



Direction des Statistiques Démographiques et Sociales

Unité des études démographiques et sociales
Division enquêtes et études démographiques

Base étude 2021 de
l'Echantillon
Démographique
Permanent



Manuel de l'utilisateur

Mise à jour : 1er juillet 2024



Sommaire général de la documentation de l'EDP

1. L'échantillon démographique permanent

1.1. Présentation

En bref

Les sources d'alimentation du panel : les bulletins statistiques de l'état civil, le recensement, le fichier général des électeurs et le panel d'actifs

Le critère d'échantillonnage : le jour de naissance (de 4 à 16 jours)

La période couverte : depuis 1967

Le champ couvert : des lieux variés selon les sources

Les informations disponibles : des informations à géométrie variable et à trous

Les fichiers mis à disposition

Schéma de la base étude 2021 de l'EDP

Tableaux des variables-clé et liens de la base étude

1.2. Bibliographie partielle

2. La base étude

2.1. La table individu (Cf. Fichier « 2.1. EDP - La table individu »)

La table Individu : EDP_BEEnnnn_INDIVIDU

Modalités de la table Individu : EDP_BEEnnnn_INDIVIDU

2.2. L'État civil (Cf. Fichier « 2.2. EDP - L'Etat civil »)

La table Naissance : EDP_BEEnnnn_NAISSANCE

Modalités de la table Naissance : EDP_BEEnnnn_NAISSANCE

La table Descendance : EDP_BEEnnnn_DESCENDANCE

Modalités de la table Descendance : EDP_BEEnnnn_DESCENDANCE

La table Mariage : EDP_BEEnnnn_MARIAGE

Modalités de la table Mariage : EDP_BEEnnnn_MARIAGE

La table Décès : EDP_BEEnnnn_DECES

Modalités de la table Décès : EDP_BEEnnnn_DECES

Annexe 1 - Départements dont les dossiers EDP ont été complétés par des bulletins d'état civil sur 2 ou 4 jours de 1982 à 1989

Annexe 2 - Présentation des bulletins d'état civil

Naissance (bulletin n°5)

Enfants sans vie (bulletin n°6)



Mariages (bulletins n°2)
Reconnaissances (bulletins n°4)
Décès (bulletin n°7)
Transcriptions et mentions en marge (bulletin n°1)
Mentions en marge

2.3. L'information électorale (Cf. Fichier « 2.3. EDP - L'information électorale »)

La table FE : EDP_BEEnnnn_FE
Modalités de la table sur l'information électorale : EDP_Bennnn_FE
La table REU : EDP_Bennnn_REU
Modalités de la table Répertoire Électoral Unique : EDP_Bennnn_REU

2.4. les RP de 1968 à 1999 (Cf. Fichier « 2.4. EDP - Les RP de 1968 à 1999 »)

Les recensements de population de 1968 à 1999

La table du recensement général de population de 1968 : EDP_BEEnnnn_RP1968
La table du recensement général de population de 1975 : EDP_BEEnnnn_RP1975
La table du recensement général de population de 1982 : EDP_BEEnnnn_RP1982
La table du recensement général de population de 1990 : EDP_BEEnnnn_RP1990
La table du recensement général de population de 1999 : EDP_BEEnnnn_RP1999

Modalités des variables brutes des tables du recensement de 1968 à 1999 :
EDP_BEEnnnn_RP1968, EDP_BEEnnnn_RP1975, EDP_BEEnnnn_RP1982,
EDP_BEEnnnn_RP1990 et EDP_BEEnnnn_RP1999

Modalités des variables redressées des tables du recensement de 1975 à 1999 :
EDP_BEEnnnn_RP1975, EDP_BEEnnnn_RP1982, EDP_BEEnnnn_RP1990 et
EDP_BEEnnnn_RP1999

Annexe 1 : Communes non dépouillées exhaustivement au RP 1975

Annexe 2 : Tableaux synoptiques des variables des recensements de population
de 1968 à 1999

Variables brutes des RP 1968, 1975, 1982, 1990 et 1999

Variables redressées des RP 1975, 1982, 1990 et 1999

2.5. Le RP à Mayotte en 2012 (Cf. Fichier « 2.5. EDP - Le RP à Mayotte en 2012 »)

La table du recensement de la population à Mayotte en 2012 :
EDP_Bennnn_RP2012_MAYOTTE

2.6. Les enquêtes annuelles de recensement à partir de 2004 (Cf. Fichier « 2.6. EDP - Les EAR »)

La table Individu des enquêtes annuelles de recensement :
EDP_BEEnnnn_EARaaaa_INDIVIDU

Modalités de la table des enquêtes annuelles de recensement pour les individus :
EDP_BEEnnnn_EARaaaa_INDIVIDU

La table Famille des enquêtes annuelles de recensement :
EDP_BEEnnnn_EARaaaa_FAMILLE



Modalités de la table des enquêtes annuelles de recensement pour les familles :
EDP_BEEnnnn_EARaaaa_FAMILLE

La table Logement des enquêtes annuelles de recensement :
EDP_BEEnnnn_EARaaaa_LOGEMENT

Modalités de la table des enquêtes annuelles de recensement pour les logements :
EDP_Bennnn_EARaaaa_LOGEMENT

La table Liens des enquêtes annuelles de recensement :
EDP_BEEnnnn_EARaaaa_LIENS

Modalités de la table des enquêtes annuelles de recensement pour les liens :
EDP_BEEnnnn_EARaaaa_LIENS

2.7. Le panel d'actifs (Cf. Fichier « 2.7. EDP - Le panel d'actifs »)

Modalités de la table du panel d'actifs : EDP_Bennnn_PANACT

2.8. Le panel des non salariés (Cf. Fichier « 2.8. EDP - Le panel non-salariés »)

Modalités de la table du panel des non salariés : EDP_BEEnnnn_PANNOSAL

2.9. Les données sociales et fiscales (Cf. Fichier « 2.9. EDP - Les données sociales et fiscales »)

La table des données fiscales individuelles : EDP_BEEnnnn_FISC_INDIVIDU

La table des données fiscales sur les logements : EDP_BEEnnnn_FISC_LOGEMENT

La table des données fiscales sur les revenus : EDP_BEEnnnn_FISC_REVENU

La table des données fiscales sur les revenus détaillés : EDP_BEEnnnn_FISC_REVDET

3. Nomenclatures et annexes (Cf. Fichier « 3. EDP - Nomenclatures et annexes »)

3.1. Géographie

3.1.1. Code des départements pour les RP 1968, 1975, 1982, 1990 et 1999

3.1.2. Code des départements ou zones frontalières du lieu de travail pour le RP 1982

3.1.3.1. Code des régions pour les RP 1968, 1975, 1982 et 1990 et zones d'emplois pour le RP 1990

3.1.3.2. Code des régions pour le RP 1999

3.1.3.3. Table de passage entre les anciens et les nouveaux codes région pour les EAR

3.1.4.1. Code des pays pour les RP 1968 et 1975

3.1.4.2. Code des pays pour le RP 1982

3.1.4.3. Code des pays pour le RP 1990

3.1.4.4. Code des pays pour le RP 1999

3.1.5 Code des lieux de vote des français à l'étranger

3.2. Nationalité

3.2.1. Nationalité pour les RP 1968, 1975 et 1982

3.2.2. Nationalité pour le RP 1990



3.2.3. Nationalité pour le RP 1999

3.2.4.1. Nationalité pour l'état civil par ordre alphabétique

3.2.4.2. Nationalité pour l'état civil par ordre ISO croissant

3.3. Taille de commune détaillée, pour les RP 1990 et 1999

3.4. Catégorie de commune détaillée : urbaine / rurale, pour les RP 1990 et 1999

3.5. Catégorie de migrants (urbain / rural / ZPIU) au RP 1982

3.6. Typologie socio-économique Tabard

3.6.1. Typologie socio-économique Tabard pour le RP 1990

3.6.2. Typologie socio-économique Tabard pour le RP 1999

3.7. Diplôme

3.7.1. Diplôme pour le RP 1975

3.7.2. Diplôme pour le RP 1982

3.7.3. Nomenclature harmonisée des diplômes depuis 1968

3.8. Catégorie sociale

3.8.1. Catégorie sociale pour les variables brutes des RP 1968 et 1975, les variables brutes et redressées du RP 1982 et l'état civil jusqu'en 1981

