichier historicisé. Pour ces niveaux, la difficulté à apparier provient principalement des cas d'indisponibilités de l'INE dans la base de sondage. En revanche, le défaut de couverture du fichier historicisé SISE joue aussi pour les niveaux bac+2, dont les BTS, et pour les formations du secteur santé/social.

De façon générale, sans se limiter au champ des répondants de l'enquête Génération, le dispositif SISE permet de suivre le parcours de l'étudiant dans l'enseignement supérieur pour une proportion très importante des étudiants. Le nombre d'inscriptions repérées dans SISE sur la période 2010-2016 et la proportion d'étudiants non retrouvés chaque année est cohérent avec un parcours crédible sur la base du niveau atteint en 2016, même si, pour une proportion peu nombreuse, mais significative, les durées d'observations et les années d'absence ne semblent pas cohérentes avec le degré d'études de l'inscription en 2016-2017. Par exemple, de 14 % à 15 % des étudiants inscrits à l'université ou dans une école de commerce en 5^e année post-baccalauréat en 2016-2017 ne sont pas retrouvés une seule fois dans les fichiers SISE de la période 2010-2016. Quatre facteurs peuvent jouer : la réalisation d'une partie de la scolarité à l'étranger, l'existence de parcours heurtés ou atypiques, des défauts dans le suivi de l'INE, les trous dans le champ couvert par SISE. Nous n'avons pas cherché à faire la part de ces facteurs, une telle analyse fine n'étant d'ailleurs pas forcément réalisable avec les seules informations disponibles. Les évolutions déjà mentionnées sur le champ couvert par SISE et dans la gestion de l'INE devraient réduire l'effet des deux derniers facteurs à l'avenir.

Si l'on s'intéresse maintenant aux répondants de l'enquête Génération 2017 :

- Les sortants ayant un doctorat comme plus haut diplôme sont bien couverts par SISE-Inscriptions et leurs trajectoires semblent plutôt bien suivies sur la période. La profondeur historique retenue de six années avant l'année de sortie ne permet pas de retracer l'ensemble du parcours dans le supérieur pour la plupart d'entre eux. Pour atteindre ce but, un recul significativement plus important serait nécessaire, de l'ordre d'une dizaine d'années supplémentaires.
- Une proportion importante de sortants ayant un diplôme d'école comme plus haut diplôme n'est pas retrouvée dans SISE, pour une part importante, en raison de l'absence d'INE dans la base de sondage de l'enquête Génération (33 % de la cohorte), mais également, pour une part significative, malgré la disponibilité d'un INE dans la base de sondage. Pour ceux qui sont retrouvés dans SISE, la trajectoire synthétique observée semble cohérente avec le plus haut diplôme atteint. La profondeur historique retenue gagnerait à être repoussée d'une ou deux années supplémentaires.
- Les sortants ayant un diplôme bac+5 hors diplômes d'école comme plus haut diplôme sont bien mieux couverts par SISE que les diplômés d'école et d'autant plus quand l'INE est disponible. Toutefois, entre un cinquième et un sixième des jeunes concernés ayant un INE ne sont pas retrouvés dans les années précédant la sortie. Ce constat suggère une difficulté à retracer avec SISE l'intégralité du parcours dans l'enseignement supérieur pour une proportion significative de jeunes ayant atteint ce niveau de diplôme, même lorsque l'on dispose d'un INE. Les scolarités à l'étranger peuvent expliquer une partie de ces trous. La profondeur historique retenue gagnerait à être repoussée de quatre à cinq années supplémentaires.
- Les sortants ayant une licence professionnelle comme plus haut diplôme sont aussi bien couverts par SISE que les bac+5 hors diplômes d'école l'année de sortie, mais la capacité à retracer l'intégralité du parcours dans l'enseignement supérieur semble moins bonne. La profondeur historique retenue semble suffisante pour ce plus haut niveau de diplôme, même si une année supplémentaire pourrait apporter une légère amélioration.

- Les sortants ayant un diplôme bac+3 comme plus haut diplôme (hors licence professionnelle et hors secteur santé/social) sont aussi bien couverts par SISE que les bac+5 hors diplôme d'école l'année de sortie, avec une capacité à retracer l'intégralité du parcours dans l'enseignement supérieur qui semble analogue, donc meilleur que pour les licences professionnelles. La profondeur historique retenue n'est pas suffisante pour ce plus haut niveau de diplôme. Elle gagnerait à être complétée de quelques années.
- Les sortants ayant un diplôme bac+2 comme plus haut diplôme (hors secteur santé/social) sont largement couverts par SISE mais une part significative de ceux ayant un INE restent non couverts. La profondeur historique retenue est suffisante pour suivre le parcours dans l'enseignement supérieur.
- Pour les répondants ayant un baccalauréat comme plus haut diplôme, le nombre de ceux identifiés dans SISE-Résultats comme ayant réussi un diplôme supérieur est également non nul, mais négligeable.

Par ailleurs, lorsque les répondants à l'enquête Génération 2017 ont pu être retrouvés dans SISE, les deux sources divergent en partie sur le plus haut niveau de diplôme atteint. Ces divergences sont non nulles, mais marginales pour ceux qui atteignent au mieux la licence professionnelle, un bac+5 ou une thèse. Pour les autres diplômes de bac+3 à bac+4 et pour les bac+2, la proportion d'individus retrouvés dans SISE-Inscriptions, mais pas dans SISE-Résultats est plus significative, ce qui amoindrit la capacité à leur attribuer un plus haut diplôme avec SISE. L'importance de cet écart devrait être réévaluée quand la couverture de SISE se sera améliorée et l'INE certifié sera en œuvre depuis suffisamment longtemps.

En guise de conclusion

Travailler sur les parcours d'inscription et de diplomation dans l'enseignement supérieur à partir des fichiers d'origine administrative produits par le SIES nécessite de rapprocher de nombreuses tables, sur une grande profondeur historique. Ces rapprochements nécessitent des choix sur les informations à privilégier, en particulier en cas de présences multiples ou en cas de divergences pour un même INE de certaines caractéristiques d'identification (sexe, mois ou année de naissance, année du baccalauréat). Construire des catégorisations synthétiques des diplômes et des formations à partir des nomenclatures détaillées proposées est également indispensable pour rendre les données facilement exploitables.

Ces travaux impliquent un investissement significatif pour connaître les référentiels, les circuits de l'information de son origine à son intégration dans les fichiers disponibles. Il faut donc disposer d'un minimum de connaissances sur les modes de gestion des établissements ainsi que sur les traitements de contrôle, de redressements et d'enrichissements effectués par le SIES. Dans ce contexte, le fait qu'un acteur tel que le Céreq produise régulièrement une table historicisée de parcours et de diplomation associée à un échantillon de sortants pourrait être une forme d'investissement mutualisé au profit de la recherche et pourrait être collectivement préférable à des appariements réalisés au coup par coup, en fonction des projets.

Si cette orientation devait être mise en œuvre, il conviendrait d'approfondir la période couverte en l'augmentant d'au moins cinq années. Reconstruire l'ensemble du parcours dans l'enseignement supérieur des titulaires d'un doctorat nécessiterait une profondeur beaucoup plus grande. Ce même constat a été fait pour reconstituer leur parcours dans l'enseignement secondaire à partir des données de la DEPP. Dans ces conditions, pour limiter la charge de travail et le coût informatique, il faudrait peut-être envisager de renoncer à couvrir l'ensemble du parcours pour ce niveau de diplôme.

Pour le dispositif Génération, cette table historicisée permettrait de disposer d'informations de parcours et de diplomations beaucoup plus complètes que ce que le questionnement actuel de l'enquête permet, celui-ci étant principalement destiné à définir et caractériser le plus haut diplôme atteint.

Dans un premier temps, il serait donc intéressant d'associer aux fichiers de résultats de l'enquête Génération une table historicisée de parcours et de diplomation produite à partir des fichiers SISE. Il conviendrait aussi d'améliorer la disponibilité de l'INE pour les répondants à l'enquête lors de la phase de construction de la base de sondage, mais aussi en explorant le cadre organisationnel et juridique qui permettrait de le récupérer à partir des informations d'état civil transitoirement disponibles pendant la période de production de l'enquête. Par ailleurs, les questions visant à caractériser de façon précise le plus haut diplôme atteint pourraient être supprimées pour se limiter aux informations nécessaires pour coder le plus haut diplôme au niveau agrégé le plus utilisé, sa caractérisation plus détaillée étant récupérée à partir de SISE.

Dans un second temps, quand les effets positifs des évolutions du champ de SISE, du nouveau mode de gestion de l'INE et d'une meilleure récupération de l'INE par le Céreq seront attestés, la suppression des questions destinées à définir et préciser le plus haut diplôme atteint pourrait être envisagée. Certaines questions pourraient toutefois rester nécessaires si elles conditionnent ou s'articulent avec d'autres questions telles que, par exemple, les souhaits d'orientation. En effet, la capacité à disposer d'une information par une source externe ne doit pas remettre en cause la fluidité de la passation du questionnaire ni risquer d'introduire des interrogations chez l'enquêté sur les informations dont le service producteur dispose déjà sur lui. En contrepartie de ces suppressions, l'ajout de quelques questions sur les parcours de formation à l'étranger pourrait être envisagé.

1. Présentation succincte des données mobilisées

1.1. L'enquête Génération 2017 (Céreq)

Depuis le début des années 1990, le dispositif d'enquêtes « Génération » a permis au Centre d'études et de recherches sur les qualifications (Céreq) de **documenter régulièrement les premières années de vie active des jeunes après leur sortie de formation initiale**. Ces enquêtes interrogent de façon homogène tous les jeunes sortis du système éducatif, des jeunes arrêtant leurs études au niveau du collège sans aucun diplôme, jusqu'aux docteurs. Comme tous les enquêtés sont sortis de formation initiale au cours ou à l'issue d'une même année scolaire, il est facile de **comparer les conditions d'accès au marché du travail en fonction du niveau et de la nature de la formation initiale suivie** : niveau de diplôme, spécialités, passage par une filière professionnelle, sortie de l'apprentissage, etc. Les enquêtes Génération sont la seule source labellisée « statistique publique » qui le permettent⁸.

Encadré 1 • Le champ de l'enquête Génération 2017

La « Génération 2017 » porte sur les élèves et les étudiants considérés comme des « primo-sortants » de formation initiale entre octobre 2016 et octobre 2017. Tous les niveaux et domaines de formation sont concernés. Plus précisément, pour appartenir au champ de l'enquête, il faut :

- avoir été inscrit dans un établissement de formation en France, en métropole ou dans un département d'outre-mer, durant l'année scolaire 2016-2017 ;
- avoir quitté le système éducatif entre octobre 2016 et octobre 2017 ;
- ne pas avoir interrompu ses études durant dix-sept mois ou plus avant l'année scolaire 2016-2017, sauf pour raison de santé ;
- ne pas avoir repris ses études pendant les seize mois qui ont suivi l'entrée sur le marché du travail ;
- avoir trente-cinq ans ou moins en 2017.

Certains ministères demandent des extensions d'échantillon au Céreq, de manière à couvrir également les sortants qui ne sont pas primo-sortants pour les écoles dont ils ont la responsabilité. Pour l'enquête Génération 2017, des sortants « post-initiaux » ont ainsi été interrogés pour les établissements relevant du ministère chargé des sports. Ces sortants ne sont pas pris en compte dans les exploitations réalisées par le Céreq. Ils ne sont donc pas pris en compte dans nos travaux exploratoires, sauf mention contraire.

Enfin, dans l'enquête Génération 2017, les contrats de professionnalisation à visée diplômante sont traités comme les contrats d'apprentissage lorsqu'ils interviennent dans la continuité du parcours scolaire initial. Dans ce cas, ils sont donc considérés comme relevant de la formation initiale même s'il s'agit juridiquement de contrats de travail. Une personne repérée comme sortant de formation en 2016-2017 qui poursuit par un contrat d'apprentissage ou un contrat de professionnalisation en 2017-2018 est considérée en poursuite d'études, donc hors-champ. En revanche, les jeunes sortants de contrats d'apprentissage ou de contrat de professionnalisation au cours de l'année scolaire 2016/2017 sont dans le champ de l'enquête si ces contrats prolongent un parcours scolaire initial achevé antérieurement.

⁸ Le dispositif « Génération » a été présenté à la commission emploi du Conseil national de l'information statistique (Cnis), le 9 novembre 2017. Le diaporama de présentation et le compte-rendu de la réunion sont disponibles auprès du Cnis. <https://www.cnis.fr/evenements/emploi-qualification-et-revenus-du-travail-2017-2e-reunion-2/>

L'enquête Génération 2017 porte sur les jeunes sortis de formation initiale au cours ou à l'issue de l'année scolaire 2016-2017 (voir encadré 1). La passation du questionnaire, réalisée par internet et par téléphone, était initialement prévue au printemps 2020. La survenue de l'épidémie de Sras-Cov2 a repoussé la collecte en raison du confinement général de la population décidé par le gouvernement afin de lutter contre l'épidémie. La collecte a finalement eu lieu d'août 2020 à mars 2021. Les premiers résultats de l'enquête ont été publiés en mai 2022 (Couppié, Gaubert & Personnaz, 2022)⁹.

Le questionnaire de l'enquête aborde le parcours scolaire et les diplômes obtenus. Ce module permet d'identifier le plus haut diplôme obtenu et de recueillir d'autres informations sur la trajectoire scolaire, susceptibles de jouer sur le lien entre formation et emploi. Un calendrier mensuel d'activité collecte des informations sur la situation professionnelle mois par mois depuis la fin des études intervenue environ trois ans auparavant. D'autres modules abordent des thématiques complémentaires. Pour l'enquête Génération 2017, des questions portent ainsi les séjours à l'étranger, l'attractivité de la fonction publique, les moyens de financement des études, la multiactivité, la satisfaction dans l'emploi occupé à la date d'enquête, le sentiment de discrimination, l'exposition aux risques physiques et chimiques, ainsi que les conséquences sur la situation professionnelle du confinement décrété au printemps 2020.

Dans le questionnaire, le module relatif au parcours scolaire couvre l'enseignement secondaire jusqu'aux souhaits formulés l'année du baccalauréat pour la poursuite d'études, mais il n'aborde pas la poursuite des études au-delà du baccalauréat de façon détaillée. **Le questionnaire permet cependant d'appréhender le parcours des enquêtés dans l'enseignement supérieur à partir de deux autres modules :**

1. Le module relatif au niveau de sortie : ce module permet de s'assurer que la formation dans laquelle l'enquêté était inscrit au cours de l'année scolaire 2016/2017 d'après la base de sondage peut être considérée comme sa formation de sortie de formation initiale. Autrement dit, ce module permet de valider l'appartenance de l'enquêté au champ de l'enquête. Une question permet de savoir si la formation s'est effectuée par apprentissage. Une autre question permet de savoir si l'enquêté a obtenu son diplôme dans le cas où cette année de sortie correspond à un niveau terminal. Pour les sortants du supérieur, le module permet de repérer de façon précise la dernière formation suivie et, s'il y a lieu, l'obtention du diplôme correspondant, sans connaître la mention obtenue.
2. Le module d'identification du plus haut diplôme atteint : ce module permet de recenser les différents diplômes obtenus dans le supérieur, avant l'année de sortie. Plus précisément, l'ensemble des sortants du supérieur, ainsi que les sortants du secondaire ayant déclaré avoir obtenu un diplôme de niveau supérieur au baccalauréat, sont questionnés pour savoir s'ils ont obtenu un BTS, un DUT, un DEUG, un autre diplôme de niveau bac+2, une licence professionnelle, une licence, un autre diplôme de niveau bac+3, une maîtrise, un diplôme d'école de commerce de niveau bac+4, un autre diplôme de niveau bac+4 ; un magistère, un master, un diplôme d'école de commerce, un diplôme d'ingénieur, un autre diplôme de niveau bac+5 ou 6, un doctorat. Le nom du diplôme obtenu et sa spécialité précise ne sont ensuite demandés que pour le plus haut diplôme identifié. À la fin du module, une question de repêchage est posée aux répondants identifiés comme non diplômés pour s'assurer qu'ils n'ont bien aucun diplôme.

Au total, les informations collectées de façon standard sur le parcours dans l'enseignement supérieur sont donc limitées. Les informations les plus complètes concernent la dernière année de formation initiale et le diplôme éventuellement réussi cette année-là, puis le plus haut diplôme atteint (tableau 1). La connaissance du reste du parcours dans l'enseignement supérieur se limite à recenser les types de diplômes obtenus (les modules évoqués sont présentés dans l'annexe 2).

⁹ Couppié, T., Gaubert, G. & Personnaz, E. (2022). « Enquête 2020 auprès de la Génération 2017. Des parcours contrastés, une insertion plus favorable, jusqu'à... ». *Céreq Bref*, n° 422.

Tableau 1 • Nature des informations collectées dans l'enquête Génération 2017 sur le parcours dans l'enseignement supérieur

	Pour la formation suivie...		Pour le diplôme...		
	...l'année de sortie	...les années précédentes l'année de sortie	...obtenu l'année de sortie	... le plus élevé atteint, s'il est obtenu avant l'année de sortie	... obtenu avant l'année de sortie, sans être le plus élevé atteint
Etablissement d'inscription	Disponible	Non disponible			
Formation d'inscription	Disponible	Non disponible			
Réalisation de la formation par apprentissage ou non	Disponible	Non disponible			
Nature du diplôme obtenu			Disponible	Disponible	Disponible
Nom précis du diplôme (nomenclature Insee)			Disponible	Disponible	Non disponible
Spécialité précise du diplôme (NSF)			Disponible	Disponible	Non disponible
Année d'obtention du diplôme			Disponible	Non disponible	Non disponible

Source : auteur, à partir du dictionnaire des codes et du questionnaire de l'enquête Génération 2017.

Pour 85 % des répondants à l'enquête Génération 2017, le plus haut diplôme atteint a été obtenu l'année de sortie de formation initiale (tableau 2). Pour les 15 % ayant obtenu leur plus haut diplôme avant l'année de sortie, on ne connaît pas l'année d'obtention, ni la nature et le lieu de l'établissement d'inscription.

Tableau 2 • Répartition des répondants ayant obtenu un diplôme de l'enseignement selon la partie du questionnaire ayant permis d'identifier leur plus haut diplôme.

Plus haut diplôme atteint	Niveau de sortie		Module du plus haut diplôme atteint	Repêchage à l'issue du module du plus haut diplôme atteint	Total
Bac+2 sauf santé/social	2 370	80 %	565	13	2 948
Bac+2 à +4 santé/social	604	97 %	18	1	623
Licence professionnelle	1 291	88 %	171	1	1 463
Autre bac+3 et bac+4	962	58 %	682	5	1 649
Bac+5 sauf écoles	4 007	90 %	418	6	4 431
École (Ingénieur, commerce)	2 243	95 %	106	1	2 350
Doctorat	482	92 %	43	0	525
Ensemble	11 959	85 %	2 003	27	13 989

Note : données non pondérées.

Source : Céreq, enquête Génération 2017 – interrogation 2020.

Champ : jeunes sortis de formation initiale au cours ou à l'issue de l'année de scolaire 2016-2017 ayant au moins un diplôme de l'enseignement supérieur.

1.2. Les fichiers annuels SISE-Inscriptions (SIES)

Depuis le début des années 2000, le SIES produit des fichiers anonymisés d'élèves pour la recherche et les études. Ils sont élaborés à partir de sources administratives relevant du ministère chargé de l'enseignement supérieur et de collectes de données individuelles réalisées auprès des établissements. Ensemble, ils composent le système d'information sur le suivi de l'étudiant (SISE).

Deux groupes de fichiers SISE sont mobilisés dans ce travail : les fichiers annuels SISE « Inscriptions », relatifs aux étudiants inscrits dans l'enseignement supérieur l'année considérée, et les fichiers annuels SISE « Résultats », relatifs à la diplomation des étudiants l'année considérée. Le SIES produit chaque année une documentation détaillée pour présenter ces fichiers. L'essentiel des éléments qui suivent s'appuie sur cette documentation et plus particulièrement celle produite pour les fichiers annuels 2017.

Pour SISE « Inscriptions », il n'y a pas de fichier annuel unique, mais une collection de fichiers de données individuelles, chacun couvrant une composante particulière de l'enseignement supérieur :

- « Universités » (UNIV) : cette composante couvre l'ensemble des universités et établissements assimilés. Elle est la première composante à avoir été collectée, à partir de 1995.
- « Ingénieurs » (INGE) : cette composante est collectée depuis 2001. Elle couvre les établissements publics et privés habilités à délivrer un titre d'ingénieur, quel que soit le ministère de rattachement. Depuis 2008, 18 écoles d'ingénieurs rattachées aux universités remontent dans ce champ tandis que les écoles d'ingénieurs internes aux universités remontent dans la composante « Universités ».
- « ENS » : cette composante est collectée depuis 2004. Elle couvre les écoles normales supérieures, mais aussi les écoles vétérinaires (depuis 2009), l'école nationale supérieure du paysage de Versailles (depuis 2009), l'école nationale de la statistique et de l'administration économique, l'institut catholique d'études supérieures et d'autres grands établissements.
- « Management » (MANA) : cette composante est collectée depuis 2006. Elle couvre les écoles de management reconnues par le ministère en charge de l'enseignement supérieur et habilitées à délivrer au moins un diplôme. Cette composante couvre aussi depuis 2016-2017 d'autres écoles délivrant principalement des formations de management et commerce.
- « Enquête 26 bis » (E26B) : cette composante a été collectée pour la première fois pour l'année scolaire 2016-2017, principalement auprès des établissements qui répondaient à l'enquête 26 traditionnelle (l'enquête 26 bis récupère des données individuelles là où l'enquête 26 collectait des données agrégées). Selon la documentation du SIES, cette composante est « plus hétérogène » que les autres composantes et « vouée à s'élargir au fil des collectes ».

Les étudiants des instituts catholiques sont aussi couverts par une composante spécifique de SISE, qui n'a pas été mobilisée dans nos travaux. Ils représenteraient près de 1 % des effectifs inscrits dans les établissements d'enseignement supérieur de France.

Au total, sur la base des éléments présentés en mars 2020 par le SIES devant Conseil national de l'information statistique (tableau 3), **les composantes mobilisées par la présente étude couvriraient un peu moins de 90 % des effectifs inscrits dans l'enseignement supérieur et près des deux tiers des établissements.**

De façon générale, le SIES s'efforce d'étendre au maximum le champ des établissements d'enseignement supérieur couverts par le dispositif SISE. Les établissements dépendants du ministère de la Culture ont ainsi été intégrés dans le dispositif en 2017. À cette date, près de 7 % des effectifs restaient non couverts, notamment en raison de l'absence des établissements de la sphère médicosociale relevant du ministère chargé des affaires sociales. Leur intégration est un objectif. Au-delà, il restera toujours une petite proportion d'effectifs non couverts : certains établissements des composantes couvertes ne remontent pas leurs effectifs (le taux de réponse des établissements oscillerait entre 90 % et 100 % selon les composantes). Par ailleurs, chaque année, de nouveaux établissements privés sont créés. Ils n'intègrent pas immédiatement le dispositif de remontées de données individuelles.

Tableau 3 • Estimation du champ couvert par les données administratives mobilisées dans la présente étude (sur la base des inscrits pour l'année universitaire 2018-2019)

Composante	Disponible au CASD depuis :	Effectifs	%	Nombre d'établissements	%
Composantes couvertes dans la présente étude					
Lycées : CPGE-STTS	2005	346 000	12,9%	2431	57%
SISE-Université	2010	1 661 000	62,0 %	87	2 %
SISE-ENS	2010	15 000	0,6 %	15	0 %
SISE-Ingénieur	2010	142 000	5,3 %	121	3 %
SISE-Management	2010	181 000	6,8 %	142	3 %
SISE-26 bis	2016	25 000	0,9 %	88	2 %
			88,5 %		
Composante disponible mais non mobilisée dans la présente étude					
SISE-Instituts catholiques	2010	32 000	1,2 %	5	0 %
Composantes non disponibles pour la présente étude					
SISE-Culture	2017	80 000	3,0 %	300	7 %
Lycées agricoles	n.d.	20 000	0,7 %	275	6 %
Paramédical	n.d.	103 000	3,8 %	386	9 %
Social	n.d.	32 000	1,2%	128	3%
Enquête 26	n.d.	41 000	1,5 %	293	7 %
Ensemble		2 678 000	100 %	4 271	100 %

Source : MESRI-SIES, diaporama présenté à la réunion du 12 mars 2020 de la commission services publics du Conseil national de l'information statistique (Cnis). Présentation de l'auteur.

Les fichiers de données individuelles ne comportent pas d'informations directement nominatives, mais disposent d'un identifiant dérivé de l'identifiant national des élèves et étudiants (INE), également utilisé dans un certain nombre de fichiers de gestion de l'enseignement secondaire et de l'enseignement supérieur. Cet identifiant permet de fiabiliser l'identification des étudiants pour les opérations administratives. Pour les utilisations statistiques, il doit permettre de faciliter le suivi des parcours par appariements des fichiers annuels. L'accès à l'INE et aux données individuelles est évidemment strictement encadré.

Les collectes des données individuelles auprès des établissements ont été intégrées dans le champ des enquêtes de la statistique publique au sens de la loi du 7 juin 1951, avec une obligation de réponse, dans le but de fiabiliser les données et d'assurer notamment le meilleur taux de couverture possible¹⁰. Les efforts conduits pour fiabiliser l'identifiant national des étudiants et élèves (INE) vont dans le même sens¹¹.

¹⁰ Voir, par exemple, la présentation générale faite lors de la réunion de la commission services publics et services aux publics du CNIS, du 12 mars 2020.

¹¹ Jugnot Stéphane (2021), « Les opportunités créées par la facilitation et le développement des utilisations des données administratives à des fins de recherche et de production statistique », Céreq, Working paper, n° 9, Septembre, 66 p. Voir notamment les parties 2.1 et 4.2.

Parmi les points qui peuvent être notés (voir les guides annuels réalisés par le SIES) :

- Les établissements d'enseignement supérieur concernés doivent remonter au SIES les inscriptions de l'année universitaire $n-1/n$ enregistrées au 15 janvier de l'année n .
- Chaque enregistrement est relatif à une inscription, c'est-à-dire, un étudiant (repéré par son numéro INE), qui prépare un diplôme donné¹² (repéré par son code SISE) dans une « unité administrative immatriculée », repérée par son numéro UAI¹³.
- Toutes les inscriptions sont collectées excepté les auditeurs libres, les stagiaires suivant des cycles d'une durée inférieure à l'année universitaire et les étudiants inscrits en école de commerce en préparation à certains diplômes (préparations au brevet de techniciens supérieurs, au diplôme des métiers d'art, au diplôme supérieur d'arts appliqués...).
- Les étudiants inscrits dans le cadre d'une convention entre un institut catholique et une université publique doivent être recensés dans les deux types d'établissement.
- Par définition de l'inscription, un étudiant peut avoir plusieurs inscriptions pour une même année universitaire, y compris dans le même établissement. Dans le cas où un étudiant dispose de plusieurs inscriptions dans le même établissement, le SIES définit une inscription « principale » et des inscriptions « secondes » à partir de différents critères qui privilégient notamment les diplômes nationaux par rapport aux diplômes d'université et, au sein de chacune des deux catégories, le diplôme permettant d'aboutir au niveau de formation le plus élevé¹⁴. Dans les fichiers, les inscriptions principales sont identifiées par la variable EFFECTIF=1.
- Le SIES opère des traitements et des redressements sur les données individuelles remontées par les établissements. Il procède aussi à quelques enrichissements à partir d'autres fichiers. Toutes les variables proposées ne sont pas systématiquement disponibles pour toutes les composantes de SISE (voir annexe 3).
- Parmi ces traitements, le SIES repère les inscriptions invalides, notamment celles remontées plusieurs fois. Les enregistrements jugés valides sont repérables par la variable FLAG_SUP ≠ 1.
- Normalement, lorsqu'un étudiant s'inscrit à l'université, il doit le faire en utilisant l'INE attribué à l'étudiant dans le secondaire. Il arrive que, faute de le retrouver, l'établissement assigne un nouvel INE à l'étudiant. La variable REIMMA permet d'identifier ces réimmatriculations « abusives », susceptibles de perturber la reconstitution des parcours.
- Le SIES enrichit les données de quelques variables calculées, notamment le degré d'études (variables DEGETU), qui donne le niveau de formation atteint par l'étudiant pendant l'année d'inscription considérée, en nombre d'années après le baccalauréat.

¹² La notion de « diplôme » couvre aussi les attestations et préparations à concours.

¹³ Depuis 1996, toutes les structures publiques ou privées d'enseignement initial et d'administration publique du système éducatif, ainsi que les établissements assurant de la formation continue et/ou de la formation par apprentissage sont identifiés au répertoire académique et ministériel sur les établissements du système éducatif (RAMSESE). Un numéro d'identification « UAI » leur est attribué. Géré par la direction de l'évaluation, de la prospective et de la performance (DEPP), ce répertoire a remplacé la base nationale des établissements utilisée depuis 1970.

¹⁴ Par exception, les diplômes d'ingénieur ou certains diplômes de santé priment sur les autres diplômes nationaux et les habilitations à diriger des recherches ne priment pas sur les autres diplômes nationaux. Pour plus de précisions, il convient de se reporter aux guides de SISE-Inscriptions, qui détaillent la hiérarchie retenue.

1.3. Les fichiers annuels BPBAC (SIES)

La source BPBAC couvre les inscriptions dans les formations post-baccalauréat des établissements du second degré : les classes post-baccalauréat (principalement, sections de technicien supérieur, classes préparatoires aux grandes écoles), les sections d'éducation spécialisée, les groupes de classes-ateliers, les formations complémentaires diplômantes, les préparations diverses pré et post-baccalauréat, les mesures d'insertion des jeunes de l'éducation nationale (MOREA, FCIL ...).

Les données sont issues des fichiers de gestion gérés par la DEPP. Elles concernent les établissements publics du second degré et les établissements privés sous contrat. Les établissements hors contrat ne sont couverts que partiellement.

Les inscriptions prises en compte couvrent la voie scolaire, l'apprentissage, les contrats de professionnalisation et les stagiaires de la formation professionnelle.

Dans nos travaux, nous nous limiterons le plus souvent aux inscriptions de BPBAC relatives aux sections de technicien supérieur et aux classes préparatoires aux grandes écoles, qui peuvent être repérées à partir de la variable MEFST2 (modalités 31 et 32).

Tableau 4 • Nombre d'observations dans les fichiers annuels BPBAC, selon le type de « module élémentaire de formation »¹⁵

	31-CPEG	32-BTS	43-Formation continue	52-divers post-bac.	Ensemble
Inscrits 2016	87 142	260 584	251	8 416	356 393
Inscrits 2015	86 423	258 620	303	8 412	353 758
Inscrits 2014	84 455	257 651	319	8 404	350 829
Inscrits 2013	83 800	257 820	274	7 766	349 660
Inscrits 2012	82 533	255 810	250	8 366	346 959
Inscrits 2011	80 655	247 447	281	8 170	336 553
Inscrits 2010	79 975	242 509	520	8 478	331 482

Source : SIES – BPBAC 2010 à 2016.

Par commodité de langage, les fichiers annuels BPBAC sont considérés dans la suite comme une composante de SISE, au même titre que SISE-Universités ou SISE-Management, par exemple. Dans les programmes présentés en annexe, cette composante est désignée par le sigle « CPES ». Dans les tableaux, elle est désignée par le sigle « BPBAC ».

¹⁵ Selon la variable construite MEFST2=substr(mefst 11,1,2).

1.4. Les fichiers annuels SISE-Résultats (SIES)

Les fichiers annuels de SISE-Résultats permettent de connaître la réussite aux diplômes préparés par les étudiants. Ces fichiers s'articulent aux fichiers de SISE-Inscriptions. Il n'y a donc pas de fichier annuel unique sur la diplomation, mais une collection de fichiers de données individuelles, couvrant chacune une composante particulière de l'enseignement supérieur, comme pour SISE-Inscriptions.

Parmi les points qui peuvent être notés (voir les guides annuels réalisés par le SIES¹⁶) :

- Tous les établissements concernés par SISE-Inscriptions doivent participer à la remontée SISE-Résultats dès lors qu'ils possèdent des formations conduisant à la délivrance d'un diplôme.
- Les établissements concernés doivent remonter les informations sur les résultats aux diplômes au titre de l'année universitaire $n-1/n$ en mai de l'année $n+2$.
- Plus précisément, les établissements doivent transmettre un enregistrement par inscription préparant à un diplôme, que l'étudiant soit en année terminale ou non. Les inscriptions sur des formations préparatoires à un concours ne sont donc pas concernées. En revanche, un étudiant en deuxième année de LMD doit, par exemple, donner lieu à une remontée dans SISE-Résultats même si généralement, elle ne se traduit pas par un diplôme¹⁷.
- Les étudiants ayant des inscriptions multiples dans un même établissement doivent être présents dans SISE-Résultats autant de fois qu'ils ont d'inscriptions préparant à un diplôme.
- Les informations demandées dans cette remontée sont limitées : informations d'identifications de l'étudiant, de l'établissement, du diplôme préparé, présence aux examens et résultats obtenus. Le SIES enrichit ensuite ces données avec des informations issues de SISE-Inscriptions pour permettre des exploitations statistiques sans recourir à un appariement. Cet enrichissement n'est possible que lorsque l'INE est disponible dans les remontées sur la diplomation et si l'étudiant est retrouvé dans SISE-Inscriptions pour l'établissement et le diplôme concernés (ce n'est pas le cas si l'inscription a été enregistrée après le 15 janvier de l'année n).
- Les établissements couverts doivent remonter aussi les diplômes délivrés aux étudiants ne relevant pas du champ de SISE (par exemple, un diplôme délivré à l'issue d'une formation continue de moins d'un an).

1.5. Quelques statistiques descriptives sur les fichiers annuels des inscriptions

Tableau 5 • Nombre d'observations dans SISE-Inscriptions, selon la composante et l'année

	Composante de SISE-Inscriptions					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	356 393	14 155	16 777	131 148	141 838	1 952 819
Inscrits 2015	353 758	n.d.	15 896	124 604	126 069	1 892 941
Inscrits 2014	350 829	n.d.	15 734	120 573	125 701	1 821 088
Inscrits 2013	349 660	n.d.	15 370	115 147	124 130	1 775 295
Inscrits 2012	346 959	n.d.	15 242	112 577	120 364	1 671 097
Inscrits 2011	336 553	n.d.	14 494	108 040	114 424	1 650 200
Inscrits 2010	331 482	n.d.	14 445	102 188	107 909	1 596 772

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

¹⁶ Les variables décrites dans les guides ne sont pas systématiquement disponibles pour toutes les composantes de SISE.

¹⁷ Le SIES a fait ce choix pour faciliter le travail des établissements, dont certains ne raisonnent pas en termes d'années terminales et d'années non terminales dans leurs fichiers de gestion (par exemple, en cas de semestrialisation du cursus).

Tableau 6 • Proportion d'observations repérées comme étant en doubles comptes* selon la composante et l'année de SISE-Inscriptions

	Composantes de SISE-Inscriptions					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	0,0 %	0,0 %	0,6 %	0,0 %	0,0 %	9,7 %
Inscrits 2015	0,0 %	n.d.	0,0%	0,0%	0,0%	8,3%
Inscrits 2014	0,0 %	n.d.	0,0%	0,0%	1,0%	8,4%
Inscrits 2013	0,0 %	n.d.	0,1%	0,0%	0,3%	7,9%
Inscrits 2012	0,0 %	n.d.	0,1%	0,0%	0,1%	4,8%
Inscrits 2011	0,0 %	n.d.	0,0%	0,0%	0,4%	4,3%
Inscrits 2010	0,0 %	n.d.	0,0%	0,0%	0,0%	1,7%

* FLAG_SUP=1.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

Tableau 7 • Proportion d'observations « principales »* selon la composante et l'année de SISE-Inscriptions

	Composantes de SISE-Inscriptions					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	n.d.	99,7 %	91,8 %	99,0 %	99,4 %	95,1 %
Inscrits 2015	n.d.	n.d.	95,0 %	98,7 %	99,4 %	94,8 %
Inscrits 2014	n.d.	n.d.	95,2%	98,7%	99,1%	94,9%
Inscrits 2013	n.d.	n.d.	91,8%	98,6%	99,4%	94,8%
Inscrits 2012	n.d.	n.d.	91,0 %	98,6 %	99,2 %	95,0 %
Inscrits 2011	n.d.	n.d.	94,3 %	98,6 %	99,8 %	94,9 %
Inscrits 2010	n.d.	n.d.	92,8%	98,7%	99,7%	94,8%

* EFFECTIF=1.

Champ : Inscriptions hors doubles comptes identifiés (FLAG_SUP ne 1).

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

Tableau 8 • Proportion d'observations repérées comme des réimmatriculations abusives* selon la composante et l'année de SISE-Inscriptions

	Composantes de SISE-Inscriptions					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	n.d.	2,9 %	4,4 %	2,3 %	3,9 %	1,2 %
Inscrits 2015	n.d.	n.d.	2,4 %	0,6 %	4,9 %	1,1 %
Inscrits 2014	n.d.	n.d.	1,5%	1,1%	4,6%	1,0%
Inscrits 2013	n.d.	n.d.	0,7%	1,3%	6,5%	0,9%
Inscrits 2012	n.d.	n.d.	0,7 %	2,0 %	7,7 %	1,0 %
Inscrits 2011	n.d.	n.d.	1,3 %	1,8 %	9,3 %	0,9 %
Inscrits 2010	n.d.	n.d.	1,2%	2,7%	8,6%	1,1%

*REIMMA=1.

Champ : Inscriptions hors doubles comptes identifiés (FLAG_SUP ne 1).

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

Tableau 9 • Proportion d'observations pour lesquelles l'année de première inscription dans le supérieur est disponible* selon la composante et l'année de SISE-Inscriptions

	Composantes de SISE-Inscriptions					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	n.d.	n.d.	36 %	n.d.	n.d.	100 %
Inscrits 2015	n.d.	n.d.	45%	n.d.	n.d.	100 %
Inscrits 2014	n.d.	n.d.	48%	n.d.	n.d.	100 %
Inscrits 2013	n.d.	n.d.	52 %	n.d.	n.d.	100 %
Inscrits 2012	n.d.	n.d.	49 %	n.d.	n.d.	100 %
Inscrits 2011	n.d.	n.d.	32 %	n.d.	n.d.	100 %
Inscrits 2010	n.d.	n.d.	35 %	n.d.	n.d.	100 %

*ANINSC renseignée.

Champ : inscriptions hors doubles comptes identifiés (FLAG_SUP ne 1).

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

Tableau 10 • Proportion d'observations pour lesquelles l'année du baccalauréat est renseignée* selon la composante et l'année de SISE-Inscriptions

	Composantes de SISE-Inscriptions					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	n.d.	84 %	84 %	86 %	87 %	87 %
Inscrits 2015	n.d.	n.d.	83 %	87 %	85 %	87 %
Inscrits 2014	n.d.	n.d.	83%	87%	85%	87%
Inscrits 2013	n.d.	n.d.	82%	87%	86%	86%
Inscrits 2012	n.d.	n.d.	82 %	87 %	87 %	86 %
Inscrits 2011	n.d.	n.d.	81 %	87 %	87 %	86 %
Inscrits 2010	n.d.	n.d.	81 %	88 %	86 %	85 %

*ANBAC renseignée.

Champ : inscriptions hors doubles comptes identifiés (FLAG_SUP ne 1).

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

2. La réalisation de l'appariement expérimental

2.1. L'INE comme clef d'appariement

Les fichiers individuels sur les étudiants du SIES disposent de l'INE, identifiant national des élèves et étudiants (INE). Cet identifiant des personnes physiques joue, pour la sphère éducative, un rôle analogue au « numéro de sécurité sociale » dans les sphères de l'emploi, de la santé et des affaires sociales. En particulier, cet identifiant permet de s'assurer de la bonne identification des personnes dans les fichiers de gestion des établissements et dans les processus administratifs tels que les inscriptions à certains examens ou l'orientation postbac. Pour les services statistiques des ministères concernés, il permet d'articuler les fichiers administratifs pour construire des parcours, sur les champs couverts par les sources mobilisées. L'INE est donc disponible dans les fichiers SISE et BPBAC.

De son côté le Céreq dispose de l'INE de façon encadrée et temporaire pendant la phase d'élaboration de ses enquêtes Génération. L'INE est récupéré à chaque fois que cela est possible pour construire la base de sondage de l'enquête (encadré 2) : cet identifiant facilite le repérage des inscriptions multiples au cours de l'année scolaire d'intérêt, l'année 2016-2017 pour l'enquête Génération 2017. En utilisant aussi les informations sur les inscriptions au cours de l'année scolaire suivante (2017-2018), l'INE permet également de repérer une partie des poursuites d'études et de supprimer ainsi de la base de sondage une partie importante des personnes hors-champ de l'enquête.

Tous niveaux de sortie confondus, l'échantillon final de l'enquête Génération 2017 compte 303 573 individus, dont 266 063 appartiennent à l'échantillon principal et 37 510 à l'échantillon de réserve qu'il a fallu engager pour atteindre un nombre de répondants suffisant. Pour environ un sixième des personnes échantillonnées, le Céreq ne dispose pas d'INE valide tandis que pour la moitié des personnes échantillonnées, le Céreq dispose de deux INE ou plus (151 385 individus disposent de deux INE et 117 de trois INE). Dans 98 % des cas, les personnes disposant de plusieurs INE sont issues des composantes SYSCA et SIFA de la base de sondage, car la DEPP a transmis au Céreq les deux variables dont elle disposait pour l'INE : d'une part, l'INE dit « RNIE », correspondant à l'INE enregistré au nouveau répertoire national alors en cours de déploiement pour assurer l'unicité de l'identifiant ; d'autre part, l'INE « BEA », qui était attribué au niveau académique dans le cadre du mode de gestion existant auparavant. Les autres cas d'INE multiples appartiennent au segment « universités » de la base de sondage.

Pour effectuer le rapprochement entre l'enquête Génération 2017 et les données administratives, les personnes échantillonnées pour lesquelles le Céreq dispose de plusieurs INE ont été dupliquées en autant d'observations qu'elles ont d'INE. La table de passage contient donc 455 192 observations distinctes.

Dans cette table de passage, la qualité de l'INE a déjà été évaluée dans le cadre d'un appariement expérimental de l'enquête Génération 2017 avec des fichiers de la DEPP¹⁸ (tableau 11). Le travail réalisé conduisait à conclure qu'« au moins 11,3 % des observations de la table de passage ont manifestement un INE invalide. Les individus issus des composantes SYSCA, SIFA et Universités de la base de sondage ne sont pas concernés. À l'opposé, les individus issus des composantes Cifre, Contrats de professionnalisation et Sports ont tous des INE invalides puisque ces sources ne fournissent pas l'INE. Pour les établissements relevant des autres composantes, la situation est intermédiaire. Dans leur cas, il existe sans doute des situations isolées de méconnaissance ou de mauvaise saisie de l'INE pour certains de leurs inscrits, mais l'ampleur des informations manquantes conduit plutôt à supposer que certains établissements n'ont, systématiquement, pas transmis l'INE au Céreq, parce qu'il n'était pas présent ou facilement extractible de leur système de gestion, ou parce qu'ils ont refusé de le faire.

¹⁸ JUGNOT Stéphane (2023), « Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants 2005 à 2016 de la DEPP : des résultats prometteurs », Céreq, Working paper, n° 19, avril 2023, 144p.

Encadré 2 • Construire la base de sondage de l'enquête Génération 2017

Il n'existe pas de fichier centralisé contenant l'ensemble des élèves et étudiants inscrits dans une année scolaire donnée et moins encore, de fichiers permettant d'identifier les seuls primo-sortants qui constituent le champ d'intérêt des enquêtes Génération. Avant chaque nouvelle « Génération », le Céreq doit donc construire sa base de sondage de façon *ad hoc*, dans le respect de la loi informatique et libertés et de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

Pour l'enquête Génération 2017, le Céreq a ainsi collecté des listes d'élèves et d'étudiants inscrits au cours des années scolaires 2016-2017 et 2017-2018 dans les établissements du second degré et de l'enseignement supérieur, en s'efforçant de récupérer l'INE, quand il était disponible. Disposer de deux années scolaires consécutives permet de supprimer une partie des poursuites d'études de la base de sondage. Disposer de l'INE facilite ce travail. Il permet aussi de repérer des personnes en doubles comptes, car inscrites dans plusieurs établissements au cours de l'année scolaire de référence. Ces travaux préalables permettent ainsi de réduire la taille de la base de sondage pour cibler le mieux possible le strict champ de l'enquête. Cependant, l'appartenance à ce champ, notamment le fait que l'élève ou l'étudiant est réellement primo-sortant, ne peut être validée définitivement que lors de la passation du questionnaire. Il commence par un module dédié à cela.

Les informations concernant certains établissements ou formations sont récupérées de façon centralisée. C'est le cas pour :

- l'essentiel des établissements du second degré, grâce à la DEPP, à partir du système d'information statistique consolidé académique (SYSCA) ;
- l'essentiel des formations par apprentissage, grâce à la DEPP, à partir du système d'information de la formation des apprentis (SIFA) ;
- les lycées agricoles ou autres établissements relevant du ministère de l'Agriculture, grâce à sa direction générale de l'enseignement et de la recherche ;
- les écoles d'architecture et d'arts couverts par la plateforme TAIGA du ministère en charge de la culture ;
- les formations « jeunesse et sports » couvertes par l'outil de gestion formations et diplômes de l'animation et du sport (FORÔMES), de la direction de la jeunesse de l'éducation populaire et de la vie associative (DJEPVA) du ministère de l'Éducation nationale ;
- les inscrits dans une convention industrielle de formation par la recherche (CIFRE), grâce à l'association nationale de la recherche et de la technologie (ANRT) ;
- les contrats de professionnalisation, grâce à la direction de l'animation de la recherche, des études et des statistiques (DARES) ;

Pour les autres établissements, qui ne sont pas couverts par ces collectes centralisés, le Céreq collecte les informations en s'adressant directement à eux. C'est notamment le cas pour les universités, les écoles d'ingénieurs, les écoles de commerce, les écoles de la santé et du social. La liste des établissements à enquêter est établie à l'aide de la base centrale des établissements gérée par le ministère en charge de l'éducation.

La base de sondage des enquêtes Génération agrège ainsi plusieurs composantes, pour lesquelles la récupération de l'INE est plus ou moins facile et plus ou moins exhaustive. Dans le cas des collectes directes auprès des établissements, la disponibilité de l'INE dépend de son usage dans les systèmes de gestion, de la facilité à l'extraire avec les listes d'inscrits et de l'acceptation de l'établissement à le transmettre.

Tableau 11 • Qualité de l'INE disponible dans l'échantillon de l'enquête Génération, selon la composante de la base de sondage

Composantes de la base de sondage	% d'INE invalides	INE invalide	dont valeur vide	INE supposé valide	Ensemble
Sysca (second degré)	0 %			214 771	214 771
Sifa (apprentissage)	0 %			106 720	106 720
Universités	0 %	6		53 052	53 058
Autres établissements	35 %	10 246	9 512	18 636	28 882
Agriculture	0 %			8 865	8 865
DREES	82%	3 457	3 338	768	4 225
Sport	100 %	11 536	11 536		11 536
Culture	69 %	762	650	345	1 107
DGAFP	61 %	926	176	591	1 517
Cifre	100 %	890	890	0	890
Contrats de professiona.	100 %	23 621	23 621	0	23 621
Total	11 %	51 444	49 723	403 748	455 192

Note : la table de passage duplique les individus de l'échantillon pour lesquels deux INE ou plus ont été récupérés. Les 214 771 observations de la table de passage associées à Sysca correspondent ainsi à 110 286 personnes échantillonnées. Pour Sifa, l'échantillon comptabilise 62 783 personnes et pour les Universités, 49 861. Il n'y a pas de doublons dans les autres composantes.

Champ : ensemble des répondants et des non-répondants de l'échantillon de l'enquête Génération 2017, y compris extensions d'échantillon.

Source : Céreq – échantillon de l'enquête Génération 2017.

D'après JUGNOT (2023), « Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants 2005 à 2016 de la DEPP : des résultats prometteurs », Céreq, Working paper, n° 19, page 16.

Les différentes composantes de la base de sondage alimentent de façon différenciée les niveaux de sortie atteints par les jeunes de la « Génération 2017 » à l'issue de leur formation initiale. De ce fait, pour tous les niveaux de sortie, au niveau de détail le plus couramment utilisé par les utilisateurs de l'enquête, une part significative des répondants ne dispose pas d'un INE jugé valide et ne peut donc bénéficier d'un enrichissement par les données administratives.

Les tableaux 12a et 12b montrent la répartition des répondants selon qu'ils appartiennent à une composante de la base de sondage disposant à 100 % de l'INE, à une composante pour laquelle l'INE n'est disponible que partiellement ou à une composante pour laquelle l'INE n'est pas du tout disponible. Les résultats sont présentés selon le plus haut diplôme atteint¹⁹ plutôt que par niveaux de sortie²⁰, parce que le Céreq utilise préférentiellement ce critère pour la publication des principaux résultats de l'enquête. Les deux approches conduisent cependant à un constat analogue. Les sortants ayant atteint au mieux un bac+2 hors santé/social, une licence professionnelle, un master hors diplôme d'école ou d'un doctorat sont les mieux couverts par les composantes disposant de l'INE de façon exhaustive, mais elles ne couvrent au mieux que 70 % des répondants du niveau concerné. *A contrario*, les jeunes ayant atteint un niveau bac+2 à bac+4 dans le secteur santé/social ou avec un diplôme d'école d'ingénieurs ou de commerce sont les moins bien couverts.

¹⁹ Variable `phd=substr(phd_cereq,1,2)`. Le plus haut diplôme atteint au cours de la formation initiale est déterminé par le Céreq à partir des réponses aux questions relatives au parcours scolaire.

²⁰ Variable `nivso=substr(nivso_cereq,1,2)`. Le niveau de sortie correspond au niveau de la formation dans laquelle l'élève ou l'étudiant était inscrit lors de l'année 2016-2017, dernière année de sa scolarité.