3.8.2. Catégorie sociale pour les variables redressées du RP 1975

3.8.3. Catégorie sociale pour les RP 1990 et 1999

3.8.4. Catégorie sociale pour l'état civil à partir de 1982

3.8.5. Nomenclature harmonisée des catégories sociales depuis 1968

3.9. Profession et métier

3.9.1. Profession pour le RP 1975 (abrégé)

3.9.2.1. Profession actuelle pour les RP 1990 et 1999

3.9.2.2. Profession antérieure pour les RP 1990 et 1999

3.10. Activité économique des établissements

3.10.1. Nomenclature d'activités et de produits de 1973, niveau 100, pour les variables du RP 1990

3.10.2. Nomenclature d'activités économiques, niveau 40, pour les variables du RP 1990

3.10.3. Nomenclature d'activités économiques, niveau 600, pour les variables du RP 1990

3.10.4. Nomenclature d'activités économiques, selon la nomenclature Nes36, pour les variables du RP 1999

3.10.5. Nomenclature d'activités économiques, selon la Naf niveau 700 adaptée, pour les variables du RP 1999

3.10.6. Nomenclature d'activités économiques, selon les sous sections de la Naf niveau 30, pour les variables du RP 1999

3.11. Catégorie juridique des établissements au RP 1999

3.12. Type de ménage

3.12.1. Type de ménage : définition pour le RP 1990

3.12.2. Type de ménage : définition pour le RP 1999

3.13. Tableau synoptique des variables RP, EAR et État Civil



1. L'échantillon démographique permanent

1.1. Présentation.....	7
En bref.....	7
Les sources d'alimentation du panel.....	8
Le critère d'échantillonnage : le jour de naissance (de 4 à 16 jours).....	11
La période couverte : depuis 1967.....	12
Le champ géographique couvert : des lieux variés selon les sources.....	12
Les informations disponibles : des informations à géométrie variable et « à trous ».....	13
Les fichiers mis à disposition.....	17
Schéma de la base étude 2021 de l'EDP.....	18
Tableaux des variables-clé et liens de la base étude.....	19



1.1. Présentation

En bref

L'échantillon démographique permanent (EDP) est un panel d'individus dont la constitution par l'Insee a débuté en 1968. Il consiste à collecter et conserver des informations puisées dans plusieurs sources pour des individus nés certains jours de l'année.

C'est un outil essentiel pour étudier les comportements et trajectoires socio-démographiques, en particulier pour les études des biographies et comportements démographiques (nuptialité, fécondité, mortalité) et familiaux, de la mobilité géographique, résidentielle, professionnelle et sociale, de la manière dont les parcours dans ces différents domaines se combinent, ou encore du lien à long terme entre situations passées et plus récentes (lien entre la situation actuelle et la situation vécue dans l'enfance par exemple). Par ailleurs, l'EDP sert de support à différentes productions statistiques régulières (enquête participation électorale, écarts d'espérance de vie selon les caractéristiques sociodémographiques, etc.) et enrichit d'autres sources statistiques. C'est également une source d'intérêt pour des études méthodologiques.

L'EDP mobilise cinq sources : les bulletins statistiques d'état civil depuis 1967 ; les informations des cinq recensements de population de 1968 à 1999 et des enquêtes annuelles de recensement depuis 2004 ; les informations issues du fichier électoral sur les inscriptions actuelles et passées depuis 1990 ; les informations issues des panels d'actifs (panel « tous salariés ») depuis 1967 ; des données socio-fiscales issues des dispositifs statistiques Fidéli (Fichier démographique d'origine fiscale sur les logements et les individus) et FiLoSoFi (Fichier localisé social et fiscal) depuis 2011.

Le critère d'échantillonnage retenu est celui du jour de naissance dans l'année : tous les individus nés du 2 au 5 janvier¹, du 1^{er} au 4 avril, du 1^{er} au 4 juillet et du 1^{er} au 4 octobre font potentiellement partie du champ de l'EDP. Ne font cependant effectivement partie du champ de l'EDP diffusé que les individus pour lesquels on dispose d'au moins un bulletin de recensement, ou bien d'un acte d'état civil (autre qu'un décès), ou bien d'une observation dans le panel d'actifs, ou bien la présence dans une déclaration fiscale. Historiquement, l'échantillon ne comprenait que les quatre jours d'octobre. Les individus EDP nés en janvier, avril et juillet sont de ce fait observés et suivis à partir de dates variables selon la source.

Le champ géographique couvert par l'EDP est intrinsèquement lié aux sources qui l'alimentent. Il correspond aux personnes résidant, travaillant, ou connaissant un événement d'état-civil enregistré en France. Le critère géographique est apprécié sur la base du lieu de recensement, donc du lieu de résidence pour le recensement. Pour les événements d'état civil, il l'est sur la base du lieu de déclaration, donc le plus souvent le lieu d'événement, indépendamment du lieu de résidence (en particulier, celui-ci peut être à l'étranger). Pour le fichier électoral, le critère géographique est le lieu d'inscription, qui peut être différent du lieu de résidence. Pour le panel d'actifs, le critère géographique est apprécié au lieu de l'activité déclaré par l'employeur. Pour les données socio-fiscales, la localisation est liée à l'adresse de taxation, qui peut être différente du lieu de résidence. Les départements d'outre-mer sont inclus dans les sources, sans limite pour le fichier électoral, à partir de 2004 pour l'état civil et les enquêtes annuelles de recensement et à partir de 2002 pour le panel d'actifs. Les données socio-fiscales couvrent théoriquement les Doms depuis 2011 mais de manière imparfaite.

En raison de son mode de construction et de sa longue durée, l'EDP n'offre pas un outil clé en main pour toutes les types d'études possibles. La nature des informations disponibles n'est en particulier pas constante au cours du temps. Pour l'état civil, 1998 et 2008 constituent deux ruptures majeures. Pour le recensement, le contenu des bulletins individuels et les informations contextuelles collectées ont variées depuis le recensement de 1968. Pour le panel d'actifs, seuls les individus nés en octobre des années paires sont suivis de 1967 à 2001. Pour les données socio-fiscales, les enfants qui

¹ Le 1^{er} janvier n'a pas été retenu car c'est une date souvent prise dans l'administration pour les jours/mois inconnus, et il présente un caractère atypique en terme de naissances.



apparaissent comme personnes à charge dans les déclarations fiscales sont mal identifiés lorsqu'ils sont nés avant juillet 2010. Par ailleurs, les contraintes de moyens ont aussi entraîné sur la période 1982-1997 « des trous de collecte » pour l'état civil qui rendent impossible un suivi très long pour les personnes nées les 2 et 3 octobre. Une définition précise des cohortes choisies doit donc être faite au cas par cas par l'utilisateur, sur la base des critères qu'il souhaite étudier (personnes nées telles années, personnes recensées telle année, etc.).

Les sources d'alimentation du panel

Historiquement, l'EDP a d'abord rassemblé des informations issues, d'une part, des bulletins statistiques d'état civil reçus par l'Insee pour les besoins de gestion du répertoire national d'identification des personnes physiques et pour l'établissement des statistiques d'état civil ; et d'autre part, les informations du recensement de population, puis des enquêtes annuelles de recensement à partir de 2004. L'information issue du fichier général des électeurs, possible depuis l'origine de l'EDP, n'est disponible dans la base étude de l'EDP qu'à partir de la version millésimée 2011 (disponible en avril 2013) et donne les inscriptions actuelles et passées. A partir de la version millésimée 2013 (disponible en décembre 2014), les informations issues du panel d'actifs (panel « tous salariés ») sont intégrées de manière rétroactive. A partir du millésime 2014, les données socio-fiscales sont intégrées, en commençant la série à partir de l'année 2011.

Pour toutes les sources de données, un travail spécifique d'identification par l'état civil ou un appariement par le NIR (numéro d'identification au répertoire) des personnes est réalisé pour rattacher les informations aux bons individus.

Bulletins statistiques de l'état civil

Les communes sont responsables de la tenue des registres d'état civil. Conformément à l'instruction générale relative à l'état civil, dont la dernière date du 11 mai 1999, elles sont tenues d'informer l'Insee de l'enregistrement d'un certain nombre d'actes dans des délais prescrits, au moyen de bulletins statistiques *ad hoc*². Pour l'Insee, ces bulletins ont généralement une double finalité : d'une part, l'établissement des statistiques d'état civil ; d'autre part, la mise à jour du répertoire national d'identification des personnes physiques (RNIPP), dont l'Insee assure la gestion depuis 1946. Les bulletins statistiques, remplis par les officiers d'état civil des communes, sont transmis sur support papier ou électronique en continu et sont traités dans le cadre de la base des répertoires des personnes physiques (BRPP), qui comprend à la fois le répertoire national d'identification des personnes physiques (RNIPP) et le fichier général des électeurs qui est remplacé par le répertoire électoral unique (REU) à partir de 2019.

Depuis l'initialisation de l'échantillon démographique permanent, les bulletins statistiques de l'état civil ont connu deux révisions importantes, en 1998 et en 2008. Pour les individus et les événements entrant dans son champ, l'échantillon démographique permanent récupère les données des sept bulletins suivants³ :

- bulletin de naissance d'une personne « EDP » ou d'un enfant d'une personne « EDP »
- bulletin de transcription d'un jugement déclaratif de naissance d'une personne « EDP » ou d'un enfant d'une personne « EDP »
- bulletin de transcription d'un jugement d'adoption plénière d'une personne « EDP » ou par adoption par une personne « EDP »
- bulletin de mariage d'une personne « EDP »
- bulletin d'enfant sans vie d'un enfant d'une personne « EDP »
- bulletin de décès d'une personne « EDP »
- bulletin de transcription d'un jugement déclaratif de décès ou d'absence d'une personne « EDP ».

² Ces bulletins sont visés par le CNIS. Le dernier visa couvre la période 2011-2015 (avis de conformité n°36/D131 du 16 avril 2010 ; avis d'opportunité n°14/D130 du 25 janvier 2008).

³ Dans des versions antérieures de la base étude, l'échantillon démographique permanent intégrait également les bulletins suivants (bulletins qui ont disparu depuis pour la gestion de l'état civil) : reconnaissance d'une personne « EDP » ou par une personne « EDP », mariage légitimant une personne « EDP », mention en marge de légitimation d'une personne « EDP », autres mentions en marge. Ces bulletins ne sont plus pris en compte dans la base étude, y compris leur historique. Les utilisateurs qui souhaitent exploiter ces bulletins se reporteront à la dernière base étude 2009 produite dans l'ancien système EDP.



Parmi ces bulletins, le bulletin d'enfant sans vie est le seul à n'avoir qu'une finalité statistique. Les bulletins sont présentés en annexe 3.

Recensements de population

Le recensement est réalisé par l'Insee et les communes. Lorsque dans un logement recensé un individu déclare une date de naissance EDP, son bulletin individuel (BI) est transmis à l'application EDP pour identification au répertoire des personnes physiques. Si l'identification échoue (manque d'information ou informations erronées) le BI n'est pas conservé. Si l'identification réussit, le BI vient enrichir la trajectoire de l'individu ou permet d'en commencer une nouvelle.

L'échantillon démographique permanent intègre des informations issues des bulletins des recensements généraux de population de 1968, 1975, 1982, 1990 et 1999 puis, à partir de 2004, des enquêtes annuelles de recensement.

En 1968, seules les informations figurant dans le bulletin individuel des personnes appartenant à l'échantillon démographique permanent ont été intégrées. Depuis le recensement de 1975, des informations figurant dans la feuille de logement et dans les bulletins individuels des autres personnes du ménage sont aussi intégrées.

Au-delà du contenu du bulletin lui-même, plusieurs évolutions sont à noter :

1°- De 1968 à 1982, les informations « brutes » correspondent à une saisie des bulletins papier effectuée spécifiquement pour l'EDP ; à partir de 1990, les informations « brutes » correspondent aux informations saisies dans le cadre des chaînes de traitement du recensement.

2°- A partir de 1975, les informations « redressées » dans le cadre des chaînes du traitement du recensement (y compris les variables imputées pour corriger la non-réponse) sont intégrées dans l'EDP, de façon toutefois parcellaire pour 1982 (à ce recensement, les ménages comprenant au moins une personne « EDP » n'ont pas été systématiquement intégrés dans l'exploitation complémentaire du recensement, faite sur un échantillon au quart des bulletins).

3°- La méthode d'intégration des informations « redressées » permet également d'intégrer des informations issues des autres bulletins que le bulletin individuel de l'individu « EDP », notamment des informations sur le logement, le ménage, le « conjoint », le « chef de famille », puis le « chef de ménage ». Ces informations complémentaires ont fortement évolué d'un recensement à l'autre, allant vers un enrichissement croissant. A partir des enquêtes annuelles de recensement, les informations sur toutes les personnes des ménages d'individu « EDP » sont conservées.

4°- La mise en place d'une collecte annuelle de recensement sur un échantillon de la population modifie fortement le contexte puisque, désormais, il ne sera plus possible de connaître simultanément les caractéristiques de l'ensemble des personnes « EDP » résidant en France à la date du recensement. Il faudra également recourir aux pondérations de l'enquête annuelle de recensement pour assurer la représentativité des cohortes définies par cette source.

Fichier général des électeurs

Jusqu'en 2019 et la constitution du répertoire électoral unique (REU), les listes électorales sont gérées par les communes et ne donnent pas lieu à la constitution d'un fichier unique centralisé. Toutefois, l'Insee gère alors un fichier général des électeurs⁴ destiné notamment à éviter les inscriptions multiples. Il est actualisé à partir des avis d'inscription et des avis de radiation transmis par les communes et des décisions judiciaires.

L'intégration d'informations sur l'inscription électorale donne les inscriptions actuelles et passées des individus EDP (en stocks et en flux de manière approximative depuis le début des années 1990 et de manière correcte depuis les années 2000, certaines inscriptions datant de 1945).

Un bulletin d'état civil (hors acte de décès) ou un bulletin de recensement suffisent à commencer la trajectoire d'un individu EDP. Les données électorales n'ont pas le même statut : elles ne créent pas de trajectoire mais enrichissent l'information sur les individus ayant déjà une trajectoire.

⁴ Loi n°46-1889 du 28 août 1946, repris dans l'article L. 37 du code électoral.



Panel d'actifs :

L'agence centrale des organismes de sécurité sociale (Acos) coordonne l'ensemble des organismes participant au recouvrement du régime général de sécurité sociale (les Urssaf). Toute entreprise a l'obligation de transmettre annuellement les déclarations de données sociales (DADS) relatives à chacun de ses salariés aux Urssaf. Par le biais de l'Acoss, l'Insee reçoit les fichiers statistiques relatifs à ces déclarations. Par ailleurs, l'Insee reçoit également les fichiers de paie des agents de l'Etat (FPE). Avec ces deux sources, il réalise un panel dit "panel tous salariés" qui collecte les informations salariales des individus nés en octobre les années paires et des individus EDP les années paires et impaires.

Les informations sur l'activité salariée des individus EDP sont constituées à partir de ce panel "tous salariés", qui est au niveau entreprise*individu*année, en l'agrégeant au niveau de l'individu (individu*année). Si un même individu a travaillé dans au moins deux entreprises différentes au cours d'une année donnée, on garde les caractéristiques du poste principal de l'entreprise pour laquelle la durée de paie est la plus longue et, en cas d'ex-æquo, le salaire le plus élevé. Les salaires nets, les heures et les durées de paie (en jours) sont additionnés.