Tableau 12a • Répartition des répondants de l'enquête Génération par plus haut diplôme atteint et groupes de qualité de l'INE des composantes de la base de sondage, en nombre

Données non pondérées Plus haut diplôme atteint d'après les réponses à l'enquête Génération	Issus d'une composante de la base de sondage...			Total
	...disposant de l'INE	...ne disposant que très partiellement de l'INE	... sans l'INE	
01-Non-diplômés				(1 779)
02-CAP-BEP-MC-divers niv. 3				(2 531)
03-Bac pro-BT-divers niv.4	3 708	702	545	4 955
04-Bac technologique	801	124	26	951
05-Bac général	780	162	17	959
06-Bac+2 sauf santé/social	2 126	386	436	2 948
07-Bac+2 à +4 santé/social	89	529	5	623
08-Licence professionnelle	1 026	106	331	1 463
09-Autre bac+3 et bac+4	976	407	266	1 649
10-Bac+5 sauf écoles	2 917	867	647	4 431
11-Écoles (Ingé,commerce)	680	1 436	234	2 350
12-Doctorat	354	110	61	525
Ensemble	13 457	4 829	2 568	20 854

Champ : ensemble des répondants à l'enquête Génération 2017, hors post-initiaux du sport.

Source : Céreq – échantillon de l'enquête Génération 2017, fichiers Progedo provisoires de l'enquête Génération 2017.

D'après JUGNOT (2023), « Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants 2005 à 2016 de la DEPP : des résultats prometteurs », Céreq, Working paper, n° 19, page 16.

Tableau 12b • Répartition des répondants de l'enquête Génération par plus haut diplôme atteint et groupes de qualité de l'INE des composantes de la base de sondage, en %

Données non pondérées Plus haut diplôme atteint d'après les réponses à l'enquête Génération	Issus d'une composante de la base de sondage...			Total
	...disposant de l'INE	...ne disposant que très partiellement de l'INE	... sans l'INE	
01-Non diplômés				
02-CAP-BEP-MC-divers niv.3				
03-Bac pro-BT-divers niv.4	75 %	14 %	11 %	100 %
04-Bac technologique	84 %	13 %	3 %	100 %
05-Bac général	81 %	17 %	2 %	100 %
06-Bac+2 sauf santé/social	72 %	13 %	15 %	100 %
07-Bac+2 à +4 santé/social	14 %	85 %	1 %	100 %
08-Licence professionnelle	70 %	7 %	23 %	100 %
09-Autre bac+3 et bac+4	59 %	25 %	16 %	100 %
10-Bac+5 sauf écoles	66 %	20 %	15 %	100 %
11-Écoles (Ingé,commerce)	29 %	61 %	10 %	100 %
12-Doctorat	67 %	21 %	12 %	100 %
Ensemble	65 %	23 %	12 %	100 %

Champ : ensemble des répondants à l'enquête Génération 2017, hors post-initiaux du sport.

Source : Céreq – échantillon de l'enquête Génération 2017, fichiers Progedo provisoires de l'enquête Génération 2017.

D'après JUGNOT (2023), « Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants 2005 à 2016 de la DEPP : des résultats prometteurs », Céreq, Working paper, n° 19, page 16.

La méthode d'identification des INE « invalides » appliquée pour la base de sondage de l'enquête Génération 2017²¹ peut aussi être appliquée à l'INE disponible dans les fichiers SISE et BPBAC. Le tableau 13 présente le taux d'INE considérés comme « invalides » pour chacune des composantes de SISE pour chacune des années mobilisées dans cette étude. Les résultats suggèrent une moindre qualité des INE disponibles dans les fichiers BPBAC mais les taux d'INE « invalides » demeurent nettement inférieurs à ceux des données du Céreq. **La difficulté que le Céreq rencontre sur une partie de son champ pour récupérer un INE valide est donc une limite majeure pour l'appariement de l'enquête avec les données administratives du SIES. Son impact quantitatif est plus important que les défauts de couvertures des données administratives pour l'intégration d'informations sur les parcours dans les établissements d'enseignement supérieur en France.**

Tableau 13 • Application de l'indicateur de qualité de l'INE sur les fichiers SISE et BPBAC mobilisés

	Taux d'INE « invalides* » selon la composante de SISE-Inscriptions et l'année considérée					
	BPBAC	E26b	ENS	INGE	MANA	UNIV
Inscrits 2016	7 %	0 %	0 %	0 %	0 %	0 %
Inscrits 2015	7 %	n.d.	0%	0%	1%	0%
Inscrits 2014	7 %	n.d.	0%	0%	1%	0%
Inscrits 2013	8 %	n.d.	1%	0%	1%	0%
Inscrits 2012	8 %	n.d.	1%	0%	0%	0%
Inscrits 2011	9 %	n.d.	1%	0%	0%	0%
Inscrits 2010	9 %	n.d.	1%	0%	0%	0%

* Selon les critères de l'auteur.

Lecture : dans le fichier global des inscrits en 2016 (dont la construction est présentée dans la partie 2.3.1.), 7 % des inscriptions de la « composante » BPBAC sont associés à un INE jugé « invalide » selon notre méthode de qualification. Le taux d'INE « invalides » dans les autres composantes est nul.

Champ : ensemble des inscriptions.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

2.2. Un cadre de travail sécurisé strict

Les fichiers SISE et BPBAC sont des données individuelles au sens du règlement général sur la protection des données (RGPD). Elles sont d'autant plus sensibles que les jeunes sont identifiables par leur identifiant national des élèves et étudiants (INE), que ces données sont exhaustives sur leur champ et qu'elles permettent de suivre leurs parcours et leur réussite dans l'enseignement supérieur, de façon incomplète, mais très précise sur le champ couvert.

Le projet d'appariement expérimental de l'enquête Génération 2017 avec les données administratives retravaillées par le SIES s'inscrit dans une réflexion plus générale d'évolution du dispositif des enquêtes Génération. Ces enquêtes s'inscrivent également dans les missions du Céreq telles que définies par l'article D313-38 du Code de l'éducation. Elles relèvent de la statistique publique au sens de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. À ce titre, les fichiers individuels complets sont couverts par le secret statistique²².

²¹ Outre les données vides, l'INE a été considéré comme « invalide » dans les cas suivants : les valeurs d'une longueur inférieure à 11 positions (11^e position vide, plus exactement) ; les valeurs entièrement numériques ; les valeurs commençant par une lettre ; les valeurs comprenant une lettre mais commençant par « 99999 ».

²² La présente étude n'a pas utilisé les fichiers individuels complets mais a utilisé l'INE, disponible de façon temporaire, qui constitue aussi une variable sensible couverte par le secret.

Le projet d'appariement a fait l'objet d'un échange avec le SIES sur ses objectifs et ses conditions de faisabilité au dernier trimestre 2020. Après un accord de principe, la demande d'accès du Céreq a été formalisée en mars 2021. L'accès aux données individuelles SISE a été mis en œuvre à partir de juillet 2021 au Centre d'accès sécurisé aux données (CASD)²³ dans une « bulle informatique sécurisée ». Concrètement, seul le chercheur habilité peut accéder aux données par l'intermédiaire d'un boîtier sécurisé. Celui-ci permet de vérifier à chaque accès l'identité du chercheur à partir de son empreinte digitale. Par ailleurs, les résultats des exploitations réalisées ne peuvent être téléchargés qu'après vérification de leur contenu, de façon à s'assurer du respect des règles de secret fixées par le fournisseur des données. Les données de l'enquête Génération 2017 ont été temporairement téléchargées dans cette bulle sécurisée afin de réaliser l'appariement.

Si l'infrastructure proposée par le CASD a globalement donné satisfaction, les conditions de mise en œuvre du respect des règles du secret statistique nous semblent perfectibles : les critères d'effectifs minimum « par case » (5 individus) ont été appliqués indifféremment, que les tabulations portent sur les données exhaustives ou qu'elles portent sur les répondants de l'enquête Génération, sans tenir compte du fait que ces dernières concernent un échantillon tiré aléatoirement parmi l'ensemble des inscrits. Le traitement du secret statistique a conduit à modifier à la marge certains résultats présentés, soit en mentionnant explicitement l'existence d'un secret, soit en modifiant artificiellement le résultat pour qu'il ne se situe plus sous le seuil du secret, soit en agrégeant plusieurs cases (les 3 méthodes ont été utilisées).

Initialement prévu pour seize mois, l'accès au CASD a été prolongé jusqu'à décembre 2022 pour finaliser les exploitations encore en cours. En raison d'une évolution de la politique de diffusion du SIES, cette prolongation a nécessité un passage devant le comité du secret statistique créé par l'article 6 bis de la loi du 7 juin 1951 déjà citée. Fin décembre 2022, les données de l'enquête Génération transférées au CASD et les répertoires de travail ont été supprimés et la bulle sécurisée, fermée.

Le projet d'appariement expérimental a été déclaré au registre RGPD du Céreq le 29 juillet 2021.

2.3. Première étape : la construction d'un fichier historique des parcours dans SISE-Inscriptions

Les programmes SAS utilisés pour l'appariement sont présentés dans l'annexe 5. Ils ont été mis en œuvre en cinq étapes :

1. Création d'une table unique d'inscrits pour chaque année par compilation des différentes composantes de SISE-Inscriptions et BPBAC (« partie A »).
2. Création d'une table historique de trajectoires dans SISE-Inscriptions par rapprochement des tables uniques annuelles construites précédemment (« partie B »).
3. Rapprochement de cette table historique avec la table de l'échantillon de l'enquête Génération 2017 (« partie C »).
4. Rapprochement de cette table avec le fichier des réponses à l'enquête Génération 2017 (« partie D »).
5. Renouvellement de la démarche avec les fichiers SISE-Résultats (« partie E »).

Dans cette partie, nous allons revenir rapidement sur les deux premières étapes qui permettent d'aboutir à une table historicisée d'inscriptions dans SISE. Les autres phases seront abordées dans les parties suivantes. L'ensemble est l'occasion de préciser certains choix faits, qui sont évidemment discutables et devront être rediscutés si l'exercice est renouvelé. C'est donc aussi l'occasion d'illustrer le fait que l'exploitation des données administratives n'est ni évidente ni simple, d'autant plus quand il s'agit de suivre des personnes dans différents fichiers.

²³ Le CASD est un groupement d'intérêt public rassemblant l'État (représenté par INSEE), le CNRS, l'École polytechnique, HEC Paris et la Banque de France. Il a été créé par l'arrêté interministériel du 29 décembre 2018, pour développer et maintenir une infrastructure sécurisée qui permette aux chercheurs d'accéder de façon contrôlée à des données individuelles (données d'origine administrative, enquêtes de la statistique publique couvertes par le secret statistique...).

2.3.1. Création des fichiers globaux annuels d'inscription

Pour commencer, les différentes composantes annuelles de SISE-Inscriptions mobilisées sont fusionnées de façon à obtenir un seul fichier global d'inscriptions par année.

Pour les inscriptions de la composante PBBAC, la variable de degré d'études (DEGETU), créée par le SIES dans les autres composantes, n'existe pas. Elle a donc été créée et imputée à partir du type de formation (variable MEFST11).

Pour les composantes BPBAC et Enquête 26bis, il n'y a pas d'identification d'inscriptions « principales » et d'inscriptions « secondaires ». Toutes les inscriptions ont donc été considérées comme des inscriptions principales (imputation de la valeur 1 à la variable EFFECTIF).

De 92 % à 94 % des étudiants, repérés par leur INE, ne sont présents qu'avec une seule inscription dans SISE-Inscriptions (tableau 14) – le taux est proche des 94 % pour les années les plus anciennes et se rapproche de 92 % pour les années les plus récentes, sans doute en lien avec l'amélioration de la couverture des établissements d'enseignement supérieur par SISE. Quelques cas d'individus (INE) ayant des inscriptions très nombreuses posent question sur la qualité de leur INE. Compte tenu du faible nombre de cas, ce sujet n'a pas été investigué dans ce travail exploratoire.

Très peu d'individus (INE) ne sont associés qu'à des inscriptions « secondaires ». Elles représentent 0,3 % des inscriptions chaque année (tableau 15). Les cas d'inscriptions principales multiples sont plus significatifs, mais restent rares : environ 80 000 étudiants chaque année sur la période étudiée, soit près de 4 % des étudiants ayant au moins une inscription principale. Parmi ces étudiants, environ 20 000 n'ont que des inscriptions principales relevant de la composante « Universités » (tableau 16) ; de 8 000 à 9 000 cumulent au moins une inscription principale dans une école de commerce ou d'ingénieur (composantes MANA et INGE) et une inscription principale à l'université (tableau 17) et 3 000 cumulent une inscription principale dans une E.N.S. et au moins une inscription principale à l'université (tableau 18).

Ces résultats ont conduit à retenir deux inscriptions principales et deux inscriptions secondaires par étudiant (INE) pour élaborer l'historique des inscriptions dans les données administratives du SIES. Nous n'avons pas défini de méthode particulière pour sélectionner les inscriptions retenues quand il y avait lieu. Les inscriptions retenues sont ainsi les premières qui apparaissent dans le fichier global annuel. Naturellement, cette façon de procéder devrait être revue en cas d'appariements destinés à la réalisation d'études, même si peu très peu d'observations sont réellement concernées ; on pourrait, par exemple, procéder à un tirage aléatoire pour sélectionner les deux observations principales (ou secondaires) à retenir quand il y en a trois ou plus pour un même étudiant une même année. Une autre option serait de hiérarchiser les formations pour définir quelles inscriptions retenir (cette approche serait d'ailleurs cohérente avec la démarche présentée plus bas lorsque nous ne retiendrons qu'une seule inscription par année pour rendre compte des transitions d'une année sur l'autre).

Tableau 14 • Répartition des étudiants (INE), selon leur nombre d'inscriptions repérées en compilant toutes les composantes, par année

	1	2	3	4	5	6	7 à 11	12 à 150
Inscrits 2016	2 050 161	152 818	11 812	1 213	103	16	14	s.s.
Inscrits 2015	2 001 231	141 777	12 670	1 397	121	25	9	s.s.
Inscrits 2014	1 969 108	120 803	11 860	1 490	178	23	7	s.s.
Inscrits 2013	1 931 585	118 672	11 644	1 315	138	56	5	
Inscrits 2012	1 891 711	112 915	10 979	1 224	127	13	s.s.	
Inscrits 2011	1 859 479	112 946	10 339	1 216	164	16	s.s.	
Inscrits 2010	1 829 705	114 124	9 917	1 168	167	29		

s.s. : secret statistique (nombre d'observations inférieur à 5).

Lecture : dans le fichier global des inscrits 2016, plus de deux millions d'étudiants identifiés par leur INE sont inscrits une seule fois ; près de 153 000 ont deux inscriptions ; 14 ont de 7 à 11 inscriptions.

Champ : ensemble des inscriptions ayant un INE jugé « valide » (méthode de l'auteur), FLAG_SUP ≠ 1,

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016.

Tableau 15 • Nombre d'étudiants (INE) sans aucune inscription principale* et nombre d'étudiants (INE) avant au moins deux inscriptions principales, par année

	Étudiants sans aucune inscription « principale »*		Étudiants avec au moins deux inscriptions « principales » (parmi ceux en ayant au moins une)	
	Nombre	%	Nombre	%
Inscrits 2016	5 677	0,3 %	82 827	3,7 %
Inscrits 2015	5 629	0,3 %	85 802	4,0 %
Inscrits 2014	6 058	0,3 %	81 471	3,9 %
Inscrits 2013	5 093	0,2 %	81 372	4,0 %
Inscrits 2012	4 908	0,2 %	76 618	3,8 %
Inscrits 2011	4 628	0,2 %	75 351	3,8 %
Inscrits 2010	4 703	0,2 %	77 977	4,0 %

*L'inscription est principale si EFFECTIF=1. Pour les composantes E26B et BAC, EFFECTIF est imputée à 1.

Lecture : dans le fichier global des inscrits 2016, 5 677 des étudiants repérés par leur INE ne sont présents sans qu'une de leurs inscriptions ne soit considérée comme « principale ». Ils ne représentent que 0,3 % des inscrits. Parmi ceux ayant au moins une inscription principale, seulement 3,7 % en ont deux ou plus.

Champ : ensemble des inscriptions ayant un INE jugé « valide » (méthode de l'auteur), FLAG_SUP ≠ 1.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 16 • Répartition des étudiants (INE) ayant au moins une inscription principale à l'université (composante UNIV), selon leur nombre total d'inscriptions principales

	Une seule inscription principale		Plusieurs, toutes dans la composante UNIV	Plusieurs, non exclusivement dans la composante UNIV
Inscrits 2016	1 570 728	95 %	18 694	66 857
Inscrits 2015	1 548 855	95 %	20 842	52 864
Inscrits 2014	1 505 183	96 %	20 029	34 632
Inscrits 2013	1 474 161	97 %	20 165	33 188
Inscrits 2012	1 438 943	97 %	18 789	32 823
Inscrits 2011	1 426 843	96 %	18 651	33 514
Inscrits 2010	1 417 062	97 %	17 967	32 313

*L'inscription est principale si EFFECTIF=1. Pour les composantes E26B et BAC, EFFECTIF est imputée à 1.
 Champ : ensemble des étudiants ayant au moins une inscription principale dans la composante UNIV. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.
 Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 17 • Répartition des étudiants (INE) ayant au moins une inscription principale dans une école (composantes INGE et MANA), selon leur nombre total d'inscriptions principales

	Une seule inscription principale		Plusieurs, toutes dans la composante INGE ou MANA	Plusieurs, dont au moins une dans la composante UNIV	Plusieurs, autres cas
Inscrits 2016	255 992	95 %	1 918	8 813	1 643
Inscrits 2015	235 618	96 %	1 074	8 321	1 624
Inscrits 2014	226 456	95 %	2 136	8 593	1 782
Inscrits 2013	223 356	95 %	1 095	8 622	1 645
Inscrits 2012	217 258	95 %	982	9 028	1 535
Inscrits 2011	208 599	95 %	766	8 400	1 323
Inscrits 2010	198 076	96 %	625	7 494	1 114

*L'inscription est principale si EFFECTIF=1. Pour les composantes E26B et BAC, EFFECTIF est imputée à 1.
 Champ : ensemble des étudiants ayant au moins une inscription principale dans la composante INGE ou MANA. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.
 Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 18 • Répartition des étudiants (INE) ayant au moins une inscription principale dans une E.N.S., selon leur nombre total d'inscriptions principales

	Une seule inscription principale		Plusieurs, toutes dans la composante ENS	Plusieurs, dont au moins une dans la composante UNIV	Plusieurs, autres cas
Inscrits 2016	11 820	78 %	129	2 927	249
Inscrits 2015	11 056	76 %	421	2 776	203
Inscrits 2014	10 766	75 %	474	3 025	127
Inscrits 2013	11 047	79 %	91	2 643	116
Inscrits 2012	10 676	78 %	94	2 723	152
Inscrits 2011	10 723	80 %	92	2 515	126
Inscrits 2010	10 415	79 %	80	2 617	91

*L'inscription est principale si EFFECTIF=1. Pour les composantes E26B et BAC, EFFECTIF est imputée à 1.

Champ : ensemble des étudiants ayant au moins une inscription principale dans la composante ENS. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

2.3.2. Création d'un fichier historique de parcours avec les données administratives

Comme l'indique la conclusion de la première étape, l'examen des fichiers annuels globaux a conduit à retenir, au plus, deux inscriptions principales et deux inscriptions secondaires par étudiant identifié par leur INE pour chacune des années étudiées.

Dans une deuxième étape :

- nous avons d'abord reconstruit un fichier annuel global synthétique n'ayant plus l'inscription comme observation, mais l'étudiant (INE). Pour chaque étudiant présent une année donnée, nous avons retenu quelques variables descriptives pour chacune de ses inscriptions retenues, avec, au plus, deux inscriptions principales et deux inscriptions secondaires ;
- nous avons ensuite apparié les fichiers annuels synthétiques ainsi construits de façon à disposer d'un fichier historique unique sur le parcours des étudiants (INE) dans les données administratives mises à disposition dans SISE ;
- pour les étudiants disposant de plusieurs inscriptions une année donnée, nous avons défini une inscription de référence à partir d'une hiérarchisation des formations, afin de faciliter l'étude des transitions.

Pour les étudiants présents dans SISE avec plusieurs inscriptions une même année, il est possible de regarder si les caractéristiques d'état civil disponibles (sexe, année et mois de naissance) sont identiques pour chacune de ces inscriptions. Il n'y a cohérence assurée que pour deux tiers à 80 % des cas selon les années, ce taux étant moins favorable pour les années les plus récentes (tableau 19). Les ordres de grandeur sont similaires lorsque l'on s'intéresse à la cohérence de l'année du baccalauréat (tableau 20).

**Tableau 19 • Cohérence* des informations d'état civil (sexe, année et mois de naissance)
en cas d'inscriptions multiples dans SISE**

	Inscrits 2016	Inscrits 2015	Inscrits 2014	Inscrits 2013	Inscrits 2012	Inscrits 2011	Inscrits 2010
Nombre d'étudiants concernés (les étudiants sont identifiés par l'INE)							
1 observation	2 044 747	1 995 842	1 963 356	1 926 658	1 886 963	1 854 942	1 825 107
2 observations							
- identiques	93 827	94 548	92 011	91 465	86 855	86 132	88 530
- différentes	59 873	48 072	29 794	27 931	26 819	27 695	26 351
3 observations							
- identiques	9 638	10 356	10 001	10 265	9 283	8 807	8 554
- 2 identiques, la 3 ^e différente	1 786	2 110	1 658	1 317	1 466	1 264	1 208
- différentes			23				
4 observations	621	708	707	699	679	698	657
Répartition annuelle des étudiants entre les différents cas possibles							
1 observation	92,50 %	92,76 %	93,60 %	93,60 %	93,78 %	93,71 %	93,61 %
2 observations							
- identiques	4,24 %	4,39 %	4,39 %	4,44 %	4,32 %	4,35 %	4,54 %
- différentes	2,71 %	2,23 %	1,42 %	1,36 %	1,33 %	1,40 %	1,35 %
3 observations							
- identiques	0,44 %	0,48 %	0,48 %	0,50 %	0,46 %	0,44 %	0,44 %
- 2 identiques, la 3 ^e différente	0,08 %	0,10 %	0,08 %	0,06 %	0,07 %	0,06 %	0,06 %
- différentes							
4 observations	0,03 %	0,03 %	0,03 %	0,03 %	0,03 %	0,04 %	0,03 %
Proportion d'étudiants (INE) sans divergence en cas d'inscriptions multiples repérées							
	63 %	68 %	76 %	78 %	77 %	77 %	78 %

* La convergence des informations relatives au sexe, au mois et l'année de naissance est étudiée en ne conservant au plus que deux inscriptions principales et deux secondaires.

Champ : ensemble des étudiants inscrits dans SISE. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 20 • Cohérence* de l'année du bac en cas d'inscriptions multiples dans SISE

	Inscrits 2016	Inscrits 2015	Inscrits 2014	Inscrits 2013	Inscrits 2012	Inscrits 2011	Inscrits 2010
Nombre d'étudiants concernés (les étudiants sont identifiés par l'INE)							
1 observation	2 044 747	1 995 842	1 963 356	1 926 658	1 886 963	1 854 942	1 825 107
2 observations							
- identiques	98 657	99 508	96 079	95 448	91 041	89 627	91 270
- différentes	55 043	43 112	25 726	23 948	22 633	24 200	23 611
3 observations							
- identiques	9 652	10 724	10 174	10 379	9 471	8 886	8 678
- 2 identiques, la 3 ^e différente	1 772	1 742	1 491	1 202	1 277	1 184	1 084
- différentes	0	0	17	0	0	0	0
4 observations	621	708	707	700	680	699	657
Répartition annuelle des étudiants entre les différents cas possibles							
1 observation	92,50 %	92,76 %	93,60 %	93,60 %	93,78 %	93,71 %	93,58 %
2 observations							
- identiques	4,46 %	4,62 %	4,58 %	4,64 %	4,52 %	4,53 %	4,68 %
- différentes	2,49 %	2,00 %	1,23 %	1,16 %	1,12 %	1,22 %	1,21 %
3 observations							
- identiques	0,44 %	0,50 %	0,49 %	0,50 %	0,47 %	0,45 %	0,44 %
- 2 identiques, la 3 ^e différente	0,08 %	0,08 %	0,07 %	0,06 %	0,06 %	0,06 %	0,06 %
- différentes							
4 observations	0,03 %	0,03 %	0,03 %	0,03 %	0,03 %	0,04 %	0,03 %
Proportion d'étudiants (INE) sans divergence en cas d'inscriptions multiples repérées							
	66 %	71 %	80 %	81 %	81 %	80 %	80 %

* La convergence des informations relatives au baccalauréat est étudiée en ne conservant au plus que deux inscriptions principales et deux secondaires.

Champ : ensemble des étudiants inscrits dans SISE. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Nous n'avons pas investigué les raisons possibles des divergences pour savoir s'il y avait lieu d'écarter ou non certaines observations. Nous n'avons pas non plus étudié la cohérence intertemporelle des profils des étudiants repérés par un même INE dans les fichiers SISE de différentes années. Il ne nous a pas semblé indispensable d'approfondir ce sujet pour ce travail expérimental dans la mesure où, en tenant compte des étudiants présents avec une seule observation, la question de la cohérence ne se pose pas pour environ 97 % des étudiants (INE). Si un rapprochement plus systématique devait avoir lieu, il faudrait sans doute investiguer davantage ce point qui suppose aussi de connaître la nature des traitements faits par le SIES, de la collecte des fichiers auprès des établissements à la production des fichiers SISE.

Pour étudier les transitions d'une année sur l'autre, nous avons fait le choix de privilégier une seule inscription parmi les autres en cas d'inscriptions multiples en appliquant les critères suivants :

- les inscriptions principales priment sur les inscriptions secondaires ;
- parmi les inscriptions principales (ou parmi les inscriptions secondaires), l'inscription associée au plus haut degré d'études (variable DEGETU) prime ;
- s'il y a deux inscriptions principales (ou deux inscriptions secondaires) associées au même degré d'études, la hiérarchie suivante est appliquée : la composante ENS prime sur la composante école (MANA ou INGE), qui prime sur les classes préparatoires (partie de la composante PBBAC) ;
- si ces critères n'ont pas permis de trancher, la première observation est arbitrairement retenue.

Tableau 21 • Critères mobilisés pour la sélection de l'observation privilégiée pour le suivi des trajectoires

	Inscrits 2016	Inscrits 2015	Inscrits 2014	Inscrits 2013	Inscrits 2012	Inscrits 2011	Inscrits 2010
Répartition d'individus (INE), en nombre							
1 seule inscription principale disponible.	2 121 049	2 074 674	2 038 246	2 002 032	1 957 759	1 925 105	1 898 217
2 inscriptions principales disponibles, de degrés d'études différents. Choix de l'observation liée au plus haut degré d'étude.	14 047	13 936	14 160	13 557	12 746	12 306	11 919
2 inscriptions principales disponibles, de même degré d'études dont une dans la composante ENS. Choix de l'observation ENS.	2 433	2 112	2 708	2 013	2 107	1 859	2 040
2 inscriptions principales disponibles, de même degré d'études, dont une dans une composante école (MANA ou INGE). Choix de l'observation école.	8 951	7 678	8 513	8 155	8 717	7 964	6 837
2 inscriptions principales disponibles de même degré d'études, ni ENS, ni école, ni CPEG. Choix arbitraire de la 1 ^{re} (P1).	63 979	53 202	33 787	32 565	30 736	32 301	31 394
Pas d'inscription principale disponible, choix d'une secondaire.	31	33	136	13			
Répartition d'individus (INE), en %							
1 seule inscription principale disponible.	95,95 %	96,42 %	97,17 %	97,26 %	97,30 %	97,25 %	97,32 %
2 inscriptions principales disponibles, de degrés d'études différents. Choix de l'observation liée au plus haut degré d'étude.	0,64 %	0,65 %	0,68 %	0,66 %	0,63 %	0,62 %	0,61 %
2 inscriptions principales disponibles, de même degré d'études dont une dans la composante ENS. Choix de l'observation ENS.	0,11 %	0,10 %	0,13 %	0,10 %	0,10 %	0,09 %	0,10 %
2 inscriptions principales disponibles, de même degré d'études, dont une dans une composante école (MANA ou INGE). Choix de l'observation école.	0,40 %	0,36 %	0,41 %	0,40 %	0,43 %	0,40 %	0,35 %
2 inscriptions principales disponibles de même degré d'études, ni ENS, ni école, ni CPEG. Choix arbitraire de la première (P1).	2,89 %	2,47 %	1,61 %	1,58 %	1,53 %	1,63 %	1,61 %
Pas d'inscription principale disponible, choix d'une secondaire.	0,00 %	0,00 %	0,01 %	0,00 %	0,00 %	0,00 %	0,00 %

Champ : ensemble des étudiants inscrits dans SISE. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Cette hiérarchie des formations utilisée peut être évidemment discutée. Comme indiqué plus haut, elle pourrait aussi servir à sélectionner les inscriptions à privilégier pour les quelques étudiants (INE) associés à de nombreuses inscriptions une même année.

Le tableau 21 précise, pour chaque année, la façon dont l'inscription privilégiée pour l'analyse des trajectoires a été définie. Dans 96 % à 97 % des cas, selon les années, la question de la hiérarchisation ne se pose pas puisqu'il n'y a qu'une inscription principale disponible. Dans les autres cas, la prise en compte du degré d'études est peu discriminante et c'est le critère du choix arbitraire qui est le plus souvent mobilisé. Ce constat devrait inciter à retravailler la méthode de sélection en cas de renouvellement de l'exercice.

Les tableaux 22 et 23 illustrent la capacité à reconstruire les parcours d'inscriptions à partir des données administratives compilées dans SISE-Inscriptions. Ces parcours sont reconstruits pour les étudiants présents dans les fichiers de SISE-Inscriptions 2016 relatifs à l'année universitaire 2016-2017, dernière année de formation initiale des répondants de l'enquête Génération 2017. Ils sont élaborés de façon rétrospective, en remontant jusqu'à 2010. Sept années sont donc couvertes, en comptant l'année 2016-2017. Les résultats sont présentés par « classes » de sortie. Celles-ci sont définies en combinant la composante de SISE-Inscriptions et le « degré d'études » (variable DEGETU), c'est-à-dire le nombre d'années au-delà du baccalauréat. Le tableau 22 répartit les étudiants d'une même « classe » selon le nombre d'années de présence sur la période 2010-2016. Le tableau 23 indique quelle proportion d'étudiants d'une « classe » n'est pas retrouvée dans SISE-Inscriptions chacune des années de cette période.

Ces tableaux conduisent aux principaux constats suivants :

- **Pour la majorité des étudiants inscrits en 2016-2017, le dispositif SISE semble permettre le suivi du parcours de l'étudiant dans l'enseignement supérieur, avec un nombre d'inscriptions repérées dans SISE sur la période 2010-2016 et une proportion d'étudiants non retrouvés chaque année cohérents avec un parcours crédible.**

Par exemple, si l'on considère les jeunes inscrits en 1^{re} année de classes préparatoires aux grandes écoles en 2016-2017, 84 % ne sont inscrits qu'une fois sur la période 2010-2016 et 13 % deux fois. Aucun n'est retrouvé avant l'année scolaire 2015-2016.

À l'opposé de l'échelle, si l'on considère les jeunes inscrits en école d'ingénieur, dans leur 5^e année post-baccalauréat en 2016-2017, 76 % sont retrouvés 5 à 7 ans sur la période 2010-2016 et seulement 24 % ne sont pas retrouvés l'année scolaire 2012-2013, soit cinq années auparavant.

- **Pour une proportion peu nombreuse, mais significative, d'étudiants, les durées d'observations et les années d'absence ne semblent pas cohérentes avec le degré d'études d'inscription en 2016-2017. Par exemple, de 14 % à 15 % des étudiants inscrits à l'université ou dans une école de commerce en 5^e année post-baccalauréat en 2016-2017 ne sont pas retrouvés une seule fois dans les fichiers SISE de la période 2010-2016. Trois facteurs peuvent jouer : l'existence de parcours heurtés ou atypiques, des défauts dans le suivi de l'INE, les trous dans le champ couvert par SISE. Nous n'avons pas cherché à faire la part de ces trois facteurs, une telle analyse fine n'étant d'ailleurs pas forcément réalisable avec les seules informations disponibles.**
- **Pour les étudiants inscrits en 2016-2017 dans des établissements relevant de l'enquête 26 bis, le suivi n'est quasiment pas possible. Selon leur degré d'étude atteint, entre la moitié et les trois quarts ne sont repérés qu'une fois. L'intégration récente de cette composante dans SISE explique sans doute en partie ce mauvais taux. La qualité de l'INE sur cette période joue peut-être aussi.**

L'annexe 4 propose des résultats de parcours plus détaillés.

Tableau 22 • Répartition des étudiants (INE) selon le nombre d'années de présence dans SISE-Inscriptions, par degré d'étude et composante d'inscription en 2016

Degré* et composante** dans SISE-Inscriptions 2016		Répartition des étudiants selon le nombre d'années de présence dans SISE							
		1	2	3	4	5	6	7	
1	BPBAC-BTS	83,9 %	12,9 %	2,5 %	0,6 %	0,1 %	0,0 %	0,0 %	100 %
	BPBAC-CPEG	96,1 %	3,4 %	0,4 %	0,1 %	0,0 %	0,0 %	0,0 %	100 %
	E26b	74,1 %	14,3 %	7,5 %	2,4 %	1,1 %	0,4 %	0,1 %	100 %
	ENS	79,1 %	10,7 %	2,9 %	4,7 %	1,3 %	0,4 %	0,9 %	100 %
	INGE	82,1%	11,1%	3,8%	2,2%	0,6%	0,2%	0,1%	100%
	MANA	80,2 %	12,6 %	3,6 %	1,2 %	1,6 %	0,7 %	0,1 %	100 %
	UNIV	68,6 %	22,2 %	5,9 %	1,9 %	0,8 %	0,4 %	0,3 %	100 %
2	BPBAC-BTS	5,0 %	71,4 %	19,2 %	3,5 %	0,7 %	0,2 %	0,0 %	100 %
	BPBAC-CPEG	7,7 %	64,9 %	25,8 %	1,4 %	0,1 %	0,0 %	0,0 %	100 %
	E26b	66,8 %	14,8 %	10,7 %	4,7 %	1,9 %	1,0 %	0,2 %	100 %
	ENS	7,8 %	41,3 %	27,0 %	13,8 %	6,1 %	1,8 %	2,1 %	100 %
	INGE	1,6%	72,4%	17,9%	5,3%	1,9%	0,7%	0,2%	100%
	MANA	15,2%	64,9%	13,5%	4,3%	1,7%	0,4%	0,1%	100%
	UNIV	4,1 %	55,1 %	26,1 %	9,4 %	3,2 %	1,3 %	0,8 %	100 %
3	E26b	57,1%	11,3%	18,2%	8,6%	3,0%	1,3%	0,4%	100%
	ENS	7,4%	3,0%	54,4%	29,6%	3,5%	1,7%	0,5%	100%
	INGE	9,2%	3,7%	56,7%	22,7%	5,6%	1,6%	0,5%	100%
	MANA	17,4%	8,1%	55,1%	13,8%	4,1%	1,1%	0,4%	100%
	UNIV	12,6 %	4,6 %	43,1 %	24,0 %	9,5 %	3,9 %	2,2 %	100 %
4	E26b	55,3%	9,7%	15,8%	11,8%	5,1%	1,9%	0,5%	100%
	ENS	12,3%	5,6%	6,2%	39,3%	23,9%	8,6%	4,1%	100%
	INGE	8,2%	7,1%	4,2%	48,9%	24,0%	6,1%	1,4%	100%
	MANA	15,6 %	8,4 %	7,3 %	39,4 %	19,4 %	7,5 %	2,4 %	100 %
	UNIV	13,5 %	6,4 %	4,1 %	33,1 %	23,1 %	12,0 %	7,9 %	100 %
5	E26b	53,3 %	10,7 %	19,3 %	10,1 %	4,7 %	1,6 %	0,4 %	100 %
	ENS	6,6 %	7,9 %	6,5 %	4,8 %	28,4 %	29,9 %	15,9 %	100 %
	INGE	5,3%	5,6%	6,8%	6,4%	39,1%	25,5%	11,3%	100%
	MANA	15,0 %	9,5 %	7,1 %	6,7 %	24,7 %	24,5 %	12,6 %	100 %
	UNIV	13,6 %	10,6 %	6,5 %	5,0 %	26,7 %	22,8 %	14,7 %	100 %
6	E26b	58,2 %	28,4 %	4,3 %	2,4 %	0,0 %	3,8 %	2,9 %	100 %
	ENS	8,0 %	6,8 %	8,8 %	9,8 %	9,7 %	19,0 %	37,9 %	100 %
	INGE	20,7%	10,4%	8,8%	8,9%	8,6%	20,5%	22,1%	100%
	MANA	32,9 %	10,9 %	4,7 %	5,4 %	10,5 %	16,2 %	19,5 %	100 %
	UNIV	11,0 %	6,8 %	6,2 %	6,0 %	5,9 %	13,2 %	50,9 %	100 %

* Variable DEGETU.

** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 23 • Années d'absence dans les fichiers de SISE-Inscriptions 2010 à 2016 des étudiants repérés en 2016, par degré d'étude et composante d'inscription en 2016

Degré* et composante** dans SISE-Inscriptions 2016		% d'individus de la composante concernée qui sont absents des fichiers SISE-Inscriptions l'année considérée					
		2015	2014	2013	2012	2011	2010
1	CPES-BTS	85	96	99	100	100	100
	CPES-CPEG	96	100	100	100	100	100
	E26b	84	89	93	95	97	98
	ENS	85	92	93	94	96	96
	INGE	84	94	97	98	99	100
	MANA	82	93	96	97	99	100
	UNIV	71	91	96	98	99	99
2	CPES-BTS	7	77	95	98	99	100
	CPES-CPEG	8	73	98	100	100	100
	E26b	95	80	83	90	93	95
	ENS	12	52	77	88	93	94
	INGE	2	76	92	96	98	99
	MANA	19	80	92	97	99	99
	UNIV	6	62	85	94	97	98
3	E26b	82	77	75	85	92	95
	ENS	20	12	60	90	96	98
	INGE	13	14	68	91	97	99
	MANA	25	27	76	92	97	99
	UNIV	18	20	60	82	91	95
4	E26b	92	77	72	75	84	91
	ENS	21	22	24	59	83	92
	INGE	10	19	20	66	91	97
	MANA	25	30	30	64	86	95
	UNIV	21	24	25	55	76	86
5	E26b	97	90	74	73	74	83
	ENS	13	19	24	25	49	76
	INGE	7	13	19	24	60	86
	MANA	23	30	35	36	55	80
	UNIV	21	31	34	35	56	75
6	E26b	78	90	91	93	91	76
	ENS	15	21	28	36	38	46
	INGE	35	42	46	47	49	57
	MANA	45	53	51	51	57	66
	UNIV	19	24	27	30	31	36

* Variable DEGETU.

** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

2.4. Deuxième étape : l'appariement avec l'échantillon de l'enquête Génération 2017

À l'issue de la première étape décrite précédemment, nous disposons d'un fichier historicisé de SISE compilant les parcours des étudiants dans les fichiers administratifs sur la période 2010-2016, où :

- chaque observation correspond à une personne identifiée par un INE. Si pour une raison ou pour une autre, l'étudiant « personne physique réelle » change d'INE, il est présent comme autant d'étudiants SISE différents qu'il a d'INE différents ;
- des variables de parcours ont été créées. Elles permettent de connaître jusqu'à deux inscriptions principales et deux inscriptions secondaires par année, et de proposer une inscription privilégiée à partir des critères présentés.

Cette table historicisée est d'abord rapprochée de la table de l'échantillon de l'enquête Génération 2017, qui couvre les répondants et les non-répondants, y compris l'extension d'échantillon des post-initiaux du sport. Dans l'échantillon du Céreq, la personne est identifiée par l'identifiant non signifiant du Céreq (variable IDENT). Comme nous l'avons vu dans la partie 2.1, la table échantillon comprend 303 573 individus (IDENT), dont 151 385 disposent de deux INE et 117 de trois INE. Il convient donc d'examiner d'abord si une personne échantillonnée par le Céreq avec plusieurs INE est retrouvée dans le fichier historicisé de SISE pour un seul de ses INE ou pour plusieurs. Dans ce deuxième cas, il faudrait agréger les lignes correspondantes du fichier historicisé afin de toutes les associer à la personne échantillonnée.

En pratique, seules 44 personnes échantillonnées par le Céreq (IDENT) ont été retrouvées dans le fichier historicisé avec deux INE différents. Aucune des personnes de l'échantillon disposant de trois INE n'a été retrouvé avec ses trois INE. Les présences multiples dans le fichier historicisé sont donc très peu nombreuses, ce qui n'est pas étonnant car l'essentiel des cas d'INE multiples dans la base de sondage est associé à la bascule de l'ancien INE vers le nouvel INE dans les fichiers de la DEPP, bascule qui s'opère à la fin de notre période d'intérêt et principalement sur un champ de sortants du secondaire. Sur les 44 personnes concernées, 27 appartiennent à la composante « Universités » de la base de sondage de Génération et 17 appartiennent à la composante SIFA (apprentissage). Il n'y a qu'un cas de vrai doublon au sens de personne échantillonnée (IDENT) présent dans les fichiers SISE avec ses deux INE la même année.

Pour tenir compte de ces quelques personnes échantillonnées (IDENT) présentes deux fois dans le fichier historicisé de SISE, on reconstruit un fichier historicisé ayant comme observation la personne échantillonnée par le Céreq (IDENT). Pour cela, l'appariement de la table de l'échantillon et du fichier SISE historicisé a été effectué en deux tours de façon à retenir, pour les individus (IDENT) concernés, les deux lignes du fichier historicisé de SISE les concernant (les variables sont suffixées T1 et T2).

Grâce à ce nouveau fichier, le tableau 24 évalue la capacité à apparier l'échantillon de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, par composante de la base de sondage de Génération :

- **Pour les personnes relevant de la composante « Université », il n'y a aucune difficulté :** 100 % des personnes échantillonnées disposent d'un INE jugé valide dans la base de sondage (seules 6 sur 49 861 n'en disposent pas) et 100 % sont retrouvées dans le fichier historicisé de SISE (seules 58 ne sont pas retrouvées). Par ailleurs, 27 personnes échantillonnées sont retrouvées pour deux INE différents.
- **Pour les personnes relevant des composantes « Autres établissements » et « DGAFP », la difficulté principale est de disposer de l'INE dans la base de sondage.** Quand l'INE est disponible, les personnes sont la plupart du temps retrouvées dans le fichier historicisé de SISE.
- **Pour les personnes relevant des composantes « Culture » et « DREES », les deux difficultés se combinent, celle à disposer de l'INE dans la base de sondage et un défaut de couverture dans SISE** (les établissements de la Culture ont été intégrés dans SISE après la période mobilisée par cette étude et ceux du secteur Santé-Social ne l'étaient toujours pas au moment de la réalisation de l'étude).

- Pour les personnes relevant de la composante « CIFRE », l'INE n'est pas récupéré dans la base de sondage. Il n'est donc pas possible de les retrouver dans SISE.
- Pour les personnes relevant des composantes « SIFA » et « SYSCA », l'INE est disponible dans la base de sondage mais peu sont retrouvées dans SISE. Ce résultat est logique puisque ces composantes couvrent essentiellement des sortants du secondaire.

Tableau 24 • Capacité à appairer l'échantillon de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, par composante de la base de sondage de Génération

Composante de la base de sondage de l'enquête Génération	Nombre d'individus dans l'échantillon (ie post-initiaux du sport)	Nombre d'individus « Céreq » avec...				Nombre d'individus ayant au moins un INE valide* et...			
		des INE valides*	des INE invalides*	un INE à blanc*	retrouvés** dans SISE pour un seul INE	retrouvés** dans SISE pour deux INE différents	non retrouvés dans SISE	% retrouvés** dans SISE	
UN-Universités	49 861	49 855	100 %	6	0	49 770	27	58	100 %
G1-Autres établissements	28 882	18 636	65 %	10 246	9 512	16 485		2 151	88 %
G3-DGAFP	1 517	591	39 %	926	176	580		11	98 %
G4-DREES	4 225	768	18%	3 457	3 338	283		485	37%
CU-Culture	1 107	345	31 %	762	650	81		264	23 %
SI-Sifa	62 783	62 783	100 %	0	0	6 717	17	56 049	11 %
SY-Sysca	110 286	110 286	100 %	0	0	13 606		96 680	12 %
AG-Agriculture	8 865	8 865	100 %	0	0	51		8 814	1 %
CI-Cifre	890	0	0 %	890	890				
SP-Sport	11 536	0	0 %	11 536	11 536				
CP-Contrats de professio.	23 621	0	0 %	23 621	23 621				
Total	303 573	252 129	83 %	51 444	49 723	87 573	44	164 512	35 %

* Selon la méthode de qualification présentée dans le document de travail.

** Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble de l'échantillon de l'enquête 2020 auprès de la Génération 2017 (y compris post-initiaux du sport).

Source : appariement de Céreq – échantillon de l'enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Le tableau 25 permet de procéder à la même analyse pour les seuls répondants à l'enquête Génération.

Les composantes de la base de sondage alimentent de façon différenciée les groupes d'intérêts privilégiés dans les analyses, à savoir, d'une part, la décomposition des sortants selon leur plus haut diplôme atteint au cours de leur scolarité ; d'autre part, la décomposition des sortants selon leur niveau de sortie (formation d'inscription lors de leur dernière année d'études). De ce fait, par exemple, même si la composante « Universités » est parfaitement couverte par l'appariement avec SISE, aucun groupe de sortants du supérieur ne l'est dès que l'on passe à une approche par plus haut diplôme atteint ou par niveau de sortie. Pour mieux identifier les champs de l'enseignement supérieur bien couverts par l'appariement, le tableau 26 propose une analyse de la capacité à appairer les répondants de l'enquête avec les données administratives selon leur niveau de sortie et le tableau 27, selon leur plus haut diplôme atteint.

À gros traits, à partir du niveau bac+3, plus de 90 % des jeunes sont retrouvés dans le fichier historicisé. Pour ces niveaux, la difficulté à apparier provient principalement des cas d'indisponibilités de l'INE dans la base de sondage. En revanche, pour les niveaux bac+2, dont les BTS, et pour les formations du secteur santé/social, le défaut de couverture du fichier historicisé SISE joue aussi.

Tableau 25 • Capacité à apparier la table des répondants de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, par composante de la base de sondage de Génération

Composante de la base de sondage de l'enquête Génération	Nombre de répondants à Génération (y.c. post-initiaux du sport)	Nombre de répondants avec...			Nombre d'individus ayant au moins un INE valide* et...			
		des INE valides*	des INE invalides* ou à blanc		retrouvés** dans SISE pour un seul INE	retrouvés** dans SISE pour deux INE différents	non retrouvés dans SISE	% retrouvés** dans SISE
UN-Universités	6 984	6984	100 %	0	6 964	7	13	100 %
G1-Autres établissements	3 054	2181	71 %	873	2 040		141	94 %
G3-DGAFP	267	171	64 %	96	166		5	97 %
G4-DREES	722	163	23%	559	57		106	35%
CU-Culture	147	64	44 %	83	23		41	36 %
SI-Sifa	3 082	3082	100 %	0	749		2 333	24 %
SY-Sysca	7 493	7493	100 %	0	1 629		5 864	22 %
AG-Agriculture	810	810	100 %	0	8		802	1 %
CI-Cifre	67	0	0 %	67	0			
SP-Sport	1 998	0	0 %	1 998	0			
CP-Contrats de prof.	1 759	0	0 %	1 759	0			
Total	26 383	20 948	79 %	5 435	11 639	7	9 302	56 %

* Selon la méthode de qualification présentée dans le document de travail.

** Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble des répondants de l'enquête 2020 auprès de la Génération 2017 (y compris post-initiaux du sport).

Source : appariement de Céreq – échantillon de l'enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 26 • Capacité à appairer la table des répondants de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le niveau de sortie (NIVSOR)

Niveau de sortie (information disponible dans la base de sondage de l'enquête Génération)	Nombre de répondants à Génération	Nombre d'individus « Céreq » avec...			Nombre d'individus ayant au moins un INE valide* et...			
		des INE valides*	des INE invalides* ou à blanc		retrouvés** dans SISE pour un seul INE	retrouvés** dans SISE pour deux INE différents	non retrouvés dans SISE	% retrouvés** dans SISE
Collège	176	176	100 %	0	0	0	176	0 %
CAP-BEP-MC de niveau 3	2 897	2 648	91 %	249	57	0	2 591	2 %
Bac pro.	4 415	4 371	99 %	44	28	0	4 343	1 %
Autre diplôme de niveau 4	1 184	415	35 %	769	64	0	351	15 %
Bac techno	535	535	100 %	0	0	0	535	0 %
Bac général	162	162	100 %	0	0	0	162	0 %
BTS/DUT	3383	2989	88 %	394	2302	0	687	77 %
Autre bac+2	305	84	28 %	221	74	0	10	88 %
Bac+2à4 santé social	645	267	41 %	378	140	0	127	52 %
Licence pro	1404	1059	75 %	345	1010	7	42	96 %
L3	1967	1900	97 %	67	1878		22	99 %
Autre bac+3 ou bac+4	490	117	24 %	373	93		24	79 %
M2	3879	3381	87 %	498	3285		96	97 %
Autres bac+5	897	499	56 %	398	467	32	94 %	
École de commerce	949	500	53 %	449	451	0	49	90 %
École d'ingénieur	1335	1138	85 %	197	1093	0	45	96 %
Doctorat santé	232	232	100 %	0	232	0	0	100 %
Doctorat hors santé	300	229	76 %	71	229	0	0	100 %
Total***	25 164	20 710	82 %	4454	11 409	7	9294	55 %

* Selon la méthode de qualification présentée dans le document de travail.

** Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble des répondants de l'enquête 2020 auprès de la Génération 2017 hors post-initiaux du sport.

Source : appariement de Céreq enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 27 • Capacité à apparier la table des répondants de l'enquête Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le plus haut diplôme atteint (PHDIP)

Plus haut diplôme atteint (variable construite à partir des réponses à l'enquête Génération)	Nombre de répondants à Génération	Nombre d'individus « Céreq » avec...			Nombre d'individus ayant au moins un INE valide* et...			
		des INE valides*		des INE invalides* ou à blanc	retrouvés** dans SISE pour un seul INE	retrouvés** dans SISE pour deux INE différents	non retrouvés dans SISE	% retrouvés** dans SISE
01-Non diplômés	1 780	1 764	99 %	16	5	0	1 759	0 %
02-CAP-BEP-MC-divers niv.3	2 531	2 454	97 %	77	10	0	2 444	0 %
03-Bac pro-BT-divers niv.4	4 955	4 203	85 %	752	591	0	3 612	14 %
04-Bac technologique	951	863	91 %	88	450	0	413	52 %
05-Bac général	959	863	90 %	96	709	0	154	82 %
06-Bac+2 sauf santé/social	2 947	2 391	81 %	556	1 929	0	462	81 %
07-Bac+2 à +4 santé/social	623	265	43 %	358	137	0	128	52 %
08-Licence professionnelle	1 463	1 084	74 %	379	1 026	5	53	95 %
09-Autres bac+3 et bac+4	1 648	1 177	71 %	471	1 133	0	44	96 %
10-Bac+5 sauf écoles	4 426	3 499	79 %	927	3 374	5	120	97 %
11-Écoles (ingénieur, commerce)	2 350	1 683	72 %	667	1 585	0	98	94 %
12-Doctorat	529	464	88 %	65	459	0	5	99 %
Total***	25 164	20 710	82 %	4454	11 409	7	9 294	55 %

* Selon la méthode de qualification présentée dans le document de travail.

** Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble des répondants de l'enquête 2020 auprès de la Génération 2017 hors post-initiaux du sport.

Source : appariement de Céreq enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 28 • Données pondérées – Capacité à apparier la Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le niveau de sortie (NIVSOR)

Niveau de sortie (information disponible dans la base de sondage de l'enquête Génération)	Pas d'INE valide*	INE valides mais non retrouvés dans SISE	Retrouvés dans SISE avec un seul INE	Retrouvés dans SISE avec deux INE différents	Ensemble
Collège	0,0	99,1	0,9	0,0	100
CAP-BEP-MC de niveau 3	13,1	84,5	2,4	0,0	100
Bac pro.	1,5	98,0	0,5	0,0	100
Autre diplôme de niveau 4	46,1	45,5	8,5	0,0	100
Bac techno	0,0	100,0	0,0	0,0	100
Bac général	0,5	99,5	0,0	0,0	100
BTS/DUT	15,5	15,8	68,7	0,0	100
Autre bac+2	48,3	6,4	45,3	0,0	100
Bac+2à4 Santé Social	58,0	18,0	24,1	0,0	100
Licence pro	18,1	3,8	77,8	0,3	100
L3	3,2	1,1	95,7	0,0	100
Autre bac+3 ou bac+4	73,7	4,6	21,7	0,0	100
M2	10,8	2,8	86,4	0,1	100
Autres bac+5	39,0	3,6	57,3	0,1	100
École de commerce	53,4	5,1	41,6	0,0	100
École d'ingénieur	11,7	3,8	84,6	0,0	100
Doctorat Santé	0,0	0,0	100,0	0,0	100
Doctorat hors Santé	17,0	0,2	82,8	0,0	100

* Selon la méthode de qualification présentée dans le document de travail.

** Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble des répondants de l'enquête 2020 auprès de la Génération 2017 hors post-initiaux du sport.

Source : appariement de Céreq enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 29 • Données pondérées – Capacité à appairer la Génération 2017 avec la table historicisée de SISE-Inscriptions, selon le plus haut diplôme atteint (PHDIP)

Plus haut diplôme atteint (variable construite à partir des réponses à l'enquête Génération)	Pas d'INE valide*	INE valides mais non retrouvés** dans SISE	Retrouvés** dans SISE avec un seul INE	Retrouvés** dans SISE avec deux INE différents	Ensemble
01-Non diplômés	1,4	98,4	0,2	0,0	100
02-CAP-BEP-MC-divers niv.3	5,9	93,9	0,2	0,0	100
03-Bac pro-BT-divers niv.4	14,4	71,5	14,1	0,0	100
04-Bac technologique	12,9	39,6	47,5	0,0	100
05-Bac général	12,0	18,9	69,2	0,0	100
06-Bac+2 sauf santé/social	18,9	13,4	67,7	0,0	100
07-Bac+2 à +4 santé/social	56,6	18,8	24,6	0,0	100
08-Licence professionnelle	19,8	4,7	75,2	0,3	100
09-Autres bac+3 et bac+4	25,2	3,1	71,7	0,0	100
10-Bac+5 sauf écoles	18,4	3,0	78,6	0,1	100
11-Écoles (ingénieur, commerce)	32,7	4,4	62,9	0,0	100
12-Doctorat	8,5	0,1	91,4	0,0	100

* Selon la méthode de qualification présentée dans le document de travail.

** Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble des répondants de l'enquête 2020 auprès de la Génération 2017 hors post-initiaux du sport.

Source : appariement de Céreq enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Le tableau 30 affine le constat en montrant, par plus haut diplôme atteint, la proportion de répondants retrouvés dans le fichier historicisé chaque année, sur la période 2010-2016.

De façon cohérente, à partir du niveau bac+2, plus le niveau s'élève, plus les répondants sont retrouvés les années les plus anciennes. **En 2010, on retrouve encore 21 % des répondants ayant comme plus haut diplôme un bac+5 hors écoles et 74 % des répondants ayant réussi une thèse. Ces deux proportions inciteraient, en cas de renouvellement de l'appariement, à allonger la profondeur historique couverte par le fichier historicisé. Elles pourraient aussi conduire à renoncer à retracer tout le parcours des titulaires d'un doctorat afin de ne pas alourdir trop l'exercice d'appariement pour une minorité de sortants.**

Tableau 30 • Proportion de répondants retrouvés* dans chacun des fichiers annuels SISE, selon le plus haut diplôme atteint

	2016	2015	2014	2013	2012	2011	2010
01-Non diplômés							
02-CAP-BEP-MC-divers niv.3							
03-Bac pro-BT-divers niv.4	10 %	6 %	2 %	1 %	0 %	0 %	0 %
04-Bac technologique	45 %	24 %	11 %	32 %	2 %	1 %	0 %
05-Bac général	71 %	43 %	28 %	16 %	10 %	6 %	3 %
06-Bac+2 sauf santé/social	62 %	59 %	24 %	11 %	5 %	2 %	1 %
07-Bac+2 à +4 santé/social	12 %	11 %	10 %	9 %	9 %	5 %	2 %
08-Licence professionnelle	68 %	47 %	51 %	25 %	11 %	7 %	4 %
09-Autres bac+3 et bac+4	65 %	60 %	60 %	44 %	31 %	20 %	11 %
10-Bac+5 sauf écoles	74 %	68 %	65 %	63 %	63 %	40 %	21 %
11-Écoles (ingénieur, commerce)	65 %	64 %	63 %	60 %	59 %	32 %	10 %
12-Doctorat	86 %	82 %	82 %	79 %	77 %	76 %	74 %

* Retrouvé au moins une fois dans le fichier historicisé de SISE construit comme présenté dans ce document de travail.

Champ : ensemble des répondants de l'enquête 2020 auprès de la Génération 2017 hors post-initiaux du sport.

Source : appariement de Céreq enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Pour les 11 416 répondants à l'enquête appariés avec le fichier historicisé de SISE, la cohérence entre les deux sources a été examinée pour quelques caractéristiques du profil de l'étudiant : sexe, année de naissance, mois de naissance, année du baccalauréat.

Pour ces variables, il a fallu d'abord examiner la cohérence interne à SISE en considérant les informations fournies par toutes les inscriptions associées à un même répondant (IDENT) dans SISE sur la période 2006-2016, sachant qu'un répondant (IDENT) peut être associé à deux étudiants (INE) dans SISE et que, pour chaque année, nous avons conservé, au plus, deux inscriptions principales et deux inscriptions secondaires par étudiants dans SISE (INE).

Si l'individu n'est présent qu'avec une seule inscription administrative sur la période, il est comptabilisé parmi les cas de cohérence interne à SISE. **L'analyse de la cohérence interne à SISE** pour les 11 416 répondants à l'enquête appariés avec le fichier historicisé conduit aux résultats suivants, très satisfaisants sous réserve de la disponibilité de l'information dans SISE :

- **concernant le sexe**, il y a cohérence interne à SISE pour 11 352 répondants et divergence pour 56 cas. L'information n'est pas disponible dans SISE pour les autres ;
- **concernant l'année de naissance**, il y a cohérence interne à SISE pour 9 717 répondants et divergence pour 25 répondants. L'information n'est pas disponible dans SISE pour les autres. Sur les 25 cas de divergence, l'écart maximal entre les années de naissance disponibles dans SISE n'est que d'un an dans 14 cas et de trois ans ou plus dans 6 cas ;
- **concernant le mois de naissance**, il y a cohérence interne à SISE pour 9 270 répondants et divergence pour 12 répondants. L'information n'est pas disponible dans SISE pour les autres ;
- **concernant l'année du baccalauréat**, il y a cohérence interne à SISE pour 9 054 répondants et divergence pour 129 répondants (avec un an d'écart maximal pour 105 répondants). L'information n'est pas disponible dans SISE pour les autres.

Pour analyser la **cohérence entre les informations de l'enquête Génération et les informations de SISE** en cas d'informations divergentes dans SISE, on prend en compte la modalité la plus fréquente pour le sexe et on ne prend en compte l'information SISE qu'en cas d'absence de divergence pour les informations numériques (année de naissance, mois de naissance, année du baccalauréat). L'analyse de la cohérence entre les deux sources pour les 11 416 répondants à l'enquête appariés avec le fichier historicisé conduit alors aux résultats suivants, là encore satisfaisants sous réserve de la disponibilité de l'information dans SISE :

- **concernant la combinaison « sexe et année de naissance »**, il y a cohérence totale, interne à SISE et avec l'enquête pour 9 633 répondants (84 % des répondants appariés). Il y a cohérence entre les deux sources sur le sexe mais indisponibilité de l'année de naissance dans SISE pour 1 663 répondants (15 %). Il y a cohérence entre les deux sources pour l'année de naissance mais pas pour le sexe pour 34 répondants, tandis que 49 sont dans le cas inverse (cohérence pour le sexe mais pas pour l'année de naissance) ;
- **en tenant compte du sexe, de l'année de naissance et du mois de naissance**, il y a cohérence totale, interne à SISE et avec l'enquête pour 9 174 répondants ;
- **concernant l'année du baccalauréat**, il y a cohérence totale, interne à SISE et avec l'enquête pour 8 305 répondants (73 % des répondants appariés). L'information n'est pas disponible dans SISE pour 2 233 répondants (20 %). Il y a cohérence interne à SISE mais divergence entre les deux sources pour 749 répondants (7 %). Les 129 autres répondants sont des cas d'incohérence interne à SISE ;
- **si l'on combine le sexe, l'année de naissance, le mois de naissance et l'année du baccalauréat**, il y a cohérence totale pour 7 853 répondants (69 % des répondants appariés), une absence d'information sur l'année du baccalauréat dans SISE pour 2 233 répondants (20 %).

2.5. Troisième étape : l'appariement avec les fichiers de SISE-Résultats

L'appariement de l'enquête Génération 2017 avec les fichiers de SISE-Résultats a été réalisé avec des principes proches de ceux retenus pour le rapprochement de l'enquête avec les fichiers de SISE-Inscriptions.

Dans un premier temps, des fichiers globaux annuels SISE-Résultats ont été produits pour chaque année les composantes disponibles (UNIV, MANA, ING, ENS sur l'ensemble de la période, complétées par E26b à partir de 2016). Seul un nombre restreint de variables a été conservé. Par ailleurs :

- pour la composante UNIV sur toute la période et pour la composante ENS à partir de 2016, seuls les enregistrements associés à des réussites au diplôme²⁴ ont été conservés. Pour les autres composantes, les travaux exploratoires ont suggéré que seules les informations concernant les diplômés avaient été remontées. Tous les enregistrements ont donc été conservés ;
- la qualité de l'INE a été évaluée selon la même méthode que pour les inscriptions. Seuls les enregistrements associés à des INE jugés « valides » ont été conservés ;
- les observations considérées comme des doublons par le SIES²⁵ ont été supprimées ;
- les observations associées à un INE qui n'a pas été apparié à un répondant de l'enquête Génération lors de la deuxième étape ont également été supprimées.

Dans un deuxième temps, un fichier historicisé de résultats a été construit. Au préalable, les fichiers globaux annuels ont été retravaillés de façon à n'avoir qu'un enregistrement par répondant à l'enquête (IDENT). Pour chaque année, chaque répondant peut avoir jusqu'à trois diplômes décrits puisque l'analyse exploratoire a montré qu'il pouvait y avoir jusqu'à trois diplômes pour une même personne une année donnée. Par exemple, 6 805 répondants ont été retrouvés dans SISE-Résultats 2016 avec au moins un diplôme, dont 433 ont été repérés deux fois diplômés et 34 trois fois (tableau 31). Les fichiers annuels ont ensuite été

²⁴ Variable TOT=1.

²⁵ FLAG_RES=1.

appariés de façon à connaître la trajectoire de diplomation dans SISE-Résultats sur la période 2010-2016 pour les répondants à l'enquête Génération 2017.

Dans un troisième temps, ce fichier historicisé de résultats a été rapproché du fichier obtenu à l'issue de l'étape décrite dans la partie 2.4 (appariement de l'enquête Génération 2017 et du fichier historicisé des inscriptions).

Tableau 31 • Nombre de répondants à l'enquête Génération 2017 pour lesquels un diplôme est récupéré* dans les fichiers SISE-Résultats de l'année considérée

	2016	2015	2014	2013	2012	2011	2010
Ayant un 1 ^{er} diplôme repéré	6 805	1 221	2 387	1 663	905	360	183
Ayant un 2 ^e diplôme repéré	467	66	126	95	41	20	6
Ayant un 3 ^e diplôme repéré	34	5	10	8	5		

* Selon la méthode de qualification présentée dans ce document de travail.

Lecture : la réussite à un diplôme est repérée dans les fichiers SISE-Résultats de 2016 pour 6 805 répondants de l'enquête Génération.

Parmi eux, la même année, 467 ont un deuxième diplôme identifié dans les fichiers SISE.

Champ : ensemble des répondants de l'enquête Génération.

Source : appariement de Céreq enquête Génération 2017 avec le fichier historicisé de SISE élaboré à partir de SIES – SISE-Résultats 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b).

Une codification synthétique des diplômes a été réalisée de façon à faciliter l'exploitation des résultats de l'appariement. Cette codification devrait être réexaminée et validée en cas de renouvellement de l'exercice. En cas de renouvellement régulier, une veille serait également nécessaire pour tenir compte des évolutions des diplômes, de leur codification et des centres d'intérêt des politiques publiques. Cette codification s'appuie principalement sur la variable TYP_DIPR et, pour affiner les regroupements « autres », sur la variable DEGETU. Le tableau 32 présente les choix effectués.

Tableau 32 • Codification synthétique des diplômes

Modalité	Contenu à partir des variables SISE (voir sa documentation)
DUT	TYP_DIPR=CB
Capacité en droit	TYP_DIPR=AC
Licence	TYP_DIPR=DB, DC, DJ, LP
Licence professionnelle	TYP_DIPR=DP
Licence LMD	TYP_DIPR=XA
Master LMD	TYP_DIPR=XB
Master professionnel	TYP_DIPR=XD
Doctorat Université hors santé	TYP_DIPR=YB
Doctorat santé	TYP_DIPR=FH, FJ, IB, SA
D.E.S. santé	TYP_DIPR=JD, JC, JF
Autre diplôme santé	TYP_DIPR=MA, MB, MC, MD, GF, GG, HH, IC, ID, IX, JB, KG, FK, FL
Diplôme d'ingénieur	TYP_DIPR=FI, FN
Autres bac+5	TYP_DIPR=RA, UF, UH, UQ, UR, US, UT, UU, UV, UX
	Autres valeurs pour TYP_DIPR et...
Autres diplômes de degré 0	... DEGETU=0
Autres diplômes de degré 1	... DEGETU=1
Autres diplômes de degré 2	... DEGETU=2
Autres diplômes de degré 3	... DEGETU=3
Autres diplômes de degré 4	... DEGETU=4
Autres diplômes de degré 5	... DEGETU=5
Autres diplômes de degré 6	... DEGETU=6
Autres diplômes de degré inconnu	... DEGETU=7

3. Les parcours dans SISE des jeunes de la Génération 2017

3.1. Les parcours d'inscription

Les tableaux 33 à 41 résument le parcours d'inscription dans SISE sur la période 2010 à 2016 des jeunes de la Génération 2017, selon leur plus haut niveau de diplôme atteint dans l'enquête. Plus précisément, les jeunes ayant atteint un plus haut diplôme donné sont répartis chaque année selon leur situation d'inscription dans SISE :

- Sont-ils retrouvés ou non dans SISE ?
- S'ils ne sont pas retrouvés, est-ce parce qu'ils n'ont pas d'INE valide ou sont-ils retrouvés au moins une fois dans la période, mais pas l'année considérée ?
- S'ils sont retrouvés, dans quelle composante de SISE et à quel degré d'études sont-ils retrouvés²⁶ ?

Les données présentées sont pondérées.

Pour **les jeunes de la Génération 2017 ayant un doctorat comme plus haut diplôme** (tableau 33) :

- 89 % des jeunes concernés sont retrouvés dans le fichier historicisé des inscriptions en 2016. Le taux monte à 97 % si l'on se restreint à ceux disposant d'un INE valide.
 - Ces taux sont respectivement de 79 % et 86 % en 2010.
 - En 2016, 91 % de ceux qui sont retrouvés dans SISE le sont dans la composante « Université », en degré d'études bac+6.
 - En 2010, 20 % des jeunes concernés retrouvés dans SISE sont inscrits en niveau bac+6, 19 % en niveau bac+5, 24 % en niveau bac+4, 16 % en niveau bac+3. Rapportés à l'ensemble des jeunes ayant un doctorat comme plus haut diplôme, qu'ils soient retrouvés ou non dans SISE, ces taux sont respectivement de 16 %, 15 %, 19 % et 12 %.
- ⇒ Les sortants ayant un doctorat comme plus haut diplôme sont bien couverts par SISE-Inscriptions et leurs trajectoires semblent plutôt bien suivies sur la période. La profondeur historique retenue de six années avant l'année de sortie ne permet pas de retracer l'ensemble du parcours dans le supérieur pour la plupart d'entre eux. Pour atteindre ce but, un recul significativement plus important serait nécessaire, de l'ordre d'une dizaine d'années supplémentaires.

Pour **les jeunes de la Génération 2017 ayant un diplôme d'école comme plus haut diplôme** (tableau 34) :

- 59 % des jeunes concernés sont retrouvés dans le fichier historicisé des inscriptions en 2016. Le taux monte à 88 % si l'on se restreint à ceux disposant d'un INE valide.
- Ces taux sont respectivement de 10 % et 15 % en 2010.
- En 2016, 62 % de ceux qui sont retrouvés dans SISE le sont dans les composantes « MANA » ou « INGE » en degré d'études bac+5 et 29 % dans la composante « Université », également au niveau bac+5.
- La proportion de jeunes disposant d'un INE valide qui sont retrouvés dans SISE reste entre 54 % et 59 % de 2016 à 2012 puis baisse à 29 % en 2011, puis 10 % en 2010. Cette année-là, 81 % des jeunes retrouvés sont inscrits au niveau bac+1.

²⁶ Variable DEGETU.

- ⇒ Une proportion importante de sortants ayant un diplôme d'école comme plus haut diplôme n'est pas retrouvée dans SISE, pour une part importante, en raison de l'absence d'INE dans la base de sondage de l'enquête Génération (33 % de la cohorte), mais également, pour une part significative, malgré la disponibilité d'un INE dans la base de sondage. Pour ceux qui sont retrouvés dans SISE, la trajectoire synthétique observée semble cohérente avec le plus haut diplôme atteint. La profondeur historique retenue gagnerait à être repoussée d'une ou deux années supplémentaires.

Pour les jeunes de la Génération 2017 ayant un bac+5 comme plus haut diplôme, hors diplôme d'école (tableau 35) :

- 76 % des jeunes concernés sont retrouvés dans le fichier historicisé des inscriptions en 2016. Le taux monte à 93 % si l'on se restreint à ceux disposant d'un INE valide.
 - Ces taux sont respectivement de 25 % et 30 % en 2010.
 - En 2016, 86 % de ceux qui sont retrouvés dans SISE le sont dans la composante « Université », au niveau bac+5.
 - La proportion de jeunes disposant d'un INE valide qui sont retrouvés dans SISE reste entre autour des deux tiers de 2015 à 2012 puis baisse à 44 % en 2011, puis 25 % en 2010. Cette année-là, 54 % des jeunes retrouvés sont inscrits au niveau bac+1, 22 % au niveau bac+2, 8 % au niveau bac+3 et 6 % au-delà.
- ⇒ Les sortants ayant un diplôme bac+5 hors diplôme d'école comme plus haut diplôme sont bien mieux couverts par SISE que les diplômés d'école et d'autant plus quand l'INE est disponible. Toutefois, entre un cinquième et un sixième des jeunes concernés ayant un INE ne sont pas retrouvés dans les années précédant la sortie. Ce constat suggère une difficulté à retracer avec SISE l'intégralité du parcours dans l'enseignement supérieur d'une proportion significative de jeunes ayant atteint ce niveau de diplôme, même lorsqu'on dispose d'un INE. La profondeur historique retenue gagnerait à être repoussée de quatre à cinq années supplémentaires.

Pour les jeunes de la Génération 2017 ayant une licence professionnelle comme plus haut diplôme (tableau 36) :

- 72 % des jeunes concernés sont retrouvés dans le fichier historicisé des inscriptions en 2016. Le taux monte à 89 % si l'on se restreint à ceux disposant d'un INE valide.
 - Ces taux sont respectivement de 5 % et 6 % en 2010.
 - En 2016, 93 % de ceux qui sont retrouvés dans SISE le sont dans la composante « Université », au niveau bac+3.
 - La proportion de jeunes disposant d'un INE valide qui sont retrouvés dans SISE baisse à 65 % en 2015, 72 % en 2014, puis 37 % en 2013, 18 % en 2012.
 - Très peu d'inscriptions sont enregistrées sur la période au niveau bac+4 ou supérieur.
- ⇒ Les sortants ayant une licence professionnelle comme plus haut diplôme sont aussi bien couverts par SISE que les bac+5 hors diplôme d'école l'année de sortie, mais la capacité à retracer l'intégralité du parcours dans l'enseignement supérieur semble moins bonne. La profondeur historique retenue semble suffisante pour ce plus haut niveau de diplôme, même si une année supplémentaire pourrait apporter une légère amélioration.

Pour les jeunes de la Génération 2017 ayant un diplôme bac+3 comme plus haut diplôme, hors licence professionnelle et hors secteur santé/social (tableau 37) :

- 67 % des jeunes concernés sont retrouvés dans le fichier historicisé des inscriptions en 2016. Le taux monte à 90 % si l'on se restreint à ceux disposant d'un INE valide.

- Ces taux sont respectivement de 15 % et 20 % en 2010.
 - En 2016, 49 % de ceux qui sont retrouvés dans SISE le sont dans la composante « Université », au niveau bac+3, 31 % au niveau bac+4 et 8 % au niveau bac+5.
 - La proportion de jeunes disposant d'un INE valide qui sont retrouvés dans SISE baisse à 83 % en 2015 et 2014, 64 % en 2013, puis 47 % en 2012, 32 % en en 2012, puis 20 % en 2010. Cette année-là, 62 % des jeunes retrouvés sont inscrits en niveau bac+1, 23 % en niveau bac+2 et 10 % en niveau bac+3.
- ⇒ Les sortants ayant un diplôme bac+3 comme plus haut diplôme (hors licence professionnelle et hors secteur santé/social) sont aussi bien couverts par SISE que les bac+5 hors diplôme d'école l'année de sortie, avec une capacité à retracer l'intégralité du parcours dans l'enseignement supérieur qui semble analogue, donc meilleure que pour les licences professionnelles. La profondeur historique retenue n'est pas suffisante pour ce plus haut niveau de diplôme. Elle gagnerait à être complétée de quelques années.

Pour **les jeunes de la Génération 2017 ayant un diplôme bac+2 comme plus haut diplôme, hors secteur santé/social** (tableau 38) :

- 63 % des jeunes concernés sont retrouvés dans le fichier historicisé des inscriptions en 2016. Le taux monte à 78 % si l'on se restreint à ceux disposant d'un INE valide.
 - Ces taux sont respectivement de 1 % et 2 % en 2010.
 - En 2016, 71 % de ceux qui sont retrouvés dans SISE le sont dans la composante « BTS » au niveau bac+2, 14 % dans la composante « Université » au niveau bac+2 et 8 % dans la même composante au niveau bac+3.
 - La proportion de jeunes disposant d'un INE valide qui sont retrouvés dans SISE est encore de 74 % en 2015 puis baisse à 30 % en 2014, 15 % en 2013, puis 7 % en 2012.
 - Très peu d'inscriptions sont enregistrées sur la période au niveau bac+4 ou supérieur.
- ⇒ Les sortants ayant un diplôme bac+2 comme plus haut diplôme (hors secteur santé/social) sont largement couverts par SISE mais une part significative de ceux ayant un INE semblent rester hors-champ. La profondeur historique retenue est suffisante pour suivre le parcours dans l'enseignement supérieur.

Les tableaux 29 à 41 portent sur **les jeunes de la Génération 2017 ayant le baccalauréat comme plus haut diplôme**, qu'il soit général (tableau 39), technologique (tableau 40) ou professionnel (tableau 41) :

- Une proportion importante de ces jeunes n'est pas forcément attendue dans le supérieur. En 2016, 67 % sont retrouvés dans SISE pour les baccalauréats généraux, 45 % pour les baccalauréats technologiques et 12 % pour les baccalauréats professionnels (ces taux sont respectivement de 76 %, 52 % et 13 % si l'on se restreint aux jeunes disposant d'un INE valide).
- Pour les bacheliers généraux retrouvés dans SISE en 2016, 64 % étaient inscrits dans la composante « université » en bac+1, 20 % en bac+2 et 6 % en bac+3. Les inscriptions en niveau bac+4 ou plus sont très peu nombreuses
- Pour les bacheliers technologiques retrouvés dans SISE en 2016, 55 % étaient inscrits dans la composante « université » en bac+1, 18 % en première année de BTS, 14 % en deuxième année de BTS. Les inscriptions en niveau bac+3 ou plus sont négligeables.
- Pour les bacheliers professionnels retrouvés dans SISE en 2016, 35 % étaient inscrits en 1^{re} année de BTS, 26 % en deuxième année de BTS et 34 % dans la composante « université » en bac+1. Les inscriptions en niveau bac+3 ou plus sont négligeables.

Tableau 33 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **DOCTORAT**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	14 512	14 512	14 512	14 512	14 512	14 512	14 512
Sans INE*	1 229	1 229	1 229	1 229	1 229	1 229	1 229
Avec INE*, non retrouvé dans SISE** sur la période	17	17	17	17	17	17	17
Avec INE, retrouvé dans SISE mais absent l'année considérée	332	739	678	1 050	1 240	1 586	1 826
Retrouvé dans SISE l'année considérée	12 933	12 526	12 587	12 215	12 025	11 678	11 439
Proportion de retrouvés par rapport au total	89 %	86 %	87 %	84 %	83 %	80 %	79 %
Proportion de retrouvés par rapport à ceux ayant un INE	97 %	94 %	95 %	92 %	91 %	88 %	86 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
1 CPEG							94
1 INGE						22	26
1 UNIV	38					434	703
2 CPEG						218	377
2 INGE					22		33
2 UNIV			40	14	391	555	1 298
3 ENS					503	295	28
3 INGE				9		139	323
3 UNIV	14	14		446	468	1 015	1 497
4 ENS				503	228	77	175
4 INGE				24	139	315	187
4 MANA							26
4 UNIV			433	445	1 162	1 696	2 302
5 E26b	37						
5 ENS	13	124	503	181	100	201	140
5 INGE				177	498	155	46
5 MANA						26	38
5 UNIV	148	696	457	1 346	2 011	2 410	1 911
6 ENS	590	611	261	86	25	26	
6 INGE	309	351	393	276			
6 MANA		44					
6 UNIV	11 783	10 687	10 500	8 710	6 480	4 094	2 235

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Lecture : sur l'ensemble de la Génération 2017, 14 512 jeunes sont sortis de formation initiale avec un doctorat comme plus haut diplôme. Parmi eux, 1 229 n'avaient pas d'INE valide, 17 ont un INE valide, mais n'ont pas été retrouvés dans SISE-Inscriptions sur la période 2010-2016 et 332 ont été retrouvés au moins une fois sur la période, mais pas dans les fichiers de 2016. Au total, 12 933 jeunes sortis de formation initiale avec un doctorat comme plus haut diplôme sont présents dans les fichiers de SISE-Inscriptions 2016, soit 89 % de l'ensemble des jeunes ayant ce niveau de diplôme (le taux de personnes retrouvées s'élève à 97 % quand on se limite aux jeunes disposant d'un INE valide). Parmi ces 12 933 jeunes retrouvés dans SISE-Inscriptions 2016, 11 783 sont retrouvés dans la composante Université, avec une inscription de degré d'étude égal à 6 (variable DEGETU) tandis que 38 sont retrouvés dans la composante Université avec degré d'étude égal à 1 (bac+1).

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 34 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **DIPLÔME D'ÉCOLE (commerce ou ingénieur)**

Données pondérées	2016	2015	2014	2013	2012	2011	2010
Ensemble	46 114	46 114	46 114	46 114	46 114	46 114	46 114
Sans INE*	15 063	15 063	15 063	15 063	15 063	15 063	15 063
Avec INE*, non retrouvé dans SISE** sur la période	2 032	2 032	2 032	2 032	2 032	2 032	2 032
Avec INE, retrouvé dans SISE mais absent l'année considérée	1 618	2 072	2 307	3 801	3 914	15 477	24 454
Retrouvé dans SISE l'année considérée	27 401	26 947	26 712	25 219	25 105	13 543	4 566
Proportion de retrouvés par rapport au total	59 %	58 %	58 %	55 %	54 %	29 %	10 %
Proportion de retrouvés par rapport à ceux ayant un INE	88 %	87 %	86 %	81 %	81 %	44 %	15 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 INGE	34						
0 UNIV						60	
1 BTS_					667	605	254
1 CPEG				22	3 071	3 375	1 014
1 INGE			79		2 224	873	101
1 MANA		178	167	38	1 475	276	168
1 UNIV	153		138	255	6 068	5 138	2 172
2 BTS_			52	745	652	288	5
2 CPEG			22	4 710	4 074	1 021	182
2 INGE		351		2 525	663	88	
2 MANA		314	71	1 495	305	180	10
2 UNIV	35	25	78	6 385	4 096	1 139	498
3 INGE	394	83	8 106	2 513	324	29	
3 MANA	108	231	3 202	2 599	462	81	31
3 UNIV	120	162	7 236	3 027	715	299	86
4 INGE	365	8 557	2 112	242	29		
4 MANA	316	5 613	3 287	259	40		
4 UNIV	329	7 303	1 409	309	122	15	46
5 E26b	166						
5 INGE	9 721	1 389	191	29	5		
5 MANA	7 345	1 321	282	36			
5 UNIV	7 983	1 049	235		46	46	
6 INGE	262	373	24				
6 MANA	28		22				
6 UNIV	43						
9 BTS_				30	67	30	

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 35 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **DIPLÔME BAC+5 hors diplôme d'école**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	110 226	110 226	110 226	110 226	110 226	110 226	110 226
Sans INE*	20 301	20 301	20 301	20 301	20 301	20 301	20 301
Avec INE*, non retrouvé dans SISE** sur la période	3 310	3 310	3 310	3 310	3 310	3 310	3 310
Avec INE, retrouvé dans SISE mais absent l'année considérée	2 732	10 172	13 121	13 947	14 891	38 619	59 389
Retrouvé dans SISE l'année considérée	83 883	76 443	73 494	72 668	71 724	47 996	27 226
Proportion de retrouvés par rapport au total	76 %	69 %	67 %	66 %	65 %	44 %	25 %
Proportion de retrouvés par rapport à ceux ayant un INE	93 %	85 %	82 %	81 %	80 %	53 %	30 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0_ENS						33	
0 UNIV	17	14	127		58		58
1 BTS_				158	1 607	1 245	730
1 CPEG					581	1 372	973
1 ENS					40		
1 INGE					214	313	186
1 MANA			37	22	487	349	176
1 UNIV	617	319	687	966	29 885	26 637	14 817
2 BTS_	9		121	1 905	1 283	740	635
2 CPEG			27	739	1 099	701	445
2 E26b							
2 ENS	21			40			
2 INGE			15	193	299	67	
2 MANA		73	31	529	355	144	
2 UNIV	231	314	772	33 882	20 888	9 271	4 979
3 E26b	96						
3 ENS		58	40		68	101	84
3 INGE		61	807	529	242	94	61
3 MANA		116	799	564	494	109	53
3 UNIV	874	1 057	41 948	20 848	8 620	3 736	2 058
4 E26b	72						
4 ENS	58	498	165	97	82		
4 INGE	128	1 040	555	181	57		63
4 MANA	134	661	700	483	31	61	12
4 UNIV	4 327	55 858	19 319	7 568	3 242	1 689	829
5 E26b	403						
5 ENS	624	200	117	94			
5 INGE	1 495	520	115	14	49	63	
5 MANA	1 189	865	421	12	12	12	
5 UNIV	72 287	13 631	5 596	2 766	1 144	755	711

6 ENS	44	44	44				
6 INGE	35	55					
6 UNIV	1 014	772	591	510	424	253	42
9 BTS_	209	288	458	469	464	252	314

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 36 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme une LICENCE PROFESSIONNELLE

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	33 139	33 139	33 139	33 139	33 139	33 139	33 139
Sans INE*	6 566	6 566	6 566	6 566	6 566	6 566	6 566
Avec INE*, non retrouvé dans SISE** sur la période	1 558	1 558	1 558	1 558	1 558	1 558	1 558
Avec INE, retrouvé dans SISE mais absent l'année considérée	1 248	7 627	6 001	15 093	20 133	21 944	23 366
Retrouvé dans SISE l'année considérée	23 766	17 387	19 013	9 921	4 882	3 071	1 649
Proportion de retrouvés par rapport au total	72 %	52 %	57 %	30 %	15 %	9 %	5 %
Proportion de retrouvés par rapport à ceux ayant un INE	89 %	65 %	72 %	37 %	18 %	12 %	6 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 UNIV						59	
1 BTS_	39	31	6 159	1 847	543	339	159
1 CPEG			46	112	51	81	
1 INGE				67	13		15
1 MANA					27	24	
1 UNIV	163	421	6 437	4 506	2 023	1 841	1 102
2 BTS_	31	6 667	1 990	623	379	143	178
2 CPEG		51	48			41	
2 INGE		21					
2 MANA					24		
2 UNIV	148	7 284	2 298	1 353	1 201	334	90
3 E26b	40						
3 ENS							56
3 INGE	52		21				
3 MANA			24	24			
3 UNIV	22 088	1 995	1 316	778	327	191	49
4 INGE	45						
4 MANA	24	48	25	20			
4 UNIV	862	741	297	124	48		
5 MANA	45	20	20				
5 UNIV	197	44	15				
6 UNIV	18						
9 BTS_	5	64	319	467	246	19	
9 INGE	8						

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 37 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **DIPLOME BAC+3 (hors licence professionnelle)**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	43 623	43 623	43 623	43 623	43 623	43 623	43 623
Sans INE	10 986	10 986	10 986	10 986	10 986	10 986	10 986
Avec INE, non retrouvé dans SISE sur la période	1 360	1 360	1 360	1 360	1 360	1 360	1 360
Avec INE, retrouvé dans SISE mais absent l'année considérée	1 918	4 130	4 275	10 319	15 868	20 779	24 593
Retrouvé dans SISE l'année considérée	29 359	27 148	27 002	20 959	15 409	10 498	6 685
Proportion de retrouvés par rapport au total	67 %	62 %	62 %	48 %	35 %	24 %	15 %
Proportion de retrouvés par rapport à ceux ayant un INE	90 %	83 %	83 %	64 %	47 %	32 %	20 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 INGE					43		
0 UNIV	162	140			81	40	40
1 BTS_		235	559	583	423	599	314
1 CPEG			53	119	138	50	140
1 ENS					31	23	
1 INGE			22		31	31	
1 MANA			369	55	12		
1 UNIV	855	685	8 350	8 809	6 577	5 481	3 720
2 BTS_		635	525	295	640	333	34
2 CPEG	189	126	52	145	14	192	190
2 E26b	89						
2 ENS				31	23		
2 INGE		22		31			
2 MANA		369	90	12			
2 UNIV	1 132	9 761	7 736	5 401	4 391	2 381	1 307
3 E26b	71						
3 ENS			31	23			
3 INGE	22	81	5				
3 MANA	466	129	41	87	48		
3 UNIV	14 409	8 221	5 527	3 636	1 926	1 057	647
4 E26b	103						
4 ENS						41	
4 INGE	47						
4 MANA	154	125	87				
4 UNIV	9 019	5 542	2 829	1 319	734	183	107
5 ENS				41	41		
5 INGE	10						
5 MANA	146						
5 UNIV	2 307	878	360	52	99	55	85
6 UNIV	29						
9 BTS_	149	199	369	321	156	32	103

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 38 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **DIPLÔME BAC+2 hors santé/social**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	79 715	79 715	79 715	79 715	79 715	79 715	79 715
Sans INE*	15 092	15 092	15 092	15 092	15 092	15 092	15 092
Avec INE*, non retrouvé dans SISE** sur la période	10 653	10 653	10 653	10 653	10 653	10 653	10 653
Avec INE, retrouvé dans SISE mais absent l'année considérée	3 780	6 331	34 611	44 230	49 644	52 153	52 984
Retrouvé dans SISE l'année considérée	50 189	47 639	19 359	9 739	4 326	1 817	985
Proportion de retrouvés par rapport au total	63 %	60 %	24 %	12 %	5 %	2 %	1 %
Proportion de retrouvés par rapport à ceux ayant un INE	78 %	74 %	30 %	15 %	7 %	3 %	2 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 UNIV		22	20		44		
1 BTS_	164	30 620	6 689	1 667	546	246	215
1 CPEG			73	103	134	94	42
1 INGE			86	25	24	5	
1 MANA			13				
1 UNIV	1 939	5 554	7 057	5 178	2 582	1 074	593
2 BTS_	35 855	7 054	1 351	652	287	168	10
2 CPEG		21			15	5	
2 E26b	22						
2 INGE	36			24	5		
2 MANA	16		22				
2 UNIV	7 123	2 709	1 935	1 329	460	224	125
3 INGE	71		93	8	5		
3 MANA				42			
3 UNIV	4 161	1 035	407	341	170		
4 INGE		93					
4 MANA		70	42				
4 UNIV	57	37	26				
5 INGE	121						
5 MANA	70						
5 UNIV	63	26					
9 BTS_	494	399	1 546	370	52		

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 39 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **BACCALAURÉAT GÉNÉRAL**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	67 728	67 728	67 728	67 728	67 728	67 728	67 728
Sans INE*	8 113	8 113	8 113	8 113	8 113	8 113	8 113
Avec INE*, non retrouvé dans SISE** sur la période	12 770	12 770	12 770	12 770	12 770	12 770	12 770
Avec INE, retrouvé dans SISE mais absent l'année considérée	1 676	20 342	30 237	37 977	41 086	43 195	44 965
Retrouvé dans SISE l'année considérée	45 168	26 502	16 608	8 867	5 758	3 650	1 880
Proportion de retrouvés par rapport au total	67 %	39 %	25 %	13 %	9 %	5 %	3 %
Proportion de retrouvés par rapport à ceux ayant un INE	76 %	44 %	28 %	15 %	10 %	6 %	3 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 UNIV	166	12	53				
1 BTS_	1 323	1 169			207		
1 CPEG	21		36	249	91	372	92
1 ENS		91	129				
1 INGE				121	104	50	
1 MANA			99				
1 UNIV	29 004	17 169	11 731	5 451	3 457	2 342	1 420
2 BTS_	1 275	132					
2 CPEG		36	216	213	326	92	38
2 ENS	91	38					
2 INGE		25	42	79	154		
2 UNIV	8 964	5 005	2 529	1 688	877	402	112
3 ENS	38						
3 INGE	129	17	222	80			
3 MANA		116					
3 UNIV	2 853	1 521	789	485	286	114	
4 INGE	138	252	80				
4 UNIV	174	227	46	131	38		
5 INGE	182	50					
5 UNIV	490	271	309	38	0	140	78
6 UNIV	159	159	159	159	159	78	140
9 BTS_	162	212	166	59	59	59	

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 40 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **BACCALAURÉAT TECHNOLOGIQUE**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	39 308	39 308	39 308	39 308	39 308	39 308	39 308
Sans INE*	5 073	5 073	5 073	5 073	5 073	5 073	5 073
Avec INE*, non retrouvé dans SISE** sur la période	15 553	15 553	15 553	15 553	15 553	15 553	15 553
Avec INE, retrouvé dans SISE mais absent l'année considérée	878	9 879	14 815	17 143	18 003	18 453	18 419
Retrouvé dans SISE l'année considérée	17 805	8 804	3 867	1 540	680	230	264
Proportion de retrouvés par rapport au total	45 %	22 %	10 %	4 %	2 %	1 %	1 %
Proportion de retrouvés par rapport à ceux ayant un INE	52 %	26 %	11 %	4 %	2 %	1 %	1 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 UNIV	51						
1 BTS_	3 163	2 905	919	168			59
1 CPEG		25	85		5	5	
1 INGE	63				5		
1 MANA	34						
1 UNIV	9 864	4 621	2 450	1 324	618	172	206
2 BTS_	2 475	391	31				
2 INGE			5				
2 UNIV	1 736	722	247	13	53	53	
3 INGE	42	5	8				
3 UNIV	106	8					
4 INGE	8	8					
4 UNIV	65						
9 BTS_	199	120	124	31			

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau 41 • Parcours rétrospectif dans SISE des jeunes de la Génération 2017 ayant comme plus haut diplôme un **BACCALAURÉAT PROFESSIONNEL**

<i>Données pondérées</i>	2016	2015	2014	2013	2012	2011	2010
Ensemble	126 993	126 993	126 993	126 993	126 993	126 993	126 993
Sans INE*	18 288	18 288	18 288	18 288	18 288	18 288	18 288
Avec INE*, non retrouvé dans SISE** sur la période	90 817	90 817	90 817	90 817	90 817	90 817	90 817
Avec INE, retrouvé dans SISE mais absent l'année considérée	3 214	9 702	14 930	16 821	17 215	17 370	17 719
Retrouvé dans SISE l'année considérée	14 674	8 187	2 958	1 068	673	518	169
Proportion de retrouvés par rapport au total	12 %	6 %	2 %	1 %	1 %	0 %	0 %
Proportion de retrouvés par rapport à ceux ayant un INE	13 %	8 %	3 %	1 %	1 %	0 %	0 %
Répartition des jeunes retrouvés l'année considérée par degré d'inscription et composante SISE							
0 UNIV	17		13				
1 BTS_	5 209	4 731	1 402	119	201	79	
1 CPEG			5			25	
1 MANA			48				
1 UNIV	5 011	2 392	1 394	762	424	367	86
2 BTS_	3 862	814	33	138			
2 CPEG		5					
2 MANA	48	48	67				
2 UNIV	413	137					
3 E26b	41						
3 UNIV	73	11					
9 BTS_		50		48	48	48	82

* Selon la méthode de qualification de l'INE présentée dans ce document.

** Retrouvé au moins une fois sur la période 2010-2016, avec les restrictions de champ présentées dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017 et SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

3.2. Les parcours de diplomation dans le supérieur

Pour cette analyse, nous n'avons pas cherché une comparaison stricte et systématique des informations recueillies dans l'enquête avec les données administratives car elle aurait demandé un travail lourd de comparaison et de rapprochement des nomenclatures utilisées dans les deux sources. Nous nous sommes limités à quelques comparaisons.

Le tableau 42 résume le parcours de diplomation des répondants de l'enquête Génération 2017 à partir des informations disponibles dans SISE-Résultats sur la période 2010 à 2016. Ces parcours sont proposés selon le plus haut niveau de diplôme atteint (les données présentées ne sont pas pondérées).

Tableau 42 • Nombre de répondants à l'enquête Génération 2017 ayant des diplômes repérés* dans les fichiers SISE-Résultats, selon leur plus haut diplôme atteint dans l'enquête

Plus haut diplôme atteint selon l'enquête :	Nombre de répondants	Absent de SISE	Pour les répondants à l'enquête retrouvés au moins une fois dans SISE sur la période 2010-2016, nombre de diplômes repérés dans SISE...								
			... de niveau bac+1			... de niveau bac+2			... de niveau bac+3		
			Aucun	Un seul	2 ou +	Aucun	Un seul	2 ou +	Aucun	Un seul	2 ou +
01- Non diplômés	1 779										
02- CAP-BEP-MC-divers niv.3	2 531										
03- Bac pro-BT-divers niv.4	4 955	4 360	586	5		586	5		591	0	
04- Bac technologique	951	498	445	5		445	5		450	0	
05- Bac général	959	248	678	31		698	11		704	5	
06- Bac+2 sauf santé/social	2 948	1 018	1 884	46		1 507	404	19	1 915	15	
07- Bac+2 à +4 santé/social	623	472	129	8		131	6		103	34	
08- Licence professionnelle	1 463	433	984	46		704	312	14	59	911	60
09- Autres bac+3 et bac+4	1 649	514	1 065	69	5	945	181	8	168	891	75
10- Bac+5 sauf écoles	4 431	1 047	3 187	192	5	2 734	622	28	818	2 320	246
11- Écoles (ingé., commerce)	2 350	765	1 547	38		1 268	300	17	1 351	210	24
12- Doctorat	525	66	459	0		441	18		396	63	

Plus haut diplôme atteint selon l'enquête :	Nombre de répondants	Absent de SISE	Pour les répondants à l'enquête retrouvés au moins une fois dans SISE sur la période 2010-2016, nombre de diplômes repérés dans SISE...								
			... de niveau bac+4			... de niveau bac+5			... de niveau bac+6		
			Aucun	Un seul	2 ou +	Aucun	Un seul	2 ou +	Aucun	Un seul	2 ou +
01- Non diplômés	1 779										
02- CAP-BEP-MC-divers niv.3	2 531										
03- Bac pro-BT-divers niv.4	4 955	4 360	591			591			591		
04- Bac technologique	951	498	450			450			450		
05- Bac général	959	248	709			702	7		704	5	
06- Bac+2 sauf santé/social	2 948	1 018	1 930			1 925	5		1 930		
07- Bac+2 à +4 santé/social	623	472	137			137			137		
08- Licence professionnelle	1 463	433	1 025	5		1 030			1 030		
09- Autres bac+3 et bac+4	1 649	514	1 126	8		1 116	18		1 134		
10- Bac+5 sauf écoles	4 431	1 047	3 271	108		275	2 703	406	3 364	12	8
11- Écoles (ingé., commerce)	2 350	765	1 542	38		173	1 291	121	1 568	17	
12- Doctorat	525	66	450	9		249	178	32	36	234	189

* Selon la méthode de qualification présentée dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC), SIES – SISE-Résultats 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b).

Sans détailler l'ensemble des résultats de ce tableau, on peut notamment noter les résultats suivants :

- **Sur les 459 répondants ayant un doctorat comme plus haut diplôme** et retrouvés au moins une fois dans SISE sur la période²⁷, 423 (92 %) ont au moins un diplôme de niveau bac+6 identifié dans SISE-Résultats, 210 (46 %) ont au moins un diplôme de niveau bac+5 identifié et 43 (14 %) au moins un diplôme de niveau bac+3.
- **Sur les 1 585 répondants ayant un diplôme d'école d'ingénieurs ou de commerce comme plus haut diplôme** et retrouvés au moins une fois dans SISE sur la période²⁸, 1 412 (89 %) ont au moins un diplôme de niveau bac+5 identifié dans SISE-Résultats, 234 (7 %) ont au moins un diplôme de niveau bac+3 identifié et 317 (9 %) au moins un diplôme de niveau bac+2. *A contrario*, 17 répondants (1 %) ont réussi un diplôme de niveau bac+6 d'après SISE-Résultats.
- **Sur les 3 384 répondants ayant un diplôme bac+5 (hors diplôme d'école) comme plus haut diplôme** et retrouvés au moins une fois dans SISE sur la période²⁹, 3 109 (92 %) ont au moins un diplôme de niveau bac+5 identifié dans SISE-Résultats, 2 566 (76 %) ont au moins un diplôme de niveau bac+3 identifié et 650 (19 %) au moins un diplôme de niveau bac+2. *A contrario*, 20 répondants (0,6 %) ont réussi un diplôme de niveau bac+6 d'après SISE-Résultats.
- **Sur les 1 135 répondants ayant une licence professionnelle comme plus haut diplôme** et retrouvés au moins une fois dans SISE sur la période³⁰, 971 (86 %) ont au moins un diplôme de niveau bac+3 identifié dans SISE-Résultats et 326 (29 %) ont au moins un diplôme de niveau bac+2 identifié et 650 (19 %). *A contrario*, 5 répondants (0,4 %) ont réussi un diplôme de niveau bac+4 d'après SISE-Résultats.
- **Pour les répondants ayant un baccalauréat comme plus haut diplôme, le nombre de ceux identifiés dans SISE-Résultats comme ayant réussi un diplôme supérieur est également non nul, mais négligeable.**

Le tableau 43 précise, par niveau de plus haut diplôme atteint dans l'enquête, la proportion de répondants ayant un type de diplôme donné dans SISE-Résultats. Il montre notamment qu'en se limitant aux répondants retrouvés au moins une fois dans SISE :

- 49 % des répondants ayant un doctorat comme plus haut diplôme à l'enquête ont réussi, une thèse hors santé sur la période 2006-2016, 33 % un DES santé, 33 % un master LMD, 11 % une licence LMD, d'après SISE.
- 67 % des répondants ayant un diplôme d'école comme plus haut diplôme à l'enquête ont obtenu un diplôme d'ingénieur sur la période 2006-2016, 7 % un master LMD, 9 % une licence LMD, 19 % un DUT, d'après SISE.
- 75 % des répondants ayant un diplôme bac+5 (hors diplôme d'école) comme plus haut diplôme à l'enquête ont obtenu un master LMD sur la période 2006-2016, 7 % un master professionnel, 70 % une licence LMD, 13 % un DUT, d'après SISE.
- 93 % des répondants ayant une licence professionnelle comme plus haut diplôme à l'enquête ont obtenu une licence professionnelle sur la période 2006-2016, 5 % une licence LMD, 30 % un DUT, d'après SISE.
- 82 % des répondants ayant un autre diplôme bac+3 ou bac+4 comme plus haut diplôme à l'enquête ont obtenu une licence LMD sur la période 2006-2016, 10 % un DUT, d'après SISE.
- 21 % des répondants ayant un bac+2 hors secteur santé/social comme plus haut diplôme à l'enquête ont obtenu un DUT sur la période 2006-2016, d'après SISE.