Le panel d'actifs permet ainsi d'enrichir la base études de l'EDP avec des données relatives à l'activité salariée des individus EDP, notamment la catégorie socioprofessionnelle déclarée par les entreprises, le secteur d'activité de l'entreprise, le revenu salarial net de l'individu et le nombre d'heures travaillées durant l'année d'exercice.

Le champ suivi est, à partir de 1967, l'ensemble des salariés (sauf agricoles), y compris les salariés des entreprises publiques. La fonction publique hospitalière, les collectivités territoriales et la fonction publique d'Etat apparaissent en 1988. Il faut noter le cas particulier de France Telecom et de La Poste, issus de la privatisation de la direction des Postes et Télécommunications qui n'entraient pas dans le champ des DADS. Les salariés de ces entreprises n'apparaissent donc qu'en 1988 avec la fonction publique d'Etat. Les Doms et le secteur agricole sont couverts à partir de 2002. Enfin, les particuliers employeurs sont intégrés à partir de 2009.

Les données sociales et fiscales :

La direction générale des finances publiques (DGFiP) établit et recouvre l'impôt sur le revenu et les impôts locaux dont la taxe d'habitation. Les personnes physiques sont imposables à raison de l'ensemble de leurs revenus de source française ou étrangère, dès lors qu'elles sont domiciliées en France. Les personnes non domiciliées en France ne sont imposables que sur leurs revenus de source française.

L'Insee a la possibilité d'exploiter les fichiers fiscaux à des fins statistiques. Deux processus d'exploitation sont utilisés pour alimenter l'EDP : le fichier démographique des logements et des individus (Fidéli) et le fichier localisé social et fiscal (FiLoSoFi).

Le dispositif Fidéli produit annuellement un fichier des individus fiscaux qui liste toutes les personnes connues au sens de l'impôt sur le revenu ou de la taxe d'habitation principale au 1^{er} janvier d'une année donnée. Les individus fiscaux sont regroupés au sein de foyers fiscaux qui sont les unités de base des déclarations fiscales. Les sources qui permettent la constitution du fichier des foyers et individus fiscaux sont :

- le fichier de la taxe d'habitation,
- le fichier d'imposition des personnes (FIP),
- le fichier des déclarations de revenu (POTE⁵).

Un individu peut être présent dans plusieurs foyers fiscaux : par exemple, un étudiant rattaché au foyer fiscal de ses parents pour l'impôt sur le revenu et comme redevable de la taxe d'habitation dans son logement. Un foyer fiscal est rattaché à au plus un logement pour le paiement de la taxe d'habitation sur la résidence principale. Pour les logements, le dispositif Fidéli produit aussi une synthèse en mobilisant les sources foncières qui permettent une description du bâti :

- le fichier des propriétés bâties (FPB),
- le fichier des propriétaires,
- le fichier des propriétés non bâties.

Le dispositif Filosofi permet de recueillir annuellement des données sur les revenus et le niveau de vie. En rapprochant les fichiers fiscaux et sociaux, Filosofi donne une estimation plus précise des revenus et des prestations réellement perçues par les ménages. Filosofi collecte via la CCMSA⁶ les

⁵ Permanent des occurrences de traitement des émissions

⁶ Caisse centrale de la mutuelle sociale agricole



montants annuels des prestations légales de la branche vieillesse et de la branche famille relevant du régime agricole. La Cnaf⁷ et la Cnav⁸ fournissent le fichier exhaustif des prestations versées en décembre. Les prestations versées sur l'année sont ensuite reconstituées par imputation. Les données collectées sont agrégées au niveau du ménage fiscal : un ménage fiscal est un regroupement des foyers fiscaux rattachés à un même logement (hors logement collectif) pour lesquels coïncident une déclaration indépendante de revenus (dite déclaration n°2042) et l'occupation d'un logement connu à la taxe d'habitation (TH). Par exemple, un couple de concubins, où chacun remplit sa propre déclaration de revenus, constitue un seul ménage fiscal parce qu'ils sont répertoriés dans le même logement, même s'ils constituent deux foyers fiscaux distincts au sens de l'administration fiscale.

Les données socio-fiscales sont exhaustives sur leur champ et sont disponibles pour les déclarations fiscales de 2011 relatives aux revenus de 2010⁹, jusqu'aux déclarations fiscales du millésime de la base étude de l'EDP. Les individus EDP y sont identifiés grâce aux informations nominatives dont dispose l'administration fiscale (nom, prénom, sexe, date de naissance, lieu de naissance)¹⁰. Pour les enfants (personnes à charge), on ne dispose que de l'année de naissance. L'identification est donc réalisée à l'aide des données relatives aux parents issues des bulletins de naissance : les personnes à charge sont de ce fait moins bien suivies dans l'EDP que les personnes adultes.

Le critère d'échantillonnage : le jour de naissance (de 4 à 16 jours)

Initialement, l'échantillon démographique permanent collectait les informations concernant les personnes nées du 1^{er} au 4 octobre (on parle parfois des « quatre jours EDP »).

L'arrêté du 21 décembre 2006 a autorisé le quadruplement de la taille de l'échantillon pour passer à seize jours de naissance. Ils correspondent aux quatre premiers jours de chaque trimestre, avec une spécificité pour le mois de janvier, soit les personnes nées du 2 au 5 janvier, du 1^{er} au 4 avril, du 1^{er} au 4 juillet et du 1^{er} au 4 octobre.

Le quadruplement est mis en œuvre à partir de l'année 2004 pour l'état civil¹¹ et à partir de 2008 pour les enquêtes annuelles de recensement¹². Pour le fichier électoral, comme il s'agit d'un fichier historique pour l'ensemble des électeurs, les informations sont disponibles pour l'ensemble des jours de naissance EDP. Pour le panel d'actifs, il faut aussi tenir compte de l'année de naissance: sur la période d'activité de 1967 à 2001, les données n'ont été collectées que pour les individus nés en octobre des années paires ; à partir de 2002, l'ensemble des 16 jours EDP quelle que soit l'année de naissance sont collectés. Pour les données socio-fiscales, l'ensemble des jours EDP sont collectés.

L'échantillon démographique permanent ne consiste pas à suivre le devenir d'une cohorte de personnes initialement identifiées. Il juxtapose, pour les années où elles sont disponibles, les informations des fichiers sources relatives aux personnes nées les jours de naissance EDP. Ce mode d'échantillonnage par les jours de naissance a été choisi pour sa facilité de mise en œuvre. Cela ne revient cependant pas tout à fait à effectuer un tirage parfaitement aléatoire dans la population (certaines études suggèrent par exemple que la période de naissance dans l'année a un effet sur le parcours scolaire, du fait des écarts d'âges à l'entrée à l'école). Par ailleurs, selon les années et la présence de week-end dans les jours EDP, les effectifs des cohortes sont variables.

⁷ Caisse nationale d'assurance famille

⁸ Caisse nationale d'assurance vieillesse

⁹ Pour les années de revenus 2010 et 2011 issues de FiLoSoFi, les données n'ont pas bénéficié du même degré de validation que les années 2012 et suivantes. L'utilisateur doit donc utiliser ces années avec précaution.

¹⁰ Les informations détaillant l'état civil ne figurent pas dans les bases études.

¹¹ Dans la table EDP_BEn_decès, les bulletins de décès des personnes nées les « nouveaux » jours EDP ne sont disponibles que pour les décès survenus depuis 2007. En revanche, pour toutes les personnes de l'EDP, les informations concernant leurs dates et lieux de décès sont disponibles dans la table EDP_individu, même si on ne dispose pas de leurs bulletins de décès.

¹² L'enquête annuelle de recensement 2008 a été traitée sur les seize jours de naissance mais pour les douze « nouveaux » jours, seule l'identification automatique a été réalisée.