²⁷ 66 répondants ayant un doctorat comme plus haut diplôme n'ont pas été retrouvés dans SISE, soit 13 % des répondants concernés.

²⁸ 765 répondants ayant un diplôme d'école en plus haut diplôme n'ont pas été retrouvés dans SISE, soit 33 % des répondants concernés.

²⁹ 1 047 répondants ayant un diplôme bac+5 (hors école) en plus haut diplôme n'ont pas été retrouvés dans SISE, soit 24 % des répondants concernés.

³⁰ 433 répondants ayant une licence professionnelle en plus haut diplôme n'ont pas été retrouvés dans SISE, soit 30 % des répondants concernés.

Tableau 43 • Proportion des répondants à l'enquête Génération 2017 repérés dans SISE-Résultats comme ayant eu certains diplômes, selon leur plus haut diplôme atteint à l'enquête

Plus haut diplôme à l'enquête :	Non retrouvés dans SISE-Inscriptions	Proportion de répondants à l'enquête du Céreq, retrouvés au moins une fois dans SISE en ayant au moins :							
		un DUT	une licence pro.	une licence LMD	un master LMD	un master pro	un diplôme d'ingénieur	une thèse hors santé	un DES santé
01- Non diplômés	100 %	0	0	0	0	0	0	0	0
02- CAP-BEP-MC-divers niv.3	100 %	0	0	0	0	0	0	0	0
03- Bac pro-BT-divers niv.4	88 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
04- Bac technologique	53 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
05- Bac général	26 %	0 %	0 %	1 %	0 %	0 %	1 %	0 %	0 %
06- Bac+2 sauf santé/social	35 %	21 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
07- Bac+2 à +4 santé/social	78 %	4 %	0 %	7 %	0 %	0 %	0 %	0 %	0 %
08- Licence professionnelle	30 %	30 %	93 %	5 %	0 %	0 %	0 %	0 %	0 %
09- Autres bac+3 et bac+4	31 %	10 %	2 %	82 %	1 %	0 %	0 %	0 %	0 %
10- Bac+5 sauf écoles	24 %	13 %	6 %	70 %	75 %	7 %	3 %	0 %	0 %
11- Écoles (Ingé., commerce)	33 %	19 %	3 %	9 %	7 %	0 %	67 %	0 %	0 %
12- Doctorat	13 %	1 %	0 %	11 %	33 %	1 %	8 %	49 %	33 %

* Selon la méthode de qualification présentée dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC), SIES – SISE-Résultats 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b).

3.3. Le plus haut diplôme identifié

Pour conclure, nous comparons le plus haut diplôme atteint par les répondants d'après leur réponse à l'enquête (variable PHD) au diplôme de plus haut niveau atteint sur la période 2006-2016 dans SISE-Résultats. Cette comparaison est réalisée par plus haut diplôme atteint dans l'enquête. Le tableau 44 donne les résultats pour les répondants (données non pondérées). Le tableau 45 donne la distribution des résultats pour les données pondérées.

Ces tableaux montrent que, lorsque les individus ont pu être retrouvé dans SISE, il existe des divergences entre les deux sources sur le plus haut niveau de diplôme atteint. Elles sont globalement plutôt marginales, mais non nulles pour ceux atteignant au mieux la licence professionnelle, un bac+5 ou une thèse. Pour les autres diplômes de bac+3 à bac+4 et pour les bac+2, la proportion d'individus retrouvés dans SISE-Inscriptions mais pas dans SISE-Résultats est plus significative.

Tableau 44 • Répartition des répondants à l'enquête Génération 2017 selon leur plus haut diplôme dans l'enquête et leur plus haut diplôme repéré dans SISE-Résultats

Plus haut diplôme à l'enquête :	Absent de SISE-Inscriptions	Dans SISE-Inscriptions sans diplôme dans SISE-Résultats	Diplôme de niveau non renseigné	Plus haut diplôme calculé à partir des fichiers SISE 2010 à 2016						
				Bac+0	Bac+1	Bac+2	Bac+3	Bac+4	Bac+5	Bac+6
01- Non diplômés	1777	0	0							
02- CAP-BEP-divers niv.3	2526	5	0							
03- Bac pro-BT-divers niv.4	4364	580	5		5	6	5			
04- Bac technologique	501	442	0		5					
05- Bac général	250	654	0	5	30	9			7	5
06- Bac+2 sauf santé/social	1018	1475	5		15	416	14		5	
07- Bac+2 à +4 santé/social	486	68	26		5	0	37			
08- Licence professionnelle	433	46	5		0	6	971			
09- Autres bac+3 et bac+4	515	131	6		5	17	950	8	18	
10- Bac+5 sauf écoles	1047	90	7		8	6	137	18	3099	20
11- Écoles (Ingé., commerce)	765	68	0			18	52	31	1399	17
12- Doctorat	66	12	5						21	423

Note : certaines cases sont regroupées pour respecter le secret statistique.

* Selon la méthode de qualification présentée dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC), SIES – SISE-Résultats 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b).

Tableau 45 • Répartition de la Génération 2017 selon leur plus haut diplôme dans l'enquête et leur plus haut diplôme repéré dans SISE-Résultats (données pondérées)

Plus haut diplôme à l'enquête :	Absent de SISE-Inscriptions	Dans SISE-Inscriptions sans diplôme dans SISE-Résultats	Diplôme de niveau non renseigné	Plus haut diplôme calculé à partir des fichiers SISE 2010 à 2016						
				Bac+0	Bac+1	Bac+2	Bac+3	Bac+4	Bac+5	Bac+6
01- Non diplômés	99,8 %	0,2 %								
02- CAP-BEP-MC-divers niv.3	99,8 %	0,2 %								
03- Bac pro-BT-divers niv.4	85,9 %	13,7 %	0,2 %		0,1 %	0,1 %				
04- Bac technologique	52,5 %	47,0 %			0,3 %	0,2 %				
05- Bac général	30,8 %	64,3 %		0,2 %	2,7 %	0,7 %	0,3 %		0,5 %	0,4 %
06- Bac+2 sauf santé/social	32,3 %	57,0 %	0,2 %		0,5 %	9,5 %	0,5 %	0,0 %	0,1 %	
07- Bac+2 à +4 santé/social	75,4 %	13,3 %	3,5 %		0,9 %	0,7 %	6,1 %	0,1 %		
08- Licence professionnelle	24,5 %	3,8 %	0,5 %		0,0 %	0,4 %	70,5 %	0,1 %	0,1 %	
09- Autres bac+3 et bac+4	28,3 %	8,9 %	0,5 %	0,1 %	0,2 %	1,1 %	58,8 %	0,7 %	1,4 %	
10- Bac+5 sauf écoles	21,4 %	2,4 %	0,1 %		0,2 %	0,1 %	3,5 %	0,5 %	71,2 %	0,5 %
11- Écoles (Ingé., commerce)	37,1 %	3,3 %	0,0 %			0,7 %	2,4 %	1,3 %	54,6 %	0,6 %
12- Doctorat	8,6 %	2,0 %	0,3 %						4,8 %	84,3 %

Note : certaines cases sont regroupées pour respecter le secret statistique.

* Selon la méthode de qualification présentée dans ce document.

Champ : ensemble des répondants de l'enquête Génération (hors post-initiaux du sport).

Source : appariement Céreq – Enquête Génération 2017, SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC), SIES – SISE-Résultats 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b).

Annexe 1. Fiche programme

Rapprochement exploratoire des données administratives de scolarité avec l'enquête Génération 2017

Fiche n° 864 - Unité de rattachement : DIR - Direction scientifique

Retour
Mettre à jour

 Version 0 publiée le 28 septembre 2020 - Fiche validée

 COUPPIE Thomas (DEEVA), JUGNOT Stéphane (DIR), VIGNALE Mélanie (DEEVA)

Le projet impliquera le DEEVA.

 DONNEE STATISTIQUE FRANCE METHODOLOGIE METHODOLOGIE D'ENQUETE ENQUETE GENERATION 2017

INFORMATIONS GÉNÉRALES  **OMT 2019-2022** Entre brouillage des frontières et enjeux de mobilité, comment se construisent les parcours des individus ?

Pourquoi ? (objectifs et hypothèses)

Le recours aux données administratives sur la scolarité est envisagé pour réduire la taille du questionnaire de l'enquête Génération, à compter de la Génération 2021. Le travail exploratoire envisagé vise à expertiser la faisabilité d'un tel rapprochement et ses conséquences pour le dispositif Génération (apports et limites).

Comment ? (méthode et déroulé)

Ce travail exploratoire serait réalisé en rapprochant les données administratives détenues par la DEPP et le SIES aux réponses des enquêtés de la génération 2017, collectées fin 2020. Il implique donc un accord et une convention de mise à disposition des données avec chacun de ces deux partenaires. Il ne pourra être réalisé qu'une fois les résultats de l'enquête Génération 2017 disponibles.

Deux options seront à envisager :

- soit un rapprochement pour les seuls répondants, qui permettra de comparer les parcours déclarés aux parcours enregistrés dans les données administratives de la DEPP et du SIES.
- soit un rapprochement sur l'ensemble des enquêtés, afin d'examiner aussi l'apport possible des données administratives dans le traitement des non réponses.

Annexe 2. Extraits du questionnaire de la première interrogation de l'enquête Génération 2017

Le Céreq ne produisant plus de questionnaire papier, les éléments présentés ci-dessous sont issus des spécifications détaillées du questionnaire, réalisées à l'attention du prestataire de collecte pour les pilotes de l'enquête.

MODULE DE CONTROLE PREALABLE A LA PASSATION DU QUESTIONNAIRE

La base de sondage de l'enquête fournit des informations sur l'établissement d'inscription et la formation suivie pour l'année scolaire 2016/2017. Le questionnaire commence par valider ou corriger ces informations pour s'assurer que l'individu contacté est la bonne personne et entre dans le champ de l'enquête Génération. En particulier, il faut s'assurer que la personne n'a pas poursuivi ses études après l'année scolaire 2016/2017 et qu'elle ne les avait pas déjà interrompues avant. Dans ce module de contrôle, les jeunes inscrits dans une classe de niveau terminal d'un diplôme sont aussi interrogés pour savoir s'ils ont obtenu ce diplôme. Après le module de contact et de vérification d'appartenance au champ de l'enquête, le questionnaire aborde le parcours scolaire puis le plus haut diplôme obtenu.

=> **classe de sortie l'année scolaire 2006/2007 et obtention ou non du diplôme en fin de scolarité pour ceux inscrits dans un niveau terminal de diplôme.**

MODULE : PARCOURS SCOLAIRE

PS010 - Pour quelles raisons avez-vous arrêté vos études en <FINETU> ?

Vous étiez lassé de faire des études Oui=1/Non=0

Vous aviez atteint le niveau de diplôme souhaité..... Oui=1/Non=0

Vos résultats étaient insuffisants pour continuer dans la filière souhaitée..... Oui=1/Non=0

Vous avez été refusé dans une formation ou vous avez échoué à un concours Oui=1/Non=0

Vous n'avez pas trouvé d'employeur pour une formation en alternance

..... 0

oui=1/Non=0.....

Vous souhaitiez travailler ou vous aviez trouvé un emploi..... Oui=1/Non=0

Parce que la formation souhaitée n'existait pas à proximité..... Oui=1/Non=0

Pour des raisons personnelles ou familiales : enfant, santé, déménagement. Oui=1/Non=0

Pour des raisons financières..... Oui=1/Non=0

LOG180 - Des problèmes d'accès à un logement ont-ils conduit à l'arrêt de vos études ?

Oui 1

Non 2

LA SIXIEME (QUESTION ABORDEE POUR TOUS)

PS020 - Nous allons vous poser quelques questions sur votre parcours scolaire. Avez-vous redoublé avant d'entrer en sixième ?

- Oui 1
 Non 2
 Vous n'avez pas suivi de sixième 3
 Vous avez suivi une sixième à l'étranger 4

LA TROISIEME ET L'ORIENTATION POST TROISIEME

(PARTIE POSEE UNIQUEMENT AUX JEUNES SORTIS³¹ DU SECONDAIRE APRES LA CLASSE DE 3EME OU DE L'ENSEIGNEMENT SUPERIEUR JUSQU'AU NIVEAU BAC+3 INCLUS)

PS040 - Quelle classe de troisième avez-vous suivi ?

- Troisième générale 1
 Troisième SEGPA, ULIS 2
 Troisième prépa-pro 3
 Vous n'avez pas suivi de classe de troisième 4
 Vous avez suivi une classe de troisième à l'étranger 5
 Autre 6

-> **PS040MC**
Veillez préciser :

PS050 - Avez-vous redoublé au collège ?

(Restriction de champ : question posée à ceux ayant répondu à PS040)

- Oui 1
 Non 2

PS060 - Quelle classe avez-vous suivi après la troisième ?

(Restriction de champ : question posée à ceux ayant répondu à PS040 une modalité différente de 4)

- Une seconde générale ou technologique 1
 Une seconde professionnelle 2
 Une première année de CAP 3
 Vous n'avez pas suivi de classe après la troisième 4
 Vous avez suivi une classe de seconde à l'étranger 5
 Autre 6

-> **PS060MC**
Veillez préciser :

PS070 - Était-ce ?

(Restriction de champ : question posée à ceux ayant répondu 2 ou 3 à PS060)

- En apprentissage 1
 Par voie scolaire 2

PS080 - Au moment de cette orientation, auriez-vous préféré suivre cette formation par apprentissage ?

(Restriction de champ : question posée à ceux ayant répondu 2 à PS070)

- Oui 1
 Non 2

³¹ Le niveau de sortie correspond au niveau auquel le jeune est inscrit au cours de l'année scolaire 2016/2017, indépendamment de son parcours scolaire antérieur.

PS090 - Votre orientation correspondait-elle au premier vœu de votre dossier d'orientation ?

(Restriction de champ : question posée à ceux ayant répondu 1, 2, 3 ou 6 à PS060)

Oui 1
Non 2

PS100 - Votre premier vœu était-il une seconde générale ou technologique ?

(Restriction de champ : question posée à ceux ayant répondu 2, 3 ou 6 à PS060 et 2 à PS090)

Oui 1
Non 2

PS110 – Après la troisième, vous n'avez pas suivi de formation par apprentissage car :

(Restriction de champ : question posée à ceux ayant répondu 1 à PS080)

Vous n'avez pas trouvé d'employeur Oui=1/Non=0
Vous n'avez pas trouvé de place en CFA Oui=1/Non=0
Pour une autre raison Oui=1/Non=0

LE BACCALAUREAT ET L'ORIENTATION POST TERMINALE

PS120 - Quel bac avez-vous obtenu ?

(Restriction de champ : question posée aux jeunes sortis à un niveau Bac+1 ou plus pour lesquels le baccalauréat n'est pas connu dans la base de sondage)

Un bac général 1
Un bac technologique 2
Un bac professionnel 3
Vous n'avez pas obtenu de bac 4
Vous avez obtenu un bac à l'étranger 5

PS130T - Dans quelle série ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac général » à PS120)

L 1
ES 2
S 3
Autre 4 -> **PS130MC**
Veillez préciser :

PS140T - Dans quelle série ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac technologique » à PS120)

STMG 1
STI2D 2
STD2A 3
STL 4
ST2S 5
STAV 6
TMD 7
STHR, Hôtellerie 8
Autre 9 -> **PS140MC**
Veillez préciser :

PS150 - Dans quelle spécialité ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac professionnel » à PS120)

[ouverture d'un menu listant les bacs professionnels]

PS160 - Comment avez-vous réalisé votre bac professionnel ?

(Restriction de champ : question posée aux jeunes ayant répondu 3 à PS120)

- Par la voie scolaire classique 3
- En apprentissage 1
- En contrat de professionnalisation..... 2

PS170 - Avez-vous obtenu une mention au Bac ?

(Restriction de champ : question posée aux jeunes pour lesquels un baccalauréat est disponible dans la base de sondage ou qui ont indiqué à PS120 avoir eu le bac en répondant à la modalité 1, 2 ou 3)

- Non, aucune ou passable 1
- Oui, assez bien 2
- Oui, bien 3
- Oui, très bien 4
- Vous avez obtenu un Bac à l'étranger..... 5

PS180 - En quelle année avez-vous obtenu votre Bac ?

(Restriction de champ : question posée aux jeunes concernés par la question PS170 pour lesquels l'année du bac n'est pas connu dans la base de sondage³²)

PS190 - Avez-vous néanmoins suivi une année de Terminale ?

(Restriction de champ : question posée aux jeunes ayant répondu ne pas avoir eu le bac à PS120)

- Oui, en France..... 1
- Oui (ou équivalent), à l'étranger 2
- Non 3

PS210 – L'année du bac, avez-vous candidaté en :

(Restriction de champ : question posée aux jeunes pour lesquels un baccalauréat est disponible dans la base de sondage, ou à ceux qui ont indiqué à PS120 avoir eu le bac en répondant à la modalité 1, 2 ou 3, ou à ceux qui ont indiqué avoir suivi une année de terminale en France à PS190)

- Première année de licence universitaire..... Oui=1/Non=0
- IUT Oui=1/Non=0
- BTS ou BTS agricole Oui=1/Non=0
- Classe préparatoire aux grandes écoles..... Oui=1/Non=0
- École d'ingénieur Oui=1/Non=0
- École de commerce Oui=1/Non=0
- PACES, médecine..... Oui=1/Non=0
- Formation dans le domaine sanitaire et social Oui=1/Non=0
- École d'art ou d'architecture Oui=1/Non=0
- Autre formation.....Oui=1/Non=0 -> **PS210MC**

Veillez préciser :

³² À confirmer car ANBAC n'est pas dans la liste des variables d'import.

PS220 - Finalement, qu'avez-vous fait immédiatement après la terminale ?

(Restriction de champ : question posée aux jeunes entrant dans le champ de PS210 sauf ceux qui sont sortis de formation initiale au niveau du bac en 2006/2007)

Première année de licence universitaire.....	01
IUT	02
BTS ou BTS agricole	03
Classe préparatoire aux grandes écoles	04
École d'ingénieur	05
École de commerce	06
Première année commune aux études de santé.....	07
Formation dans le domaine sanitaire et social	08
École d'art ou d'architecture.....	09
<PS210MC>.....	10
Autre formation.....	11
Pas de formation post-bac	12 -> PS220MC

Veillez préciser :

PS230 - Vous nous avez déclaré avoir arrêté vos études car vous avez été refusé dans une formation ou échoué à un concours. Dans quelle formation aviez-vous candidaté ?

(Restriction de champ : question posée aux jeunes ayant répondu « oui » à la quatrième question PS010, première question du module, sous réserve que leur classe de sortie soit de niveau Bac+2 ou plus)

Licence générale	1
Licence professionnelle	2
Formation dans le domaine sanitaire et social	3
Master	4
École d'ingénieur	5
École de commerce	6
Autre formation.....	7 -> PS230MC

Veillez préciser :

MODULE : IDENTIFICATION DU PLUS HAUT DIPLÔME

1^{ER} GROUPE DE SORTANTS : LES SORTANTS DU SECONDAIRE ET LES SORTANTS D'UNE FORMATION DE NIVEAU BAC A BAC+3 PRESUMES AVOIR UN BAC PROFESSIONNEL

PHD010 - Quand vous avez arrêté vos études en <FINETU>, possédiez-vous :

Un CAP	Oui=1/Non=0
Un BEP	Oui=1/Non=0
Une MC au CAP.....	Oui=1/Non=0
Une MC au BEP	Oui=1/Non=0
Un autre diplôme de niveau CAP ou BEP	Oui=1/Non=0
Un bac.....	Oui=1/Non=0
Une MC au bac	Oui=1/Non=0
Un autre diplôme de niveau bac	Oui=1/Non=0
Un diplôme supérieur au bac	Oui=1/Non=0

PHD020 - Ce bac est-il le <BACLIBNEW> décrit précédemment ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac » à PHD010 et ayant déclaré avoir un bac dans le module précédent, à la question PS120)

- Oui 1
- Non 2

PHD030 - Ce bac est-il votre <DIPSOR> ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac » à PHD010, n'ayant pas déclaré avoir un bac dans le module précédent, à la question PS120, mais ayant déclaré avoir réussi leur diplôme dans le module de contrôle alors qu'ils étaient inscrits en Terminale)

- Oui 1
- Non 2

PHD040 - Ce bac est-il un <BACLIBNEW> ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac » à PHD010, « non » à PHD030 et ayant un bac connu dans la base de sondage)

- Oui 1
- Non 2

PHD050 - Avez-vous obtenu :

(Restriction de champ : question posée aux jeunes ayant répondu « bac » à PHD010 et n'ayant pas répondu « oui » à l'une ou l'autre des trois questions précédentes : PHD020, PHD030, PHD040)

Consigne : Si plusieurs bacs obtenus, sélectionner le dernier

- Un Bac général..... 1
- Un Bac technologique..... 2
- Un Bac professionnel..... 3

PHD060T - Dans quelle série ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac général » à PHD050)

- L 1
- ES 2
- S 3
- Autre 4 -> **PHD060MC**
Veillez préciser :

PHD070T - Dans quelle série ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac technologique » à PHD050)

- STMG 1
- STI2D..... 2
- STD2A 3
- STL..... 4
- ST2S 5
- STAV..... 6
- TMD 7
- STHR, Hôtellerie..... 8
- Autre 9 -> **PHD070MC**
Veillez préciser :

PHD080 - Dans quelle spécialité ?

(Restriction de champ : question posée aux jeunes ayant répondu « bac professionnel » à PHD050)

→ **Libellé de spécialité**

PHD100T - Avez-vous obtenu une mention pour ce Bac ?

(Restriction de champ : question posée à ceux ayant répondu à PHD050)

- Non, aucune ou passable 1
Oui, assez bien 2
Oui, bien 3
Oui, très bien 4

PHD110 - En quelle année avez-vous obtenu ce Bac ?

(Restriction de champ : question posée à ceux ayant répondu à PHD050)

DESCRIPTION DU PLUS HAUT DIPLOME DE L'ENSEIGNEMENT SECONDAIRE
Après détermination automatique du plus haut diplôme de l'enseignement secondaire
parmi les modalités de PHD010 (-> FIPHD120)

1. Autre diplôme de niveau CAP ou BEP
2. BEP
3. CAP
4. MC au BEP
5. MC au CAP
6. Autre diplôme de niveau bac
7. Bac
8. MC au bac
9. Autre diplôme de niveau supérieur au bac

PHD130 - Quel est ce <FIPHD120> ?

(Restriction de champ : question posée aux jeunes ayant un CAP, un BEP, une MC au CAP ou au BEP mais pas le bac ou plus, ni une MC au bac, selon FIPHD120 – modalités 1 à 6)

→ **Libellé de diplôme**

PHD140 - Quel est la spécialité de <FIPHD120> ?

(Restriction de champ : question posée aux jeunes concerné par la question précédente + ceux ayant eu une MC au bac, selon FIPHD120)

→ **Libellé de spécialité**

2E GROUPE DE SORTANTS : LES SORTANTS DE L'ENSEIGNEMENT SUPERIEUR ET LES SORTANTS DU SECONDAIRE AYANT REPONDU AVOIR UN DIPLOME DU SUPERIEUR A LA QUESTION PHD010

PHD160 - Quand vous avez arrêté vos études en <FINETU>, possédiez-vous :

- Un BTS ou un BTS agricole Oui=1/Non=0
- Un DUT Oui=1/Non=0
- Un DEUG Oui=1/Non=0
- Un autre diplôme de niveau Bac+2 Oui=1/Non=0
- Une licence professionnelle Oui=1/Non=0
- Une licence Oui=1/Non=0
- Un autre diplôme de niveau Bac+3 Oui=1/Non=0
- Un diplôme de niveau Bac+4 ou plus Oui=1/Non=0

FIPHD170 - Vous avez déclaré un diplôme de niveau bac+4 ou plus et n'avez déclaré aucun autre diplôme de l'enseignement supérieur. Confirmez-vous votre réponse ?

(question posée ou non, en fonction des résultats internes aux sous-questions de PHD160 pour les personnes qui ne sortent pas d'école d'ingénieur ou de commerce)

- Modifier 1 -> **Retour à PHD160**
- Valider..... 2

FIPHD170X - Lorsque vous avez arrêté vos études en <FINETU>, vous étiez dans une formation de niveau supérieur à bac+3, et n'avez déclaré aucun diplôme de l'enseignement supérieur. Confirmez-vous votre réponse ?

(question posée ou non, en fonction des résultats internes aux sous-questions de PHD160 pour les personnes qui ne sortent pas d'école d'ingénieur ou de commerce)

- Modifier 1 -> **Retour à PHD160**
- Valider..... 2

PHD170 – Quand vous avez arrêté vos études en <FINETU>, possédiez-vous :

(restriction de champ : question posée à ceux ayant répondu « un diplôme de niveau bac+4 ou plus » à la question PHD180)

- Une maîtrise Oui=1/Non=0
- Un diplôme d'école de commerce Bac+4..... Oui=1/Non=0
- Un autre diplôme de niveau Bac+4 Oui=1/Non=0
- Un magistère Oui=1/Non=0
- Un master Oui=1/Non=0
- Un diplôme d'école de commerce Bac+5 Oui=1/Non=0
- Un diplôme d'ingénieur Oui=1/Non=0
- Un autre diplôme de niveau Bac+5 ou 6 Oui=1/Non=0
- Un doctorat Oui=1/Non=0

DESCRIPTION DU PLUS HAUT DIPLOME DE L'ENSEIGNEMENT SUPERIEUR
Après détermination automatique du plus haut diplôme de l'enseignement supérieur
parmi les modalités suivantes (-> FIPHD180)

1. Autre diplôme de niveau Bac+2
2. DEUG
3. BTS ou BTS agricole
4. DUT
5. Autre diplôme de niveau Bac+3
6. Licence
7. Licence professionnelle
8. Autre diplôme de niveau Bac+4
9. Maîtrise
10. Diplôme d'école de commerce Bac+4
11. Diplôme de niveau Bac+5 ou 6
12. Magistère
13. Master
14. Diplôme d'école de commerce Bac+5
15. Diplôme d'ingénieur
16. Doctorat

PHD190 - Quel est ce <FIPHD180> ?

(Restriction de champ : question posée aux jeunes ayant un plus haut diplôme du supérieur mal déterminé – modalité 1, 5, 8 ou 11 pour FIPHD180)

→ **Libellé de diplôme**

PHD200 - Quel est la spécialité de <FIPHD180> ?

(pour tous ceux ayant un plus haut diplôme du supérieur)

→ **Libellé de spécialité**

3E GROUPE DE SORTANTS (QUESTION BALAI) : JEUNES N'AYANT PAS INDIQUE AVOIR OBTENU DE DIPLOME LORS DE LEUR ANNEE DE SORTIE DANS LE MODULE DE CONTROLE ET N'AYANT AUCUN PLUS HAUT DIPLOME REPERE, NI DANS LE SECONDAIRE (FIPHD120 VIDE), NI DANS LE SUPERIEUR (FIPHD180 VIDE)

PHD220 - Lorsque vous avez arrêté vos études en <FINETU>, possédiez-vous cependant un diplôme ?

Oui 1
Non 2

PHD230 - Quel était ce diplôme ?

(Restriction de champ : si PHD220=1)

→ **Libellé de diplôme**

PHD240 – Dans quelle spécialité ?

(Restriction de champ : si PHD220=1)

→ **Libellé de spécialité**

Annexe 3. Les informations de SISE pouvant intéresser le Céreq

Deux ensembles de fichiers annuels SISE sont disponibles : les fichiers relatifs aux étudiants inscrits au 1^{er} janvier de l'année considérée (SISE-Inscriptions) et les fichiers relatifs à la diplomation des étudiants l'année considérée (SISE-Diplomation).

Pour un domaine donné (inscriptions ou diplomation) et une année donnée, les données sont réparties entre plusieurs fichiers, chacun portant sur un champ spécifique : les universités publiques, les écoles de management ou de gestion, les écoles d'ingénieurs, les écoles non rattachées aux universités, les instituts catholiques, les grands établissements (écoles normales supérieures, écoles vétérinaires, école de paysage de Versailles...). Les informations et les noms de variables sont normés et harmonisés dans les fichiers mis à disposition, mais certaines informations peuvent être manquantes pour certains champs.

Les deux tableaux ci-dessous listent les variables susceptibles d'intéresser le Céreq dans les deux domaines couverts. Les informations contenues dans SISE-Diplomation qui sont issues de SISE-Inscriptions ne sont pas reprises dans le tableau SISE-Diplomation.

SISE-Inscriptions

Variable	Contenu	Observations
Caractéristiques de l'établissement fréquenté		
ETABLI	Numéro d'identification de l'établissement dans la BCE	
UNIV	Indicatrice permettant d'isoler les universités	Certains établissements de SISE-Université ne sont pas des universités (exemple : IEP, INALCO, Dauphine. Voir guide)
COMETA	Code INSEE de la commune d'implantation de l'établissement	Code INSEE. Information issue de la BCE
COMINS	Code INSEE de la commune d'implantation de la composante	Code INSEE. Information issue de la BCE
Caractéristiques de l'étudiant		
IDETU	Numéro INE	
JONAI	Jour de naissance	Non diffusée en dehors du SIES Absent de SISE-Ingénieurs
MONAI	Mois de naissance	Absent de SISE-Ingénieurs
ANNAI	Année de naissance	Si l'âge est inférieur à 14 ans ou supérieur à 99 ans, l'année de naissance est redressée en fonction du niveau et du type de formation suivi. Une variable AGE en âge atteint dans l'année de référence du fichier est aussi proposée.
SEXE	Sexe	En cas de non-réponse, imputation aléatoire
NATION	Nationalité de l'étudiant	La variable détaillée n'est pas diffusée en dehors du SIES. Les fichiers proposent aussi une variable agrégée par continent, avec distinction de l'UE au sein de l'Europe (NATRG), ainsi qu'une variable distinguant plus globalement français et étrangers (FR_ETR)
COMETU	Code INSEE de la commune de la résidence de l'étudiant	Les établissements peuvent fournir le libellé de la commune ou le code INSEE ou les deux. Le code postal est aussi demandé. Si le code commune n'est pas renseigné, il est codé à partir du libellé et du code postal. D'après le guide, « il est important d'utiliser cette variable avec précautions dans la mesure où parfois le code postal est parfois utilisé à la place du code commune ». Absent de SISE-Management et SISE-26bis
LCOMETU	Libellé de la commune de résidence de l'étudiant	
CP_ETU	Code postal de l'étudiant	
PCSPAR	PCS du premier parent	Pas de variable pour isoler père et mère.
PCSPAR2	PCS du deuxième parent	Pas de variable pour isoler père et mère.
COMREF	Code INSEE de la commune des parents de l'étudiant	La documentation parle tantôt d'adresse des parents, tantôt d'adresse de référence. Il s'agit de l'adresse permanente, à laquelle l'étudiant peut être contacté à tout moment, à laquelle on peut par exemple lui expédier du courrier pendant les périodes de vacances scolaires ou après la fin de l'année universitaire. Même remarque sur le codage que pour la commune de l'étudiant Absent de SISE-Management et SISE-26bis
LCOMREF	Libellé de l'adresse des parents	
CP_PAR	Code postal des parents	

Scolarité antérieure de l'étudiant		
BAC	Série du bac ou équivalence	Des redressements sont effectués sur le nom du bac en fonction de l'année d'obtention pour se caler sur les dénominations en vigueur alors
BAC_RGPR	Série du bac regroupé	7 modalités (littéraires, économiques, scientifiques, technologiques STT, autres technologiques, professionnels, dispenses)
ANBAC	Année d'obtention du baccalauréat	Renseigné pour le bac, pas pour les équivalences
DEPBAC	Département d'obtention du bac	Département de l'établissement de scolarisation lors du passage du bac ou du centre d'examen pour les non scolarisés Absent de SISE-Ingénieurs
ANETAB	Année d'entrée dans l'établissement	Collecté uniquement sur SISE-Ingénieurs, SISE-Management et SISE-26bis
ANINSC	Année de première inscription dans le système universitaire français	Absent de SISE-Ingénieurs, SISE-Management et SISE-26bis
ANSUP	Année d'entrée dans l'enseignement supérieur	Absent de SISE-Management et SISE-26bis
SITUPRE	Situation de l'étudiant l'année précédente	Indique plutôt le type d'établissements de l'année précédente pour les scolarisés et la non-scolarisation
TYPREPA	Classe préparatoire fréquentée par l'étudiant	Concerne uniquement les nouveaux entrants ayant fréquenté l'année précédente une classe préparatoire aux grandes écoles (CPGE). Dans SISE-Université, n'est renseigné que pour les formations d'ingénieur.
Formation en cours de l'étudiant (rappel : en cas d'inscriptions multiples, une observation par inscription)		
FONCTIO	Fonction de l'étudiant dans la formation considérée	Indicateur des élèves fonctionnaires
DIPLOM	Code du diplôme d'inscription	La codification couvre les diplômes nationaux et les diplômes d'établissement (voir guide de l'utilisateur)
TYP_DIPL	Type de diplôme	Nomenclature fine en deux positions du diplôme préparé. Les variables CYCLE et CURSUS_LMD peuvent se déduire de cette variable.
SECTDIS	Secteur disciplinaire du diplôme d'inscription (52 modalités)	52 secteurs, codés à partir de DIPLOM
DISCIPLI	Discipline associée au diplôme d'inscription (16 modalités)	Regroupement en 16 modalités des 52 secteurs disciplinaires
GROUPE	Regroupement disciplinaire (8 modalités)	Regroupement des disciplines. Cette variable construite permet aussi de distinguer les formations secondaires et les formations tertiaires pour les IUT (selon une méthode présentée dans le guide utilisateur)
CURSUS_LMD	Cursus d'appartenance dans le schéma LMD	3 modalités : L, M, D

CYCLE	Cycle d'appartenance dans le diplôme préparé	Exemple pour la licence : 1 ^{er} cycle pour les deux premières années, 2 ^e cycle pour la troisième. Exemple pour le master : 2 ^e cycle pour la première année, 3 ^e cycle pour la seconde.
VOIE	Voie du diplôme pour les diplômes LMD	3 modalités : générale, professionnelle, recherche
NIVEAU_D	Niveau théorique atteint au début de formation du diplôme d'inscription	En nombre d'années après le bac. Variable construite par le SIES
NIVEAU_F	Niveau théorique atteint en fin de formation du diplôme d'inscription	En nombre d'années après le bac. Variable construite par le SIES
DEGETU	Degré d'étude théorique atteint pendant l'année d'inscription en cours	En nombre d'années après le bac. Variable construite par le SIES
NATURE	Nature juridique du diplôme préparé	Permet, par exemple, de distinguer les diplômes homologués nationalement des diplômes d'université, les diplômes des écoles, les préparations à des concours administratifs, etc.
Information complémentaire sur la scolarité de l'étudiant		
AMENA	Indicateur de l'existence ou non d'un parcours aménagé	Cette variable permet d'isoler des cas qui ne se conforment pas au parcours théorique, sans pour autant qu'il n'y ait redoublement. Par exemple, licence en 4 ans, master en 3 ans, année de césure. Absent de SISE-Management et SISE-26bis
CONV	Indicateur signalant si les enseignements sont dispensés dans le cadre d'une convention	Permet d'isoler les inscriptions relevant d'une convention entre une université et un institut catholique. Depuis 2015, repère aussi les inscriptions liées à une convention entre une classe préparatoire aux grandes écoles et un EPSCP Absent de SISE-Ingénieur, SISE Management et SISE-26bis
CURPAR	Indicateur signalant l'existence d'un cursus parallèle dans un autre établissement supérieur	Par exemple, en cas de cursus dans une école et une université, dans une classe préparatoire et une université (avec ou sans convention), etc. Absent de SISE-Management et de SISE-26bis
EXOINS	Type d'exonération des frais d'inscription (s'il y a lieu)	Absent de SISE-Management et de SISE-26bis
ECHANG	Indicateur signalant les inscriptions dans le cadre d'un programme d'échange international	Absent de SISE-Management et de SISE-26bis
INSPR	Indicateur de repérage des inscriptions principales	L'inscription principale est définie par le SIES selon différents critères détaillés dans le guide utilisateur

SISE-Diplomation

Variable	Contenu	Observations
Rappel : en cas d'inscriptions multiples, une observation par inscription		
IDETU IDETU1	Identifiant INE	Certaines universités ayant modifié le numéro INE de leurs étudiants, IDETU propose le dernier INE disponible et IDETU1, l'INE contenu dans le fichier de SISE-Inscriptions de l'année correspondante. Le SIES dispose en effet du numéro d'inscription de l'étudiant dans l'établissement, ce qui lui permet de repérer les changements d'INE.
NONAPPAR	Résultat de l'appariement entre SISE-Résultats et SISE-Inscriptions	Permet par exemple de repérer notamment les inscrits ayant un diplôme attendu (« appariement total »), ceux ne correspondant pas à celui attendu (« appariement partiel »), etc.
COMPOSR	Unité de rattachement	
DIPLOMR	Diplôme SISE	Diplôme obtenu (variable de référence pour exploiter SISE-Résultats, notamment pour repérer les réussites et les échecs)
ETATINS	État de l'inscription	Permet de distinguer les différents types de situation au regard de l'inscription et de l'appartenance au champ SISE
RESDIP	Obtention du diplôme	
DIPINT	Code diplôme intermédiaire	Vise notamment les DEUG et maîtrises que les universités peuvent délivrer en niveau intermédiaire depuis la mise en place du LMD. Depuis 2013, la remontée des DEUG intermédiaires n'est cependant plus obligatoire.
RESINT	Indicatrice de réussite du diplôme	1=diplômé. 0 intègre les échecs, la non-présence, les réorientations. La variable reste à blanc quand elle n'a pas été renseignée.
NIVEAUR	Niveau dans le diplôme	Voir Guide.
PREXAM	Présence à l'examen	Demandé à toutes les universités, mais pas renseigné par tous les établissements dans SISE. Non exploitée par le SIES

Annexe 4. Quelques statistiques descriptives de suivi des parcours dans les fichiers

SISE-Inscriptions mobilisés

Tableau A4.1. • Où les inscrits de 2016 sont-ils inscrits en 2015, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2015																	
		ABSENT	0		1		1		1		2		2		2		3		3
		ENS	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV
1	CPES-BTS	85,3		4,1	0,1				7,7	0,3					0,4				0,1
	CPES-CPEG	96,2			1,4		0,1		2,0	0,2									0,0
	E26b	84,1	0,1	0,6			0,1	0,1	11,7	0,9	0,1				1,1				0,9
	ENS	85,1		0,2	0,2	2,7	0,2		5,6	0,2					0,5	2,5			1,1
	INGE	83,6			0,9		6,3		4,3	0,5	1,0				1,7		0,5		0,7
	MANA	82,2		0,4	0,5		0,3	4,7	8,2	0,2					0,6			0,1	0,2
	UNIV	71,5	0,5	0,5	0,3		0,1		24,7	0,6	0,1				0,8				0,4
2	CPES-BTS	6,8		84,2					0,9	7,4					0,1				
	CPES-CPEG	8,2			46,1		0,2		19,8	2,5	16,6		0,1		6,5				0,1
	E26b	94,6		0,2					1,5	1,2	0,1		0,1		0,8				0,8
	ENS	12,4			0,3	46,4			1,2		8,7	16,5			9,3				1,8
	INGE	2,2		0,1	1,0		78,7		3,5	0,2	1,0		6,6		1,1		4,9		0,2
	MANA	19,2		0,1			0,1	71,9	1,1	0,5	0,1			4,2	0,9			0,9	0,4
	UNIV	5,7			1,1				78,1	0,7	0,5				12,5				0,7
3	E26b	81,8							0,4	9,3	0,1				5,1		0,1	0,1	2,6
	ENS	19,7							0,1		29,9	15,3			24,8	2,2	0,3	0,1	7,2
	INGE	12,8					0,1	0,4	0,1	1,6	25,4		24,8		20,6		6,8	0,4	3,8
	MANA	25,5						1,4	0,2	4,8	15,4			31,0	13,5		0,1	6,2	1,3
	UNIV	18,4							0,6	5,9	1,7		0,1		59,9		0,1		11,8
4	E26b	91,8													0,1			0,1	6,8
	ENS	20,6							0,4		2,0	4,3			1,1	27,6		0,4	29,6
	INGE	9,9					0,3				0,1				0,1		75,4		0,7
	MANA	24,8						2,7	0,2					1,0	0,1			45,0	7,4
	UNIV	20,6	0,1						0,5		0,1				0,3	0,1		0,1	58,3
5	E26b	97,4							0,1										0,6
	ENS	12,8							0,1			0,1			0,1	7,3			0,3
	INGE	7,1									0,9		0,2		0,3		2,7		0,1
	MANA	23,5												3,0				0,9	0,2
	UNIV	21,0													0,1				0,5
6	E26b	78,4								1,9					1,0				1,0
	ENS	15,3							0,1			0,1			0,1				0,1
	INGE	34,7							0,1							0,1	0,2		0,1
	MANA	45,1							0,1					0,1				0,4	0,4
	UNIV	18,9							0,1						0,1				0,1

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2015														
		4				5				6				?		
		ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	CPES-BTS	ENS	MANA
1	CPES-BTS													2,0		
	CPES-CPEG													0,0		
	E26b				0,1			0,1						0,1		
	ENS				0,5	0,5		0,5				0,2		0,2		
	INGE				0,1			0,0						0,0		
	MANA			2,3		0,0		0,2						0,1		
	UNIV				0,1		0,0	0,0	0,2				0,0	0,1		
2	CPES-BTS													0,7		
	CPES-CPEG															
	E26b				0,4			0,2				0,0				
	ENS	2,1					0,2	0,5								
	INGE		0,4				0,0	0,0								
	MANA			0,2			0,0	0,1			0,4			0,0		
	UNIV				0,2		0,0	0,0	0,2			0,1		0,1		
3	E26b				0,3			0,2								
	ENS				0,3	0,1		0,1								
	INGE		2,0	0,7	0,1		0,2	0,0	0,0							
	MANA			0,1	0,1		0,1	0,0	0,0		0,2		0,0		0,0	
	UNIV				0,7	0,0	0,0	0,0	0,5			0,1		0,1		
4	E26b				0,5		0,2	0,0	0,4							
	ENS	7,0		0,2	3,0	1,7	0,3	0,1	1,4	0,0		0,2	0,0	0,0		
	INGE		9,5	0,1	0,3	0,0	3,2	0,0	0,1		0,1					
	MANA		0,1	14,7	1,5	0,0	0,3	1,8	0,3		0,1		0,1		0,0	
	UNIV				16,2	0,0	0,0	0,0	3,3			0,0	0,1	0,2		
5	E26b		0,1	0,4	0,9		0,1	0,1	0,4							
	ENS	47,4	0,4	0,2	7,5	15,7	1,2	0,1	5,9	0,0	0,1		0,3	0,1	0,2	
	INGE	0,1	68,4	0,1	0,8	0,1	16,5	0,2	0,5		2,1		0,1	0,0		
	MANA		0,2	44,8	0,9	0,0	0,8	23,0	1,1		0,0	0,4	0,1			
	UNIV	0,1	0,1	0,1	59,5	0,2	0,3	0,1	17,2	0,0	0,0		0,7	0,0		
6	E26b		1,0		7,2		3,9	0,5	4,8				0,5			
	ENS	3,0			0,2	14,5	0,8	0,0	2,5	63,1	0,1		0,3			
	INGE	0,1	1,0	0,2	0,4	0,6	9,9	0,4	7,9		43,8	0,3	0,3			
	MANA		0,1	4,1	2,5	0,0	3,8	2,9	8,0		0,1	31,1	1,3			
	UNIV				0,3	0,3	0,4	0,0	13,7	0,1	0,0	0,0	65,9			

* Variable DEGETU. ** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau A4.2. • Où les inscrits de 2016 sont-ils inscrits en 2014, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2014																		
		ABSENT	0		1		1		1		2		2		2		3		3	
			ENS	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV
1	CPES-BTS	96,3	0,0	0,6	0,0		0,0	0,0	2,6	0,1	0,0				0,2				0,0	
	CPES-CPEG	99,5		0,2	0,0		0,0		0,3	0,0										
	E26b	88,5	0,0	1,0	0,2		0,1	0,1	6,9	0,5					1,6				0,6	
	ENS	92,2		0,2	0,2				2,2		0,5	2,2			0,5	0,5				
	INGE	93,6	0,0	0,5	0,9		0,1	0,0	2,8	0,5	0,3		0,1		1,0		0,0		0,3	
	MANA	92,8	0,0	0,3	0,1		0,1	0,3	3,4	0,1	0,0	0,0	0,0	0,1	0,4			1,4	0,3	
	UNIV	91,1	0,3	0,7	0,2		0,0	0,0	6,1	0,3	0,0		0,0	0,0	0,6		0,0	0,0	0,2	
2	CPES-BTS	77,1	0,0	9,8	0,2		0,0	0,0	8,8	0,5	0,0				0,4				0,1	
	CPES-CPEG	72,6	0,0	2,3	16,7	0,0	0,2	0,0	7,7	0,2	0,1		0,0	0,0	0,2		0,0		0,0	
	E26b	79,6	0,1	1,3	0,3		0,4	0,0	11,7	1,0	0,3		0,0		3,0		0,1		1,3	
	ENS	51,9			5,7	13,0		0,2	10,7	0,2	4,4	6,4			4,3	1,5			0,5	
	INGE	75,6	0,0	0,0	0,3	2,4		11,0	0,0	5,7	0,5	1,2		0,3	1,9		0,2		0,8	
	MANA	80,2	0,0	0,0	0,8	0,8	0,0	0,3	6,1	8,8	0,5	0,0		0,0	1,4			0,1	0,3	
	UNIV	61,6	0,3	0,9	1,4		0,2	0,0	30,1	0,5	0,2		0,0	0,0	3,4		0,0	0,0	0,6	
3	E26b	76,8	0,1	8,1	0,3		0,0	0,0	5,8	2,4	0,1		0,0	0,1	3,7	0,0	0,1	0,0	1,7	
	ENS	11,9			22,3	12,7			17,4	0,1	20,0	1,7	0,1		11,7	0,3	0,1	0,0	0,8	
	INGE	14,2		1,4	19,6		21,0		17,1	1,3	10,1	0,0	2,8	0,0	7,5		2,5	0,4	0,8	
	MANA	27,0		4,4	13,7		0,0	27,3	13,1	1,6	3,5		0,0	3,0	3,5		0,0	1,8	0,5	
	UNIV	20,0	0,0	5,3	1,8	0,0	0,1	0,0	50,5	1,9	0,8	0,0	0,1	0,0	14,9		0,1	0,0	3,3	
4	E26b	77,3		0,0				0,0	0,3	10,0	0,2		0,1	0,1	6,6		0,1	0,1	3,3	
	ENS	22,3			1,0				0,5	0,2	27,6	5,3			22,4	4,8	0,2	0,3	8,7	
	INGE	18,6			0,1		0,1		0,2	1,3	26,5	0,0	21,7		14,8	0,0	11,8	0,1	3,4	
	MANA	29,6		0,0	0,0			0,6	0,3	3,4	12,1	0,0	0,0	15,1	13,3	0,0	0,2	14,9	5,4	
	UNIV	24,3	0,1	0,0				0,0	0,6	1,4	2,0	0,1	0,1	0,1	44,8	0,0	0,1	0,1	17,0	
5	E26b	89,9							0,2						0,6		0,1	0,3	7,4	
	ENS	19,4	0,1			0,0			0,3	0,0	0,5				0,2	30,4	0,3	0,0	14,5	
	INGE	13,0			0,8		0,2		0,2	0,0	1,6		0,5		0,2	0,0	59,0	0,0	1,2	
	MANA	30,4		0,0				2,3	0,2	0,0	0,0			1,4	0,2	0,0	0,2	21,1	5,5	
	UNIV	31,0	0,1						0,5	0,0	0,1		0,0	0,0	0,3	0,1	0,1	0,2	41,6	
6	E26b	89,9		1,4					1,0						1,0					
	ENS	21,2	0,0			0,1			0,2			0,0			0,1	0,4			0,2	
	INGE	41,7							0,2				0,0		0,2		0,8	0,2	0,3	
	MANA	53,4						0,1	0,1	0,0				0,4	0,4		0,0	3,9	1,8	
	UNIV	23,8	0,0						0,1	0,0					0,2	0,0			0,2	

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2014													
		4				5				6				?	
		ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	CPES-BTS	
1	CPES-BTS				0,0									0,1	100
	CPES-CPEG														100
	E26b				0,2				0,0					0,1	100
	ENS	0,2			0,2	0,2			0,2				0,5	0,2	100
	INGE				0,0				0,0					0,0	100
	MANA	0,0		0,6	0,1				0,0					0,0	100
UNIV		0,0	0,0	0,2		0,0	0,0	0,1				0,0	0,1	100	
2	CPES-BTS				0,0									3,0	100
	CPES-CPEG													0,0	100
	E26b	0,0			0,4				0,2				0,0	0,2	100
	ENS	0,2			0,2	0,3	0,2			0,3			0,2		100
	INGE		0,0		0,1		0,0							0,0	100
	MANA		0,0	0,1	0,1			0,0	0,0			0,0		0,1	100
UNIV		0,0	0,0	0,2		0,0	0,0	0,1				0,1	0,2	100	
3	E26b			0,0	0,6				0,1					0,0	100
	ENS	0,1			0,3	0,2	0,1		0,3					0,1	100
	INGE		0,2	0,5	0,1		0,0	0,6	0,0						100
	MANA		0,1	0,1	0,1		0,0	0,0	0,0			0,2		0,0	100
	UNIV	0,0	0,0	0,0	0,7		0,0	0,0	0,3				0,1	0,1	100
4	E26b		0,2		1,1	0,0	0,0		0,3					0,0	100
	ENS	2,7	0,1	0,3	1,6	1,0	0,1	0,0	0,7	0,0			0,2	0,0	100
	INGE	0,0	1,1	0,0	0,2		0,0	0,1	0,1						100
	MANA	0,0	0,2	2,9	0,9	0,0	0,1	0,4	0,2			0,1	0,0	0,2	100
	UNIV	0,0	0,1	0,0	6,9	0,0	0,0	0,0	1,9				0,2	0,2	100
5	E26b		0,1	0,1	0,9		0,1		0,3						100
	ENS	17,4	1,1	0,2	6,7	4,7	0,6	0,2	3,0	0,0			0,3	0,1	100
	INGE	0,1	19,6	0,1	0,6	0,0	2,4	0,2	0,2		0,0		0,0	0,0	100
	MANA	0,0	0,8	29,0	2,4	0,0	0,4	5,0	0,7		0,0	0,1	0,1	0,1	100
	UNIV	0,1	0,2	0,1	18,7	0,1	0,1	0,1	5,5	0,0	0,0		0,8	0,2	100
6	E26b		3,4	0,5	1,0		1,4		0,5						100
	ENS	13,8	0,5	0,0	1,8	14,2	0,9	0,1	3,9	42,0	0,2		0,7	0,0	100
	INGE	0,1	8,1	0,3	6,3	0,9	8,8	0,4	6,9	0,0	24,0	0,0	0,8		100
	MANA	0,1	3,8	5,0	9,0	0,2	2,5	3,7	8,2	0,0		6,7	0,7		100
	UNIV	0,1	0,3	0,0	11,1	0,5	0,7	0,0	13,9	0,1	0,1	0,0	48,9		100

* Variable DEGETU. ** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau A4.3. • Où les inscrits de 2016 sont-ils inscrits en 2013, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2013																		
		ABSENT	0		1		1		1		2		2		2		3		3	
		ENS	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	
1	CPES-BTS	98,7		0,0	0,2	0,0		0,0		0,9	0,0	0,0			0,1				0,0	
	CPES-CPEG	99,9			0,0	0,0			0,0											
	E26b	92,9			0,7				4,3	0,3					1,0				0,6	
	ENS	92,9				0,7	2,7		1,1	0,7		0,2								0,5
	INGE	96,7	0,0		0,4	0,3		0,2	1,6	0,2	0,1		0,0		0,4					0,1
	MANA	95,8		0,0	0,1	0,1	0,0	0,0	0,1	1,4	0,1	0,1		0,0	1,0	0,5	0,0		0,4	0,2
	UNIV	95,9		0,1	0,3	0,1		0,0	0,0	2,5	0,1	0,0			0,0	0,3		0,0	0,0	0,2
2	CPES-BTS	94,6			1,0	0,1		0,0	0,0	3,6	0,1	0,0		0,0	0,3					0,1
	CPES-CPEG	98,2	0,0	0,0	0,2	0,4		0,1	0,0	1,0	0,0				0,0					0,0
	E26b	83,0		0,1	1,2	0,5		0,4		9,9	0,5	0,1		0,0	2,1	0,0	0,0	0,0		1,2
	ENS	77,2			0,3	3,4	3,7		5,2	1,4	0,2	4,9			1,7				0,2	1,1
	INGE	92,2	0,0	0,0	0,5	1,2		0,7	0,0	3,2	0,4	0,4		0,1	0,9			0,1		0,3
	MANA	92,1		0,0	0,4	0,2	0,0	0,1	1,5	4,5	0,3	0,0		0,0	0,1	0,6		0,0	0,1	0,1
	UNIV	85,2		0,2	0,6	0,4		0,1	0,0	10,7	0,3	0,1		0,0	0,0	1,6		0,0	0,0	0,4
3	E26b	75,4		0,1	2,5	0,5		0,3	0,2	14,2	1,3	0,1	0,0	0,0	3,2		0,1			1,2
	ENS	59,6			0,2	16,0	2,1	0,1	0,1	14,0	0,2	3,3	0,1		2,8	0,1				0,8
	INGE	68,4	0,0	0,0	1,2	9,8		0,0	3,9	0,0	11,0	0,6	1,7		2,1			0,3		0,5
	MANA	76,4			1,7	3,6		0,2	3,6	10,1	0,5	0,9		0,0	0,5	1,7		0,1	0,2	0,3
	UNIV	60,1		0,2	2,0	1,3	0,0	0,2	0,0	26,7	0,9	0,3		0,0	0,0	5,3	0,0	0,0	0,0	1,8
	E26b	71,6			9,1	0,4		0,2	0,1	7,9	2,6	0,2		0,0	0,1	4,1		0,3	0,0	1,8
ENS	23,5		0,0	0,2	17,6	0,7			21,6	0,4	14,9	0,3			11,0	1,9	0,1	0,0	4,2	
INGE	20,1			1,2	17,7	0,0	18,5		15,2	1,3	13,2	0,0	3,0	0,0	7,2	0,0	1,4		1,0	
MANA	30,0			3,0	9,9	0,0	0,0	13,9	13,5	1,9	7,0	0,0	0,0	4,0	8,9	0,0	0,3	3,4	2,3	
UNIV	25,3		0,0	1,3	1,6	0,1	0,1	0,1	39,3	1,1	1,0	0,0	0,1	0,1	17,0	0,0	0,1	0,0	7,7	
5	E26b	74,5				0,1			0,7	9,9	0,2		0,1		10,2		0,1	0,1	3,4	
	ENS	24,3		0,1					0,5	0,5	18,9	4,1	0,1		15,1	8,2	1,0	0,1	14,8	
	INGE	19,5			0,0	1,2		0,0	0,3	1,1	20,7	0,0	17,0	0,0	12,3	0,1	19,9	0,1	3,0	
	MANA	34,6			0,0			0,9	0,3	2,3	7,0	0,0	0,1	7,4	8,5	0,0	0,8	21,0	6,7	
	UNIV	33,8		0,1	0,0	0,0		0,0	0,0	0,5	1,2	1,9	0,1	0,1	0,1	32,4	0,1	0,3	0,2	17,8
6	E26b	90,9							1,9								3,4	0,5	1,4	
	ENS	28,0					0,0		0,2		0,0	0,1			0,1	12,2	0,4	0,0	2,0	
	INGE	45,9		0,1				0,0	0,1	0,1	0,0		0,7	0,1	0,3	0,4	7,0	0,2	5,3	
	MANA	51,1			0,0			0,4	0,3	0,4	0,1			3,5	1,7		3,3	3,8	7,8	
	UNIV	27,3		0,0					0,2	0,0					0,1	0,1	0,2		9,8	

Dans SISE Inscriptions 2016															
		4	4	4	4	5	5	5	5	6	6	6	6	?	
		ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	CPES-BTS	
1	CPES-BTS													0,0	100
	CPES-CPEG														100
	E26b				0,1									0,1	100
	ENS				0,5	0,2							0,7		100
	INGE				0,0				0,0						100
	MANA			0,0	0,0				0,0						100
	UNIV		0,0	0,0	0,2		0,0	0,0	0,1				0,0	0,0	100
2	CPES-BTS				0,0									0,2	100
	CPES-CPEG														100
	E26b				0,7			0,1					0,1	0,1	100
	ENS	0,5	0,2			0,2				0,2					100
	INGE		0,0		0,0		0,0		0,0			0,0			100
	MANA			0,0	0,0				0,0					0,0	100
	UNIV		0,0	0,0	0,2		0,0	0,0	0,1				0,1	0,1	100
3	E26b			0,0	0,3				0,2				0,0	0,2	100
	ENS	0,1	0,1		0,3	0,1			0,3					0,1	100
	INGE		0,0		0,0		0,0		0,0					0,0	100
	MANA		0,0	0,1	0,0			0,0	0,0					0,2	100
	UNIV		0,0	0,0	0,5		0,0	0,0	0,2				0,2	0,3	100
4	E26b	0,0	0,0		0,8				0,5					0,2	100
	ENS	1,3	0,1	0,1	0,8	0,1		0,0	0,6	0,0			0,2	0,0	100
	INGE		0,2		0,1				0,0						100
	MANA	0,0	0,1	1,1	0,4		0,0	0,1	0,0					0,3	100
	UNIV	0,0	0,0	0,0	3,6	0,0	0,0	0,0	1,1				0,2	0,2	100
5	E26b		0,1		0,5		0,1		0,2					0,1	100
	ENS	4,2	0,6	0,3	3,7	1,1	0,2	0,1	1,8	0,0			0,4	0,1	100
	INGE	0,0	3,5	0,1	0,3	0,0	0,5	0,0	0,1		0,0	0,0	0,1	0,0	100
	MANA	0,0	0,5	5,9	2,0	0,0	0,2	1,1	0,5			0,0	0,0	0,2	100
	UNIV	0,1	0,1	0,1	7,0	0,1	0,1	0,1	3,0	0,0	0,0		1,0	0,2	100
6	E26b		1,4		0,5										100
	ENS	11,6	0,7	0,1	3,0	8,8	1,0	0,0	4,4	26,2	0,3		0,9	0,0	100
	INGE	0,5	8,0	0,3	5,5	0,9	7,5	0,2	6,5	0,0	9,8	0,0	0,8		100
	MANA	0,1	2,7	4,0	9,7	0,1	0,6	1,8	3,9	0,0		4,2	0,5	0,1	100
	UNIV	0,3	0,6	0,0	12,2	0,6	0,7	0,0	14,3	0,0	0,1	0,0	33,6		100

* Variable DEGETU. ** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en comptes que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau A4.4. • Où les inscrits de 2016 sont-ils inscrits en 2012, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2012																		
		ABSENT	0		1		1		1		2		2		2		3		3	
		ENS	UNIV	CPES- BTS	CPES- CPEG	ENS	INGE	MANA	UNIV	CPES- BTS	CPES- CPEG	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	
1	CPES- BTS	99,5		0,1	0,0				0,3	0,0					0,0				0,0	
	CPES- CPEG	100,0							0,0											
	E26b	95,4		0,3			0,0		2,7	0,2	0,0				0,8				0,3	
	ENS	93,5		0,9	0,2	0,9			1,6	0,2					0,2			0,2	1,1	
	INGE	98,5		0,2	0,1		0,0		0,8	0,0	0,0				0,1				0,1	
	MANA	97,1		0,0	0,1	0,0		0,0	1,1	0,9	0,1	0,1	0,0		0,3	0,2			0,0	0,1
	UNIV	97,7		0,1	0,1	0,0		0,0	0,0	1,2	0,1	0,0			0,0	0,3		0,0	0,0	0,2
2	CPES- BTS	98,3		0,0	0,2	0,0		0,0	1,2	0,0					0,1				0,0	
	CPES- CPEG	99,8			0,0	0,0			0,2	0,0					0,0				0,0	
	E26b	90,2		0,6	0,2				5,6	0,2	0,1	0,0			1,6				0,8	
	ENS	88,2		1,2	0,8	2,8			2,0	0,2	0,2	2,3		0,2	0,9		0,2		0,6	
	INGE	96,4		0,4	0,4		0,2	0,0	1,7	0,2	0,2		0,0		0,4		0,0	0,0	0,1	
	MANA	96,6		0,0	0,3	0,1		0,1	0,3	2,1	0,1	0,0		0,0	0,2			0,0	0,0	
	UNIV	93,5		0,1	0,3	0,2		0,0	0,0	4,2	0,1	0,0		0,0	0,9		0,0	0,0	0,3	
3	E26b	84,8		1,5	0,1	0,0	0,0	0,1	9,5	0,5	0,1		0,0		1,9		0,0		0,7	
	ENS	89,5		0,1	3,1	0,2			4,2	0,2	0,5	0,1		1,0		0,1		0,4		
	INGE	90,7	0,0	0,6	1,9		0,6		4,3	0,3	0,4		0,1		0,8		0,1	0,1	0,2	
	MANA	91,5		0,6	0,9		0,1	1,0	4,5	0,2	0,1		0,0	0,1	0,7		0,0	0,1	0,1	
	UNIV	82,0		0,2	0,9	0,4		0,1	11,3	0,4	0,2		0,0	0,0	2,6		0,0	0,0	1,1	
4	E26b	74,9		2,6	0,5		0,3	0,1	14,2	0,9	0,3		0,3	0,0	2,8	0,0	0,0	0,0	1,5	
	ENS	59,2		0,3	12,0	0,1	0,0		14,0	0,2	4,1	0,1	0,0		5,6	0,4	0,1	0,1	2,0	
	INGE	66,4	0,0	1,3	11,2	0,0	4,0		10,8	0,3	2,8		0,3	0,0	1,9		0,3	0,1	0,3	
	MANA	64,4		1,9	6,6	0,0	0,1	4,5	12,7	0,6	2,5		0,1	0,7	3,8		0,1	0,6	0,8	
	UNIV	54,7		1,1	1,4	0,0	0,1	0,1	25,5	0,6	0,4	0,0	0,0	0,0	8,5	0,0	0,1	0,0	4,0	
5	E26b	73,0		8,5	0,2				9,1	2,4	0,1		0,1	0,1	4,1		0,1	0,1	1,5	
	ENS	24,9		0,4	10,8	0,5	0,1		14,5	0,3	16,6	0,3	0,1		16,0	1,6	0,8	0,1	7,6	
	INGE	23,7		1,0	12,7	0,0	14,9		12,5	1,0	16,7	0,0	3,6		7,4	0,0	4,0	0,1	1,2	
	MANA	35,6		2,0	5,8		0,1	7,7	8,1	1,8	10,7	0,0	0,3	4,4	11,2	0,0	0,5	5,0	3,7	
	UNIV	34,8		1,1	1,4	0,0	0,1	0,1	28,8	0,9	1,4	0,0	0,1	0,1	16,8	0,1	0,2	0,1	7,1	
6	E26b	92,8									1,0		1,4	0,5	1,0		1,4		1,4	
	ENS	36,0							0,2	0,1	5,7	2,4	0,1		2,7	9,8	0,7	0,1	3,7	
	INGE	47,1		0,0	0,0	0,1		0,3	0,4	0,5	3,5	0,0	2,4	0,1	4,7	0,8	7,0	0,2	5,2	
	MANA	51,1		0,4	0,1			3,4	1,9	0,5	1,7		1,4	3,4	7,6	0,0	2,7	3,4	9,8	
	UNIV	29,9		0,0	0,0				0,2	0,0	0,3	0,0	0,1		8,7	0,3	0,5	0,0	10,9	

Dans SISE Inscriptions 2016																
		4	4	4	4	5	5	5	5	6	6	6	6	?		
		ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	CPES- BTS		
1	CPES- BTS				0,0									0,0	100	
	CPES- CPEG														100	
	E26b				0,1			0,0						0,1	100	
	ENS	0,2												0,9	100	
	INGE				0,0				0,0						100	
	MANA														100	
	UNIV		0,0	0,0	0,1		0,0	0,0	0,1				0,0	0,0	100	
2	CPES- BTS													0,1	100	
	CPES- CPEG														100	
	E26b				0,4		0,0		0,1				0,1	0,0	100	
	ENS					0,2			0,2	0,3					100	
	INGE		0,0		0,0			0,0	0,0		0,0			0,0	100	
	MANA				0,0									0,1	100	
	UNIV			0,0	0,1			0,0	0,1				0,1	0,1	100	
3	E26b				0,2				0,2					0,1	0,4	100
	ENS	0,1			0,4			0,1	0,2						100	
	INGE		0,0	0,0	0,0			0,0	0,0		0,0			0,0	100	
	MANA		0,0		0,0				0,0					0,1	100	
	UNIV		0,0	0,0	0,3		0,0	0,0	0,2				0,2	0,1	100	
4	E26b				1,0				0,3					0,3	100	
	ENS	0,4	0,0		0,6			0,0	0,4				0,2	0,1	100	
	INGE		0,1	0,0	0,1		0,0	0,0	0,0					0,0	100	
	MANA		0,0	0,2	0,2			0,0	0,0				0,0	0,2	100	
	UNIV	0,0	0,0	0,0	2,0	0,0	0,0	0,0	0,7				0,2	0,2	100	
5	E26b		0,1		0,3			0,1	0,1						100	
	ENS	1,2	0,2	0,2	2,2	0,3	0,2	0,0	0,6	0,0			0,4	0,1	100	
	INGE	0,0	0,6	0,1	0,1		0,2	0,0	0,1				0,1		100	
	MANA	0,0	0,2	0,8	1,1		0,1	0,5	0,3			0,0	0,0	0,3	100	
	UNIV	0,1	0,1	0,1	3,6	0,0	0,1	0,0	1,9	0,0	0,0		1,0	0,2	100	
6	E26b							0,5							100	
	ENS	5,9	0,9		3,3	6,6	0,9	0,1	4,4	15,8			0,8		100	
	INGE	0,6	6,7	0,4	5,1	0,7	6,5	0,1	5,4	0,0	1,7		0,7		100	
	MANA	0,1	0,8	2,1	4,3	0,1	0,3	0,7	1,8	0,0		2,1	0,4	0,2	100	
	UNIV	0,4	0,7	0,0	13,0	0,5	0,6	0,0	12,6	0,0	0,0		21,3	0,0	100	

* Variable DEGETU. ** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en comptes que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau A4.5. • Où les inscrits de 2016 sont-ils inscrits en 2011, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?

Dans SISE Inscriptions 2016	Degré d'études* et composante** dans SISE Inscription 2011																		
	ABSENT	0		1		1		1		2		2		2		3		3	
		ENS	UNIV	CPES- BTS	CPES- CPEG	ENS	INGE	MANA	UNIV	CPES- BTS	CPES- CPEG	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV
1	CPES- BTS	99,8		0,0				0,1	0,0					0,0				0,0	
	CPES- CPEG	100,0						0,0											
	E26b	97,3	0,1	0,3				1,6	0,0					0,4				0,1	
	ENS	96,4		0,2				0,5	0,2	0,2				0,9				0,5	
	INGE	99,3	0,0	0,0	0,1		0,0		0,3	0,0				0,1				0,1	
	MANA	98,8	0,0	0,1	0,1		0,0	0,3	0,5	0,1	0,0	0,0		0,1				0,0	
	UNIV	98,5	0,1	0,1	0,0		0,0	0,0	0,7	0,0	0,0			0,2		0,0	0,0	0,1	
2	CPES- BTS	99,4		0,1	0,0				0,5	0,0					0,1			0,0	
	CPES- CPEG	100,0		0,0					0,0	0,0									
	E26b	93,4		0,4	0,2	0,0	0,1	0,0	3,4	0,4	0,0			1,2				0,6	
	ENS	92,7		0,2	0,3	1,1		0,2	2,0	0,3		0,8		0,6	0,5			0,6	
	INGE	98,4		0,2	0,1		0,1	0,0	0,7	0,1	0,1		0,0	0,2		0,0		0,1	
	MANA	98,7	0,0	0,1	0,0		0,0	0,0	0,9	0,0				0,0	0,1			0,1	
	UNIV	96,5	0,1	0,1	0,1		0,0	0,0	2,1	0,1	0,0			0,5		0,0	0,0	0,2	
3	E26b	91,9		0,6	0,2		0,1		5,0	0,3	0,1			0,0	0,9			0,5	
	ENS	96,0		0,1	0,7	0,2			1,7		0,1		0,1	0,4	0,1			0,3	
	INGE	96,8		0,3	0,4		0,1		1,7	0,1	0,1		0,0	0,3		0,0		0,1	
	MANA	96,7		0,2	0,2		0,0	0,2	2,0	0,1	0,1			0,3		0,0	0,0	0,1	
	UNIV	90,9	0,1	0,5	0,2		0,0	0,0	5,3	0,2	0,1		0,0	0,0	1,4		0,0	0,7	
4	E26b	84,2		1,1	0,3		0,3	0,0	9,5	0,5	0,2		0,1	0,0	1,9		0,0	1,4	
	ENS	83,2	0,1	0,2	3,1		0,0		7,4	0,1	1,0	0,1	0,1	2,4	0,1	0,0		1,0	
	INGE	90,6	0,0	0,3	2,6	0,0	0,6		4,3	0,1	0,4		0,1	0,5		0,1		0,2	
	MANA	85,9	0,0	0,6	2,4		0,1	0,9	6,9	0,2	0,7		0,0	1,4		0,0	0,1	0,3	
	UNIV	75,7	0,1	0,6	0,5	0,0	0,1	0,0	13,4	0,3	0,2		0,0	0,0	4,5	0,0	0,0	2,3	
5	E26b	73,9	0,1	2,8	0,3		0,7	0,1	15,2	1,1	0,2		0,1	0,1	3,2			1,2	
	ENS	48,5	0,1	0,2	11,9	0,0	0,2		17,9	0,2	5,5	0,1	0,1	0,0	8,2	0,4	0,2	3,6	
	INGE	60,5	0,0	0,9	11,7	0,0	4,3		11,4	0,3	5,9		0,7	0,0	2,6		0,7	0,4	
	MANA	55,5	0,0	1,7	8,9	0,0	0,3	4,5	13,9	0,7	3,6	0,0	0,1	1,0	5,8	0,0	0,2	1,6	
	UNIV	56,3	0,1	0,9	1,5	0,0	0,2	0,1	23,1	0,5	0,6	0,0	0,1	0,0	7,5	0,0	0,1	3,8	
6	E26b	90,9		0,5			1,4	0,5	1,9		1,0		0,5		1,4			0,5	
	ENS	38,5	0,0	0,1	2,3	0,0	0,1		2,7	0,1	10,6	0,1	0,1		6,1	3,7	0,7	4,7	
	INGE	49,0		0,3	2,1	0,0	2,1	0,1	4,7	0,4	5,6	0,0	1,7	0,1	5,5	0,3	5,6	5,1	
	MANA	56,9		0,4	1,0		1,3	3,4	7,8	0,4	2,2		1,0	1,9	10,5	0,1	1,0	5,2	
	UNIV	31,5	0,0	0,0	0,2		0,1		8,4	0,1	0,7	0,0	0,1		10,3	0,2	0,6	11,3	

Dans SISE Inscriptions 2016	Degré d'études* et composante** dans SISE Inscription 2011														
	4		4		4		5		5		6		6		?
	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	CPES- BTS		
1	CPES- BTS														100
	CPES- CPEG														100
	E26b				0,1									0,1	100
	ENS								0,2				0,9		100
	INGE				0,0				0,0						100
	MANA			0,0										0,0	100
	UNIV		0,0		0,1			0,0	0,1				0,0	0,0	100
2	CPES- BTS				0,0									0,0	100
	CPES- CPEG														100
	E26b		0,0		0,1				0,1				0,0		100
	ENS	0,3							0,3	0,3					100
	INGE		0,0		0,0		0,0	0,0	0,0					0,0	100
	MANA				0,0									0,0	100
	UNIV				0,1			0,0	0,1				0,1	0,0	100
3	E26b				0,3		0,0		0,1						100
	ENS		0,1		0,2				0,2				0,1		100
	INGE		0,0		0,0				0,0					0,0	100
	MANA				0,0				0,0					0,0	100
	UNIV		0,0	0,0	0,2		0,0	0,0	0,2				0,2	0,1	100
4	E26b				0,3				0,1					0,1	100
	ENS	0,1		0,0	0,5	0,1	0,0		0,3				0,2	0,0	100
	INGE		0,0		0,0		0,0		0,0				0,0	0,0	100
	MANA			0,0	0,1		0,0	0,0	0,0			0,0	0,0	0,1	100
	UNIV	0,0	0,0	0,0	1,3	0,0	0,0	0,0	0,5				0,2	0,1	100
5	E26b		0,2	0,1	0,2		0,1		0,1					0,4	100
	ENS	0,5	0,2	0,1	0,8	0,1	0,1	0,0	0,6	0,0			0,2	0,2	100
	INGE		0,2	0,0	0,1	0,0	0,1	0,0	0,0				0,1	0,0	100
	MANA		0,1	0,4	0,4		0,1	0,2	0,1				0,0	0,3	100
	UNIV	0,0	0,1	0,0	2,2	0,0	0,1	0,1	1,6	0,0	0,0		0,9	0,2	100
6	E26b				0,5		0,5	0,5							100
	ENS	4,6	0,9	0,0	3,7	5,6	0,4	0,1	3,1	11,0			0,9		100
	INGE	0,5	5,3	0,1	4,3	0,4	3,0	0,0	2,3	0,0	0,5		0,5	0,1	100
	MANA	0,0	0,4	0,6	1,6	0,2	0,1	0,4	1,0			0,4	0,2	0,3	100
	UNIV	0,3	0,6	0,0	11,3	0,3	0,3	0,0	10,2	0,0	0,0		13,5	0,0	100

* Variable DEGETU. ** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues.

Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Tableau A4.6. • Où les inscrits de 2016 sont-ils inscrits en 2010, selon leur degré et composante d'inscription dans de SISE-Inscriptions 2016 (en % ligne) ?

Dans SISE Inscriptions 2016		Degré d'études* et composante** dans SISE Inscription 2010																		
		ABSENT	0		1		1		1		2		2		2		3		3	
			ENS	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	CPES-BTS	CPES-CPEG	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV
1	CPES-BTS	99,9			0,0				0,0						0,0				0,0	
	CPES-CPEG	100,0																	0,0	
	E26b	98,0		0,1	0,1				1,3	0,1					0,1			0,0	0,1	
	ENS	96,4			0,2				1,3				0,2						0,7	
	INGE	99,6			0,0	0,0			0,2	0,0	0,0				0,0		0,0		0,1	
	MANA	99,6		0,0	0,0	0,0	0,0		0,3	0,0					0,1			0,0		
	UNIV	99,0		0,0	0,0	0,0		0,0	0,5	0,0	0,0				0,1		0,0		0,1	
2	CPES-BTS	99,7			0,0				0,2	0,0					0,0				0,0	
	CPES-CPEG	100,0			0,0				0,0											
	E26b	95,3			0,4	0,1			2,3	0,1					0,8		0,0		0,4	
	ENS	94,5			0,3		0,8		2,0	0,2		0,5			0,3	0,3			0,5	
	INGE	99,0			0,1	0,1		0,0	0,4	0,1	0,0		0,0		0,1		0,0		0,1	
	MANA	99,5		0,0	0,0			0,0	0,4	0,0					0,1				0,0	
	UNIV	97,9		0,1	0,1	0,0			1,2	0,0	0,0				0,3		0,0		0,1	
3	E26b	94,8			0,4	0,2		0,1	2,9	0,2	0,0			0,0	0,7				0,4	
	ENS	97,8				0,1		0,1	0,9		0,1				0,4		0,1		0,2	
	INGE	98,6			0,1	0,1		0,0	0,7	0,0	0,1				0,2		0,0		0,1	
	MANA	98,6			0,1	0,1		0,0	0,9	0,0	0,0			0,0	0,2			0,0	0,0	
	UNIV	94,7		0,1	0,3	0,1		0,0	2,8	0,1	0,0		0,0		0,8		0,0	0,0	0,5	
4	E26b	91,1			0,6	0,2		0,2	4,8	0,4	0,0		0,0	0,0	1,5				0,6	
	ENS	92,1		0,1	0,1	1,1	0,0	0,1	3,5		0,3				0,9	0,0		0,0	0,8	
	INGE	97,2		0,0	0,1	0,6		0,1	1,4	0,0	0,1		0,0		0,2		0,1		0,1	
	MANA	94,9		0,0	0,3	0,6		0,1	2,9	0,1	0,2		0,0	0,0	0,5		0,0	0,0	0,1	
	UNIV	86,1		0,1	0,3	0,2		0,0	7,4	0,2	0,1		0,0	0,0	2,4		0,0	0,0	1,4	
5	E26b	83,0			1,3	0,3		0,2	11,9	0,6	0,2				1,3			0,1	0,8	
	ENS	76,1		0,1	0,2	4,5		0,1	10,3	0,1	1,7	0,1	0,0		3,7	0,1	0,2	0,1	1,1	
	INGE	85,9		0,0	0,3	5,0	0,0	0,9	5,0	0,1	1,1		0,1		0,8		0,2	0,0	0,2	
	MANA	80,0		0,0	0,8	3,4	0,0	0,2	9,1	0,3	0,9	0,0	0,1	0,2	2,3	0,0	0,1	0,3	0,7	
	UNIV	75,3		0,1	0,6	0,7	0,0	0,1	12,2	0,3	0,3	0,0	0,0	0,0	3,8	0,0	0,1	0,0	2,3	
6	E26b	76,4		0,5		0,5		0,5	19,2		0,5				1,0				0,5	
	ENS	46,0		0,0	0,1	6,1	0,0	0,1	5,9	0,0	7,4	0,1	0,0	0,0	5,4	1,3	0,7	0,0	5,2	
	INGE	57,4		0,0	0,3	3,6	0,0	1,6	6,5	0,2	5,4	0,0	1,5	0,0	5,4	0,3	4,6	0,1	4,4	
	MANA	65,9			0,4	1,9		0,9	13,0	0,4	1,7	0,0	0,4	0,9	5,9	0,0	0,4	0,6	2,0	
	UNIV	36,3		0,0	0,1	0,5	0,0	0,1	14,5	0,1	1,0	0,0	0,1	0,0	10,7	0,2	0,5	0,0	10,2	

Dans SISE Inscriptions 2016	Degré d'études* et composante** dans SISE Inscription 2010														
	4	4	4	4	5	5	5	5	6	6	6	6	?		
	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	ENS	INGE	MANA	UNIV	CPES-BTS		
1	CPES-BTS													100	
	CPES-CPEG													100	
	E26b			0,1									0,0	100	
	ENS							0,7				0,5		100	
	INGE			0,0				0,0						100	
	MANA													0,0	100
	UNIV			0,0	0,1			0,1					0,0	0,0	100
2	CPES-BTS													100	
	CPES-CPEG													100	
	E26b			0,2		0,1		0,1		0,0		0,0		100	
	ENS			0,3					0,3			0,2		100	
	INGE		0,0	0,0	0,0		0,0		0,0					100	
	MANA													100	
	UNIV			0,0	0,1		0,0	0,0	0,1				0,1	0,0	100
3	E26b		0,0	0,1				0,2				0,0		100	
	ENS			0,2				0,1				0,1	0,1	100	
	INGE			0,0				0,0						100	
	MANA			0,0				0,0					0,0	100	
	UNIV		0,0	0,0	0,2		0,0	0,0	0,1				0,2	0,0	100
4	E26b		0,0	0,1				0,3	0,0				0,0	100	
	ENS	0,1	0,0	0,3	0,1			0,3				0,1		100	
	INGE		0,0	0,0		0,0		0,0				0,0	0,0	100	
	MANA		0,0	0,0		0,0		0,0					0,0	100	
	UNIV	0,0	0,0	0,0	0,9	0,0	0,0	0,0	0,4				0,2	0,1	100
5	E26b		0,1	0,1		0,1							0,1	100	
	ENS	0,1	0,2	0,0	0,7		0,0		0,2	0,0			0,2	0,2	100
	INGE		0,1	0,0	0,1		0,0		0,0		0,0		0,1	0,0	100
	MANA		0,0	0,2	0,2		0,0	0,0	0,1				0,0	0,2	100
	UNIV	0,0	0,1	0,0	1,7	0,0	0,0	0,0	1,3	0,0	0,0		0,7	0,1	100
6	E26b		0,5	0,5										100	
	ENS	3,4	0,3	0,1	3,0	3,6	0,3	0,0	3,0	6,8			0,8	0,1	100
	INGE	0,4	2,9	0,1	2,1	0,2	1,0	0,1	1,3	0,0	0,2		0,4	0,1	100
	MANA	0,1	0,1	0,4	0,8	0,1	0,2	0,4	0,9			0,2	0,2	0,1	100
	UNIV	0,2	0,3	0,0	9,1	0,2	0,1	0,0	7,5	0,0	0,0		8,3	0,0	100

* Variable DEGETU. ** En cas d'inscriptions multiples en 2016, la composante retenue résulte de l'application de la méthode présentée dans le document de travail, parmi au plus, les deux inscriptions principales et les deux inscriptions secondaires retenues. Champ : ensemble des étudiants inscrits dans SISE 2016. Ne sont prises en compte que les inscriptions ayant un INE jugé « valide » (méthode de l'auteur), avec FLAG_SUP ≠ 1.

Source : fichier historicisé de SISE-Inscriptions élaboré à partir de SIES – SISE-Inscriptions 2010 à 2016 (composantes UNIV, MANA, INGE, ENS, E26b), BPBAC 2010 à 2016 (seules les inscriptions en BTS et CPEG sont conservées pour BPBAC).

Annexe 5. Programmes SAS

Partie A. Création d'une table unique d'inscrits par année (INSCRITSaa, aa=2010 à 2017) et exploitations exploratoires

- Création d'un fichier annuel INSCRITSaa compilant tous les fichiers des inscrits. Une variable SOURCEaa précise la table d'origine de l'information (UNIV, MANA, INGE...).
- Des PROC FREQ exploratoires simples sont faites sur ces tables annuelles pour quelques variables.
- Comme un même individu, repéré par son identifiant IDETU (lié de manière unique à l'INE), peut avoir plusieurs inscriptions, on repère ensuite le nombre d'inscriptions multiples et quelques PROC FREQ essaient de voir quels secteurs sont les plus concernés (ENS ? Écoles ? Université ?).

```
/* **** */
/* **** */
/** Projet CÉREQ de rapprochement experimental entre Génération 2017
/** et les données administratives du SIES sur les inscrits dans l'enseignement supérieur
/** Début des travaux en juillet 2021
/* **** */
/* **** */

/** Le repertoires des tables du Céreq et des tables en sortie ***/

libname SOR "C:\Users\Public\Documents\Bases de travail";run;

/** Les repertoires des inscrits en entrée ***/

libname Ins17uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE inscrits UNIV_2017";
libname Ins16uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2016";
libname Ins15uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2015";
libname Ins14uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2014";
libname Ins13uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2013";
libname Ins12uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2012";
libname Ins11uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2011";
```

```
libname Ins10uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits UNIV_2010";

libname Ins17man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE inscrits MANA_2017";
libname Ins16man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2016";
libname Ins15man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2015";
libname Ins14man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2014";
libname Ins13man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2013";
libname Ins12man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2012";
libname Ins11man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2011";
libname Ins10man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits MANA_2010";

libname Ins17ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE inscrits INGE_2017";
libname Ins16ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2016";
libname Ins15ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2015";
libname Ins14ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2014";
libname Ins13ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2013";
libname Ins12ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2012";
libname Ins11ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2011";
libname Ins10ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits INGE_2010";

libname Ins17ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE inscrits ENS_2017";
libname Ins16ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2016";
libname Ins15ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2015";
libname Ins14ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2014";
libname Ins13ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2013";
libname Ins12ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2012";
libname Ins11ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2011";
libname Ins10ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits ENS_2010";

libname Ins17e26 "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE inscrits Enq26bis_2017";
libname Ins16e26 "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Inscrits Enq26bis_2016";

/*** Les répertoires des lycées ***/

libname Lyc17PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2017";
libname Lyc16PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2016";
libname Lyc15PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2015";
libname Lyc14PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2014";
libname Lyc13PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2013";
libname Lyc12PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2012";
libname Lyc11PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2011";
```

```

libname Lyc10PB "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_BPAC_2010";

libname Lyc17e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2017";
libname Lyc16e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2016";
libname Lyc15e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2015";
libname Lyc14e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2014";
libname Lyc13e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2013";
libname Lyc12e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2012";
libname Lyc11e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2011";
libname Lyc10e24 "\\casd.fr\casdfs\Projets\APPSISE\Data\Suivi_lycées_Enquête 24_2010";
run;

/*****
/*****
/*****
/**** construction d'un fichier unique des inscrits par année */
/**** par compilation simple -> SOR.INSCRITS&AN. ****/
/**** (qques corrections simples pour permettre la compil ****/
/*****
/*****
/*****

%macro Inscrits;
/* Import des tables par champ */
%do an = 10 %to 17;
Data Univ ; set Ins&an.uni.inscri&an.;source="UNIV";run;
Data Mana ; set Ins&an.man.mana&an.;source="MANA";run;
Data Inge ; set Ins&an.ing.inge&an.;source="INGE";run;
Data Ens ; set Ins&an.ens.ens&an.;source="ENS ";run;

Data Prepa ; set Lyc&an.PB.bpbac20&an.; length Degetu $3.;
source="CPES";
mefst2=substr(mefst11,1,2);
IDENTU=IDENT;
if substr(mefst11,1,4) in ("3111","3211","3221","3131") then degetu="1";
else if substr(mefst11,1,4) in ("3112","3222","3232","3233") then degetu="2";
else degetu="9";
flag_sup=0;
run;

%if &an. ge 11 %then %do;

```

```

Data BTS ; set Lyc&an.e24.bts&an.;source="BTS ";drop age; run;
%end;

%if &an. ge 16 %then %do;
Data Enq26 ; set Ins&an.e26.enq26bis&an.;source="E26b";run;
%end;

/* creation des tables globales annuelles */
%if &an. = 10 %then %do; /* pour 2010 */
Data sor.inscrits&an.; set Univ Mana Inge Ens Prepa;
niveau&an.=source!!"-!!degetu;
if anbac=. or anbac="9999" then do ; anbac= "9999"; annee_connue="NON";end;
else annee_connue="OUI";
If effectif="" then effectif="9";
If nbach="" then nbach="9";
If net="" then net="9";
If flag_sup="" then flag_sup="9";
rename Source=Source&an. Degetu=Degetu&an.;run;
Data Univ Mana Inge Ens Prepa; set _null_; run;
%end;

%else %if &an. lt 16 %then %do; /* pour 2011 à 2016 */
Data sor.inscrits&an.; set Univ Mana Inge Ens Prepa BTS;
niveau&an.=source!!"-!!degetu;
if anbac=. or anbac="9999" then do ; anbac= "9999"; annee_connue="NON";end;
else annee_connue="OUI";
If effectif="" then effectif="9";
If nbach="" then nbach="9";
If net="" then net="9";
If flag_sup="" then flag_sup="9";
rename Source=Source&an. Degetu=Degetu&an.;run;
Data Univ Mana Inge Ens Prepa BTS; set _null_; run;
%end;

%else %if &an. ge 16 %then %do; /* pour 2017 */
Data sor.inscrits&an.; set Univ Mana Inge Ens Prepa BTS Enq26;
niveau&an.=source!!"-!!degetu;
if anbac=. or anbac="9999" then do ; anbac= "9999"; annee_connue="NON";end;
else annee_connue="OUI";
If effectif="" then effectif="9";
If nbach="" then nbach="9";
If net="" then net="9";

```

```

If flag_sup="" then flag_sup="9";
rename Source=Source&an. Degetu=Degetu&an.;run;
Data Univ Mana Inge Ens Prepa BTS Enq26; set _null_; run;
%end;
%end;
%mend;

%Inscrits;
run;

/*****
/*****
/*****
/***** EXPLORATION DES DONNEES DU SIES *****/
/**** A partir de la compilation simple des fichiers ****/
/**** en un seul fichier par année (SOR.INSCRITS&AN.) ****/
/*****
%let an=10; /* choix de l'année à examiner */
/*****
/*****

Data exploratoire; set sor.inscrits&an;

*** indicateur de qualité de l'INE;
If substr(IDETU,11,1)="" then Qualite_INE="INVALIDE";
  *repérage des valeurs trop courtes;
Else if IDETU*1 ge 0 then Qualite_INE="INVALIDE";
  *repérage des valeurs sans lettre;
Else if substr(IDETU,1,1) in ("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O",
  "P","Q","R","S","T","U","V","W","X","Y","Z")
  then Qualite_INE="INVALIDE";   *repérage des valeurs commençant par une lettre;
Else if substr(IDETU,1,5)="99999" then Qualite_INE="INVALIDE";
Else Qualite_INE ="VALIDE ?";
***;

If DEGETU&an.="" then DEGETU&an.="9";
If Aninsc="" then aninsc="9999";

run;

Title "Exploratoire sur la compilation des fichiers d'inscriptions (valeur 9 pour valeurs vides)";

```

```
Title3 "Source d'origine de l'inscription";
Proc freq data=exploratoire;
tables source&an.;
run;

Title3 "Qualité de l'INE (calculs SJ)";
Proc freq data=exploratoire;
tables Qualite_INE*source&an./nopercent nocol norow;
run;

Title3 "FLAG_SUP=1 pour les lignes en doublons repérées par le SIES (A SUPPRIMER ENSUITE)";
Proc freq data=exploratoire;
tables Flag_sup*source&an./nopercent nocol norow;
run;

Title "Exploratoire sur la compilation des fichiers d'inscriptions (valeur 9 pour valeurs vides)";
Title2 "CHAMP : FLAG_SUP ne 1";
Title3 "EFFECTIF=1 pour l'inscription principale (calculs sans double compte)";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables effectif*source&an./nopercent nocol norow;
run;

Title3 "NBACH=1 pour les bacheliers de l'année en première inscription à l'université";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables nbach*source&an./nopercent nocol norow;
run;

Title3 "NET=1 pour les 1ère année, 1er cycle qui sont des 1ères inscriptions à l'université";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables net*source&an. net*nbach/nopercent nocol norow;
run;

Title3 "REIMMA=P (?) pour repérer les réimmatriculations (donc nouvel INE) abusives";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables reimma*source&an./nopercent nocol norow;
run;

Title3 "DEGETU&an pour repérer le niveau d'études (en années post-bac)";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables DEGETU&an.*source&an. DEGETU&an.*effectif/nopercent nocol norow;
run;
```

```

Title3 "ANNEE_CONNUE (SJ) : connaissance ou non de ANBAC, année du bac";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables (annee_connue anbac)*source&an./nopercent nocol norow;
run;

Title3 "ANNEE_INS : année de 1ère inscription dans le sup.";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables aninsc*source&an./nopercent nocol norow;
run;

Title3 "variable MEFST2 du fichier depp sur les CPES : CPEG(31),BTS(32),FC(43),Divers post bas (52)";
Proc freq data=exploratoire (where=(flag_sup ne 1));
tables MEFST2*source&an./nopercent nocol norow;
run;

/*****/
/* recherche d'identifiants en doublon (Supérieur ie EFFECTIF = 0 ou 1) */
/*****/

Data doublon_INE;
set exploratoire (keep=IDETU Flag_SUP Source&an. qualite_ine effectif MEFST2) ;
If Flag_SUP ne 1; * suppression des doublons repérés par le SIES;
If source&an. ne "BTS "; * suppression des BTS car pas d'INE valide pour eux;
If qualite_ine = "VALIDE ?"; *conservation des INE supposés possiblement valides;

* On impute les sources CPES (CPEG et BTS) et Enq26 à EFFECTIF=1;
If source&an.="E26b" then effectif=1;
If source&an.="CPES" and MEFST2 in ("31","32")then effectif=1;

*remise à zéro pour les comptages des effectifs mis à 9 pour dire "non renseigné";
If effectif=9 then effectif=0;
run;

proc sort data=doublon_INE; by Idetu;run;
Data Doublon_INE; set doublon_INE;
retain nb&an. 0 eff&an. 0 Univ 0 Ecole 0 ENS 0 Id;
If idetu ne id then do;
    nb&an.=1;
    eff&an.=effectif;
    Univ=(source&an.="UNIV")*effectif;
    Ecole=(source&an. in ("MANA","INGE"))*effectif;

```

```

    ENS=(source&an.="ENS ")*effectif;
  end;
else do;
  nb&an.=nb&an.+1;
  eff&an.=effectif+eff&an.;
  Univ=Univ+(source&an.="UNIV")*effectif;
  Ecole=Ecole+(source&an. in ("MANA","INGE"))*effectif;
  ENS=ENS+(source&an.="ENS ")*effectif;
  end;
id=idetu;
ENS_Ecole=ENS+Ecole;
run;
Data Doublon_INE; set doublon_INE;by idetu; if last.idetu;
Sans_ins_princ=(eff&an.=0);
Tout_princ=(eff&an. gt 0 and eff&an.=nb&an.);
InsRetenu=(eff&an. le 2 or nb&an.-eff&an. le 2);*on gardera 2 principales et 2 secondaires;

* Les étudiants ayant une inscription principale à l'université;
If Univ = 1 and Univ=eff&an.      then Princ_Univ="Une seule (qui est UNIV)";
Else If Univ gt 1 and Univ=eff&an. then Princ_Univ="Plusieurs, ttes UNIV ";
Else If Univ ge 1 and Univ lt eff&an. then Princ_Univ="Plusieurs, pas ttes UNIV";
Else                               Princ_Univ="Aucune dans UNIV ";

* Les étudiants ayant une inscription principale dans une Ecole;
If ECOLE = 1 and ECOLE=eff&an.      then Princ_ECOL="Une seule (qui est ECOLE) ";
Else If ECOLE gt 1 and ECOLE=eff&an. then Princ_ECOL="Plusieurs, ttes ECOLE ";
Else If ECOLE ge 1 and UNIV ge 1     then Princ_ECOL="Plusieurs, dont ECOL & UNIV";
Else If ECOLE ge 1 and ECOLE lt eff&an. then Princ_ECOL="Autres plusieurs dont ECOLE";
Else                               Princ_ECOL="Aucune dans ECOLE ";

* Les étudiants ayant une inscription principale dans une ENS;
If ENS = 1 and ENS=eff&an.      then Princ_ENS="Une seule (qui est ENS) ";
Else If ENS gt 1 and ENS=eff&an. then Princ_ENS="Plusieurs, ttes ENS ";
Else If ENS ge 1 and UNIV ge 1     then Princ_ENS="Plusieurs, dont ENS & UNIV";
Else If ENS ge 1 and ENS lt eff&an. then Princ_ENS="Autres plusieurs dont ENS ";
Else                               Princ_ENS="Aucune dans ENS ";
run;

Title "Repérage des doublons dans la compilation des fichiers des inscrits de l'année";
Title2 "CHAMP : FLAG_SUP ne 1 // Source ne BTS // Qualité_INE=VALIDE ?";
Title3 "Nombre d'identifiants présents n fois dans la table annuelle (inscription principale ou non)";
Proc freq data=Doublon_INE;tables nb&an.;

```



```
run;

Title3 "Nombre de présences de l'identifiant et nbre d'inscriptions principales (EFFECTIF=1)";
Proc freq data=Doublon_INE; tables nb&an.*eff&an./nopercnt norow nocol;
run;

Title3 "Nbre d'inscriptions principales selon nature des établissements fréquentés ";
Proc freq data=Doublon_INE; tables Univ*(Ecole ENS)/nopercnt norow nocol;
run;

Title3 "Nombre d'étudiants sans inscriptions principales";
Proc freq data=Doublon_INE; tables Sans_ins_princ/nocum;
run;

Title3 "Les étudiants n'ayant que des inscriptions principales parmi ceux en ayant au moins une";
Proc freq data=Doublon_INE (where=(eff&an gt 0)); tables Tout_princ/nocum;
run;

Title3 "Où sont les principales ? ";
Proc freq data=Doublon_INE (where=(eff&an gt 0)); tables Princ_Univ Princ_ECOL Princ_ENS/nocum;
run;

Title3 "Les inscriptions retenues (2 principales max et 2 secondaires max)";
Proc freq data=Doublon_INE; tables InsRetenu/nocum;
run;

Data Doublon_INE Exploratoire; set _null_;
run;
```

Partie B. Création d'une table de trajectoires 2010 à 2017 (table HISTORIQUE SISE) et exploitations exploratoires

- Comme un même individu (IDETU) peut avoir plusieurs inscriptions une même année, on crée d'abord pour chaque année, une table INDIVaa qui a l'IDETU comme observation.
Pour cela, on conserve au plus deux inscriptions « principales » au sens du SIES (variable EFFECTIF=1) et deux inscriptions secondaires (EFFECTIF=0). Cette variable n'étant pas produite sur les segments E26b et CPES, on impute dans ce cas EFFECTIF=1.
On ne garde qu'une sélection de variables pour décrire les inscriptions. Ces variables sont suffixées par « _P1 » pour la 1^{re} inscription principale retenue, « _P2 » pour la 2^e inscription principale, « _S1 » pour la première inscription secondaire et « _S2 » pour la seconde inscription secondaire.
- On crée ensuite une table de trajectoire unique HISTORIQUE SISE par simple « merge » des tables annuelles INDIVaa
- Pour ces deux étapes, des analyses exploratoires sont faites avec des PROC FREQ simples. En particulier, des contrôles de cohérence sur les caractéristiques sociodémographiques sont réalisés en cas d'inscriptions multiples une même année pour s'assurer qu'il s'agit bien de la même personne (examen du sexe, année et mois de naissance, indicatrice d'une nationalité étrangère, année et type de bac).

```
/**/ Le repertoire des tables du Céreq et des tables en sortie /**/
```

```
libname SOR "C:\Users\Public\Documents\Bases de travail";run;
```

```

/*****
/*****
/*****
/**/ Constitution des tables annuelles      ***/
/**/ AVEC UNE SEULE LIGNE PAR IDETU      ***/
/**/ en conservant au plus 2 inscriptions principales ***/
/**/ et 2 inscriptions secondaires -> SOR.INDIV&AN. ***/
/**/ AVEC Macro ELABOR_INDIV(&an.);
/*****
/*****