La période couverte : depuis 1967

L'échantillon démographique permanent rassemble les bulletins statistiques de naissances de personnes « EDP » à partir de l'année 1967. Les autres informations d'état civil et les informations issues des recensements sont intégrées à partir de l'année 1968. Les informations relatives à l'activité salariée sont intégrées à partir de l'année 1967. Il existe toutefois des trous de collecte :

- Les informations issues de l'exploitation complémentaire du recensement de 1982 ne sont disponibles que pour un quart des individus « EDP » recensés en 1982, ceux qui appartenaient à l'échantillon au quart du recensement.
- La plupart des bulletins statistiques de décès couvrant la période 1974 à 1989 n'ont pas été intégrés dans l'EDP. La plupart des décès de la période a été repérée ultérieurement par rapprochement avec le répertoire des personnes physiques (la date et le lieu de décès sont alors connus, mais pas les autres données des bulletins).
- Les informations des bulletins statistiques relatifs aux mariages, aux naissances d'enfants de personnes « EDP » (y compris les enfants mort-nés) ne sont pas disponibles pour les personnes nées les 2 ou 3 octobre pour les événements survenus au cours des années 1982 à 1989¹³.
- L'ensemble des bulletins d'état civil de la période 1990-1997 fait défaut pour les personnes nées les 2 ou 3 octobre. La mise à jour des décès réalisée par rapprochement de l'échantillon démographique permanent avec le répertoire des personnes physiques permet toutefois d'avoir la date et le lieu de décès.
- Les données sur les salariés nés les quatre premiers jours d'octobre des années paires sont disponibles à partir de 1967. Les données sur les salariés nés les quatre premiers jours d'octobre des années impaires et des 12 autres jours EDP sont disponibles à partir de 2002.
- Les données socio-fiscales sont disponibles pour l'ensemble des individus EDP, depuis l'année de déclaration fiscale 2011 (revenus 2010).

Le champ géographique couvert : des lieux variés selon les sources

Le champ géographique est défini à partir d'un critère qui varie selon la source considérée. En particulier, il ne s'agit pas systématiquement du lieu de résidence, même si l'essentiel des événements intégrés à l'EDP ont eu lieu en France pour des personnes résidant en France :

- **Pour le recensement** : le champ géographique est défini à partir de la commune de recensement, donc la commune de résidence habituelle à la date du recensement.
- **Pour les événements d'état civil** : pour les événements survenus en France, l'enregistrement de l'acte s'effectue normalement dans la commune où l'événement a eu lieu. Toutefois, les décès survenus à l'étranger peuvent être enregistrés en France et intégrés dans la base des répertoires des personnes physiques¹⁴. Par ailleurs, les événements survenus et enregistrés en France peuvent concerner des personnes résidant à l'étranger. Pour les statistiques de l'état civil, c'est le lieu de l'événement qui fait référence. Les bulletins statistiques relatifs à un événement survenu hors de France sont donc exclus des statistiques de l'état civil. L'EDP retient par contre ceux d'entre eux qui ont été intégrés dans la BRPP.

¹³ Seules quelques directions régionales ont intégré ces bulletins (Cf. annexe 2).

¹⁴ Selon les cas, les événements d'état civil concernant des ressortissants français à l'étranger (naissances, mariages, reconnaissances, décès) peuvent être déclarés et enregistrés directement à l'ambassade, l'ambassadeur agissant en tant qu'officier d'état civil. Dans ce cas, ils sont enregistrés par le service central d'état civil du ministère des Affaires étrangères qui centralise depuis 1965 l'ensemble des registres consulaires. Sinon, les événements d'état civil sont enregistrés selon les procédures du pays concerné. Dans ce cas, les actes d'état civil étrangers peuvent ensuite être transcrits sur les registres d'état civil français. Le service central du ministère des Affaires étrangères n'informe l'Insee que des décès. Certains actes donnent également lieu à une mention en marge dans la mairie du lieu de naissance si la personne concernée est née en France, mention en marge qui peut alors donner lieu à l'établissement d'un bulletin de mention en marge transmis à l'Insee.



- **Pour le fichier général des électeurs**, seule la commune d'inscription est disponible et fait donc référence. Celle-ci peut différer de l'adresse de résidence puisque toute personne majeure jouissant des droits civiques et répondant à la condition de nationalité¹⁵ peut s'inscrire, au choix, dans sa commune de domicile, sa commune de résidence¹⁶ ou toute commune où elle est assujettie à des contributions communales directes depuis cinq ans au moins. De plus, l'inscription est permanente. En cas de déménagement, la radiation n'est pas automatique tant que l'électeur n'entreprend pas de lui-même les démarches de réinscription dans sa nouvelle commune de résidence. Enfin, les électeurs inscrits au registre des Français établis hors de France peuvent s'inscrire dans une commune française sans y résider¹⁷.
- **Pour le panel d'actifs**, deux localisations peuvent être disponibles : le lieu de résidence et le lieu de travail. Le lieu de résidence est le lieu déclaré par l'entreprise concernant le lieu d'habitation du salarié, mais il est souvent mal renseigné. Le lieu de travail n'est pas forcément le lieu d'exercice de l'activité du salarié. En effet, certaines entreprises regroupent leurs déclarations de données sociales sur un seul établissement : l'établissement payeur des salaires. L'Insee réalise alors un travail de dégroupement afin de localiser les effectifs salariés le mieux possible.
- **Pour les données socio-fiscales** : si l'individu est rattaché à un foyer fiscal déclarant une taxe d'habitation principale, il est localisé en France à l'adresse de cette habitation. S'il est rattaché à un foyer fiscal connu dans les données fiscales que par sa déclaration de revenu, il peut être localisé en France mais aussi à l'étranger. Dans tous les cas, la localisation est « fiscale » dans le sens où l'individu peut ne pas résider effectivement à cette adresse. C'est le cas par exemple des enfants majeurs rattachés au foyer fiscal de leurs parents, qui peuvent résider ailleurs que chez leur parents (dans un internat par exemple).

Dans ce contexte, plutôt que de travailler sur le champ de l'EDP de façon générale, il peut être plus pertinent de définir le champ adapté au cas par cas, étude par étude, en fonction des critères retenus pour définir la cohorte étudiée. Par exemple, « personnes recensées à l'enquête annuelle de recensement de telle année », « personnes nées en France métropolitaine de telle année à telle année », « personnes inscrites sur les listes électorales en France métropolitaine telle année », etc.

Les informations disponibles : des informations à géométrie variable et « à trous »

De façon générale, la nature des informations disponibles varie dans le temps, d'une part en fonction de l'évolution des sources et des nomenclatures pour les appréhender ; d'autre part, en fonction des choix de gestion réalisés. Les variables créées spécifiquement pour l'EDP sont peu nombreuses. En particulier, il n'y a pas de variables qui permettraient de sélectionner de façon homogène dans le temps certaines populations selon leurs caractéristiques sociodémographiques, par exemple selon leur origine nationale, leur origine sociale ou leur niveau de diplôme au recensement. En construire ne va cependant pas de soi puisque l'évolution des questions ou des nomenclatures reflète aussi l'évolution des réalités qu'il s'agit d'appréhender, par exemple pour les diplômes ou la catégorie socioprofessionnelle.

Par ailleurs, des trous d'observation peuvent perturber l'appréhension des trajectoires démographiques individuelles que l'EDP permet d'étudier : des mariages ou des naissances d'enfants ne sont pas repérés pour les études de nuptialité ou de fécondité différentielle ; la catégorie socioprofessionnelle ou le lieu de résidence ne sont pas toujours connus pour les études de mobilité sociale ou de mobilité résidentielle, etc. Certains de ces trous résultent du champ couvert par les

¹⁵ Le fichier général des électeurs couvre également les ressortissants étrangers de l'Union européenne qui s'inscrivent sur les listes complémentaires pour voter aux élections municipales ou européennes (ils sont alors traités pour l'EDP comme les nationaux, sans distinction particulière). Il couvre aussi les inscriptions des Français de l'étranger sur les listes consulaires (ils ne sont pas intégrés à l'EDP).