```

```
%Macro Elabor_Indiv (an);
```

```
Data Elabor_Indiv;
```

```
set sor.inscrits&an (keep=IDETU Flag_SUP Source&an. effectif
```

```

degetu&an. TYP_DIPL anbac sexe annais monais etabli PARIPA
FR_ETR MEFST2 BAC_RGRP ANSUP ANETAB TYPREPA INSPR ECHANG REGIME
      CURPAR CONV AMENA VOIE);

      /*****
      /* Construction de QUALITE_INE nécessaire au filtrage */
      *****/
If substr(IDETU,11,1)="" then Qualite_INE="INVALIDE";
  *repérage des valeurs trop courtes;
Else if IDETU*1 ge 0 then Qualite_INE="INVALIDE";
  *repérage des valeurs sans lettre;
Else if substr(IDETU,1,1) in ("A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O",
  "P", "Q", "R", "S", "T", "U", "V", "W", "X", "Y", "Z")
  then Qualite_INE="INVALIDE";      *repérage des valeurs commençant par une lettre;
Else if substr(IDETU,1,5)="99999" then Qualite_INE="INVALIDE";
Else Qualite_INE ="VALIDE ?";

* On impute les sources CPES (CPEG et BTS) et Enq26 à EFFECTIF=1;
If source&an.="E26b" then effectif=1;
If source&an.="CPES" and MEFST2="31" then do; source&an.="CPEG"; effectif=1;end;
If source&an.="CPES" and MEFST2="32" then do; source&an.="BTS_"; effectif=1;end;
  *In fine, pas de BTS car pas d INE valide;

      /*****
      /* FILTRAGE */
      *****/
If Flag_SUP= 1 then delete; * suppression des doublons repérés par le SIES;
If source&an. = "BTS " then delete; * suppression des BTS car pas d INE valide pour eux;
If qualite_ine ne "VALIDE ?" then delete ; *conservation des INE supposés possiblement valides;
If effectif =9 then delete ;*suppression des observations à 9 (= de fait, aux "CPES" non retenus)

      /*****
      /* Modification et construction de Variables */
      *****/

*imputation à 9 des valeurs manquantes (pour avoir des comptages exhaustifs);
If DEGETU&an.=" " then DEGETU&an.="9";
If Aninsc="" then aninsc="9999";
run;

      /*****

```

```

/* Numérotation des inscriptions principales et secondaires */
/*****/

Proc Freq data=Elabor_Indiv; tables source&an.*Effectif/nopercent norow nocol;run;

Proc sort data=Elabor_Indiv; by Idetu;run;

Data Elabor_Indiv; set Elabor_Indiv;
retain Sec&an. 0 Princ&an. 0 Id;
If idetu ne id then do;
    Princ&an.=effectif;
    Sec&an=(effectif=0);
end;
else do;
    Princ&an.=Princ&an.+effectif;
    Sec&an.=Sec&an.+(effectif=0);
end;
id=idetu;
run;

Title "CONSTRUCTION DE LA TABLE SOR.INDIV&an.";
Title2 "Contrôle sur les inscriptions multiples";
Title3 "(Eff&an. : Nbre d inscriptions principales // Nb&an. : nbre d inscriptions secondaires)";
Data verif; set Elabor_Indiv;by idetu; if last.idetu;
Keep idetu Sec&an. Princ&an.;
Proc freq data=verif; tables Sec&an.*Princ&an./nopercent norow nocol;
run;

/*****/
/* Table des lères inscriptions principales (P1) */
/*****/
Data P1; set Elabor_Indiv (keep= Sec&an. Princ&an. IDETU Source&an. effectif
degetu&an. TYP_DIPL anbac sexe annais monais etabli
FR_ETR PARIPA
BAC_RGRP ANSUP ANETAB TYPREPA INSPR ECHANG REGIME
CURPAR CONV AMENA VOIE);

If effectif=1 and Princ&an.=1; *selection de la première inscription principale observée;

Sociodemo&an._P1=sexe!!annais!!Monais;

```

```

Rename sexe=sexe&an._P1 annais=annais&an._P1 monais=monais&an._P1 FR_ETR=FR_ETR&an._P1
    etabli=etabli&an._P1 degetu&an.=degetu&an._P1 anbac=anbac&an._P1 BAC_RGRP=BAC_RGRP&an._P1
    PARIPA=PARIPA&an._P1
    ANSUP=ANSUP&an._P1 ANETAB=ANETAB&an._P1 TYPREPA=TYPREPA&an._P1 INSPR=INSPR&an._P1
    ECHANG=ECHANG&an._P1 REGIME=REGIME&an._P1 CURPAR=CURPAR&an._P1 CONV=CONV&an._P1
    AMENA=AMENA&an._P1 VOIE=VOIE&an._P1 Source&an.=Source&an._P1 TYP_DIPL=TYP_DIPL&an._P1;
Drop effectif ;
run;
Title2 "TABLE P1 (1ere inscription principale)";
proc freq data=P1; tables Sec&an.*Princ&an./nopercent norow nocol;
run;
    /*****/
    /* Table des 2e inscriptions principales (P2) */
    /*****/
Data P2; set Elabor_Indiv (keep= Sec&an. Princ&an. IDETU Source&an. effectif
    degetu&an. TYP_DIPL anbac sexe annais monais etabli
    FR_ETR PARIPA
        BAC_RGRP ANSUP ANETAB TYPREPA INSPR ECHANG REGIME
        CURPAR CONV AMENA VOIE);

If effectif=1 and Princ&an.=2; *selection de la première inscription principale observée;

Sociodemo&an._P2=sexe!!annais!!Monais;

Rename sexe=sexe&an._P2 annais=annais&an._P2 monais=monais&an._P2 FR_ETR=FR_ETR&an._P2
    etabli=etabli&an._P2 degetu&an.=degetu&an._P2 anbac=anbac&an._P2 BAC_RGRP=BAC_RGRP&an._P2
    PARIPA=PARIPA&an._P2
    ANSUP=ANSUP&an._P2 ANETAB=ANETAB&an._P2 TYPREPA=TYPREPA&an._P2 INSPR=INSPR&an._P2
    ECHANG=ECHANG&an._P2 REGIME=REGIME&an._P2 CURPAR=CURPAR&an._P2 CONV=CONV&an._P2
    AMENA=AMENA&an._P2 VOIE=VOIE&an._P2 Source&an.=Source&an._P2 TYP_DIPL=TYP_DIPL&an._P2;
Drop effectif ;
run;
Title2 "TABLE P2 (2e inscription principale)";
proc freq data=P2; tables Sec&an.*Princ&an./nopercent norow nocol;
run;
    /*****/
    /* Table des 1ères inscriptions secondaires (S1) */
    /*****/
Data S1; set Elabor_Indiv (keep= Sec&an. Princ&an. IDETU Source&an. effectif
    degetu&an. TYP_DIPL anbac sexe annais monais etabli
    FR_ETR PARIPA

```

```

BAC_RGRP ANSUP ANETAB TYPREPA INSPR ECHANG REGIME
CURPAR CONV AMENA VOIE);

If effectif=0 and Sec&an.=1; *selection de la première inscription principale observée;

Sociodemo&an._S1=sexe!!annais!!Monais;

Rename sexe=sexe&an._S1 annais=annais&an._S1 monais=monais&an._S1 FR_ETR=FR_ETR&an._S1
    etabli=etabli&an._S1 degetu&an.=degetu&an._S1 anbac=anbac&an._S1 BAC_RGRP=BAC_RGRP&an._S1
    PARIPA=PARIPA&an._S1
    ANSUP=ANSUP&an._S1 ANETAB=ANETAB&an._S1 TYPREPA=TYPREPA&an._S1 INSPR=INSPR&an._S1
    ECHANG=ECHANG&an._S1 REGIME=REGIME&an._S1 CURPAR=CURPAR&an._S1 CONV=CONV&an._S1
    AMENA=AMENA&an._S1 VOIE=VOIE&an._S1 Source&an.=Source&an._S1 TYP_DIPL=TYP_DIPL&an._S1;

Drop effectif ;
run;
Title2 "TABLE S1 (1ere inscription secondaire)";
proc freq data=S1; tables Sec&an.*Princ&an./nopercnt norow nocol;
run;
    /*****/
    /* Table des 2e inscriptions secondaires (S2)          */
    /*****/
Data S2; set Elabor_Indiv (keep= Sec&an. Princ&an. IDETU Source&an. effectif
    degetu&an. TYP_DIPL anbac sexe annais monais etabli
    FR_ETR PARIPA
    BAC_RGRP ANSUP ANETAB TYPREPA INSPR ECHANG REGIME
    CURPAR CONV AMENA VOIE);

If effectif=0 and Sec&an.=2; *selection de la première inscription principale observée;

Sociodemo&an._S2=sexe!!annais!!Monais;

Rename sexe=sexe&an._S2 annais=annais&an._S2 monais=monais&an._S2 FR_ETR=FR_ETR&an._S2
    etabli=etabli&an._S2 degetu&an.=degetu&an._S2 anbac=anbac&an._S2 BAC_RGRP=BAC_RGRP&an._S2
    PARIPA=PARIPA&an._S2
    ANSUP=ANSUP&an._S2 ANETAB=ANETAB&an._S2 TYPREPA=TYPREPA&an._S2 INSPR=INSPR&an._S2
    ECHANG=ECHANG&an._S2 REGIME=REGIME&an._S2 CURPAR=CURPAR&an._S2 CONV=CONV&an._S2
    AMENA=AMENA&an._S2 VOIE=VOIE&an._S2 Source&an.=Source&an._S2 TYP_DIPL=TYP_DIPL&an._S2;

Drop effectif ;
run;
Title2 "TABLE S2 (1e inscription secondaire)";
proc freq data=S2; tables Sec&an.*Princ&an./nopercnt norow nocol;
run;

```

```

/*****/
/* Appariement pour créer la table des individus */
/* avec au plus 2 inscriptions principales et 2 second*/
/* -> SOR.INDIV&AN. */
/*****/

Proc sort data= verif; by idetu;run;
Proc sort data= P1; by idetu;run;
Proc sort data= P2; by idetu;run;
Proc sort data= S1; by idetu;run;
Proc sort data= S2; by idetu;run;

Data INDIV&AN.; merge verif P1 P2 S1 S2; by IDETU;

** selection de 1 inscription privilégiée pour 1 annee;
** Plus haut degré d etude et ecole/ENS prime sur univ;
** Sinon,on garde P1;

If Source&an._P1 ne "" and Source&an._P2=""
then do; Inscrit&an.=degetu&an._P1!!Source&an._P1;Selec_Ins&an.="P1-1P ";end;

Else if Source&an._P1 ne "" and Source&an._P2 ne ""
then do;

If degetu&an._P1*1 gt degetu&an._P2*1
then do ;Inscrit&an.=degetu&an._P1!!Source&an._P1;Selec_Ins&an.="P1-2P/a";end;

else If degetu&an._P1*1 lt degetu&an._P2*1
then do; Inscrit&an.=degetu&an._P2!!Source&an._P2;Selec_Ins&an.="P2-2P/a";end;

else If degetu&an._P1*1=degetu&an._P2*1 and source&an._P1="ENS "
then do ;Inscrit&an.=degetu&an._P1!!Source&an._P1;Selec_Ins&an.="P1-2P/b";end;

else If degetu&an._P1*1=degetu&an._P2*1 and source&an._P2="ENS "
then do; Inscrit&an.=degetu&an._P2!!Source&an._P2;Selec_Ins&an.="P2-2P/b";end;

else If degetu&an._P1*1=degetu&an._P2*1 and source&an._P1 in ("MANA","INGE")
then do ;Inscrit&an.=degetu&an._P1!!Source&an._P1;Selec_Ins&an.="P1-2P/c";end;

else If degetu&an._P1*1=degetu&an._P2*1 and source&an._P2 in ("MANA","INGE")
then do; Inscrit&an.=degetu&an._P2!!Source&an._P2;Selec_Ins&an.="P2-2P/c";end;

```

```

else If degetu&an._P1*1=degetu&an._P2*1 and source&an._P1 in ("CGPE")
    then do ;Inscrit&an.=degetu&an._P1!!Source&an._P1;Selec_Ins&an.="P1-2P/d";end;

else If degetu&an._P1*1=degetu&an._P2*1 and source&an._P2 in ("CGPE")
    then do ;Inscrit&an.=degetu&an._P2!!Source&an._P2;Selec_Ins&an.="P2-2P/d";end;

else If degetu&an._P1*1=degetu&an._P2*1
    then do ;Inscrit&an.=degetu&an._P1!!Source&an._P1;Selec_Ins&an.="P1-2P/e";end;

end;

Else If Source&an._S1 ne ""
    then do; Inscrit&an.=degetu&an._S1!!Source&an._S1;Selec_Ins&an.="S1-0P ";end;
Else do; Selec_Ins&an.="ABSENT "; Inscrit&an.="ABSEN";end;

Drop Sec&an. Princ&an.;
run;

Title "SELECTION D UNE INSCRIPTION PRIVILEGIEE";
Proc freq data=INDIV&AN.;
tables Inscrit&an. Selec_Ins&an. /nocum;
run;

Data sor.INDIV&AN.;set INDIV&AN.;run;
Data INDIV&AN. Elabor_Indiv verif P1 P2 S1 S2; set _null_;
run;
%mend;

*****;
* FIN MACRO Elabor_Indiv;
*****;

/*****/
/* MACRO COHERENCE pour */
/* Examen de la coéherence des infos dans SOR.INDIV&AN. */
/* en cas d'inscriptions multiples */
/*****/

%Macro Coherence (VAR,an);
Data INDIV&AN.;set sor.INDIV&AN.;

```



```

If &VAR.&an._P1 ne "" then do;
  If &VAR.&an._P2 ne "" and &VAR.&an._P1=&VAR.&an._P2 then Coher1_&VAR.="OK/P1-P2";
  If &VAR.&an._P2 ne "" and &VAR.&an._P1 ne &VAR.&an._P2 then Coher1_&VAR.="NO/P1-P2";

  If &VAR.&an._S1 ne "" and &VAR.&an._P1=&VAR.&an._S1 then Coher2_&VAR.="OK/P1-S1";
  If &VAR.&an._S1 ne "" and &VAR.&an._P1 ne &VAR.&an._S1 then Coher2_&VAR.="NO/P1-S1";

  If &VAR.&an._S2 ne "" and &VAR.&an._P1=&VAR.&an._S2 then Coher3_&VAR.="OK/P1-S2";
  If &VAR.&an._S2 ne "" and &VAR.&an._P1 ne &VAR.&an._S2 then Coher3_&VAR.="NO/P1-S2";
end;

If &VAR.&an._P2 ne "" then do;
  If &VAR.&an._S1 ne "" and &VAR.&an._P2=&VAR.&an._S1 then Coher4_&VAR.="OK/P2-S1";
  If &VAR.&an._S1 ne "" and &VAR.&an._P2 ne &VAR.&an._S1 then Coher4_&VAR.="NO/P2-S1";

  If &VAR.&an._S2 ne "" and &VAR.&an._P2=&VAR.&an._S2 then Coher5_&VAR.="OK/P2-S2";
  If &VAR.&an._S2 ne "" and &VAR.&an._P2 ne &VAR.&an._S2 then Coher5_&VAR.="NO/P2-S2";
end;

If &VAR.&an._S1 ne "" then do;
  If &VAR.&an._S2 ne "" and &VAR.&an._S1=&VAR.&an._S2 then Coher6_&VAR.="OK/S1-S2";
  If &VAR.&an._S2 ne "" and &VAR.&an._S1 ne &VAR.&an._S2 then Coher6_&VAR.="NO/S1-S2";
end;

If Coher1_&VAR.=""" then Coher1_&VAR.="SANS OBJ";
If Coher2_&VAR.=""" then Coher2_&VAR.="SANS OBJ";
If Coher3_&VAR.=""" then Coher3_&VAR.="SANS OBJ";
If Coher4_&VAR.=""" then Coher4_&VAR.="SANS OBJ";
If Coher5_&VAR.=""" then Coher5_&VAR.="SANS OBJ";
If Coher6_&VAR.=""" then Coher6_&VAR.="SANS OBJ";

Nb_&VAR._OK=(substr(Coher1_&VAR.,1,2)="OK")+(substr(Coher2_&VAR.,1,2)="OK")+(substr(Coher3_&VAR.,1,2)="OK")
+(substr(Coher4_&VAR.,1,2)="OK")+(substr(Coher5_&VAR.,1,2)="OK")+(substr(Coher6_&VAR.,1,2)="OK");

Nb_&VAR._NO=(substr(Coher1_&VAR.,1,2)="NO")+(substr(Coher2_&VAR.,1,2)="NO")+(substr(Coher3_&VAR.,1,2)="NO")
+(substr(Coher4_&VAR.,1,2)="NO")+(substr(Coher5_&VAR.,1,2)="NO")+(substr(Coher6_&VAR.,1,2)="NO");run;
Data sor.Coher_&VAR._&an.; set INDIV&AN.;
Keep idetu Nb_&VAR._OK Nb_&VAR._NO;
run;

```

```

Title "CONSTRUCTION DE LA TABLE SOR.INDIV&an.";
Title2 "Cohérence dans la table INDIV finale";
Title3 "P1 et P2= inscription principales // S1 et S2= inscription secondaires";
Proc freq data=INDIV&AN.;
tables Nb_&VAR._OK Nb_&VAR._NO /nocum;
*tables Coher1_&VAR. Coher2_&VAR. Coher3_&VAR. Coher4_&VAR. Coher5_&VAR. Coher6_&VAR./nocum;
run;
%mend;

*****;
* FIN MACRO Coherence;
*****;

*****;
*Creation de la composante 2017 de l'historique et verif de sa coherence;
%Elabor_Indiv (17);run;
%Coherence (SocioDemo,17);
%Coherence (Sexe,17);
%Coherence (annais,17); %Coherence (monais,17);run;
%Coherence (FR_ETR,17);run;
%Coherence (anbac,17);%Coherence (BAC_RGRP,17);run;
***;
*Creation de la composante 2016 de l'historique et verif de sa coherence;
%Elabor_Indiv (16);run;
%Coherence (SocioDemo,16);
%Coherence (Sexe,16);
%Coherence (annais,16); %Coherence (monais,16);run;
%Coherence (FR_ETR,16);run;
%Coherence (anbac,16);%Coherence (BAC_RGRP,16);run;
***;
*Creation de la composante 2015 de l'historique et verif de sa coherence;
%Elabor_Indiv (15);run;
%Coherence (SocioDemo,15);
%Coherence (Sexe,15);
%Coherence (annais,15); %Coherence (monais,15);run;
%Coherence (FR_ETR,15);run;
%Coherence (anbac,15);%Coherence (BAC_RGRP,15);run;
***;
*Creation de la composante 2014 de l'historique et verif de sa coherence;
%Elabor_Indiv (14);run;
%Coherence (SocioDemo,14);
%Coherence (Sexe,14);

```

```
%Coherence (annais, 14) ; %Coherence (monais, 14) ; run;
%Coherence (FR_ETR, 14) ; run;
%Coherence (anbac, 14) ; %Coherence (BAC_RGRP, 14) ; run;
***;
*Creation de la composante 2013 de l'historique et verif de sa coherence;
%Elabor_Indiv (13) ; run;
%Coherence (SocioDemo, 13) ;
%Coherence (Sexe, 13) ;
%Coherence (annais, 13) ; %Coherence (monais, 13) ; run;
%Coherence (FR_ETR, 13) ; run;
%Coherence (anbac, 13) ; %Coherence (BAC_RGRP, 13) ; run;
***;
*Creation de la composante 2012 de l'historique et verif de sa coherence;
%Elabor_Indiv (12) ; run;
%Coherence (SocioDemo, 12) ;
%Coherence (Sexe, 12) ;
%Coherence (annais, 12) ; %Coherence (monais, 12) ; run;
%Coherence (FR_ETR, 12) ; run;
%Coherence (anbac, 12) ; %Coherence (BAC_RGRP, 12) ; run;
***;
*Creation de la composante 2011 de l'historique et verif de sa coherence;
%Elabor_Indiv (11) ; run;
%Coherence (SocioDemo, 11) ;
%Coherence (Sexe, 11) ;
%Coherence (annais, 11) ; %Coherence (monais, 11) ; run;
%Coherence (FR_ETR, 11) ; run;
%Coherence (anbac, 11) ; %Coherence (BAC_RGRP, 11) ; run;
***;
*Creation de la composante 2010 de l'historique et verif de sa coherence;
%Elabor_Indiv (10) ; run;
%Coherence (SocioDemo, 10) ;
%Coherence (Sexe, 10) ;
%Coherence (annais, 10) ; %Coherence (monais, 10) ; run;
%Coherence (FR_ETR, 10) ; run;
%Coherence (anbac, 10) ; %Coherence (BAC_RGRP, 10) ; run;
***;

/*****/
/*****/
/*** CREATION DE LA TABLE HISTORIQUE UNIQUE
```

```
/** par appariement des tables INDIV&an créées
/** ci-dessus
/**
/**

Proc sort data=sor.indiv17; by idetu;run;
Proc sort data=sor.indiv16; by idetu;run;
Proc sort data=sor.indiv15; by idetu;run;
Proc sort data=sor.indiv14; by idetu;run;
Proc sort data=sor.indiv13; by idetu;run;
Proc sort data=sor.indiv12; by idetu;run;
Proc sort data=sor.indiv11; by idetu;run;
Proc sort data=sor.indiv10; by idetu;run;

Proc sort data=sor.coher_sociodemo_17; by idetu;run;
Proc sort data=sor.coher_sociodemo_16; by idetu;run;
Proc sort data=sor.coher_sociodemo_15; by idetu;run;
Proc sort data=sor.coher_sociodemo_14; by idetu;run;
Proc sort data=sor.coher_sociodemo_13; by idetu;run;
Proc sort data=sor.coher_sociodemo_12; by idetu;run;
Proc sort data=sor.coher_sociodemo_11; by idetu;run;
Proc sort data=sor.coher_sociodemo_10; by idetu;run;

Data sor.HistoriqueSise;
merge sor.indiv17 sor.indiv16 sor.indiv15 sor.indiv14
      sor.indiv13 sor.indiv12 sor.indiv11 sor.indiv10
      sor.coher_sociodemo_17 sor.coher_sociodemo_16 sor.coher_sociodemo_15 sor.coher_sociodemo_14
      sor.coher_sociodemo_13 sor.coher_sociodemo_12 sor.coher_sociodemo_11 sor.coher_sociodemo_10;
by idetu;

If Inscrit17="" then Inscrit17="ABSEN";
If Inscrit16="" then Inscrit16="ABSEN";
If Inscrit15="" then Inscrit15="ABSEN";
If Inscrit14="" then Inscrit14="ABSEN";
If Inscrit13="" then Inscrit13="ABSEN";
If Inscrit12="" then Inscrit12="ABSEN";
If Inscrit11="" then Inscrit11="ABSEN";
If Inscrit10="" then Inscrit10="ABSEN";
run;
```

```
/*  
/*  
*** EXAMEN DES TRAJECTOIRES  
/*  
/*
```

```
Data HistoriqueSise; set sor.HistoriqueSise;  
Nb_annee=(Inscrit15 ne "ABSEN")+ (Inscrit14 ne "ABSEN")+ (Inscrit13 ne "ABSEN")  
+ (Inscrit12 ne "ABSEN")+ (Inscrit11 ne "ABSEN")+ (Inscrit10 ne "ABSEN");  
run;
```

```
Title "Où retrouve-t-on les inscrits de 2016 les années précédentes ?";  
Proc freq data=HistoriqueSise  
(where=(Inscrit16 ne "ABSEN" and substr(Inscrit16,1,1) in ("1","2","3","4","5","6")));  
Tables Inscrit16*Nb_annee/norow nocol nopercnt;  
Tables Inscrit16*(Inscrit15 Inscrit14 Inscrit13 Inscrit12 Inscrit11 Inscrit10)/nofreq nocol nopercnt;  
run;
```

Partie C. Rapprochement de la table HISTORIQUE SISE et de la table de l'échantillon de l'enquête Génération (table SISE_ECHANT) et exploitations exploratoires sur la capacité à faire ce rapprochement

- La table de l'échantillon de Génération (« ine_gene2017 »-> ECHANTILLON) est la table de passage qui permet de relier l'identifiant SISE (IDETU) à l'identifiant non signifiant de l'enquête (IDENT).

Plus exactement, seuls les deux premiers caractères de l'IDENT ont un sens : il désigne la composante de sondage, celle-ci étant construite par récupération d'informations auprès de différents acteurs (« UN » pour les universités, « SY » pour le système d'information Scolarité de la DEPP, « SI » pour le système d'information des formations par apprentissage de la DEPP, etc.).

La table ECHANTILLON comporte aussi une variable REPONDANT, variable indicatrice permettant de repérer les personnes échantillonnées ayant répondu à l'enquête.

- Certains établissements n'ont pas pu transmettre d'INE/IDETU au Céreq. Cette information n'est pas disponible dans la table de passage du Céreq. Une variable QUALITE_INE est donc construite pour identifier sur la base du format si l'INE disponible dans la table échantillon est « valide » ou « invalide ».
- Pour certaines personnes (IDENT), le Céreq a collecté plusieurs INE/IDETU. Avant l'appariement avec HISTORIQUE SISE, on crée une table ayant une seule observation par IDENT et contenant tous les INE/IDETU qui lui sont associés (3 au plus après analyse exploratoire) => table Echant_Céreq
- On apparie ensuite cette table Echant_Céreq avec la table HISTORIQUE SISE. Cet appariement est effectué en deux tours (car très peu d'IDENT sont associés à 3 INE/IDETU) => tables SISE_ECHANT, avec des variables suffixées par « T1 » et « T2 ».
- À ce stade, il s'agit uniquement de voir si on retrouve des informations SISE pour les individus de l'échantillon et de voir si, quand un IDENT a plusieurs INE/IDETU associés, il s'agit bien de la même personne. De ce fait on ne retient que quelques variables sociodémographiques dans HISTORIQUE SISE (année de naissance, mois de naissance, sexe, indicatrice d'une nationalité étrangère).
- Quelques analyses exploratoires de contrôles sont faites aux différentes étapes.

```
/** Le repertoire des tables du Céreq et des tables en sortie **/
```

```
libname SOR "C:\Users\Public\Documents\Bases de travail";run;
```

```
/***/  
/***/  
/***/  
/** Repérage des doublons */  
/** dans la table de passage du Céreq "IDENT-INE" */
```

```
/** Et création d'une table avec une seule ligne */
/** par IDENT (bcp de doublons et qqes triplettes) */
/** Surtout lié à SISE et SYCA et un peu UNIV */
/*****
/*****
/*****
/*****
/*****/

Data Echantillon; set sor.ine_gene2017;
run;
Proc sort data=Echantillon; by Ident;
run;

Data Echantillon; set Echantillon; retain n 0 Ident1;
*comptage des ident;
If Ident ne Ident1 then do;Ident1=Ident; n=1; end;
Else do; Ident1=Ident; n=n+1; end;
*Identification de la composante de la base de sondage;
Prefix_Ident=substr(ident,1,2);

*** indicateur de qualité de l'INE;
If substr(IDETU,11,1)="" then Qualite_INE="INVALIDE";
*repérage des valeurs trop courtes;
Else if IDETU*1 ge 0 then Qualite_INE="INVALIDE";
*repérage des valeurs sans lettre;
Else if substr(IDETU,1,1) in ("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O",
" P","Q","R","S","T","U","V","W","X","Y","Z")
then Qualite_INE="INVALIDE"; *repérage des valeurs commençant par une lettre;
Else if substr(IDETU,1,5)="99999" then Qualite_INE="INVALIDE";
Else Qualite_INE ="VALIDE ?";
run;

run;
Title "Examen des IDENT en doublon dans la table de passage IDENT/INE du CÉREQ";
Title2 "SUR TOUT L ECHANTILLON";
Proc freq data=Echantillon;
Table n;
Tables Prefix_Ident*n/nopercent norow nocol;
Tables Qualite_INE*n/nopercent norow nocol;
Tables Repondant*n/nopercent norow nocol;
run;
Title2 "Sur champ des répondants (ie Sport)";
```

```
Proc freq data=Echantillon (where=(repondant="1"));
Tables Prefix_Ident*n/nopercent norow nocol;
run;

/*****/
* Creation de la table avec une ligne par IDENT ;
* et les différents IDETU associés;
**** ATTENTION !!!! SUPPRESSION DES INE SUPPOSES INVALIDES;
/*****/

** conservation des observations avec n=1;
Data Echant1; set Echantillon; if n=1;
IDETU1=IDETU; rep1=repondant;
If Qualite_INE="INVALIDE" then delete;
keep IDENT IDETU1 rep1;run;
Proc sort data=Echant1; by ident; run;

** conservation des observations avec n=2;
Data Echant2; set Echantillon; if n=2;
IDETU2=IDETU; rep2=repondant;
If Qualite_INE="INVALIDE" then delete;
keep IDENT IDETU2 rep2;run;
Proc sort data=Echant2; by ident; run;

** conservation des observations avec n=3;
Data Echant3; set Echantillon; if n=3;
IDETU3=IDETU; rep3=repondant;
If Qualite_INE="INVALIDE" then delete;
keep IDENT IDETU3 rep3;run;
Proc sort data=Echant3; by ident; run;

** creation de la table globale;
Data sor.Echant_Céreq; Merge Echant1 Echant2 Echant3; by ident;
repondant=rep1;
*Identification de la composante de la base de sondage;
Prefix_Ident=substr(ident,1,2);
run;

Data Echantillon Echant1 Echant2 Echant3; set _null_;
run;
```



```

/*****
/***** APPARIEMENT AVEC LES IDENTIFIANTS GENE 2017 */
/***** (RECONDANTS ET NON RECONDANTS) */
**** ATTENTION !!!! SUPPRESSION DES INE SUPPOSES INVALIDES;
**** Et rapprochement en plusieurs tours pour tester les;
**** différents IDETU des IDENT en ayant plusieurs;
/*****/
Proc sort data=sor.HistoriqueSise;
by Idetu;
run;

/*****/
/*****/
**** 1er tour;
/*****/
/*****/

Data Tour1; set sor.Echant_Céreq;
Idetu=idetu1; If Idetu="" then delete;run;
Proc sort data=Tour1;by idetu;
run;

Data Tour1;
merge Tour1 (in=a)
sor.HistoriqueSise (keep=IDETU Source17_P1 Source17_S1 Source16_P1 Source16_S1
Source15_P1 Source15_S1 Source14_P1 Source14_S1
Source13_P1 Source13_S1 Source12_P1 Source12_S1
Source11_P1 Source11_S1 Source10_P1 Source10_S1

Sexe17_P1 Sexe17_S1 Sexe16_P1 Sexe16_S1
Sexe15_P1 Sexe15_S1 Sexe14_P1 Sexe14_S1
Sexe13_P1 Sexe13_S1 Sexe12_P1 Sexe12_S1
Sexe11_P1 Sexe11_S1 Sexe10_P1 Sexe10_S1

annais17_P1 annais17_S1 annais16_P1 annais16_S1
annais15_P1 annais15_S1 annais14_P1 annais14_S1
annais13_P1 annais13_S1 annais12_P1 annais12_S1

```

```

annais11_P1 annais11_S1 annais10_P1 annais10_S1

FR_ETR17_P1 FR_ETR17_S1 FR_ETR16_P1 FR_ETR16_S1
FR_ETR15_P1 FR_ETR15_S1 FR_ETR14_P1 FR_ETR14_S1
FR_ETR13_P1 FR_ETR13_S1 FR_ETR12_P1 FR_ETR12_S1
FR_ETR11_P1 FR_ETR11_S1 FR_ETR10_P1 FR_ETR10_S1

Inscrit17 Inscrit16 Inscrit15 Inscrit14 Inscrit13
Inscrit12 inscrist11 Inscrit10

Selec_Ins17 Selec_Ins16 Selec_Ins15 Selec_Ins14 Selec_Ins13
Selec_Ins12 Selec_Ins11 Selec_Ins10);

by idetu ; if a;

Tour1=1;

Sise17_T1=(Source17_P1 ne "" or Source17_S1 ne "");
Sise16_T1=(Source16_P1 ne "" or Source16_S1 ne "");
Sise15_T1=(Source15_P1 ne "" or Source15_S1 ne "");
Sise14_T1=(Source14_P1 ne "" or Source14_S1 ne "");
Sise13_T1=(Source13_P1 ne "" or Source13_S1 ne "");
Sise12_T1=(Source12_P1 ne "" or Source12_S1 ne "");
Sise11_T1=(Source11_P1 ne "" or Source11_S1 ne "");
Sise10_T1=(Source10_P1 ne "" or Source10_S1 ne "");
Nb_SISE_T1=Sise17_T1+Sise16_T1+Sise15_T1+Sise14_T1+Sise13_T1+Sise12_T1+Sise11_T1+Sise10_T1;

SocioDemo17_T1=Sexe17_P1!!annais17_P1!!FR_ETR17_P1;
SocioDemo16_T1=Sexe16_P1!!annais16_P1!!FR_ETR16_P1;
SocioDemo15_T1=Sexe15_P1!!annais15_P1!!FR_ETR15_P1;
SocioDemo14_T1=Sexe14_P1!!annais14_P1!!FR_ETR14_P1;
SocioDemo13_T1=Sexe13_P1!!annais13_P1!!FR_ETR13_P1;
SocioDemo12_T1=Sexe12_P1!!annais12_P1!!FR_ETR12_P1;
SocioDemo11_T1=Sexe11_P1!!annais11_P1!!FR_ETR11_P1;
SocioDemo10_T1=Sexe10_P1!!annais10_P1!!FR_ETR10_P1;

Inscrit17_T1=inscrist17; Selec_Ins17_T1=Selec_Ins17;
Inscrit16_T1=inscrist16; Selec_Ins16_T1=Selec_Ins16;
Inscrit15_T1=inscrist15; Selec_Ins15_T1=Selec_Ins15;
Inscrit14_T1=inscrist14; Selec_Ins14_T1=Selec_Ins14;
Inscrit13_T1=inscrist13; Selec_Ins13_T1=Selec_Ins13;
Inscrit12_T1=inscrist12; Selec_Ins12_T1=Selec_Ins12;

```

```
Inscrit11_T1=inscrit11; Selec_Ins11_T1=Selec_Ins11;  
Inscrit10_T1=inscrit10; Selec_Ins10_T1=Selec_Ins10;
```

```
Drop Source17_P1 Source17_S1 Source16_P1 Source16_S1 Source15_P1 Source15_S1 Source14_P1 Source14_S1  
Source13_P1 Source13_S1 Source12_P1 Source12_S1 Source11_P1 Source11_S1 Source10_P1 Source10_S1
```

```
Sexe17_P1 Sexe17_S1 Sexe16_P1 Sexe16_S1 Sexe15_P1 Sexe15_S1 Sexe14_P1 Sexe14_S1  
Sexe13_P1 Sexe13_S1 Sexe12_P1 Sexe12_S1 Sexe11_P1 Sexe11_S1 Sexe10_P1 Sexe10_S1
```

```
annais17_P1 annais17_S1 annais16_P1 annais16_S1 annais15_P1 annais15_S1 annais14_P1 annais14_S1  
annais13_P1 annais13_S1 annais12_P1 annais12_S1 annais11_P1 annais11_S1 annais10_P1 annais10_S1
```

```
FR_ETR17_P1 FR_ETR17_S1 FR_ETR16_P1 FR_ETR16_S1 FR_ETR15_P1 FR_ETR15_S1 FR_ETR14_P1 FR_ETR14_S1  
FR_ETR13_P1 FR_ETR13_S1 FR_ETR12_P1 FR_ETR12_S1 FR_ETR11_P1 FR_ETR11_S1 FR_ETR10_P1 FR_ETR10_S1
```

```
Inscrit17 Inscrit16 Inscrit15 Inscrit14 Inscrit13 Inscrit12 inscrit11 Inscrit10  
Selec_Ins17 Selec_Ins16 Selec_Ins15 Selec_Ins14 Selec_Ins13 Selec_Ins12 Selec_Ins11 Selec_Ins10;  
run;
```

```
/*  
/*  
**** 2e tour;  
/*  
/*
```

```
Data Tour2; set sor.Echant_Céreq;  
Idetu=idetu2;if idetu="" then delete; run;  
Proc sort data=Tour2;by idetu;  
run;
```

```
Data Tour2;  
merge Tour2 (in=a)  
sor.HistoriqueSise (keep=IDETU Source17_P1 Source17_S1 Source16_P1 Source16_S1  
Source15_P1 Source15_S1 Source14_P1 Source14_S1  
Source13_P1 Source13_S1 Source12_P1 Source12_S1  
Source11_P1 Source11_S1 Source10_P1 Source10_S1  
  
Sexe17_P1 Sexe17_S1 Sexe16_P1 Sexe16_S1  
Sexe15_P1 Sexe15_S1 Sexe14_P1 Sexe14_S1  
Sexe13_P1 Sexe13_S1 Sexe12_P1 Sexe12_S1
```

```

Sexe11_P1 Sexe11_S1 Sexe10_P1 Sexe10_S1

annais17_P1 annais17_S1 annais16_P1 annais16_S1
annais15_P1 annais15_S1 annais14_P1 annais14_S1
annais13_P1 annais13_S1 annais12_P1 annais12_S1
annais11_P1 annais11_S1 annais10_P1 annais10_S1

FR_ETR17_P1 FR_ETR17_S1 FR_ETR16_P1 FR_ETR16_S1
FR_ETR15_P1 FR_ETR15_S1 FR_ETR14_P1 FR_ETR14_S1
FR_ETR13_P1 FR_ETR13_S1 FR_ETR12_P1 FR_ETR12_S1
FR_ETR11_P1 FR_ETR11_S1 FR_ETR10_P1 FR_ETR10_S1

Inscrit17 Inscrit16 Inscrit15 Inscrit14 Inscrit13
Inscrit12 inscrite11 Inscrite10

Selec_Ins17 Selec_Ins16 Selec_Ins15 Selec_Ins14 Selec_Ins13
Selec_Ins12 Selec_Ins11 Selec_Ins10);

by idetu ; if a;

Tour2=1;

Sise17_T2=(Source17_P1 ne "" or Source17_S1 ne "");
Sise16_T2=(Source16_P1 ne "" or Source16_S1 ne "");
Sise15_T2=(Source15_P1 ne "" or Source15_S1 ne "");
Sise14_T2=(Source14_P1 ne "" or Source14_S1 ne "");
Sise13_T2=(Source13_P1 ne "" or Source13_S1 ne "");
Sise12_T2=(Source12_P1 ne "" or Source12_S1 ne "");
Sise11_T2=(Source11_P1 ne "" or Source11_S1 ne "");
Sise10_T2=(Source10_P1 ne "" or Source10_S1 ne "");
Nb_SISE_T2=Sise17_T2+Sise16_T2+Sise15_T2+Sise14_T2+Sise13_T2+Sise12_T2+Sise11_T2+Sise10_T2;

SocioDemo17_T2=Sexe17_P1!!annais17_P1!!FR_ETR17_P1;
SocioDemo16_T2=Sexe16_P1!!annais16_P1!!FR_ETR16_P1;
SocioDemo15_T2=Sexe15_P1!!annais15_P1!!FR_ETR15_P1;
SocioDemo14_T2=Sexe14_P1!!annais14_P1!!FR_ETR14_P1;
SocioDemo13_T2=Sexe13_P1!!annais13_P1!!FR_ETR13_P1;
SocioDemo12_T2=Sexe12_P1!!annais12_P1!!FR_ETR12_P1;
SocioDemo11_T2=Sexe11_P1!!annais11_P1!!FR_ETR11_P1;
SocioDemo10_T2=Sexe10_P1!!annais10_P1!!FR_ETR10_P1;

Inscrit17_T1=inscrite17; Selec_Ins17_T1=Selec_Ins17;

```

```
Inscrit16_T1=inscrit16; Selec_Ins16_T1=Selec_Ins16;  
Inscrit15_T1=inscrit15; Selec_Ins15_T1=Selec_Ins15;  
Inscrit14_T1=inscrit14; Selec_Ins14_T1=Selec_Ins14;  
Inscrit13_T1=inscrit13; Selec_Ins13_T1=Selec_Ins13;  
Inscrit12_T1=inscrit12; Selec_Ins12_T1=Selec_Ins12;  
Inscrit11_T1=inscrit11; Selec_Ins11_T1=Selec_Ins11;  
Inscrit10_T1=inscrit10; Selec_Ins10_T1=Selec_Ins10;
```

```
Drop Source17_P1 Source17_S1 Source16_P1 Source16_S1 Source15_P1 Source15_S1 Source14_P1 Source14_S1  
Source13_P1 Source13_S1 Source12_P1 Source12_S1 Source11_P1 Source11_S1 Source10_P1 Source10_S1
```

```
Sexe17_P1 Sexe17_S1 Sexe16_P1 Sexe16_S1 Sexe15_P1 Sexe15_S1 Sexe14_P1 Sexe14_S1  
Sexe13_P1 Sexe13_S1 Sexe12_P1 Sexe12_S1 Sexe11_P1 Sexe11_S1 Sexe10_P1 Sexe10_S1
```

```
annais17_P1 annais17_S1 annais16_P1 annais16_S1 annais15_P1 annais15_S1 annais14_P1 annais14_S1  
annais13_P1 annais13_S1 annais12_P1 annais12_S1 annais11_P1 annais11_S1 annais10_P1 annais10_S1
```

```
FR_ETR17_P1 FR_ETR17_S1 FR_ETR16_P1 FR_ETR16_S1 FR_ETR15_P1 FR_ETR15_S1 FR_ETR14_P1 FR_ETR14_S1  
FR_ETR13_P1 FR_ETR13_S1 FR_ETR12_P1 FR_ETR12_S1 FR_ETR11_P1 FR_ETR11_S1 FR_ETR10_P1 FR_ETR10_S1
```

```
Inscrit17 Inscrit16 Inscrit15 Inscrit14 Inscrit13 Inscrit12 inscrit11 Inscrit10  
Selec_Ins17 Selec_Ins16 Selec_Ins15 Selec_Ins14 Selec_Ins13 Selec_Ins12 Selec_Ins11 Selec_Ins10;  
run;
```

```
/*  
/*  
**** Fusion des deux tours;  
/*  
/*
```

```
Proc sort data=Tour1; by ident;run;  
Proc sort data=Tour2; by ident;run;
```

```
Data SISE_Echant; merge Tour1 (in=a) Tour2;  
by ident; if a;  
tour1=(tour1=1);  
tour2=(tour2=1);
```

```
If NB_SISE_T2=. then NB_SISE_T2=0;
```

```
If NB_SISE_T1=0 and NB_SISE_T2=0 then APPSISE="Absent SISE";
```

```

else if NB_SISE_T1 ge 1 and NB_SISE_T2=0 then APPSISE="Present SISE avec Idetu1";
else if NB_SISE_T1=0 and NB_SISE_T2 ge 1 then APPSISE="Present SISE avec Idetu2";
else if NB_SISE_T1 ge 1 and NB_SISE_T2 ge 1 then APPSISE="Present sur Idetu1 et 2 ";
else
                                APPSISE="ZZZZZ                ";

AppSise17=Sise17_T1+Sise17_T2;
AppSise16=Sise16_T1+Sise16_T2;
AppSise15=Sise15_T1+Sise15_T2;
AppSise14=Sise14_T1+Sise14_T2;
AppSise13=Sise13_T1+Sise13_T2;
AppSise12=Sise12_T1+Sise12_T2;
AppSise11=Sise11_T1+Sise11_T2;
AppSise10=Sise10_T1+Sise10_T2;

If AppSise17=2 then Controle17=(SocioDemo17_T1=SocioDemo17_T2); else Controle17=9;
If AppSise16=2 then Controle16=(SocioDemo16_T1=SocioDemo16_T2); else Controle16=9;
If AppSise15=2 then Controle15=(SocioDemo15_T1=SocioDemo15_T2); else Controle15=9;
If AppSise14=2 then Controle14=(SocioDemo14_T1=SocioDemo14_T2); else Controle14=9;
If AppSise13=2 then Controle13=(SocioDemo13_T1=SocioDemo13_T2); else Controle13=9;
If AppSise12=2 then Controle12=(SocioDemo12_T1=SocioDemo12_T2); else Controle12=9;
If AppSise11=2 then Controle11=(SocioDemo11_T1=SocioDemo11_T2); else Controle11=9;
If AppSise10=2 then Controle10=(SocioDemo10_T1=SocioDemo10_T2); else Controle10=9;
run;

/*****
*** Examen de l'apport des deux premiers tours d'appariement;
*****/

Title "Contrôle des tours d'appariements de l'échantillon Céreq avec SISE";
Title2 "SUR TOUT L ECHANTILLON";
Proc freq data=SISE_Echant;
tables tour1*tour2/nopercent nocol norow;
tables Nb_SISE_T1*Nb_SISE_T2/nopercent nocol norow;
tables APPSISE;
tables prefix_ident*APPSISE/nopercent nocol norow;
tables repondant*APPSISE/nopercent nocol norow;
run;

Title "Contrôle des tours d'appariements de l'échantillon Céreq avec SISE";
Title2 "CHAMP DES REpondant (ie SPORT)";
Proc freq data=SISE_Echant (where=(repondant="1"));
tables prefix_ident*APPSISE/nopercent nocol norow;

```

run;

```
Title "En cas de presence avec plusieurs IDETU, est-ce la même année ?";
Title2 "EXAMEN DES SEULS IDENT RETROUVES DANS SISE AVEC DEUX IDETU";
Proc freq data=SISE_Echant (where=(APPSISE="Present sur Idet1 et 2 "));
tables prefix_ident*(AppSise17 AppSise16 AppSise15 AppSise14 AppSise13
                    AppSise12 AppSise11 AppSise10)/nopercnt nocol norow;
```

run;

```
Title "En cas de presence avec plusieurs IDETU la même année, est-ce le même sexe, age, FR-ETR ?";
Title2 "EXAMEN DES SEULS IDENT RETROUVES DANS SISE AVEC DEUX IDETU";
Proc freq data=SISE_Echant (where=(APPSISE="Present sur Idet1 et 2 "));
tables Controle17 Controle16 Controle15 Controle14 Controle13 Controle12 Controle11 Controle10;
run;
```

```
Data sor.SISE_Echant; set SISE_Echant;
run;
```

Partie D. Rapprochement des répondant l'enquête et des informations issues de SISE-Incriptions (appariement du fichier « Progedo » Génération 2017 et HISTORIQUE SISE)

- Dans un premier temps, on rapproche le fichier « Progedo » des répondants à l'enquête Génération 2017 (table g17individus) et SISE_ECHANT pour savoir qui est apparié avec SISE et si c'est le cas, si c'est avec l'identifiant du « 1^{er} tour », du « 2^e tour » ou des deux. On récupère aussi les IDETU concernés.
- On apparie ensuite la table avec HSITORIQUE SISE pour récupérer les informations sur l'inscription issue de SISE.
- On étudie ensuite la cohérence des informations sociodémographiques (sexe, année et mois de naissance) et du bac (réussite au bac et année du bac) entre ce que dit l'enquête et ce que disent les données administratives.
- On regarde enfin à grosse maille, qui est inscrit où dans SISE de 2010 à 2016. Les analyses se font principalement par tabulation (PROC FREQ) selon la variable du plus haut diplôme atteint pour l'enquête Génération, en douze modalités (PHD_QEEF2). Quelques tabulations sont faites selon le niveau de sortie à l'enquête Génération, en 18 modalités (NIVSOR), et selon la composante de la base de sondage, en 12 modalités.

```
/** Le repertoire des tables du Céreq et des tables en sortie **/
```

```
libname SOR "C:\Users\Public\Documents\Bases de travail";run;
```

```
/*  
/*  
/*  
***** APPARIEMENT DES REpondANTS DE GENERATION 2017  
***** AVEC SISE_HISTORIQUE  
/*  
/*  
*****
```

```
*****;  
*** Etape 1 : on associe Generation avec ;  
*** la table issue de la phase C, qui liste les IDETU;  
*** de SISE associé à un IDENT donné;  
*****;
```

```
Proc sort data=sor.SISE_Echant; by ident;run;  
Proc sort data=sor.g17individus; by ident;run;
```



```

Data SISE_GENE; merge sor.g17individus (in=a) sor.SISE_Echant ;
by ident; if a;
Prefix_Ident=substr(ident,1,2);

Nivsor=substr(nisor_Céreq,1,2);

*** Codification du plus haut diplôme agrégé,
en partant de la nomenclature du Céreq pour Génération 2017;

If substr(phd_Céreq,1,2)="01" then          phd_qeef2="01-Non diplômés          ";
If substr(phd_Céreq,1,2)in ("02") then      phd_qeef2="02-CAP-BEP-MC-Divers Niv.3";
If substr(phd_Céreq,1,2)in ("03","04") then  phd_qeef2="03-Bac Pro-BT-Divers Niv.4";
If substr(phd_Céreq,1,2)="05" then          phd_qeef2="04-Bac technologique          ";
If substr(phd_Céreq,1,2)="06" then          phd_qeef2="05-Bac général              ";
If substr(phd_Céreq,1,2)in ("07","08") then  phd_qeef2="06-Bac+2 sauf Santé/Social";
If substr(phd_Céreq,1,2)="09" then          phd_qeef2="07-Bac+2 à +4 Santé/Social";
If phd_Céreq in ("10L","10M") then          phd_qeef2="08-Licence professionnelle";
If phd_Céreq in ("11L","11M","12L","12M") then phd_qeef2="09-Autres Bac+3 et Bac+4 ";
If phd_Céreq in ("13L","13M","14L","14M") then phd_qeef2="10-Bac+5 sauf Ecoles          ";
If substr(phd_Céreq,1,2)in ("15","16") then  phd_qeef2="11-Ecoles (Ingé,commerce) ";
If substr(phd_Céreq,1,2)in ("17","18") then  phd_qeef2="12-Doctorat              ";
If phd_qeef2="" then phd_qeef2= "ZZZZZZZZZZZ          ";

If APPSISE="" Then APPSISE="Absent de APPSISE_ECHANT";
* cas des IDENT invalides qui sont absents de APPSISE_ECHANT;
run;

Title "Etape 1- Verification construction de PHD_QEEF2";
Proc freq data=SISE_GENE;
Tables phd_qeef2*phd_Céreq/nopercent norow nocol;
run;

Title "Etape 1- Exploratoire du bilan global de l'appariement Génération/SISE historique";
Proc freq data=SISE_GENE;
Tables (Prefix_Ident phd_qeef2 Nivsor)*APPSISE/nopercent norow nocol;
run;

Proc freq data=SISE_GENE;
Tables (Prefix_Ident phd_qeef2 Nivsor)*APPSISE/nopercent nofreq nocol;
weight pondef;

```

```

run;

*****;
*** Etape 2 : selection de l'IDETU retenue ;
*** puis appariement avec la table historique de SISE;
*** faite dans le programme B;
*****;

Data SISE_GENE2; set SISE_GENE;
If APPSISE="Present SISE avec Idetu1" then IDETU=IDETU1;
else If APPSISE="Present SISE avec Idetu2" then IDETU=IDETU2;
else If APPSISE="Present sur Idetu1 et 2 " then IDETU=IDETU1;
    *choix arbitraire mais trop peu de cas pour examiner une façon optimale de choisir;
else IDETU="";
run;

Proc sort data=SISE_GENE2;by idetu; run;
Proc sort data=sor.HistoriqueSise;by idetu; run;

Data sor.SISE_GENE2; merge SISE_GENE2 (in=a) sor.HistoriqueSise;
By idetu; if a;
run;

Title "Etape 2- Exploratoire du bilan global de l'appariement Génération/SISE historique";
Proc freq data=sor.SISE_GENE2;
Tables (Prefix_Ident phd_qeef2)*APPSISE/nopercent norow nocol;
run;

*****;
*** Etape 3 : Analyses ;
*** ;
*** ;
*****;

Data SISE_GENE3; set sor.SISE_GENE2;

princl6=(source16_P1 ne "")+(source16_P2 ne "");sec16=(source16_S1 ne "")+(source16_S2 ne "");
princl5=(source15_P1 ne "")+(source15_P2 ne "");sec15=(source15_S1 ne "")+(source15_S2 ne "");
princl4=(source14_P1 ne "")+(source14_P2 ne "");sec14=(source14_S1 ne "")+(source14_S2 ne "");
princl3=(source13_P1 ne "")+(source13_P2 ne "");sec13=(source13_S1 ne "")+(source13_S2 ne "");

```

```

princl2=(source12_P1 ne "")+(source12_P2 ne "");sec12=(source12_S1 ne "")+(source12_S2 ne "");
princl1=(source11_P1 ne "")+(source11_P2 ne "");sec11=(source11_S1 ne "")+(source11_S2 ne "");
princl0=(source10_P1 ne "")+(source10_P2 ne "");sec10=(source10_S1 ne "")+(source10_S2 ne "");

*Nombre d'années de présence avec au moins une inscription principale;

nb_presence_princ=(princl6 ge 1)+(princl5 ge 1)+(princl4 ge 1)+(princl3 ge 1)+
                    (princl2 ge 1)+(princl1 ge 1)+(princl0 ge 1);

*Nombre d'années de présence avec au moins une inscription secondaire;
nb_presence_sec=(sec17 ge 1)+(sec16 ge 1)+(sec15 ge 1)+(sec14 ge 1)+(sec13 ge 1)+
                    (sec12 ge 1)+(sec11 ge 1)+(sec10 ge 1);

*Indicateur de presence dans SISE;
If nb_presence_princ ge 1 then I_AppSise="P";
    * Dans Sise avec au moins une inscription principale;
Else if nb_presence_sec ge 1 then I_AppSise="S";
    * Dans Sise sans inscription principale mais avec au moins une inscription secondaire;
Else If substr(APPSISE,1,6)="Absent" then do ;
    I_AppSise="Z";* Hors SISE;
    If substr(APPSISE,1,6)="Absent" then nb_presence_princ=999;
    If substr(APPSISE,1,6)="Absent" then nb_presence_sec=999;
end;
Else I_AppSise="W";

*variable Princ_Sec&an qui agrège le nbre d'inscriptions principales (PRINC&an) et secondaire (SEC&an);
If princ16="" then princ16=9; If sec16="" then sec16=9; Princ_Sec16=princl6!!"_"!!sec16;
If princ15="" then princ15=9; If sec15="" then sec15=9; Princ_Sec15=princl5!!"_"!!sec15;
If princ14="" then princ14=9; If sec14="" then sec14=9; Princ_Sec14=princl4!!"_"!!sec14;
If princ13="" then princ13=9; If sec13="" then sec13=9; Princ_Sec13=princl3!!"_"!!sec13;
If princ12="" then princ12=9; If sec12="" then sec12=9; Princ_Sec12=princl2!!"_"!!sec12;
If princ11="" then princ11=9; If sec11="" then sec11=9; Princ_Sec11=princl1!!"_"!!sec11;
If princ10="" then princ10=9; If sec10="" then sec10=9; Princ_Sec10=princl10!!"_"!!sec10;

*****;
*****;
*** Pour analyse coherence Sexe;
*** repérage du sexe le plus fréquent dans SISE et comparaison avec Généré;

SexeSiseH=(sexel6_P1="1")+(sexel6_P2="1")+(sexel6_S1="1")+(sexel6_S2="1")

```

```

+ (sexel5_P1="1")+ (sexel5_P2="1")+ (sexel5_S1="1")+ (sexel5_S2="1")
+ (sexel4_P1="1")+ (sexel4_P2="1")+ (sexel4_S1="1")+ (sexel4_S2="1")
+ (sexel3_P1="1")+ (sexel3_P2="1")+ (sexel3_S1="1")+ (sexel3_S2="1")
+ (sexel2_P1="1")+ (sexel2_P2="1")+ (sexel2_S1="1")+ (sexel2_S2="1")
+ (sexel1_P1="1")+ (sexel1_P2="1")+ (sexel1_S1="1")+ (sexel1_S2="1")
+ (sexel0_P1="1")+ (sexel0_P2="1")+ (sexel0_S1="1")+ (sexel0_S2="1");

SexeSiseF=(sexel6_P1="2")+ (sexel6_P2="2")+ (sexel6_S1="2")+ (sexel6_S2="2")
+ (sexel5_P1="2")+ (sexel5_P2="2")+ (sexel5_S1="2")+ (sexel5_S2="2")
+ (sexel4_P1="2")+ (sexel4_P2="2")+ (sexel4_S1="2")+ (sexel4_S2="2")
+ (sexel3_P1="2")+ (sexel3_P2="2")+ (sexel3_S1="2")+ (sexel3_S2="2")
+ (sexel2_P1="2")+ (sexel2_P2="2")+ (sexel2_S1="2")+ (sexel2_S2="2")
+ (sexel1_P1="2")+ (sexel1_P2="2")+ (sexel1_S1="2")+ (sexel1_S2="2")
+ (sexel0_P1="2")+ (sexel0_P2="2")+ (sexel0_S1="2")+ (sexel0_S2="2");

If SexeSiseH ge 1 and SexeSiseH ge SexeSiseF+1 then SexeSise="MASC";
Else if SexeSiseH ge 1 and SexeSiseH = SexeSiseF then SexeSise="N.D.";
Else if SexeSiseF ge 1 and SexeSiseF ge SexeSiseH+1 then SexeSise="FEMI";
Else SexeSise="XXXX";

CoherenceSexe=((Sexe="1" and SexeSise="MASC") or (Sexe="2" and SexeSise="FEMI"));

*****;
*****;
*** Pour analyse coherence de l'année de naissance;
*** interne à SISE et comparaison avec Génè;
AnnaisGene=2020-AGE20;

NbAnSise=(Annais16_P1 gt 0)+(Annais16_P2 gt 0)+(Annais16_S1 gt 0)+(Annais16_S2 gt 0)
+ (Annais15_P1 gt 0)+(Annais15_P2 gt 0)+(Annais15_S1 gt 0)+(Annais15_S2 gt 0)
+ (Annais14_P1 gt 0)+(Annais14_P2 gt 0)+(Annais14_S1 gt 0)+(Annais14_S2 gt 0)
+ (Annais13_P1 gt 0)+(Annais13_P2 gt 0)+(Annais13_S1 gt 0)+(Annais13_S2 gt 0)
+ (Annais12_P1 gt 0)+(Annais12_P2 gt 0)+(Annais12_S1 gt 0)+(Annais12_S2 gt 0)
+ (Annais11_P1 gt 0)+(Annais11_P2 gt 0)+(Annais11_S1 gt 0)+(Annais11_S2 gt 0)
+ (Annais10_P1 gt 0)+(Annais10_P2 gt 0)+(Annais10_S1 gt 0)+(Annais10_S2 gt 0);

MinAnSise=Min((Annais16_P1 gt 0)*Annais16_P1, (Annais16_P2 gt 0)*Annais16_P2,
(Annais16_S1 gt 0)*Annais16_S1, (Annais16_S2 gt 0)*Annais16_S2,
(Annais15_P1 gt 0)*Annais15_P1, (Annais15_P2 gt 0)*Annais15_P2,

```

```

(Annais15_S1 gt 0)*Annais15_S1, (Annais15_S2 gt 0)*Annais15_S2,
(Annais14_P1 gt 0)*Annais14_P1, (Annais14_P2 gt 0)*Annais14_P2,
(Annais14_S1 gt 0)*Annais14_S1, (Annais14_S2 gt 0)*Annais14_S2,
(Annais13_P1 gt 0)*Annais13_P1, (Annais13_P2 gt 0)*Annais13_P2,
(Annais13_S1 gt 0)*Annais13_S1, (Annais13_S2 gt 0)*Annais13_S2,
(Annais12_P1 gt 0)*Annais12_P1, (Annais12_P2 gt 0)*Annais12_P2,
(Annais12_S1 gt 0)*Annais12_S1, (Annais12_S2 gt 0)*Annais12_S2,
(Annais11_P1 gt 0)*Annais11_P1, (Annais11_P2 gt 0)*Annais11_P2,
(Annais11_S1 gt 0)*Annais11_S1, (Annais11_S2 gt 0)*Annais11_S2,
(Annais10_P1 gt 0)*Annais10_P1, (Annais10_P2 gt 0)*Annais10_P2,
(Annais10_S1 gt 0)*Annais10_S1, (Annais10_S2 gt 0)*Annais10_S2);
If NbAnSise=0 then MinAnSise=-1;

MaxAnSise=Max((Annais16_P1 gt 0)*Annais16_P1, (Annais16_P2 gt 0)*Annais16_P2,
(Annais16_S1 gt 0)*Annais16_S1, (Annais16_S2 gt 0)*Annais16_S2,
(Annais15_P1 gt 0)*Annais15_P1, (Annais15_P2 gt 0)*Annais15_P2,
(Annais15_S1 gt 0)*Annais15_S1, (Annais15_S2 gt 0)*Annais15_S2,
(Annais14_P1 gt 0)*Annais14_P1, (Annais14_P2 gt 0)*Annais14_P2,
(Annais14_S1 gt 0)*Annais14_S1, (Annais14_S2 gt 0)*Annais14_S2,
(Annais13_P1 gt 0)*Annais13_P1, (Annais13_P2 gt 0)*Annais13_P2,
(Annais13_S1 gt 0)*Annais13_S1, (Annais13_S2 gt 0)*Annais13_S2,
(Annais12_P1 gt 0)*Annais12_P1, (Annais12_P2 gt 0)*Annais12_P2,
(Annais12_S1 gt 0)*Annais12_S1, (Annais12_S2 gt 0)*Annais12_S2,
(Annais11_P1 gt 0)*Annais11_P1, (Annais11_P2 gt 0)*Annais11_P2,
(Annais11_S1 gt 0)*Annais11_S1, (Annais11_S2 gt 0)*Annais11_S2,
(Annais10_P1 gt 0)*Annais10_P1, (Annais10_P2 gt 0)*Annais10_P2,
(Annais10_S1 gt 0)*Annais10_S1, (Annais10_S2 gt 0)*Annais10_S2);
If NbAnSise=0 then MaxAnSise=-2;

DiffMinMax=MaxAnSise-MinAnSise;

If NbAnSise ge 1 and DiffMinmax=0 and MaxAnSise=AnnaisGene then CoherenceAnais="1-CoherenceSise & Gene ";
Else if NbAnSise ge 1 and DiffMinmax=0 then CoherenceAnais="2-CoherenceSise-pasGene";
Else if NbAnSise ge 1 and DiffMinmax ne 0 then CoherenceAnais="3-Incoherence Sise ";
Else CoherenceAnais="4-Pas d'info Sise ";
";

*** Pour analyse coherence du mois de naissance;
*** interne à SISE ;

NbMoisSise=(Monais16_P1 gt 0)+(Monais16_P2 gt 0)+(Monais16_S1 gt 0)+(Monais16_S2 gt 0)

```

```

+ (Monais15_P1 gt 0) + (Monais15_P2 gt 0) + (Monais15_S1 gt 0) + (Monais15_S2 gt 0)
+ (Monais14_P1 gt 0) + (Monais14_P2 gt 0) + (Monais14_S1 gt 0) + (Monais14_S2 gt 0)
+ (Monais13_P1 gt 0) + (Monais13_P2 gt 0) + (Monais13_S1 gt 0) + (Monais13_S2 gt 0)
+ (Monais12_P1 gt 0) + (Monais12_P2 gt 0) + (Monais12_S1 gt 0) + (Monais12_S2 gt 0)
+ (Monais11_P1 gt 0) + (Monais11_P2 gt 0) + (Monais11_S1 gt 0) + (Monais11_S2 gt 0)
+ (Monais10_P1 gt 0) + (Monais10_P2 gt 0) + (Monais10_S1 gt 0) + (Monais10_S2 gt 0);

MinMoisSise=Min((Monais16_P1 gt 0)*Monais16_P1, (Monais16_P2 gt 0)*Monais16_P2,
(Monais16_S1 gt 0)*Monais16_S1, (Monais16_S2 gt 0)*Monais16_S2,
(Monais15_P1 gt 0)*Monais15_P1, (Monais15_P2 gt 0)*Monais15_P2,
(Monais15_S1 gt 0)*Monais15_S1, (Monais15_S2 gt 0)*Monais15_S2,
(Monais14_P1 gt 0)*Monais14_P1, (Monais14_P2 gt 0)*Monais14_P2,
(Monais14_S1 gt 0)*Monais14_S1, (Monais14_S2 gt 0)*Monais14_S2,
(Monais13_P1 gt 0)*Monais13_P1, (Monais13_P2 gt 0)*Monais13_P2,
(Monais13_S1 gt 0)*Monais13_S1, (Monais13_S2 gt 0)*Monais13_S2,
(Monais12_P1 gt 0)*Monais12_P1, (Monais12_P2 gt 0)*Monais12_P2,
(Monais12_S1 gt 0)*Monais12_S1, (Monais12_S2 gt 0)*Monais12_S2,
(Monais11_P1 gt 0)*Monais11_P1, (Monais11_P2 gt 0)*Monais11_P2,
(Monais11_S1 gt 0)*Monais11_S1, (Monais11_S2 gt 0)*Monais11_S2,
(Monais10_P1 gt 0)*Monais10_P1, (Monais10_P2 gt 0)*Monais10_P2,
(Monais10_S1 gt 0)*Monais10_S1, (Monais10_S2 gt 0)*Monais10_S2);
If NbMoisSise=0 then MinMoisSise=-1;

MaxMoisSise=Max((Monais16_P1 gt 0)*Monais16_P1, (Monais16_P2 gt 0)*Monais16_P2,
(Monais16_S1 gt 0)*Monais16_S1, (Monais16_S2 gt 0)*Monais16_S2,
(Monais15_P1 gt 0)*Monais15_P1, (Monais15_P2 gt 0)*Monais15_P2,
(Monais15_S1 gt 0)*Monais15_S1, (Monais15_S2 gt 0)*Monais15_S2,
(Monais14_P1 gt 0)*Monais14_P1, (Monais14_P2 gt 0)*Monais14_P2,
(Monais14_S1 gt 0)*Monais14_S1, (Monais14_S2 gt 0)*Monais14_S2,
(Monais13_P1 gt 0)*Monais13_P1, (Monais13_P2 gt 0)*Monais13_P2,
(Monais13_S1 gt 0)*Monais13_S1, (Monais13_S2 gt 0)*Monais13_S2,
(Monais12_P1 gt 0)*Monais12_P1, (Monais12_P2 gt 0)*Monais12_P2,
(Monais12_S1 gt 0)*Monais12_S1, (Monais12_S2 gt 0)*Monais12_S2,
(Monais11_P1 gt 0)*Monais11_P1, (Monais11_P2 gt 0)*Monais11_P2,
(Monais11_S1 gt 0)*Monais11_S1, (Monais11_S2 gt 0)*Monais11_S2,
(Monais10_P1 gt 0)*Monais10_P1, (Monais10_P2 gt 0)*Monais10_P2,
(Monais10_S1 gt 0)*Monais10_S1, (Monais10_S2 gt 0)*Monais10_S2);
If NbMoisSise=0 then MaxMoisSise=-2;

DiffMinMax2=MaxMoisSise-MinMoisSise;

*** Indicateur de coherence d'identité;

```

```

If CoherenceSexe=1 and CoherenceAnais="1-CoherenceSise & Gene " and DiffMinMax2=0
  then CoherenceGlobal=1;
Else if I_AppSise in ("S","P") then CoherenceGlobal=0;
Else CoherenceGlobal=9;

*****;
*****;
*** Pour analyse coherence de l'année du baccalauréat;
*** interne à SISE et comparaison avec Géné;
AnbacGene=Anbac*1;

If Anbac16_P1="9999" then Anbac16_P1=""; If Anbac16_P2="9999" then Anbac16_P2="";
If Anbac16_S1="9999" then Anbac16_S1=""; If Anbac16_S2="9999" then Anbac16_S2="";
If Anbac15_P1="9999" then Anbac15_P1=""; If Anbac15_P2="9999" then Anbac15_P2="";
If Anbac15_S1="9999" then Anbac15_S1=""; If Anbac15_S2="9999" then Anbac15_S2="";
If Anbac14_P1="9999" then Anbac14_P1=""; If Anbac14_P2="9999" then Anbac14_P2="";
If Anbac14_S1="9999" then Anbac14_S1=""; If Anbac14_S2="9999" then Anbac14_S2="";
If Anbac13_P1="9999" then Anbac13_P1=""; If Anbac13_P2="9999" then Anbac13_P2="";
If Anbac13_S1="9999" then Anbac13_S1=""; If Anbac13_S2="9999" then Anbac13_S2="";
If Anbac12_P1="9999" then Anbac12_P1=""; If Anbac12_P2="9999" then Anbac12_P2="";
If Anbac12_S1="9999" then Anbac12_S1=""; If Anbac12_S2="9999" then Anbac12_S2="";
If Anbac11_P1="9999" then Anbac11_P1=""; If Anbac11_P2="9999" then Anbac11_P2="";
If Anbac11_S1="9999" then Anbac11_S1=""; If Anbac11_S2="9999" then Anbac11_S2="";
If Anbac10_P1="9999" then Anbac10_P1=""; If Anbac10_P2="9999" then Anbac10_P2="";
If Anbac10_S1="9999" then Anbac10_S1=""; If Anbac10_S2="9999" then Anbac10_S2="";

NbBacSise=(Anbac16_P1 gt 0)+(Anbac16_P2 gt 0)+(Anbac16_S1 gt 0)+(Anbac16_S2 gt 0)
+(Anbac15_P1 gt 0)+(Anbac15_P2 gt 0)+(Anbac15_S1 gt 0)+(Anbac15_S2 gt 0)
+(Anbac14_P1 gt 0)+(Anbac14_P2 gt 0)+(Anbac14_S1 gt 0)+(Anbac14_S2 gt 0)
+(Anbac13_P1 gt 0)+(Anbac13_P2 gt 0)+(Anbac13_S1 gt 0)+(Anbac13_S2 gt 0)
+(Anbac12_P1 gt 0)+(Anbac12_P2 gt 0)+(Anbac12_S1 gt 0)+(Anbac12_S2 gt 0)
+(Anbac11_P1 gt 0)+(Anbac11_P2 gt 0)+(Anbac11_S1 gt 0)+(Anbac11_S2 gt 0)
+(Anbac10_P1 gt 0)+(Anbac10_P2 gt 0)+(Anbac10_S1 gt 0)+(Anbac10_S2 gt 0);

MinBacSise=Min((Anbac16_P1 gt 0)*Anbac16_P1, (Anbac16_P2 gt 0)*Anbac16_P2,
(Anbac16_S1 gt 0)*Anbac16_S1, (Anbac16_S2 gt 0)*Anbac16_S2,
(Anbac15_P1 gt 0)*Anbac15_P1, (Anbac15_P2 gt 0)*Anbac15_P2,
(Anbac15_S1 gt 0)*Anbac15_S1, (Anbac15_S2 gt 0)*Anbac15_S2,
(Anbac14_P1 gt 0)*Anbac14_P1, (Anbac14_P2 gt 0)*Anbac14_P2,
(Anbac14_S1 gt 0)*Anbac14_S1, (Anbac14_S2 gt 0)*Anbac14_S2,
(Anbac13_P1 gt 0)*Anbac13_P1, (Anbac13_P2 gt 0)*Anbac13_P2,

```

```

(Anbac13_S1 gt 0)*Anbac13_S1, (Anbac13_S2 gt 0)*Anbac13_S2,
(Anbac12_P1 gt 0)*Anbac12_P1, (Anbac12_P2 gt 0)*Anbac12_P2,
(Anbac12_S1 gt 0)*Anbac12_S1, (Anbac12_S2 gt 0)*Anbac12_S2,
(Anbac11_P1 gt 0)*Anbac11_P1, (Anbac11_P2 gt 0)*Anbac11_P2,
(Anbac11_S1 gt 0)*Anbac11_S1, (Anbac11_S2 gt 0)*Anbac11_S2,
(Anbac10_P1 gt 0)*Anbac10_P1, (Anbac10_P2 gt 0)*Anbac10_P2,
(Anbac10_S1 gt 0)*Anbac10_S1, (Anbac10_S2 gt 0)*Anbac10_S2);
If NbBacSise=0 then MinBacSise=-1;

MaxBacSise=Max((Anbac16_P1 gt 0)*Anbac16_P1, (Anbac16_P2 gt 0)*Anbac16_P2,
(Anbac16_S1 gt 0)*Anbac16_S1, (Anbac16_S2 gt 0)*Anbac16_S2,
(Anbac15_P1 gt 0)*Anbac15_P1, (Anbac15_P2 gt 0)*Anbac15_P2,
(Anbac15_S1 gt 0)*Anbac15_S1, (Anbac15_S2 gt 0)*Anbac15_S2,
(Anbac14_P1 gt 0)*Anbac14_P1, (Anbac14_P2 gt 0)*Anbac14_P2,
(Anbac14_S1 gt 0)*Anbac14_S1, (Anbac14_S2 gt 0)*Anbac14_S2,
(Anbac13_P1 gt 0)*Anbac13_P1, (Anbac13_P2 gt 0)*Anbac13_P2,
(Anbac13_S1 gt 0)*Anbac13_S1, (Anbac13_S2 gt 0)*Anbac13_S2,
(Anbac12_P1 gt 0)*Anbac12_P1, (Anbac12_P2 gt 0)*Anbac12_P2,
(Anbac12_S1 gt 0)*Anbac12_S1, (Anbac12_S2 gt 0)*Anbac12_S2,
(Anbac11_P1 gt 0)*Anbac11_P1, (Anbac11_P2 gt 0)*Anbac11_P2,
(Anbac11_S1 gt 0)*Anbac11_S1, (Anbac11_S2 gt 0)*Anbac11_S2,
(Anbac10_P1 gt 0)*Anbac10_P1, (Anbac10_P2 gt 0)*Anbac10_P2,
(Anbac10_S1 gt 0)*Anbac10_S1, (Anbac10_S2 gt 0)*Anbac10_S2);
If NbBacSise=0 then MaxBacSise=-2;

DiffMinMax3=MaxBacSise-MinBacSise;

If NbBacSise ge 1 and DiffMinmax3=0 and MaxBacSise=AnbacGene then CoherenceBac="1-CoherenceSise 10 Gene ";
Else if NbBacSise ge 1 and DiffMinmax3=0 then CoherenceBac="2-CoherenceSise-pasGene";
Else if NbBacSise ge 1 and DiffMinmax3 ne 0 then CoherenceBac="3-Incoherence Sise ";
Else CoherenceBac="4-Pas d'info Sise ";
";

*****;
*****;
*** Traitement des valeurs manquantes pour l'étude des parcours;
*** d'isnction;
If APPSISE="Absent de APPSISE_ECHANT" then do;
    Inscrit16="0 INE";Inscrit15="0 INE";Inscrit14="0 INE";Inscrit13="0 INE";
    Inscrit12="0 INE";Inscrit11="0 INE";Inscrit10="0 INE";
end;
Else if APPSISE="Absent SISE " then do;

```



```

    Inscrit16="0SISE";Inscrit15="0SISE";Inscrit14="0SISE";Inscrit13="0SISE";
    Inscrit12="0SISE";Inscrit11="0SISE";Inscrit10="0SISE";
    end;
*****/
*****;
Data sor.Sise_Gene3; set SISE_GENE3;
run;
*****/
*****;

Title "Etape 3a- Nbre d'inscriptions principales et secondaires retrouvées dans chaque SISE annuel";
Proc freq data=SISE_GENE3;
Tables phd_qeef2*(Princ_Sec16 Princ_Sec15 Princ_Sec14 Princ_Sec13
                 Princ_Sec12 Princ_Sec11 Princ_Sec10)/nopercnt norow nocol;
run;

*****;
*****;

Title2 " Nbre d'années de présence avec un moins une inscription principale";
Proc freq data=SISE_GENE3;
Tables (Prefix_Ident phd_qeef2 NivSor)*nb_presence_princ/nopercnt norow nocol;
run;
Title2 " Nbre d'années de présence avec un moins une inscription secondaire";
Proc freq data=SISE_GENE3;
Tables (Prefix_Ident phd_qeef2 NivSor)*nb_presence_sec/nopercnt norow nocol;
run;

*****;
*****;

Title "Etape 3d- Cohérence du sexe SISE (SexeSise) et Génération (Sexe)";
Title2 "";
Proc freq data=SISE_GENE3;
tables SexeSiseH*SexeSiseF/nopercnt norow nocol;
tables SexeSise*(SexeSiseH SexeSiseF)/nopercnt norow nocol;
Tables Sexe*SexeSise/nopercnt norow nocol;
Tables CoherenceGlobal*CoherenceSexe/nopercnt norow nocol;
run;

Title "Etape 3d- Cohérence de l'année de naissance SISE et Génération";

```

```
Proc freq data=SISE_GENE3;
tables NbAnSise DiffMinMax;
tables NbAnSise*DiffMinMax/norow nocol nopercnt;
tables CoherenceAnais*CoherenceSexe/norow nocol nopercnt;
Tables CoherenceGlobal*CoherenceAnais/nopercnt norow nocol;
run;

Title "Etape 3d- Cohérence du mois de naissance interne à SISE";
Proc freq data=SISE_GENE3;
tables NbMoisSise DiffMinMax2;
tables NbMoisSise*DiffMinMax2/norow nocol nopercnt;
tables CoherenceGlobal*DiffMinMax2/norow nocol nopercnt;
run;

Title "Etape 3d- Cohérence pour les cas où on n'a au moins une inscription principale";
Proc freq data=SISE_GENE3 (where=(I_appSise ="P"));
tables CoherenceGlobal*(CoherenceSexe CoherenceAnais DiffMinMax2)/norow nocol nopercnt;
run;

Title "Etape 3d- Cohérence pour les cas où on n'a que des inscriptions secondaire";
Proc freq data=SISE_GENE3 (where=(I_appSise ="S"));
tables CoherenceGlobal*(CoherenceSexe CoherenceAnais DiffMinMax2)/norow nocol nopercnt;
run;

*****;
*****;
Title "Etape 3e- Cohérence de l'année du baccalauréat interne à SISE et Génération";
Proc freq data=SISE_GENE3;
tables NbBacSise DiffMinMax3;
tables NbBacSise*DiffMinMax3/norow nocol nopercnt;
tables CoherenceBac*CoherenceGlobal/norow nocol nopercnt;
run;

Title2 "champ restreint aux CoherenceGlobal=1";
Proc freq data=SISE_GENE3;
tables Prefix_Ident*CoherenceBac/norow nocol nopercnt;
run;

*****;
```

```
*****;
```

```
Title "Etape 3f-Inscriptions successives dans SISE selon le Plus haut diplôme";  
Proc freq data=SISE_GENE3;  
tables (Inscrit16 Inscrit15 Inscrit14 Inscrit13 Inscrit12 Inscrit11 Inscrit10)*phd_qeef2  
/norow nocol nopercnt;  
run;
```

Partie E. Rapprochement des répondant l'enquête et des informations issues de SISE-Résultats

- On procède de façon analogue à celle pour les inscriptions.
- On crée d'abord une table annuelle unique par compilation des différentes tables disponible dans SISE-Résultats pour l'année considérée. Une variable SOURCEaa est aussi créée pour préciser la table d'origine de l'information (UNIV, MANA, INGE...), ainsi que DIPLOMEaa qui indique si le diplôme a été obtenu ou non (variable TOT pour la composante UNIV, mise à 1 pour les autres composantes).
- Après analyse exploratoire, on se restreint aux observations associées aux diplômes obtenus (DIPLOMEaa=1).
- Pour alléger les calculs, on se restreint ensuite aux IDETU des seuls répondants à l'enquête Génération.
- On procède à un recodage des diplômes de façon à limiter le nombre de modalités => NatDiplo&an (construite principalement à partir de la variable de Sise TYP_DIPR) et NatDiplobis&an (qui indique aussi le degré d'études et la table source de l'information).
- De nouvelles analyses exploratoires sont effectuées, notamment pour voir une année donnée, combien de diplômes sont disponibles pour un répondant.
- À la vue des comptages exploratoires, un fichier historique compilant les fichiers annuels est créé (HISTDIPLOM) en conservant les informations pour 3 diplômes au plus par année au plus, avec une sélection réduite de variables, suffixée respectivement, « _L1 », « _L2 », « _L3 ».
- On apparie la table HISTDIPLOM avec la table des répondants à Génération 2017. Différentes variables de comptage sont ensuite construites.
- Une première série de variables vise à voir si certains grands types de diplômes que les répondants disent avoir dans l'enquête sont repérés dans les données administratives (pour le DUT, les licences, les licences pro, les masters, les diplômes d'ingénieurs...).
- Une deuxième série de variables vise à voir pour quels répondants, selon leur plus haut diplôme atteint, il est possible de repérer un diplôme dans SISE chaque année sur la période 2010 à 2016, en tenant compte uniquement du degré d'études du diplôme (en nombre d'années post-bac).

```
/** Le répertoires des tables du Céreq et des tables en sortie ***/
```

```
libname SOR "C:\Users\Public\Documents\Bases de travail";run;
```

```
/** Les répertoires des diplômés en entrée ***/
```

```
libname Dip17uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2017";  
libname Dip16uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2016";  
libname Dip15uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2015";  
libname Dip14uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2014";
```

```

libname Dip13uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2013";
libname Dip12uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2012";
libname Dip11uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2011";
libname Dip10uni "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés UNIV_2010";

libname Dip17man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2017";
libname Dip16man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2016";
libname Dip15man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2015";
libname Dip14man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2014";
libname Dip13man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2013";
libname Dip12man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2012";
libname Dip11man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2011";
libname Dip10man "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés MANA_2010";

libname Dip17ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2017";
libname Dip16ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2016";
libname Dip15ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2015";
libname Dip14ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2014";
libname Dip13ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2013";
libname Dip12ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2012";
libname Dip11ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2011";
libname Dip10ing "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés INGE_2010";

libname Dip17ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2017";
libname Dip16ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2016";
libname Dip15ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2015";
libname Dip14ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2014";
libname Dip13ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2013";
libname Dip12ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2012";
libname Dip11ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2011";
libname Dip10ens "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés ENS_2010";

libname Dip17e26 "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés Enq26bis_2017";
libname Dip16e26 "\\casd.fr\casdfs\Projets\APPSISE\Data\SISE_SISE Diplômés Enq26bis_2016";
run;

```

```

/*****/
/*****/
/*****/

```

```
/** FORMAT *  
/*****  
/*****  
/*****
```

Proc format;

Value \$NONAPPAR

"A"="Présent dans INSCRIT au même diplôme que dans RESULTAT"
"B"="Présent dans INSCRIT avec autre diplôme"
"C"="Présent dans INSCRIT et attendu mais absent de RESULTAT"
"D"="Présent uniquement dans RESULTAT (chgt INE ?)"
"E"="Resultat pour des enregistrements secondaires d'INSCRIT";

Value \$FLAG_DIP

"1"="Diplôme hors champ de l'enquête du SIES"
"0"="Dans le champ";

Value \$FLAG_RES

"1"="Enregistrement supprimé car en doublon"
"0"="Conservé";

Value TOT

1="Diplômé pour SISE"
0="Non diplômé";

Value \$ETATINS

"10"="Inscription du champ SISE et transmise au 15 janv."
"11"="Champ SISE, non transmise au 15 janv., formation initiale"
"12"="Champ SISE, non transmise au 15 janv., apprentissage"
"21"="Champ SISE, non transmise au 15 janv., formation continue"
"22"="Champ SISE, non transmise au 15 janv., reprise non financée"
"30"="Inscription n'appartenant pas au champ de SISE (formation < 1 an)"
"31"="Pas d'inscription directement relative au diplôme ou titre décerné"
"41"="Cas porticulier santé"
"98"="sans objet";

Value \$CURPAR

"00"="SANS OBJET"
"01"="BTS"
"02"="CPGE OU PREPARATION INTEGREE"
"03"="ECOLE DE COMMERCE, GESTION,..."
"04"="ECOLE D'INGENIEURS"

```
"05"="ETAB. PRIVE D'ENSEIGNEMENT UNIVERSITAIRE"  
"06"="ETAB. D'ENSEIGNEMENT ARTISTIQUE OU CULTUREL"  
"10"="ETAB. ETRANGER"  
"11"="ECOLE NORMALE SUPERIEURE"  
"13"="ECOLE D'ARCHITECTURE"  
"14"="IUFM"  
"15"="AUTRE ECOLE OU CURSUS"  
"16"="UNIVERSITE"  
"17"="ENSEIGNEMENT PAR CORRESPONDANCE"  
"18"="ETAB. PARAMEDICAL OU SOCIAL";
```

```
run;
```

```
/*  
/*  
/*  
/** construction d'un fichier unique de diplômés par année */  
/** par compilation -> SOR.DIP_GENE&AN. ***/  
/** 1° Ques corrections simples pour permettre la compil ***/  
/** 2° Ajout d'un n° de ligne lié au nbre de présence ***/  
/** des IDENT (variable N) ***/  
/** 3° Codification des diplômes (variable NatDiplo&an. ***/  
/** ***/  
/** ATTENTION : restreint aux répondants à Génération ***/  
/*  
/*  
/*  
/*
```

```
%macro Diplomes (an);
```

```
/* repérage des idetu des répondants à Génération */  
Data Identifiants;set sor.Sise_Gene3 (Keep=IDENT IDETU PHD_qeef2);  
IDETU1=IDETU;  
Drop Idetu;  
run;
```

```

Proc sort data=Identifiants; by idetul;
run;

/* Import des tables par champ */
/* attention, certaines info des tables UNIV et ENS ne sont pas dans les autres sources.
Il faut donc les supprimer des keep (TOT, CURPAR, ETATINS, ECTS
=> plus variation selon les années (gros travail avant de fusionner)
Rque : REDISP ET TOT ne sont disponibles sur toute la période que pour la source UNIV.
Rque : par exception, TOT dispo pour la source ENS en 2016
Rque : Vu les effectifs, on postule qu'on a que les diplômés pour les autres sources
Rque : La variable RESDIP et TOT sont dans l'essentiel des cas convergentes pour
la source UNIV (REDISP=0 et TOT=1) On retiendra TOT
*/

Data Univ ; set Dip&an.uni.result&an.; sourcedip&an.="UNIV";
If resdip ="" then resdip="Z";
Keep sourcedip&an. idetul ETATINS nonappar TOT FLAG_DIP FLAG_RES
regime situpre CURPAR DIPLOM DIPLOMR ECTS GROUPEUR NIVEAU DEGETU
RESDIP TYP_DIPR VOIER DISCIPLR voier;
run;

Data Ens ; set Dip&an.ens.dipens&an.; sourcedip&an.="ENS ";
If resdip ="" then resdip="Z";
If tot=. then tot=9;
Keep sourcedip&an. idetul nonappar TOT FLAG_DIP FLAG_RES DEGETU
regime situpre CURPAR CURSUS_LMDR CYCLER DIPLOM DIPLOMR GROUPEUR NIVEAU
RESDIP TYP_DIPR VOIER DISCIPLR voier;
run;

Data Mana ; set Dip&an.man.dipmana&an.; sourcedip&an.="MANA";
If resdip ="" then resdip="Z";
If tot=. then tot=9;
Keep sourcedip&an. idetul nonappar TOT FLAG_DIP FLAG_RES DEGETU
regime situpre CURSUS_LMDR CYCLER DIPLOM DIPLOMR GROUPEUR NIVEAU
RESDIP TYP_DIPR VOIER DISCIPLR voier;
run;

Data Inge ; set Dip&an.ing.dipinge&an.; sourcedip&an.="INGE";
If resdip ="" then resdip="Z";
If tot=. then tot=9;
Keep sourcedip&an. idetul nonappar TOT FLAG_DIP FLAG_RES DEGETU
regime situpre CURSUS_LMDR CYCLER DIPLOM DIPLOMR GROUPEUR NIVEAU

```



```

RESDIP TYP_DIPR VOIER DISCIPLR voier;
run;

%if &an. ge 16 %then %do;
Data Enq26 ; set Dip&an.e26.dipenq26bis&an.; sourcedip&an.="E26b";
If resdip="" then resdip="Z";
If tot=. then tot=9;
Keep sourcedip&an. idetu1 nonappar TOT FLAG_DIP FLAG_RES DEGETU
regime situpre CURSUR_LMDR CYCLER DIPLOM DIPLOMR GROUPER NIVEAUR
RESDIP TYP_DIPR VOIER DISCIPLR voier;
run;
%end;

/*****
/* creation des tables globales annuelles pour 2010 à 2015*/
*****/

%if &an. lt 16 %then %do;

Data dip&an.; set Univ Mana Inge Ens ;

If sourcedip&an.="UNIV" then Diplome&an.=tot; else Diplome&an.=1;

/* Construction de QUALITE_INE nécessaire au filtrage */
If substr(IDETU1,11,1)="" then Qualite_INE="INVALIDE";
*repérage des valeurs trop courtes;
Else if IDETU1*1 ge 0 then Qualite_INE="INVALIDE";
*repérage des valeurs sans lettre;
Else if substr(IDETU1,1,1) in ("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O",
"P","Q","R","S","T","U","V","W","X","Y","Z")
then Qualite_INE="INVALIDE"; *repérage des valeurs commençant par une lettre;
Else if substr(IDETU1,1,5)="99999" then Qualite_INE="INVALIDE";
Else Qualite_INE ="VALIDE ?";

run;

Data Univ Mana Inge Ens Prepa BTS; set _null_; run;
%end;

```

```

/*****/
/* creation des tables globales annuelles à partir de 2016*/
/*****/
%else %if &an. ge 16 %then %do;

Data dip&an.; set Univ Mana Inge Ens Enq26;

If sourcedip&an. in ("UNIV","ENS ") then Diplome&an.=tot; else Diplome&an.=1;

    /* Construction de QUALITE_INE nécessaire au filtrage */
If substr(IDETU1,11,1)="" then Qualite_INE="INVALIDE";
    *repérage des valeurs trop courtes;
Else if IDETU1*1 ge 0 then Qualite_INE="INVALIDE";
    *repérage des valeurs sans lettre;
Else if substr(IDETU1,1,1) in ("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O",
    "P","Q","R","S","T","U","V","W","X","Y","Z")
    then Qualite_INE="INVALIDE";    *repérage des valeurs commençant par une lettre;
Else if substr(IDETU1,1,5)="99999" then Qualite_INE="INVALIDE";
Else Qualite_INE ="VALIDE ?";
run;

Data Univ Mana Inge Ens Prepa BTS Enq26; set _null_; run;
%end;

/*****/
/* Exploratoire sur la table annuelle crée*/
/*****/

Title "Structure de la table annuelle &an. compilant les fichiers des diplômés";
Proc freq data=dip&an.;
tables Qualite_INE*tot/nopercent nocol norow;
tables sourcedip&an.*(resdip tot)/nopercent nocol norow;
tables sourcedip&an.*Diplome&an./nopercent nocol norow;
Format tot tot.;
run;

/*****/
/*Restriction aux diplômés et suppression des INE supposés invalides*/

```

```

/*****/
Data dip&an.; set dip&an.;
if Diplome&an.=1;
If qualite_ine ne "VALIDE ?" then delete ;
If curpar="" then curpar="ZZ";
If Etatins="" then etatins="ZZ";
run;

Title "Structure de la table annuelle &an. RESTREINTE aux diplômés et INE valide";
Proc freq data=dip&an.;
tables sourcedip&an.*Nonappar/nopercent nocol norow;
tables sourcedip&an.*Flag_dip/nopercent nocol norow;
tables sourcedip&an.*Flag_res/nopercent nocol norow;
tables sourcedip&an.*curpar/nopercent nocol norow;
tables sourcedip&an.*etatins/nopercent nocol norow;
tables Flag_dip*etatins/nopercent nocol norow;
Format Nonappar $nonappar. Flag_dip $flag_dip. Flag_res $Flag_res.
curpar $curpar. etatins $etatins.;
run;

/*****/
/* Suppression des doublons */
/*****/

Data dip&an.; set dip&an.;
If flag_res="1" then delete ; * suppression des observations jugés en doublons par le SIES;
run;

/*****/
/*Restriction aux identifiants des répondants de Génération*/
/*****/

Proc sort data=dip&an.; by idetu1; run;

Data Dip_Gene&an.; merge Identifiants (in=a) dip&an.;
by idetu1; if a ;

If Diplome&an.=. then Diplome&an.=9;
If degetu="" then degetu&an.="9";
else degetu&an.=degetu;
```

```

If sourcedip&an.="" then sourcedip&an.="ABST";

If typ_dipr="CB" then NatDiplo&an.="DUT ";
If typ_dipr="AC" then NatDiplo&an.="Capacité Droit";
If typ_dipr in ("DB", "DC", "DJ") then NatDiplo&an.="Licence ";
If typ_dipr="LP" then NatDiplo&an.="Licence ";
If typ_dipr="DP" then NatDiplo&an.="Licence pro. ";
If typ_dipr="XA" then NatDiplo&an.="Licence LMD ";
If typ_dipr="XB" then NatDiplo&an.="Master LMD ";
If typ_dipr="XD" then NatDiplo&an.="Master Prof ";
If typ_dipr="YB" then NatDiplo&an.="Doctorat Univ.";
If typ_dipr="FH" then NatDiplo&an.="Doctorat Santé";*Pharmacie;
If typ_dipr="FJ" then NatDiplo&an.="Doctorat Santé";*Chirurgie dentaire;
If typ_dipr="IB" then NatDiplo&an.="Doctorat Santé";*Docteur en santé;
If typ_dipr="SA" then NatDiplo&an.="Doctorat Santé";*Vétérinaire;
If typ_dipr in ("JD", "JC", "JF") then NatDiplo&an.="DES Santé ";*Pharmacie/Médecine/Odontologie;
If typ_dipr in ("MA", "MB", "MC", "MD")
    then NatDiplo&an.="Aut Dip Santé ";*DESC;
If typ_dipr in ("GF", "GG", "HH") then NatDiplo&an.="Aut Dip Santé ";*CES;
If typ_dipr in ("IC", "ID", "IX") then NatDiplo&an.="Aut Dip Santé ";
If typ_dipr in ("JB", "KG") then NatDiplo&an.="Aut Dip Santé ";
If typ_dipr in ("FK", "FL") then NatDiplo&an.="Aut Dip Santé ";
If typ_dipr="FI" then NatDiplo&an.="Ingénieur ";
If typ_dipr="FN" then NatDiplo&an.="Ingénieur ";*Ingénieurs en partenariat;
If typ_dipr in ("RA", "UF", "UQ", "UH", "UR", "US", "UT", "UU", "UV", "UX")
    then NatDiplo&an.="Autres Bac+5 ";
If sourcedip&an.="ABST" then NatDiplo&an.="Absent de Dipl";
If NatDiplo&an.="" then NatDiplo&an.="Autre Diplome ";

If NatDiplo&an.="Autre Diplome " and degetu&an.="0" then NatDiplo&an.="Autre Diplome0";
If NatDiplo&an.="Autre Diplome " and degetu&an.="1" then NatDiplo&an.="Autre Diplome1";
If NatDiplo&an.="Autre Diplome " and degetu&an.="2" then NatDiplo&an.="Autre Diplome2";
If NatDiplo&an.="Autre Diplome " and degetu&an.="3" then NatDiplo&an.="Autre Diplome3";
If NatDiplo&an.="Autre Diplome " and degetu&an.="4" then NatDiplo&an.="Autre Diplome4";
If NatDiplo&an.="Autre Diplome " and degetu&an.="5" then NatDiplo&an.="Autre Diplome5";
If NatDiplo&an.="Autre Diplome " and degetu&an.="6" then NatDiplo&an.="Autre Diplome6";
If NatDiplo&an.="Autre Diplome " and degetu&an.="9" then NatDiplo&an.="Autre Diplome9";

*Autre variable de diplo^me agrégée;
If NatDiplo&an. ne "Absent de Dipl"
    then NatDiplobis&an.=Sourcedip&an.!!"-!!degetu&an.!!"-!!NatDiplo&an.;

```

```
Else NatDiplobis&an.="ZZZZ-9-Absent de Dipl";

*restriction de champ sur les diplômés à conserver;
If Nonappar ne "A" or etatins in ("21","22","23","30","31","41") then selecDip&an.=0;
Else selecDip&an.=1;

run;

Title "Structure de la table &an. RESTREINTE AUX REpondANTS DE GENERATION";
Proc freq data=Dip_Gene&an.;
tables Nonappar/nocum;
tables Flag_dip/nocum;
tables Flag_res/nocum;
tables curpar/nocum;
tables etatins/nocum;

Format Nonappar $nonappar. Flag_dip $flag_dip. Flag_res $Flag_res.
curpar $curpar. etatins $etatins.;
run;

Title "Diplômes identifiés pour les répondants de Génération";
Title2 "Tous";
Proc freq data=Dip_Gene&an.;
tables NatDiplo&an.*degetu&an./nopercent nocol norow;
tables NatDiplobis&an.* selecDip&an./nopercent nocol norow;
tables phd_qeef2*sourcedip&an./nopercent nocol norow;
tables phd_qeef2*NatDiplo&an./nopercent nocol norow;
run;
Title2 "Cas des Diplômes non codés (AUTRES)";
Proc freq data=Dip_Gene&an. (where=(NatDiplo&an.="Autre Diplome "));
tables sourcedip&an. Typ_dipr*degetu&an./nopercent nocol norow;
run;

*** Ajout d une numerotation des lignes pour un même IDENT;
Proc sort data= Dip_Gene&an.; by ident; run;
Data Sor.Dip_Gene&an.; set Dip_Gene&an.;
retain n 0 nbis 0 ident1;
* Numerotation globale;
If Ident=Ident1 then n=n+1;
```

```

Else n=1;
* Numerotation des diplômes selecbis=1;
If n=1 and selecdip&an.=1 then nbis=1;
Else If Ident=Ident1 and selecdip&an.=1 then nbis=nbis+1;
Else nbis=0;
*;
Ident1=Ident;
run;

*** Nombre de lignes par IDENT;
Data verif; set sor.Dip_Gene&an.; by ident;
If last.ident;
run;
Title "Nombre de lignes par IDENT ";
Title2 "Table &an.";
Proc freq data=verif;
tables n*nbis/norow nocol nopercent;
run;
Data verif; set _null_;
run;

*** Nature du diplôme par ligne d observations;
Title "Nature du diplôme par numero de ligne des IDENT ";
Title2 "Table &an.";
Proc freq data=sor.Dip_Gene&an.;
tables NatDiplo&an.*n/nopercent nocol norow;
run;
%mend;
run;

%Diplomes (16); run;
%Diplomes (15); run;
%Diplomes (14); run;
%Diplomes (13); run;
%Diplomes (12); run;
%Diplomes (11); run;
%Diplomes (10); run;
%Diplomes (17);
run;

```

```

/*****
/*****
/*****
/**** construction d'un fichier unique historique          */
/**** de diplômés avec une ligne par IDENT                ****/
/**** Conservation de 3 diplomes par an max              ****/
/**** -> sor.HistDiplom                                    ****/
/** ATTENTION : restreint aux répondants à Génération    ****/
/*****
/*****
/*****

***** DEBUT MACRO;
%Macro Histdiplom;

%do an = 10 %to 16;

*Table des 1ère observations associées à un IDENT;
Data Temp_L1; set sor.Dip_Gene&an.
(keep=IDENT n NatDiplo&an. degetu&an. sourcedip&an. Diplome&an. selecdip&an. typ_dipr);
if n=1;
L1_Diplome&an.= Diplome&an.;
L1_NatDiplo&an.=NatDiplo&an.;
L1_degetu&an.=degetu&an.;
L1_sourcedip&an.=sourcedip&an.;
L1_selecdip&an.=selecdip&an.;
L1_typ_dipr&an.=typ_dipr;
Drop n NatDiplo&an. degetu&an. sourcedip&an. selecdip&an. typ_dipr Diplome&an.;
run;

*Table des 2e observations associées à un IDENT;
Data Temp_L2;
set sor.Dip_Gene&an.
(keep=IDENT n NatDiplo&an. degetu&an. sourcedip&an. Diplome&an. selecdip&an. typ_dipr);
if n=2;
L2_Diplome&an.= Diplome&an.;
L2_NatDiplo&an.=NatDiplo&an.;
L2_degetu&an.=degetu&an.;
L2_sourcedip&an.=sourcedip&an.;