¹⁶ Le « domicile » est une notion juridique qui, notamment, n'implique pas une résidence habituelle depuis un temps minimum, à la différence du lieu de « résidence ». Le plus souvent, lieu de résidence et lieu de domicile sont identiques. Exemples de situation où ils peuvent différer : un actif travaillant et résidant la semaine dans une autre commune que le domicile familial ; un étudiant domicilié chez ses parents et résidant dans la commune du lieu d'études.

¹⁷ Les textes listent les communes possibles : leur commune de naissance, leur dernier domicile ou leur dernière résidence ; la commune où est né, est inscrit ou à été inscrit un de leurs ascendants ; la commune où est inscrit ou a été inscrit un de leur parent jusqu'au 4^{ème} degré ; la commune où est inscrit leur conjoint.



sources alimentant l'EDP (non couverture des évènements d'état civil survenus à l'étranger par exemple). D'autres résultent d'arbitrage sur les moyens disponibles pour intégrer les informations dans l'EDP (interruption de l'intégration des bulletins de décès de 1973 à 1989 et de tous les bulletins d'état civil des personnes nées les 2 et 3 octobre de 1982 à 1997). D'autres enfin résultent d'évolutions des sources alimentant l'EDP (exploitation « lourde » du recensement de 1982 réalisée sur un sondage au quart ; fin de l'exhaustivité de la collecte du recensement avec la mise en place des enquêtes annuelles de recensement). A noter que pour les études de mortalité, la récupération des dates de décès permet, malgré les trous d'observation sur les bulletins de décès¹⁸, des analyses pertinentes.

A propos des recensements de population

Pour les recensements de 1968, 1975 et 1982, les bulletins individuels des personnes nées un « jour EDP » ont fait l'objet d'une saisie spécifique pour l'EDP. Les variables concernées sont dites « brutes ». Les informations disponibles concernent les caractéristiques sociodémographiques de la personne (sexe, année de naissance, lieu de naissance, nationalité, lieu de résidence au recensement antérieur, niveau de diplôme, année d'installation en France pour les immigrés), la situation d'activité, le statut, la catégorie socioprofessionnelle, le lieu de travail, le secteur d'activité, le lieu de résidence (commune, département, région).

Pour les recensements de 1975 et 1982, la saisie du numéro du bulletin individuel du recensement des personnes de l'EDP a permis de récupérer les informations issues des travaux d'exploitation du recensement. Les variables concernées sont dites « redressées ». Certaines variables « redressées » sont redondantes avec les variables « brutes ». Les différences peuvent résulter de différences de saisie entre les deux opérations, mais aussi des corrections d'anomalies et des imputations de non-réponse effectuées pour l'exploitation du recensement. Le rapprochement avec le recensement permet surtout de récupérer des informations sur le logement, la composition de la famille et du ménage. Il permet aussi de récupérer une variable de diplôme synthétique, qui isole notamment les « sans diplôme ».

Pour le recensement de 1975, les variables « redressées » sont disponibles pour la quasi-totalité des personnes « EDP » recensées cette année là puisque le recensement a été traité de façon exhaustive. Il l'a été en deux temps : un sondage au cinquième a d'abord été traité pour permettre la publication de premiers résultats nationaux et départementaux ; les quatre cinquièmes suivants ont été exploités ensuite. Seules 130 communes n'ont pas été traitées par l'exploitation exhaustive du recensement : 55 communes dans trois départements de Provence-Alpes-Côte d'Azur (dont Nice, Menton, Grasse, Antibes, Arles, Fréjus) ; 70 en Île-de-France (principalement dans les Hauts-de-Seine, la Seine-Saint-Denis, le Val-de-Marne et sept arrondissements de Paris), ainsi qu'à Toulouse, Nantes, Rezé et, de façon partielle, Lyon (Cf. annexe 1). Pour ces communes, seule l'exploitation au cinquième a été réalisée. Les informations issues du rapprochement avec les fichiers du recensement de 1975 sont donc manquantes pour une partie des personnes « EDP » recensées en 1975 (pour 11,3 % des cas, ce taux combinant le trou d'exploitation et les échecs d'appariement).

Pour le recensement de 1982, les variables « redressées », donc les variables de niveau logement, ménage et famille, ne sont disponibles que pour un quart des personnes « EDP » recensées cette année là, celle relevant de l'exploitation complémentaire du recensement.

A partir du recensement de 1990, l'ensemble des informations est récupéré directement des chaînes d'exploitation du recensement. Les variables « brutes » correspondent désormais aux valeurs issues de la saisie réalisée pour le recensement. Les variables « redressées » restent celles utilisées pour les résultats diffusés du recensement, après détection des anomalies, redressements et imputations de la non-réponse. Les informations intégrées continuent de porter sur les quatre niveaux : individu, ménage, famille et logement.

A partir de 2004, avec la mise en place d'un recensement en continu sur un échantillon rotatif, les informations du recensement sont intégrées chaque année dans l'EDP pour une partie des individus EDP. L'utilisation de la pondération de l'enquête annuelle (variables POIDS_EA ou POIDS_EA_CALE) est donc nécessaire pour assurer la représentativité de cohortes définies à partir

¹⁸ La table individu dispose néanmoins de la date de décès pour la majorité des individus décédés, des études sur la mortalité sont donc bien possibles.



d'informations issues d'une EAR donnée. Pour les cohortes définies à partir d'autres critères mais utilisant des informations d'une ou plusieurs EAR, une pondération spécifique devrait être utilisée.

	1968	1975	1982	1990	1999	EAR
Caractéristiques de l'individu "EDP"						
Saisies spécifiquement pour l'EDP (variables "brutes")	Oui	Oui	Oui	Non	Non	Non
Saisies pour l'exploitation du recensement (variables "brutes")	Non	Non	Non	Oui	Oui	Oui
Corrigées et imputées pour l'exploitation du recensement (variables "redressées")	Non	Oui	Sur 1/4	Oui	Oui	Oui
Caractéristiques du logement de l'individu "EDP"						
Saisies pour l'exploitation du recensement (variables "brutes")	Non	Non	Non	Oui	Oui	Oui
Corrigées et imputées pour l'exploitation du recensement (variables "redressées")	Non	Oui	Sur 1/4	Oui	Oui	Oui
Structure du ménage de l'individu "EDP"						
Corrigée et imputée pour l'exploitation du recensement (variables "redressées")	Non	Oui	Sur 1/4	Oui	Oui	Oui
Caractéristiques des autres personnes du ménage						
Saisies pour l'exploitation du recensement (variables "brutes")	Non	Non	Non	Oui (partiel*)	Oui (partiel*)	Oui (sur toutes)
Corrigées et imputées pour l'exploitation du recensement (variables "redressées")	Non	Oui	Oui (partiel*)	Oui (partiel*)	Oui (partiel*)	Oui (sur toutes)
* Informations plus ou moins nombreuses sur :						
Structure de la famille		Oui	Sur 1/4	Oui	Oui	Oui
Membres de la "famille" au sens du recensement :						
- la "mère présumée" de l'individu EDP, s'ils cohabitent	Non	Non	Non	Oui	Oui	Oui
- le "père présumé" de l'individu EDP, s'ils cohabitent	Non	Non	Non	Non	Oui	Oui
- les "enfants présumés" de l'individu EDP, s'ils cohabitent	Non	Non	Non	Oui	Oui	Oui
- le "conjoint présumé" de l'individu EDP, s'ils cohabitent	Non	Non	Sur 1/4	Non	Oui	Oui
- le "chef de famille" (ou "personne de référence de la famille")	Non	Oui	Non	Non	Oui	Oui
- le "conjoint du chef de famille" (ou de la "personne de référence de la famille")	Non	Oui	Non	Non	Non	Oui
- le "père de famille"	Non	Non	Sur 1/4	Oui	Oui	Oui
- la "mère de famille"	Non	Non	Sur 1/4	Oui	Oui	Oui
Membres du "ménage"						
- la "personne de référence du ménage"	Non	Non	Non	Non	Oui	Oui
- le conjoint de la "personne de référence du ménage"	Non	Non	Non	Non	Oui	Oui

A propos de l'état civil

Les bulletins d'état civil sont présentés de façon détaillée en annexe 2 du 2.2 . Parmi les informations disponibles :

- Les bulletins de naissances permettent de connaître pour les personnes « EDP » nées en France depuis 1967 : l'année de naissance, la commune de naissance et le type de lieu de naissance de la personne (hôpital, domicile, etc.) ; la profession, la nationalité, l'âge et le lieu de naissance de chacun de ses parents et le fait qu'ils soient mariés ou non ; le lieu de résidence).
- Les bulletins de mariage permettent de connaître le nombre de mariages successifs contractés en France à partir de 1967 par les personnes « EDP », leur date, leur lieu ; la date de naissance, le lieu de naissance et la nationalité de chacun des conjoints, ainsi que leur situation matrimoniale antérieure et leur lieu de résidence . Les informations sont cependant lacunaires pour les personnes nées les 2 ou 3 octobre en raison du trou de collecte pour l'EDP de la période 1982-1997 (du fait de ce trou de collecte, le rang des mariages à nouveau observés à partir de 1998 est aussi biaisé). Par ailleurs, la remontée des bulletins statistiques de mariage vers l'Insee est celle qui connaît le plus de difficultés, notamment parce qu'alors que les naissances et les décès sont de plus en plus localisés dans un nombre réduit de communes, les mariages restent très diffus sur le territoire, pouvant concerner des petites communes où des oublis de transmission peuvent avoir lieu. Il n'est alors pas facile de les repérer du fait du peu nombre d'actes enregistrés dans ces communes. Le fichier des



statistiques d'état civil des mariages est ainsi le seul à faire l'objet d'un redressement en exhaustivité (ces dix dernières années la correction a oscillé entre 3 % et 6 %). Jusqu'en 1998, la catégorie socioprofessionnelle des conjoints ou le nombre d'enfants légitimés par le mariage étaient disponibles. Les utilisateurs qui veulent étudier ces variables non migrées doivent se reporter à la dernière base étude produite par l'ancienne chaîne EDP (voir avis de mise à disposition n° 2011/42 des bases AAA50090.DEA99SAS et AAA50090.DEB99SAS).

- Les bulletins de naissance et d'enfants sans vie de personnes « EDP » permettent d'approcher le nombre d'enfants des personnes « EDP » et de connaître la date de naissance, la commune de naissance et le sexe de l'enfant ; la date et le lieu de naissance, la nationalité et la profession de chacun des parents quand les deux parents sont indiqués dans l'acte ; le fait que les parents soient mariés et, dans ce cas, la date du mariage. Les informations sont cependant lacunaires pour les personnes nées les 2 ou 3 octobre en raison du trou de collecte de la période 1982-1997 (du fait de ce trou de collecte, le rang des naissances à nouveau observées à partir de 1998 est biaisé). Ne sont pas non plus pris en compte les bulletins de naissance effectuée sous X. En revanche, les jugements d'adoption plénière figurent dans la base étude.
- Les bulletins de décès permettent de connaître la commune de décès et la date du décès, le lieu de résidence, la situation matrimoniale et la profession à la date du décès. Toutefois, pour les décès survenus de 1974 à 1989, et pour les décès des personnes nées les 2 ou 3 octobre et décédés de 1990 à 1997), les informations disponibles se limitent à la date et au lieu du décès (i.e. la commune de déclaration du décès). Les décès sont aussi très mal repérés pour les personnes nées avant 1891 (lors du passage du répertoire manuel au répertoire magnétique du RNIPP, seules les personnes nées à partir de 1891 ont été prises en compte).

Lorsque le bulletin a été traité pour les besoins des statistiques d'état civil, l'information diffusée dans la base étude est celle redressée par les traitements de l'état civil.

Jours de naissances intégrés dans l'EDP selon l'année et le type de bulletin

	Naissance de l'individu EDP	Mariages d'individus EDP	Naissances d'enfants d'individu EDP	Décès d'individus EDP
1967	Du 1er au 4 octobre. Aucun les autres mois	Aucun		
1968-1973	Du 1er au 4 octobre. Aucun les autres mois			
1974-1981	Du 1er au 4 octobre. Aucun les autres mois			Aucun mais connaissance des décès des jours d'octobre par la BRPP (bulletins « fantômes »)
1982-1989	Complet pour les 1er et 4 octobre. Très lacunaire pour les nés le 2 ou 3 octobre. Aucun pour les autres mois			
1990-1997	Aucun pour les nés le 2 ou 3 octobre ou les autres mois			Aucun pour les nés le 2 ou 3 octobre ou les autres mois (mais « fantôme » pour les 2 et 3 octobre)
1998-2003	Du 1er au 4 octobre. Aucun les autres mois			
2004-millésime	Tous les jours EDP (à partir de 2007 pour les décès des nés en janvier, avril et juillet)			



Les fichiers mis à disposition

La base étude est éclatée en différents fichiers thématiques :

- une table INDIVIDU « noyau », avec les principales caractéristiques de l'individu et des indicatrices pour renvoyer vers les autres tables. Les informations démographiques disponibles dans cette table sont les plus fiables et à privilégier (date de naissance, lieu de naissance, date de décès, lieu de décès).
- une table par type d'évènements d'état civil, soit quatre tables au total : naissance d'une personne « EDP » (table NAISSANCE), descendance d'une personne « EDP » (sans nombre maximum) (table DESCENDANCE), mariage d'une personne « EDP » (sans nombre maximum) (table MARIAGE), décès d'une personne « EDP » (table DECES);
- une table FE sur l'historique de la situation d'inscription électorale ;
- une table RP par recensement général de population (1968, 1975, 1982, 1990 et 1999), soit cinq tables au total ;trois tables (INDIVIDU, FAMILLE et LOGEMENT) par enquête annuelle de recensement(EAR) à partir de l'année 2004 : une table de niveau « logement », une table de niveau « famille » et une table de niveau « individu » (la table « individu » décrit tous les individus des logements comportant au moins une personne « EDP » ; la table « famille » décrit les familles des personnes vivant dans un logement « EDP » comportant au moins une personne « EDP »). S'y ajoute une table pour le recensement de la population de Mayotte en 2012.
- une table PANACT sur l'historique de l'activité salariale des individus.
- une table FISC_INDIVIDU par année fiscale (à partir de l'année 2011 jusqu'à l'année millésime de la base étude) qui compile les informations sociales et fiscales des individus EDP et des autres individus de son logement.
- une table FISC_LOGEMENT par année fiscale (à partir de l'année 2011 jusqu'à l'année millésime de la base étude) qui compile les informations sur les logements à partir des données fiscales.
- une table FISC_REVENU par année fiscale (à partir de l'année 2011 jusqu'à l'année millésime de la base étude) qui compile les informations sur les revenus des ménages constitués d'au moins un individu EDP.
- une table FISC_REVDET par année fiscale (à partir de l'année 2011 jusqu'à l'année millésime de la base étude) qui compile les informations sur les revenus individuels.