```

```
L2_selecdip&an.=selecdip&an.;
L2_typ_dipr&an.=typ_dipr;
Drop n NatDiplo&an. degetu&an. sourcedip&an. selecdip&an. typ_dipr Diplome&an.;
run;

*Table des 3e observations associées à un IDENT;
Data Temp_L3;
set sor.Dip_Gene&an.
(keep=IDENT n NatDiplo&an. degetu&an. sourcedip&an. Diplome&an. selecdip&an. typ_dipr);
if n=3;
L3_Diplome&an.= Diplome&an.;
L3_NatDiplo&an.=NatDiplo&an.;
L3_degetu&an.=degetu&an.;
L3_sourcedip&an.=sourcedip&an.;
L3_selecdip&an.=selecdip&an.;
L3_typ_dipr&an.=typ_dipr;
Drop n NatDiplo&an. degetu&an. sourcedip&an. selecdip&an. typ_dipr Diplome&an.;
run;

*Table compilant les 3 observations (1 ligne par IDENT);
Proc sort data=Temp_L1;by ident;run;
Proc sort data=Temp_L2;by ident;run;
Proc sort data=Temp_L3;by ident;run;
Data Temp&an. ; merge Temp_L1 Temp_L2 Temp_L3;by ident;
run;

Data Temp_L1 Temp_L2 Temp_L3; set _null_; run;
%end;
%mend;
run;
***** FIN MACRO;

%Histdiplom;
run;

Proc sort data=Temp16; by ident;run;
Proc sort data=Temp15; by ident;run;
Proc sort data=Temp14; by ident;run;
Proc sort data=Temp13; by ident;run;
Proc sort data=Temp12; by ident;run;
Proc sort data=Temp11; by ident;run;
Proc sort data=Temp10; by ident;run;
```



```

Data Hist ; merge Temp16 Temp15 Temp14 Temp13 Temp12 Temp11 Temp10 ;
by ident;
*variables de comptage à des fins de contrôle sur le nbre de diplôme L1, L2, L3 par an;
NbDip16_L1=(L1_Diplome16=1); NbDip16_L2=(L2_Diplome16=1); NbDip16_L3=(L3_Diplome16=1);
NbDip15_L1=(L1_Diplome15=1); NbDip15_L2=(L2_Diplome15=1); NbDip15_L3=(L3_Diplome15=1);
NbDip14_L1=(L1_Diplome14=1); NbDip14_L2=(L2_Diplome14=1); NbDip14_L3=(L3_Diplome14=1);
NbDip13_L1=(L1_Diplome13=1); NbDip13_L2=(L2_Diplome13=1); NbDip13_L3=(L3_Diplome13=1);
NbDip12_L1=(L1_Diplome12=1); NbDip12_L2=(L2_Diplome12=1); NbDip12_L3=(L3_Diplome12=1);
NbDip11_L1=(L1_Diplome11=1); NbDip11_L2=(L2_Diplome11=1); NbDip11_L3=(L3_Diplome11=1);
NbDip10_L1=(L1_Diplome10=1); NbDip10_L2=(L2_Diplome10=1); NbDip10_L3=(L3_Diplome10=1);

*variables de repérage des diplômés au moins une fois une année donnée;
Dip16=(NbDip16_L1+NbDip16_L2+NbDip16_L3 ge 1);
Dip15=(NbDip15_L1+NbDip15_L2+NbDip15_L3 ge 1);
Dip14=(NbDip14_L1+NbDip14_L2+NbDip14_L3 ge 1);
Dip13=(NbDip13_L1+NbDip13_L2+NbDip13_L3 ge 1);
Dip12=(NbDip12_L1+NbDip12_L2+NbDip12_L3 ge 1);
Dip11=(NbDip11_L1+NbDip11_L2+NbDip11_L3 ge 1);
Dip10=(NbDip10_L1+NbDip10_L2+NbDip10_L3 ge 1);

run;

Title "Comptages de contrôle sur le nbre de diplômés par année pour chacune des ligne L1 à L3";
Title2 "Champ des répondants à Génération";
Proc freq data=hist;
tables NbDip16_L1 NbDip16_L2 NbDip16_L3;
tables NbDip15_L1 NbDip15_L2 NbDip15_L3;
tables NbDip14_L1 NbDip14_L2 NbDip14_L3;
tables NbDip13_L1 NbDip13_L2 NbDip13_L3;
tables NbDip12_L1 NbDip12_L2 NbDip12_L3;
tables NbDip11_L1 NbDip11_L2 NbDip11_L3;
tables NbDip10_L1 NbDip10_L2 NbDip10_L3;
run;

Data sor.HistDiplom; set Hist; run;

Data Hist Temp16 Temp15 Temp14 Temp13 Temp12 Temp11 Temp10 ;
set _null_ ;run;

```

```

/*****
/*****
/*****
/**** Appariement avec      SISE_GENE3          */
/*****
/*****
/*****

Proc sort data=sor.SISE_GENE3;by ident;; run;
Proc sort data=sor.HistDiplom;by ident; run;

Data TableGlobale;
merge sor.SISE_GENE3 sor.HistDiplom;
by ident;

*** variables de comptages de diplômes présents dans SISE - PAR TYPE;
DUT=(L1_NatDiplo16="DUT      ")+(L2_NatDiplo16="DUT      ")+(L3_NatDiplo16="DUT      ")
  +(L1_NatDiplo15="DUT      ")+(L2_NatDiplo15="DUT      ")+(L3_NatDiplo15="DUT      ")
  +(L1_NatDiplo14="DUT      ")+(L2_NatDiplo14="DUT      ")+(L3_NatDiplo14="DUT      ")
  +(L1_NatDiplo13="DUT      ")+(L2_NatDiplo13="DUT      ")+(L3_NatDiplo13="DUT      ")
  +(L1_NatDiplo12="DUT      ")+(L2_NatDiplo12="DUT      ")+(L3_NatDiplo12="DUT      ")
  +(L1_NatDiplo11="DUT      ")+(L2_NatDiplo11="DUT      ")+(L3_NatDiplo11="DUT      ")
  +(L1_NatDiplo10="DUT      ")+(L2_NatDiplo10="DUT      ")+(L3_NatDiplo10="DUT      ");
If substr (APPSISE,1,6)="Absent" then DUT=999;

LicPro=(L1_NatDiplo16="Licence pro.  ")+(L2_NatDiplo16="Licence pro.  ")+(L3_NatDiplo16="Licence pro.  ")
  +(L1_NatDiplo15="Licence pro.  ")+(L2_NatDiplo15="Licence pro.  ")+(L3_NatDiplo15="Licence pro.  ")
  +(L1_NatDiplo14="Licence pro.  ")+(L2_NatDiplo14="Licence pro.  ")+(L3_NatDiplo14="Licence pro.  ")
  +(L1_NatDiplo13="Licence pro.  ")+(L2_NatDiplo13="Licence pro.  ")+(L3_NatDiplo13="Licence pro.  ")
  +(L1_NatDiplo12="Licence pro.  ")+(L2_NatDiplo12="Licence pro.  ")+(L3_NatDiplo12="Licence pro.  ")
  +(L1_NatDiplo11="Licence pro.  ")+(L2_NatDiplo11="Licence pro.  ")+(L3_NatDiplo11="Licence pro.  ")
  +(L1_NatDiplo10="Licence pro.  ")+(L2_NatDiplo10="Licence pro.  ")+(L3_NatDiplo10="Licence pro.  ");
If substr (APPSISE,1,6)="Absent" then LicPro=999;

Lic=(L1_NatDiplo16="Licence      ")+(L2_NatDiplo16="Licence      ")+(L3_NatDiplo16="Licence      ")
  +(L1_NatDiplo15="Licence      ")+(L2_NatDiplo15="Licence      ")+(L3_NatDiplo15="Licence      ")
  +(L1_NatDiplo14="Licence      ")+(L2_NatDiplo14="Licence      ")+(L3_NatDiplo14="Licence      ")
  +(L1_NatDiplo13="Licence      ")+(L2_NatDiplo13="Licence      ")+(L3_NatDiplo13="Licence      ")
  +(L1_NatDiplo12="Licence      ")+(L2_NatDiplo12="Licence      ")+(L3_NatDiplo12="Licence      ")

```

```

+ (L1_NatDiplo11="Licence      ") + (L2_NatDiplo11="Licence      ") + (L3_NatDiplo11="Licence      ")
+ (L1_NatDiplo10="Licence      ") + (L2_NatDiplo10="Licence      ") + (L3_NatDiplo10="Licence      ");
If substr (APPSISE,1,6)="Absent" then Lic=999;

LicLMD=(L1_NatDiplo16="Licence LMD  ") + (L2_NatDiplo16="Licence LMD  ") + (L3_NatDiplo16="Licence LMD  ")
+ (L1_NatDiplo15="Licence LMD  ") + (L2_NatDiplo15="Licence LMD  ") + (L3_NatDiplo15="Licence LMD  ")
+ (L1_NatDiplo14="Licence LMD  ") + (L2_NatDiplo14="Licence LMD  ") + (L3_NatDiplo14="Licence LMD  ")
+ (L1_NatDiplo13="Licence LMD  ") + (L2_NatDiplo13="Licence LMD  ") + (L3_NatDiplo13="Licence LMD  ")
+ (L1_NatDiplo12="Licence LMD  ") + (L2_NatDiplo12="Licence LMD  ") + (L3_NatDiplo12="Licence LMD  ")
+ (L1_NatDiplo11="Licence LMD  ") + (L2_NatDiplo11="Licence LMD  ") + (L3_NatDiplo11="Licence LMD  ")
+ (L1_NatDiplo10="Licence LMD  ") + (L2_NatDiplo10="Licence LMD  ") + (L3_NatDiplo10="Licence LMD  ");
If substr (APPSISE,1,6)="Absent" then LicLMD=999;

MasterLMD=(L1_NatDiplo16="Master  LMD  ") + (L2_NatDiplo16="Master  LMD  ") + (L3_NatDiplo16="Master  LMD  ")
+ (L1_NatDiplo15="Master  LMD  ") + (L2_NatDiplo15="Master  LMD  ") + (L3_NatDiplo15="Master  LMD  ")
+ (L1_NatDiplo14="Master  LMD  ") + (L2_NatDiplo14="Master  LMD  ") + (L3_NatDiplo14="Master  LMD  ")
+ (L1_NatDiplo13="Master  LMD  ") + (L2_NatDiplo13="Master  LMD  ") + (L3_NatDiplo13="Master  LMD  ")
+ (L1_NatDiplo12="Master  LMD  ") + (L2_NatDiplo12="Master  LMD  ") + (L3_NatDiplo12="Master  LMD  ")
+ (L1_NatDiplo11="Master  LMD  ") + (L2_NatDiplo11="Master  LMD  ") + (L3_NatDiplo11="Master  LMD  ")
+ (L1_NatDiplo10="Master  LMD  ") + (L2_NatDiplo10="Master  LMD  ") + (L3_NatDiplo10="Master  LMD  ");
If substr (APPSISE,1,6)="Absent" then MasterLMD=999;

MasterPro=(L1_NatDiplo16="Master  Prof  ") + (L2_NatDiplo16="Master  Prof  ") + (L3_NatDiplo16="Master  Prof  ")
+ (L1_NatDiplo15="Master  Prof  ") + (L2_NatDiplo15="Master  Prof  ") + (L3_NatDiplo15="Master  Prof  ")
+ (L1_NatDiplo14="Master  Prof  ") + (L2_NatDiplo14="Master  Prof  ") + (L3_NatDiplo14="Master  Prof  ")
+ (L1_NatDiplo13="Master  Prof  ") + (L2_NatDiplo13="Master  Prof  ") + (L3_NatDiplo13="Master  Prof  ")
+ (L1_NatDiplo12="Master  Prof  ") + (L2_NatDiplo12="Master  Prof  ") + (L3_NatDiplo12="Master  Prof  ")
+ (L1_NatDiplo11="Master  Prof  ") + (L2_NatDiplo11="Master  Prof  ") + (L3_NatDiplo11="Master  Prof  ")
+ (L1_NatDiplo10="Master  Prof  ") + (L2_NatDiplo10="Master  Prof  ") + (L3_NatDiplo10="Master  Prof  ");
If substr (APPSISE,1,6)="Absent" then MasterPro=999;

Ingenieur=(L1_NatDiplo16="Ingénieur      ") + (L2_NatDiplo16="Ingénieur      ") + (L3_NatDiplo16="Ingénieur      ")
+ (L1_NatDiplo15="Ingénieur      ") + (L2_NatDiplo15="Ingénieur      ") + (L3_NatDiplo15="Ingénieur      ")
+ (L1_NatDiplo14="Ingénieur      ") + (L2_NatDiplo14="Ingénieur      ") + (L3_NatDiplo14="Ingénieur      ")
+ (L1_NatDiplo13="Ingénieur      ") + (L2_NatDiplo13="Ingénieur      ") + (L3_NatDiplo13="Ingénieur      ")
+ (L1_NatDiplo12="Ingénieur      ") + (L2_NatDiplo12="Ingénieur      ") + (L3_NatDiplo12="Ingénieur      ")
+ (L1_NatDiplo11="Ingénieur      ") + (L2_NatDiplo11="Ingénieur      ") + (L3_NatDiplo11="Ingénieur      ")
+ (L1_NatDiplo10="Ingénieur      ") + (L2_NatDiplo10="Ingénieur      ") + (L3_NatDiplo10="Ingénieur      ");
If substr (APPSISE,1,6)="Absent" then Ingenieur=999;

DoctUniv=(L1_NatDiplo16="Doctorat Univ.") + (L2_NatDiplo16="Doctorat Univ.") + (L3_NatDiplo16="Doctorat Univ.")

```

```

+ (L1_NatDiplo15="Doctorat Univ.")+(L2_NatDiplo15="Doctorat Univ.")+(L3_NatDiplo15="Doctorat Univ.")
+ (L1_NatDiplo14="Doctorat Univ.")+(L2_NatDiplo14="Doctorat Univ.")+(L3_NatDiplo14="Doctorat Univ.")
+ (L1_NatDiplo13="Doctorat Univ.")+(L2_NatDiplo13="Doctorat Univ.")+(L3_NatDiplo13="Doctorat Univ.")
+ (L1_NatDiplo12="Doctorat Univ.")+(L2_NatDiplo12="Doctorat Univ.")+(L3_NatDiplo12="Doctorat Univ.")
+ (L1_NatDiplo11="Doctorat Univ.")+(L2_NatDiplo11="Doctorat Univ.")+(L3_NatDiplo11="Doctorat Univ.")
+ (L1_NatDiplo10="Doctorat Univ.")+(L2_NatDiplo10="Doctorat Univ.")+(L3_NatDiplo10="Doctorat Univ.");
If substr (APPSISE,1,6)="Absent" then DoctUniv=999;

DoctSante=(L1_NatDiplo16="Doctorat Santé")+(L2_NatDiplo16="Doctorat Santé")+(L3_NatDiplo16="Doctorat Santé")
+ (L1_NatDiplo15="Doctorat Santé")+(L2_NatDiplo15="Doctorat Santé")+(L3_NatDiplo15="Doctorat Santé")
+ (L1_NatDiplo14="Doctorat Santé")+(L2_NatDiplo14="Doctorat Santé")+(L3_NatDiplo14="Doctorat Santé")
+ (L1_NatDiplo13="Doctorat Santé")+(L2_NatDiplo13="Doctorat Santé")+(L3_NatDiplo13="Doctorat Santé")
+ (L1_NatDiplo12="Doctorat Santé")+(L2_NatDiplo12="Doctorat Santé")+(L3_NatDiplo12="Doctorat Santé")
+ (L1_NatDiplo11="Doctorat Santé")+(L2_NatDiplo11="Doctorat Santé")+(L3_NatDiplo11="Doctorat Santé")
+ (L1_NatDiplo10="Doctorat Santé")+(L2_NatDiplo10="Doctorat Santé")+(L3_NatDiplo10="Doctorat Santé");
If substr (APPSISE,1,6)="Absent" then DoctSante=999;

DES_Sante=(L1_NatDiplo16="DES Santé ")+(L2_NatDiplo16="DES Santé ")+(L3_NatDiplo16="DES Santé ")
+ (L1_NatDiplo15="DES Santé ")+(L2_NatDiplo15="DES Santé ")+(L3_NatDiplo15="DES Santé ")
+ (L1_NatDiplo14="DES Santé ")+(L2_NatDiplo14="DES Santé ")+(L3_NatDiplo14="DES Santé ")
+ (L1_NatDiplo13="DES Santé ")+(L2_NatDiplo13="DES Santé ")+(L3_NatDiplo13="DES Santé ")
+ (L1_NatDiplo12="DES Santé ")+(L2_NatDiplo12="DES Santé ")+(L3_NatDiplo12="DES Santé ")
+ (L1_NatDiplo11="DES Santé ")+(L2_NatDiplo11="DES Santé ")+(L3_NatDiplo11="DES Santé ")
+ (L1_NatDiplo10="DES Santé ")+(L2_NatDiplo10="DES Santé ")+(L3_NatDiplo10="DES Santé ");
If substr (APPSISE,1,6)="Absent" then DES_Sante=999;

Aut_Sante=(L1_NatDiplo16="Aut Dip Santé ")+(L2_NatDiplo16="Aut Dip Santé ")+(L3_NatDiplo16="Aut Dip Santé ")
+ (L1_NatDiplo15="Aut Dip Santé ")+(L2_NatDiplo15="Aut Dip Santé ")+(L3_NatDiplo15="Aut Dip Santé ")
+ (L1_NatDiplo14="Aut Dip Santé ")+(L2_NatDiplo14="Aut Dip Santé ")+(L3_NatDiplo14="Aut Dip Santé ")
+ (L1_NatDiplo13="Aut Dip Santé ")+(L2_NatDiplo13="Aut Dip Santé ")+(L3_NatDiplo13="Aut Dip Santé ")
+ (L1_NatDiplo12="Aut Dip Santé ")+(L2_NatDiplo12="Aut Dip Santé ")+(L3_NatDiplo12="Aut Dip Santé ")
+ (L1_NatDiplo11="Aut Dip Santé ")+(L2_NatDiplo11="Aut Dip Santé ")+(L3_NatDiplo11="Aut Dip Santé ")
+ (L1_NatDiplo10="Aut Dip Santé ")+(L2_NatDiplo10="Aut Dip Santé ")+(L3_NatDiplo10="Aut Dip Santé ");
If substr (APPSISE,1,6)="Absent" then Aut_Sante=999;

*** variables de comptages de diplômes présents dans SISE AVANT 2016/2017 - PAR TYPE;
DUT_old=(L1_NatDiplo15="DUT ")+(L2_NatDiplo15="DUT ")+(L3_NatDiplo15="DUT ")
+ (L1_NatDiplo14="DUT ")+(L2_NatDiplo14="DUT ")+(L3_NatDiplo14="DUT ")
+ (L1_NatDiplo13="DUT ")+(L2_NatDiplo13="DUT ")+(L3_NatDiplo13="DUT ")
+ (L1_NatDiplo12="DUT ")+(L2_NatDiplo12="DUT ")+(L3_NatDiplo12="DUT ")
+ (L1_NatDiplo11="DUT ")+(L2_NatDiplo11="DUT ")+(L3_NatDiplo11="DUT ")

```

```

+ (L1_NatDiplo10="DUT          ") + (L2_NatDiplo10="DUT          ") + (L3_NatDiplo10="DUT          ");
If substr (APPSISE,1,6)="Absent" then DUT_old=999;

LicPro_old=(L1_NatDiplo15="Licence pro.  ") + (L2_NatDiplo15="Licence pro.  ") + (L3_NatDiplo15="Licence pro.  ")
+ (L1_NatDiplo14="Licence pro.  ") + (L2_NatDiplo14="Licence pro.  ") + (L3_NatDiplo14="Licence pro.  ")
+ (L1_NatDiplo13="Licence pro.  ") + (L2_NatDiplo13="Licence pro.  ") + (L3_NatDiplo13="Licence pro.  ")
+ (L1_NatDiplo12="Licence pro.  ") + (L2_NatDiplo12="Licence pro.  ") + (L3_NatDiplo12="Licence pro.  ")
+ (L1_NatDiplo11="Licence pro.  ") + (L2_NatDiplo11="Licence pro.  ") + (L3_NatDiplo11="Licence pro.  ")
+ (L1_NatDiplo10="Licence pro.  ") + (L2_NatDiplo10="Licence pro.  ") + (L3_NatDiplo10="Licence pro.  ");
If substr (APPSISE,1,6)="Absent" then LicPro_old=999;

LicLMD_old=(L1_NatDiplo15="Licence LMD   ") + (L2_NatDiplo15="Licence LMD   ") + (L3_NatDiplo15="Licence LMD   ")
+ (L1_NatDiplo14="Licence LMD   ") + (L2_NatDiplo14="Licence LMD   ") + (L3_NatDiplo14="Licence LMD   ")
+ (L1_NatDiplo13="Licence LMD   ") + (L2_NatDiplo13="Licence LMD   ") + (L3_NatDiplo13="Licence LMD   ")
+ (L1_NatDiplo12="Licence LMD   ") + (L2_NatDiplo12="Licence LMD   ") + (L3_NatDiplo12="Licence LMD   ")
+ (L1_NatDiplo11="Licence LMD   ") + (L2_NatDiplo11="Licence LMD   ") + (L3_NatDiplo11="Licence LMD   ")
+ (L1_NatDiplo10="Licence LMD   ") + (L2_NatDiplo10="Licence LMD   ") + (L3_NatDiplo10="Licence LMD   ");
If substr (APPSISE,1,6)="Absent" then LicLMD_old=999;

Master_old=(L1_NatDiplo15="Master  LMD   ") + (L2_NatDiplo15="Master  LMD   ") + (L3_NatDiplo15="Master  LMD   ")
+ (L1_NatDiplo14="Master  LMD   ") + (L2_NatDiplo14="Master  LMD   ") + (L3_NatDiplo14="Master  LMD   ")
+ (L1_NatDiplo13="Master  LMD   ") + (L2_NatDiplo13="Master  LMD   ") + (L3_NatDiplo13="Master  LMD   ")
+ (L1_NatDiplo12="Master  LMD   ") + (L2_NatDiplo12="Master  LMD   ") + (L3_NatDiplo12="Master  LMD   ")
+ (L1_NatDiplo11="Master  LMD   ") + (L2_NatDiplo11="Master  LMD   ") + (L3_NatDiplo11="Master  LMD   ")
+ (L1_NatDiplo10="Master  LMD   ") + (L2_NatDiplo10="Master  LMD   ") + (L3_NatDiplo10="Master  LMD   ")
+ (L1_NatDiplo15="Master  Prof  ") + (L2_NatDiplo15="Master  Prof  ") + (L3_NatDiplo15="Master  Prof  ")
+ (L1_NatDiplo14="Master  Prof  ") + (L2_NatDiplo14="Master  Prof  ") + (L3_NatDiplo14="Master  Prof  ")
+ (L1_NatDiplo13="Master  Prof  ") + (L2_NatDiplo13="Master  Prof  ") + (L3_NatDiplo13="Master  Prof  ")
+ (L1_NatDiplo12="Master  Prof  ") + (L2_NatDiplo12="Master  Prof  ") + (L3_NatDiplo12="Master  Prof  ")
+ (L1_NatDiplo11="Master  Prof  ") + (L2_NatDiplo11="Master  Prof  ") + (L3_NatDiplo11="Master  Prof  ")
+ (L1_NatDiplo10="Master  Prof  ") + (L2_NatDiplo10="Master  Prof  ") + (L3_NatDiplo10="Master  Prof  ");
If substr (APPSISE,1,6)="Absent" then Master_old=999;

Ingenieur_old=(L1_NatDiplo15="Ingénieur      ") + (L2_NatDiplo15="Ingénieur      ") + (L3_NatDiplo15="Ingénieur      ")
+ (L1_NatDiplo14="Ingénieur      ") + (L2_NatDiplo14="Ingénieur      ") + (L3_NatDiplo14="Ingénieur      ")
+ (L1_NatDiplo13="Ingénieur      ") + (L2_NatDiplo13="Ingénieur      ") + (L3_NatDiplo13="Ingénieur      ")
+ (L1_NatDiplo12="Ingénieur      ") + (L2_NatDiplo12="Ingénieur      ") + (L3_NatDiplo12="Ingénieur      ")

```

```

+ (L1_NatDiplo11="Ingénieur   ") + (L2_NatDiplo11="Ingénieur   ") + (L3_NatDiplo11="Ingénieur   ")
+ (L1_NatDiplo10="Ingénieur   ") + (L2_NatDiplo10="Ingénieur   ") + (L3_NatDiplo10="Ingénieur   ");
If substr (APPSISE,1,6)="Absent" then Ingenieur_old=999;

Doct_old=(L1_NatDiplo15="Doctorat Univ.")+(L2_NatDiplo15="Doctorat Univ.")+(L3_NatDiplo15="Doctorat Univ.")
+ (L1_NatDiplo14="Doctorat Univ.")+(L2_NatDiplo14="Doctorat Univ.")+(L3_NatDiplo14="Doctorat Univ.")
+ (L1_NatDiplo13="Doctorat Univ.")+(L2_NatDiplo13="Doctorat Univ.")+(L3_NatDiplo13="Doctorat Univ.")
+ (L1_NatDiplo12="Doctorat Univ.")+(L2_NatDiplo12="Doctorat Univ.")+(L3_NatDiplo12="Doctorat Univ.")
+ (L1_NatDiplo11="Doctorat Univ.")+(L2_NatDiplo11="Doctorat Univ.")+(L3_NatDiplo11="Doctorat Univ.")
+ (L1_NatDiplo10="Doctorat Univ.")+(L2_NatDiplo10="Doctorat Univ.")+(L3_NatDiplo10="Doctorat Univ.")
+ (L1_NatDiplo15="Doctorat Santé")+ (L2_NatDiplo15="Doctorat Santé")+ (L3_NatDiplo15="Doctorat Santé")
+ (L1_NatDiplo14="Doctorat Santé")+ (L2_NatDiplo14="Doctorat Santé")+ (L3_NatDiplo14="Doctorat Santé")
+ (L1_NatDiplo13="Doctorat Santé")+ (L2_NatDiplo13="Doctorat Santé")+ (L3_NatDiplo13="Doctorat Santé")
+ (L1_NatDiplo12="Doctorat Santé")+ (L2_NatDiplo12="Doctorat Santé")+ (L3_NatDiplo12="Doctorat Santé")
+ (L1_NatDiplo11="Doctorat Santé")+ (L2_NatDiplo11="Doctorat Santé")+ (L3_NatDiplo11="Doctorat Santé")
+ (L1_NatDiplo10="Doctorat Santé")+ (L2_NatDiplo10="Doctorat Santé")+ (L3_NatDiplo10="Doctorat Santé");
If substr (APPSISE,1,6)="Absent" then Doct_old=999;

*** variables de comptages de diplômes présents dans SISE - PAR DEGRE D'ETUDES;
BACplus0=(L1_Degetu16="0")+ (L2_Degetu16="0")+ (L3_Degetu16="0")
+ (L1_Degetu15="0")+ (L2_Degetu15="0")+ (L3_Degetu15="0")
+ (L1_Degetu14="0")+ (L2_Degetu14="0")+ (L3_Degetu14="0")
+ (L1_Degetu13="0")+ (L2_Degetu13="0")+ (L3_Degetu13="0")
+ (L1_Degetu12="0")+ (L2_Degetu12="0")+ (L3_Degetu12="0")
+ (L1_Degetu11="0")+ (L2_Degetu11="0")+ (L3_Degetu11="0")
+ (L1_Degetu10="0")+ (L2_Degetu10="0")+ (L3_Degetu10="0");
If substr (APPSISE,1,6)="Absent" then BACplus0=999;

BACplus1=(L1_Degetu16="1")+ (L2_Degetu16="1")+ (L3_Degetu16="1")
+ (L1_Degetu15="1")+ (L2_Degetu15="1")+ (L3_Degetu15="1")
+ (L1_Degetu14="1")+ (L2_Degetu14="1")+ (L3_Degetu14="1")
+ (L1_Degetu13="1")+ (L2_Degetu13="1")+ (L3_Degetu13="1")
+ (L1_Degetu12="1")+ (L2_Degetu12="1")+ (L3_Degetu12="1")
+ (L1_Degetu11="1")+ (L2_Degetu11="1")+ (L3_Degetu11="1")
+ (L1_Degetu10="1")+ (L2_Degetu10="1")+ (L3_Degetu10="1");
If substr (APPSISE,1,6)="Absent" then BACplus1=999;

```

```
BACplus2=(L1_Degetu16="2")+ (L2_Degetu16="2")+ (L3_Degetu16="2")
+ (L1_Degetu15="2")+ (L2_Degetu15="2")+ (L3_Degetu15="2")
+ (L1_Degetu14="2")+ (L2_Degetu14="2")+ (L3_Degetu14="2")
+ (L1_Degetu13="2")+ (L2_Degetu13="2")+ (L3_Degetu13="2")
+ (L1_Degetu12="2")+ (L2_Degetu12="2")+ (L3_Degetu12="2")
+ (L1_Degetu11="2")+ (L2_Degetu11="2")+ (L3_Degetu11="2")
+ (L1_Degetu10="2")+ (L2_Degetu10="2")+ (L3_Degetu10="2");
If substr (APPSISE,1,6)="Absent" then BACplus2=999;
```

```
BACplus3=(L1_Degetu16="3")+ (L2_Degetu16="3")+ (L3_Degetu16="3")
+ (L1_Degetu15="3")+ (L2_Degetu15="3")+ (L3_Degetu15="3")
+ (L1_Degetu14="3")+ (L2_Degetu14="3")+ (L3_Degetu14="3")
+ (L1_Degetu13="3")+ (L2_Degetu13="3")+ (L3_Degetu13="3")
+ (L1_Degetu12="3")+ (L2_Degetu12="3")+ (L3_Degetu12="3")
+ (L1_Degetu11="3")+ (L2_Degetu11="3")+ (L3_Degetu11="3")
+ (L1_Degetu10="3")+ (L2_Degetu10="3")+ (L3_Degetu10="3");
If substr (APPSISE,1,6)="Absent" then BACplus3=999;
```

```
BACplus4=(L1_Degetu16="4")+ (L2_Degetu16="4")+ (L3_Degetu16="4")
+ (L1_Degetu15="4")+ (L2_Degetu15="4")+ (L3_Degetu15="4")
+ (L1_Degetu14="4")+ (L2_Degetu14="4")+ (L3_Degetu14="4")
+ (L1_Degetu13="4")+ (L2_Degetu13="4")+ (L3_Degetu13="4")
+ (L1_Degetu12="4")+ (L2_Degetu12="4")+ (L3_Degetu12="4")
+ (L1_Degetu11="4")+ (L2_Degetu11="4")+ (L3_Degetu11="4")
+ (L1_Degetu10="4")+ (L2_Degetu10="4")+ (L3_Degetu10="4");
If substr (APPSISE,1,6)="Absent" then BACplus4=999;
```

```
BACplus5=(L1_Degetu16="5")+ (L2_Degetu16="5")+ (L3_Degetu16="5")
+ (L1_Degetu15="5")+ (L2_Degetu15="5")+ (L3_Degetu15="5")
+ (L1_Degetu14="5")+ (L2_Degetu14="5")+ (L3_Degetu14="5")
+ (L1_Degetu13="5")+ (L2_Degetu13="5")+ (L3_Degetu13="5")
+ (L1_Degetu12="5")+ (L2_Degetu12="5")+ (L3_Degetu12="5")
+ (L1_Degetu11="5")+ (L2_Degetu11="5")+ (L3_Degetu11="5")
+ (L1_Degetu10="5")+ (L2_Degetu10="5")+ (L3_Degetu10="5");
If substr (APPSISE,1,6)="Absent" then BACplus5=999;
```

```
BACplus6=(L1_Degetu16="6")+ (L2_Degetu16="6")+ (L3_Degetu16="6")
+ (L1_Degetu15="6")+ (L2_Degetu15="6")+ (L3_Degetu15="6")
+ (L1_Degetu14="6")+ (L2_Degetu14="6")+ (L3_Degetu14="6")
```

```

+ (L1_Degetu13="6")+ (L2_Degetu13="6")+ (L3_Degetu13="6")
+ (L1_Degetu12="6")+ (L2_Degetu12="6")+ (L3_Degetu12="6")
+ (L1_Degetu11="6")+ (L2_Degetu11="6")+ (L3_Degetu11="6")
+ (L1_Degetu10="6")+ (L2_Degetu10="6")+ (L3_Degetu10="6");
If substr (APPSISE,1,6)="Absent" then BACplus6=999;

*** Repérage des titulaire d'un diplôme dans la base MANA;
DipMANA3=(L1_sourcedip16="MANA" and L1_Degetu16="3")+ (L2_sourcedip16="MANA" and L2_Degetu16="3")
+ (L3_sourcedip16="MANA" and L3_Degetu16="3")+
  (L1_sourcedip15="MANA" and L1_Degetu15="3")+ (L2_sourcedip15="MANA" and L2_Degetu15="3")
+ (L3_sourcedip15="MANA" and L3_Degetu15="3")+
  (L1_sourcedip14="MANA" and L1_Degetu14="3")+ (L2_sourcedip14="MANA" and L2_Degetu14="3")
+ (L3_sourcedip14="MANA" and L3_Degetu14="3")+
  (L1_sourcedip13="MANA" and L1_Degetu13="3")+ (L2_sourcedip13="MANA" and L2_Degetu13="3")
+ (L3_sourcedip13="MANA" and L3_Degetu13="3")+
  (L1_sourcedip12="MANA" and L1_Degetu12="3")+ (L2_sourcedip12="MANA" and L2_Degetu12="3")
+ (L3_sourcedip12="MANA" and L3_Degetu12="3")+
  (L1_sourcedip11="MANA" and L1_Degetu11="3")+ (L2_sourcedip11="MANA" and L2_Degetu11="3")
+ (L3_sourcedip11="MANA" and L3_Degetu11="3")+
  (L1_sourcedip10="MANA" and L1_Degetu10="3")+ (L2_sourcedip10="MANA" and L2_Degetu10="3")
+ (L3_sourcedip10="MANA" and L3_Degetu10="3");
If substr (APPSISE,1,6)="Absent" then DipMANA3=999;

DipMANA4=(L1_sourcedip16="MANA" and L1_Degetu16="4")+ (L2_sourcedip16="MANA" and L2_Degetu16="4")
+ (L3_sourcedip16="MANA" and L3_Degetu16="4")+
  (L1_sourcedip15="MANA" and L1_Degetu15="4")+ (L2_sourcedip15="MANA" and L2_Degetu15="4")
+ (L3_sourcedip15="MANA" and L3_Degetu15="4")+
  (L1_sourcedip14="MANA" and L1_Degetu14="4")+ (L2_sourcedip14="MANA" and L2_Degetu14="4")
+ (L3_sourcedip14="MANA" and L3_Degetu14="4")+
  (L1_sourcedip13="MANA" and L1_Degetu13="4")+ (L2_sourcedip13="MANA" and L2_Degetu13="4")
+ (L3_sourcedip13="MANA" and L3_Degetu13="4")+
  (L1_sourcedip12="MANA" and L1_Degetu12="4")+ (L2_sourcedip12="MANA" and L2_Degetu12="4")
+ (L3_sourcedip12="MANA" and L3_Degetu12="4")+
  (L1_sourcedip11="MANA" and L1_Degetu11="4")+ (L2_sourcedip11="MANA" and L2_Degetu11="4")
+ (L3_sourcedip11="MANA" and L3_Degetu11="4")+
  (L1_sourcedip10="MANA" and L1_Degetu10="4")+ (L2_sourcedip10="MANA" and L2_Degetu10="4")
+ (L3_sourcedip10="MANA" and L3_Degetu10="4");
If substr (APPSISE,1,6)="Absent" then DipMANA4=999;

DipMANA5=(L1_sourcedip16="MANA" and L1_Degetu16="5")+ (L2_sourcedip16="MANA" and L2_Degetu16="5")
+ (L3_sourcedip16="MANA" and L3_Degetu16="5")+

```



```

(L1_sourcedip15="MANA" and L1_Degetu15="5")+ (L2_sourcedip15="MANA" and L2_Degetu15="5")
+ (L3_sourcedip15="MANA" and L3_Degetu15="5")+
(L1_sourcedip14="MANA" and L1_Degetu14="5")+ (L2_sourcedip14="MANA" and L2_Degetu14="5")
+ (L3_sourcedip14="MANA" and L3_Degetu14="5")+
(L1_sourcedip13="MANA" and L1_Degetu13="5")+ (L2_sourcedip13="MANA" and L2_Degetu13="5")
+ (L3_sourcedip13="MANA" and L3_Degetu13="5")+
(L1_sourcedip12="MANA" and L1_Degetu12="5")+ (L2_sourcedip12="MANA" and L2_Degetu12="5")
+ (L3_sourcedip12="MANA" and L3_Degetu12="5")+
(L1_sourcedip11="MANA" and L1_Degetu11="5")+ (L2_sourcedip11="MANA" and L2_Degetu11="5")
+ (L3_sourcedip11="MANA" and L3_Degetu11="5")+
(L1_sourcedip10="MANA" and L1_Degetu10="5")+ (L2_sourcedip10="MANA" and L2_Degetu10="5")
+ (L3_sourcedip10="MANA" and L3_Degetu10="5");
If substr (APPSISE,1,6)="Absent" then DipMANA5=999;

```

*** Repérage des titulaire d'un diplôme dans la base INGE;

```

DipINGE3=(L1_sourcedip16="INGE" and L1_Degetu16="3")+ (L2_sourcedip16="INGE" and L2_Degetu16="3")
+ (L3_sourcedip16="INGE" and L3_Degetu16="3")+
(L1_sourcedip15="INGE" and L1_Degetu15="3")+ (L2_sourcedip15="INGE" and L2_Degetu15="3")
+ (L3_sourcedip15="INGE" and L3_Degetu15="3")+
(L1_sourcedip14="INGE" and L1_Degetu14="3")+ (L2_sourcedip14="INGE" and L2_Degetu14="3")
+ (L3_sourcedip14="INGE" and L3_Degetu14="3")+
(L1_sourcedip13="INGE" and L1_Degetu13="3")+ (L2_sourcedip13="INGE" and L2_Degetu13="3")
+ (L3_sourcedip13="INGE" and L3_Degetu13="3")+
(L1_sourcedip12="INGE" and L1_Degetu12="3")+ (L2_sourcedip12="INGE" and L2_Degetu12="3")
+ (L3_sourcedip12="INGE" and L3_Degetu12="3")+
(L1_sourcedip11="INGE" and L1_Degetu11="3")+ (L2_sourcedip11="INGE" and L2_Degetu11="3")
+ (L3_sourcedip11="INGE" and L3_Degetu11="3")+
(L1_sourcedip10="INGE" and L1_Degetu10="3")+ (L2_sourcedip10="INGE" and L2_Degetu10="3")
+ (L3_sourcedip10="INGE" and L3_Degetu10="3");
If substr (APPSISE,1,6)="Absent" then DipINGE3=999;

```

```

DipINGE4=(L1_sourcedip16="INGE" and L1_Degetu16="4")+ (L2_sourcedip16="INGE" and L2_Degetu16="4")
+ (L3_sourcedip16="INGE" and L3_Degetu16="4")+
(L1_sourcedip15="INGE" and L1_Degetu15="4")+ (L2_sourcedip15="INGE" and L2_Degetu15="4")
+ (L3_sourcedip15="INGE" and L3_Degetu15="4")+
(L1_sourcedip14="INGE" and L1_Degetu14="4")+ (L2_sourcedip14="INGE" and L2_Degetu14="4")
+ (L3_sourcedip14="INGE" and L3_Degetu14="4")+
(L1_sourcedip13="INGE" and L1_Degetu13="4")+ (L2_sourcedip13="INGE" and L2_Degetu13="4")

```

```

+ (L3_sourcedip13="INGE" and L3_Degetu13="4")+
  (L1_sourcedip12="INGE" and L1_Degetu12="4")+ (L2_sourcedip12="INGE" and L2_Degetu12="4")
+ (L3_sourcedip12="INGE" and L3_Degetu12="4")+
  (L1_sourcedip11="INGE" and L1_Degetu11="4")+ (L2_sourcedip11="INGE" and L2_Degetu11="4")
+ (L3_sourcedip11="INGE" and L3_Degetu11="4")+
  (L1_sourcedip10="INGE" and L1_Degetu10="4")+ (L2_sourcedip10="INGE" and L2_Degetu10="4")
+ (L3_sourcedip10="INGE" and L3_Degetu10="4");
If substr (APPSISE,1,6)="Absent" then DipINGE4=999;

```

```

DipINGE5=(L1_sourcedip16="INGE" and L1_Degetu16="5")+ (L2_sourcedip16="INGE" and L2_Degetu16="5")
+ (L3_sourcedip16="INGE" and L3_Degetu16="5")+
  (L1_sourcedip15="INGE" and L1_Degetu15="5")+ (L2_sourcedip15="INGE" and L2_Degetu15="5")
+ (L3_sourcedip15="INGE" and L3_Degetu15="5")+
  (L1_sourcedip14="INGE" and L1_Degetu14="5")+ (L2_sourcedip14="INGE" and L2_Degetu14="5")
+ (L3_sourcedip14="INGE" and L3_Degetu14="5")+
  (L1_sourcedip13="INGE" and L1_Degetu13="5")+ (L2_sourcedip13="INGE" and L2_Degetu13="5")
+ (L3_sourcedip13="INGE" and L3_Degetu13="5")+
  (L1_sourcedip12="INGE" and L1_Degetu12="5")+ (L2_sourcedip12="INGE" and L2_Degetu12="5")
+ (L3_sourcedip12="INGE" and L3_Degetu12="5")+
  (L1_sourcedip11="INGE" and L1_Degetu11="5")+ (L2_sourcedip11="INGE" and L2_Degetu11="5")
+ (L3_sourcedip11="INGE" and L3_Degetu11="5")+
  (L1_sourcedip10="INGE" and L1_Degetu10="5")+ (L2_sourcedip10="INGE" and L2_Degetu10="5")
+ (L3_sourcedip10="INGE" and L3_Degetu10="5");
If substr (APPSISE,1,6)="Absent" then DipINGE5=999;

```

*** Calcul de PHD_SISE (plus haut degré d'étude dans SISE);

```

If BACplus6 ge 1 and BACplus6 lt 100 then PHD_SISE="Bac+6 ";
Else if BACplus5 ge 1 and BACplus5 lt 100 then PHD_SISE="Bac+5 ";
Else if BACplus4 ge 1 and BACplus4 lt 100 then PHD_SISE="Bac+4 ";
Else if BACplus3 ge 1 and BACplus3 lt 100 then PHD_SISE="Bac+3 ";
Else if BACplus2 ge 1 and BACplus2 lt 100 then PHD_SISE="Bac+2 ";
Else if BACplus1 ge 1 and BACplus1 lt 100 then PHD_SISE="Bac+1 ";
Else if BACplus0 ge 1 and BACplus0 lt 100 then PHD_SISE="Bac+0 ";
Else if substr (APPSISE,1,6)="Absent" then PHD_SISE="AbsSIS";
Else if dip16+dip15+dip14+dip13+dip12+dip11+dip10=0 then PHD_SISE="AbsDIP";
Else if dip17=1 then PHD_SISE="DIP17 ";
Else PHD_SISE="?????";

```

** Variable PHD distinguant école de commerce et écoles d'ingénieurs;

```
Phd_qeef3=phd_qeef2;
```

```

If substr(phd_Céreq,1,2)="15" then phd_qeef3="11-Ecoles Commerce           ";
If substr(phd_Céreq,1,2)="16" then phd_qeef3="11-Ecoles Ingénieurs           ";

*** Codification des questions non posées dans Géné ;
If phd010M9="" then phd010M9="9";

If phd160M1="" then phd160M1="9";
If phd160M2="" then phd160M2="9";
If phd160M3="" then phd160M3="9";
If phd160M4="" then phd160M4="9";
If phd160M5="" then phd160M5="9";
If phd160M6="" then phd160M6="9";
If phd160M7="" then phd160M7="9";
If phd160M8="" then phd160M8="9";

If phd170M1="" then phd170M1="9";
If phd170M2="" then phd170M2="9";
If phd170M3="" then phd170M3="9";
If phd170M4="" then phd170M4="9";
If phd170M5="" then phd170M5="9";
If phd170M6="" then phd170M6="9";
If phd170M7="" then phd170M7="9";
If phd170M8="" then phd170M8="9";
If phd170M9="" then phd170M9="9";

run;

*****;
*** les diplômes dans SISE selon le PHD de Génération;
*****;
Title "Nombre de diplômes par type retrouvés, selon le plus haut diplôme atteint";
Title2 "Champ : répondants de génération";
Proc freq data=TableGlobale;
Tables phd_qeef2*( DUT Licpro Lic LicLMD MasterLMD MasterPro Ingenieur
                  DoctUniv DoctSante DES_Sante Aut_sante)/nopercent nocol norow;
run;

Title "Nombre de diplômes par niveau postbac, selon le plus haut diplôme atteint";
Title2 "Champ : répondants de génération";
Proc freq data=TableGlobale;

```

```
Tables phd_qeef2*( BACplus1 BACplus2 BACplus3 BACplus4 Bacplus5 Bacplus6)/nopercent nocol norow;  
run;  
Proc freq data=TableGlobale;  
Tables phd_qeef3*( Ingenieur DipMana3 DipMana4 DipMana5 DipInge3 DipInge4 DipInge5)/nopercent nocol norow;  
run;  
  
Title "Nombre de diplômes par niveau postbac, selon le plus haut diplôme atteint";  
Title2 "Champ : répondants de génération DE LA COMPOSANTE UNIV";  
Proc freq data=TableGlobale(where=(prefix_Ident = "UN"));  
Tables phd_qeef2*(BACplus3 BACplus4 Bacplus5 Bacplus6)/nopercent nocol norow;  
run;  
  
*****;  
*** Comparaison Génération et SISE;  
*****;  
  
Title "COMPARAISON : PHD calculé dans Génération et avec SISE";  
Title2 "Champ : répondants de génération";  
Proc freq data=TableGlobale;  
Tables phd_qeef2*PHD_SISE/nopercent nocol norow;run;  
  
Title "COMPARAISON : Déclarer des diplômes antérieurs dans Génération et constats dans SISE";  
Title2 "Champ : répondants de génération";  
Proc freq data=TableGlobale;  
Tables phd010M9*PHD_SISE/nopercent nocol norow;  
Tables phd160M2*DUT_old/nopercent nocol norow;  
Tables phd160M5*LicPro_old/nopercent nocol norow;  
Tables phd160M6*LicLMD_old/nopercent nocol norow;  
Tables phd170M5*Master_old/nopercent nocol norow;  
Tables phd170M7*Ingenieur_old/nopercent nocol norow;  
Tables phd170M9*Doct_old/nopercent nocol norow;run;  
  
*****;  
*** Comparaison Génération et SISE RESTREINTE A LA COMPOSANTE "UNIV";  
*****;  
  
Title "COMPARAISON : PHD calculé dans Génération et avec SISE";  
Title2 "Champ : répondants de génération DE LA COMPOSANTE UNIV";  
Proc freq data=TableGlobale (where=(prefix_Ident = "UN"));;  
Tables phd_qeef2*PHD_SISE/nopercent nocol norow;run;
```

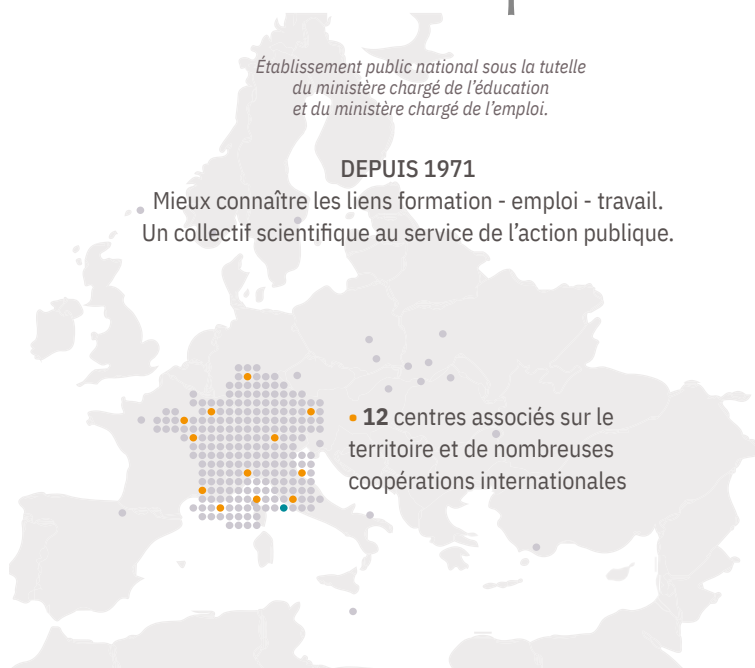
```
Title "COMPARAISON : Déclarer des diplômes antérieurs dans Génération et constats dans SISE";
Title2 "Champ : répondants de génération DE LA COMPOSANTE UNIV";
Proc freq data=TableGlobale (where=(prefix_Ident = "UN"));
Tables phd010M9*PHD_SISE/nopercent nocol norow;
Tables phd160M2*DUT_old/nopercent nocol norow;
Tables phd160M5*LicPro_old/nopercent nocol norow;
Tables phd160M6*LicLMD_old/nopercent nocol norow;
Tables phd170M5*Master_old/nopercent nocol norow;
Tables phd170M7*Ingenieur_old/nopercent nocol norow;
Tables phd170M9*Doct_old/nopercent nocol norow;run;
```

Céreq

*Établissement public national sous la tutelle
du ministère chargé de l'éducation
et du ministère chargé de l'emploi.*

DEPUIS 1971

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



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

↓ **+ d'infos**
et tous les travaux

À explorer
www.cereq.fr

🔓 **+ de 600 publications**
Accessibles librement