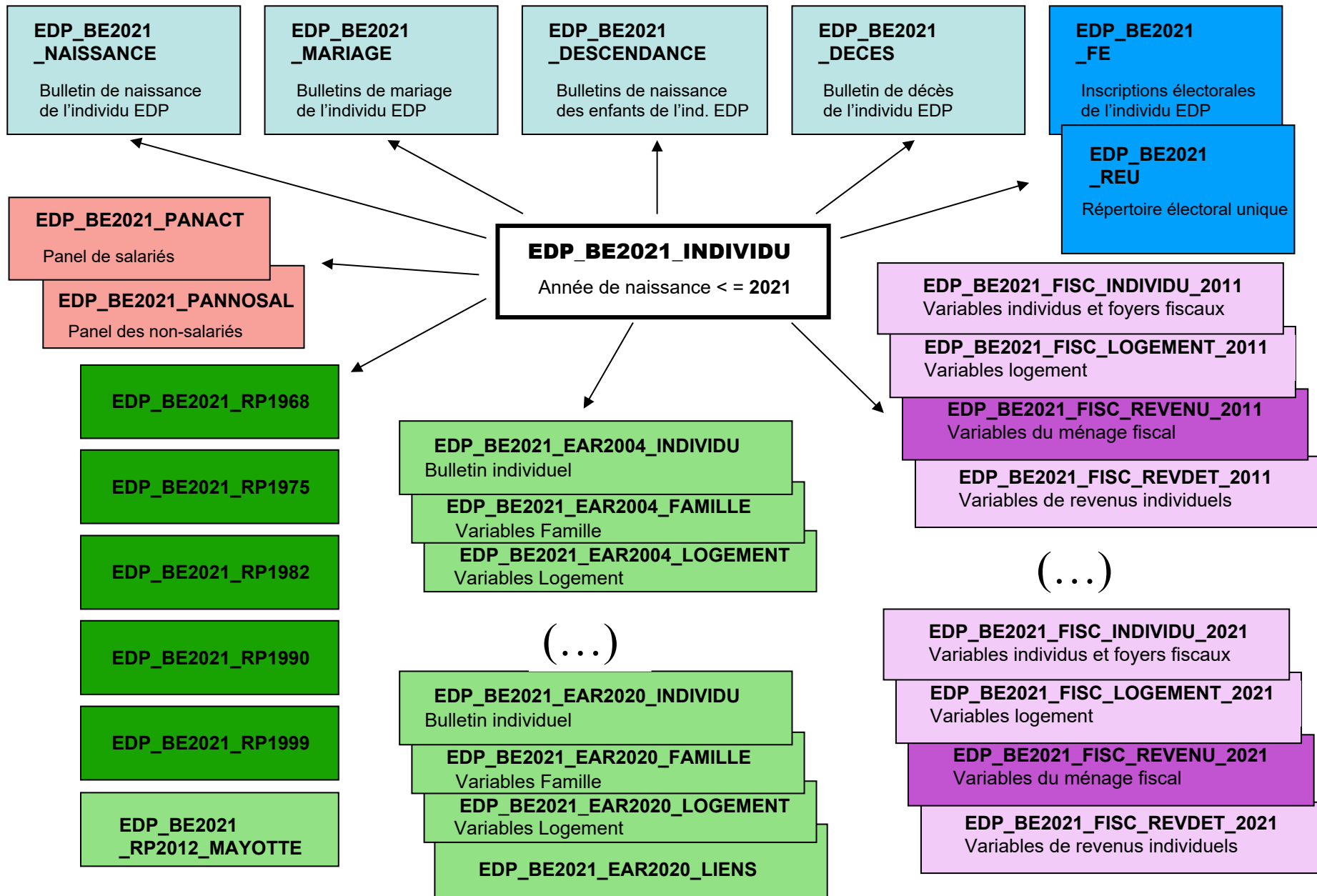
Les identifiants utilisés dans la base étude de diffusion sont non significatifs et attribués aléatoirement¹⁹. En particulier, ils sont non constants d'une version de la base étude à l'autre. Lors de l'initialisation d'un nouvel environnement « BRPP » en 2010, certains individus présents dans les anciennes bases EDP n'ont pu être identifiés. Les trajectoires de ces individus ne sont plus enrichies de nouvelles informations mais sont cependant incorporées à la base étude et peuvent être repérées grâce à la présence d'un « X » en début d'identifiant.

Depuis la base d'étude 2021, les tables ne sont plus en SAS mais en format Parquet. L'ordre des variables du dictionnaire des variables ne suit pas l'ordre des variables dans les fichiers. Certaines modalités de variables dans ce dictionnaire ne sont pas renseignées (nous n'avons pu trouver ces informations dans les fichiers sources).

¹⁹ Variable ID_DIFF (pour l'individu), variable ID_EVT_DIFF pour les évènements



Schéma de la base étude 2021 de l'EDP



Tableaux des variables-clé et liens de la base étude

VARIABLES-CLÉ DES TABLES DE LA BASE ÉTUDE

TABLE	VARIABLE CLÉ
EDP_BEEnnnn_INDIVIDU	ID_DIFF
EDP_BEEnnnn_NAISSANCE	ID_EVT_DIFF ID_DIFF (*)
EDP_BEEnnnn_DESCENDANCE	ID_EVT_DIFF, OPE_TYPE ID_DIFF, ID_EVT_DIFF (*)
EDP_BEEnnnn_MARIAGE	ID_DIFF, ID_EVT_DIFF
EDP_BEEnnnn_DECES	ID_EVT_DIFF ID_DIFF (*)
EDP_BEEnnnn_FE	ID_DIFF, INS_DATE, INS_LIEU, LISTE_TYPE
EDP_BEEnnnn_RPaaaa	ID_EVT_DIFF
EDP_BEEnnnn_EARaaaa_INDIVIDU	ID_BI_DIFF
EDP_BEEnnnn_EARaaaa_FAMILLE	ID_FAM_DIFF
EDP_BEEnnnn_EARaaaa_LOGEMENT	ID_LOG_DIFF
EDP_BEEnnnn_PANACT	ID_DIFF, AN
EDP_BEEnnnn_FISC_INDIVIDU_aaaa	ID_FISC_FOY_DIFF, ORDREFIP, TYPE_FISC
EDP_BEEnnnn_FISC_LOGEMENT_aaaa	ID_FISC_LOG_DIFF, AN_FISC
EDP_BEEnnnn_FISC_REVENU_aaaa	ID_FISC_LOG_DIFF, AN_FISC
EDP_BEEnnnn_FISC_REVDET_aaaa	ID_FISC_FOY_DIFF, ORDREFIP, TYPE_FISC

* deuxième possibilité

VARIABLES À UTILISER POUR RAPPROCHER LES TABLES DE LA BASE ÉTUDE

TABLE1	TABLE2	VARIABLE LIEN
EDP_BEEnnnn_INDIVIDU	EDP_BEEnnnn_NAISSANCE	ID_DIFF
	EDP_BEEnnnn_DESCENDANCE	ID_DIFF
	EDP_BEEnnnn_MARIAGE	ID_DIFF
	EDP_BEEnnnn_DECES	ID_DIFF
	EDP_BEEnnnn_FE	ID_DIFF
	EDP_BEEnnnn_RPaaaa	ID_DIFF
	EDP_BEEnnnn_EARaaaa_INDIVIDU	ID_DIFF
	EDP_BEEnnnn_EARaaaa_FAMILLE	Pas de lien direct
	EDP_BEEnnnn_EARaaaa_LOGEMENT	Pas de lien direct
	EDP_BEEnnnn_PANACT	ID_DIFF
	EDP_BEEnnnn_FISC_INDIVIDU_aaaa	ID_DIFF
	EDP_BEEnnnn_FISC_LOGEMENT_aaaa	Pas de lien direct
	EDP_BEEnnnn_FISC_REVENU_aaaa	Pas de lien direct
	EDP_BEEnnnn_FISC_REVDET_aaaa	Pas de lien direct
EDP_BEEnnnn_NAISSANCE	EDP_BEEnnnn_DESCENDANCE	ID_EVT_DIFF
EDP_BEEnnnn_EARaaaa_INDIVIDU	EDP_BEEnnnn_EARaaaa_FAMILLE	ID_FAM_DIFF
	EDP_BEEnnnn_EARaaaa_LOGEMENT	ID_LOG_DIFF
	EDP_BEEnnnn_FISC_INDIVIDU_aaaa	EDP_BEEnnnn_FISC_LOGEMENT_aaaa
EDP_BEEnnnn_FISC_INDIVIDU_aaaa	EDP_BEEnnnn_FISC_REVENU_aaaa	ID_FISC_LOG_DIFF
	EDP_BEEnnnn_FISC_REVDET_aaaa	ID_FISC_FOY_DIFF
		+ORDREFIP
		+TYPEFISC

