

L'ÉCHANTILLON DÉMOGRAPHIQUE PERMANENT

EN 50 ANS, L'EDP A BIEN GRANDI !

Isabelle Robert-Bobée* et Natacha Gualbert**

Depuis plus de cinquante ans, l'Insee rassemble des informations socio-démographiques sur un échantillon d'individus représentatifs de la population résidant en France : l'échantillon démographique permanent. Pour chacun de ces individus, l'EDP s'enrichit chaque année de données issues du recensement, de l'état civil, du fichier électoral, et plus récemment, de données d'emploi pour les salariés, et de données fiscales (déclaration de revenus et taxe d'habitation). Suivront prochainement l'ajout de données sur les non salariés.

Actuellement, l'EDP retrace 3,7 millions de trajectoires individuelles, dont plus de 200 000 sur 50 ans. C'est une source unique pour l'étude des mobilités géographiques et sociales sur longue période, mais aussi des évolutions du niveau de vie en lien avec les événements familiaux ou professionnels vécus, comme une séparation ou le passage à la retraite. L'introduction du niveau de vie dans l'EDP, en plus des caractéristiques socio-démographiques habituelles (âge, sexe, catégorie socio-professionnelle, diplôme, situation familiale), a encore étendu le champ des études avec ce panel.

Rare panel d'individus en population générale, l'EDP a su s'adapter aux changements des sources qui l'alimentent. Le passage du recensement exhaustif aux enquêtes de recensement au milieu des années 2000 a été l'occasion d'élargir la taille de l'échantillon et d'intégrer de nouvelles sources exhaustives (données socio-fiscales) ou en panel (déclarations de salaires).

 *For more than fifty years, the permanent demographic sample (known as "EDP") has been accumulating socio-demographic information for a sample of individuals representative of the population living in France. For each of these individuals, the EDP is enriched each year with data from the census, civil status, the electoral register and, more recently, employment data for salaried employees and tax data (income tax return and housing tax). This will soon be followed by the addition of data on the self-employed, and progress on the health side to evaluate the national health strategy.*

The EDP currently tracks 3.7 million individual trajectories, including more than 200,000 over 50 years. It is a unique source for the study of geographical mobility over a long period, but also of changes in living standards in connection with family or professional events experienced, such as couples breakdowns or the transition to retirement. The inclusion of the standard of living in the EDP, in addition to the usual socio-demographic characteristics (age, gender, socio-professional category, diploma, family situation) has further extended the field of study with this panel.

As a rare panel of individuals in the general population, the EDP has been able to adapt to changes in the sources it is based on. The transition from comprehensive census to census surveys in the mid-2000s was an opportunity to broaden the sample size and integrate new comprehensive (tax data) or panel (employees) sources.

* Cheffe de la division Enquêtes et études démographiques, Insee, isabelle.robert-bobee@insee.fr

** Responsable statistique de l'échantillon démographique permanent, Insee, natacha.gualbert@insee.fr

L'échantillon démographique permanent (EDP) a été créé en 1968 à l'Insee à partir de la compilation de données des recensements de la population et d'état civil. Il s'agissait de mettre en place un nouvel outil pour l'analyse des mobilités géographiques et des trajectoires sociales (Sautory, 1988), comme les différentiels sociaux de mortalité, les parcours professionnels sur longue période ou la trajectoire des immigrés (mobilité professionnelle, acquisition de la nationalité française par exemple).

Plusieurs approches sont possibles pour constituer des données de trajectoires : enquêter des personnes à un moment donné en leur posant des questions sur leur parcours passé (enquête rétrospective, faisant appel à la mémoire des enquêtés) ; interroger plusieurs fois les mêmes personnes pour recueillir des informations au fil des années (mais avec la difficulté de retrouver les personnes pour les réinterroger, et une attrition qui augmente donc au fil des années) ; ou, comme le panel EDP, rassembler au fil des années des données recueillies par ailleurs.

Dès sa création, l'EDP mobilise des données administratives et des données du recensement de la population, ce qui en fait un dispositif particulièrement économe : pas de coût de collecte, pas de charge de réponse auprès d'enquêtés. L'échantillon peut donc être de grande taille, le tout sans attrition ni biais de mémoire. Un critère d'échantillonnage très simple a été retenu : le jour de naissance. Cela simplifie la mise en œuvre du panel et donc sa pérennité.

Sans effet de mémoire ni attrition, l'EDP est unique également par sa taille (3,7 millions de personnes actuellement), sa profondeur historique (50 ans de données pour plus de 200 000 personnes), et par la diversité de ses sources (état civil, recensement, puis fichier électoral et données socio-fiscales).

En 50 ans, l'EDP a dû s'adapter aux évolutions parfois importantes de ses sources historiques ; le panel a su aussi intégrer de nouvelles données, qui ont enrichi les études, mais ont aussi rendu plus complexe son exploitation.

● À L'ORIGINE, ÉTAT CIVIL ET RECENSEMENT POUR LES NATIFS DES 4 PREMIERS JOURS D'OCTOBRE

Les personnes faisant partie de l'EDP sont celles nées certains jours de l'année, dits « jours EDP » et fixés par arrêté depuis 2014¹. Au départ, il s'agissait des personnes nées les 4 premiers jours d'octobre. La première source retenue pour démarrer le panel a été le recensement de 1968, complétée par des données d'état civil (naissances, décès, mariages).

En 1968, année de la création de l'EDP, 500 000 personnes sont ainsi entrées dans ce panel (*figure 1*) :

- 484 200 personnes recensées en 1968 et nées un 1, 2, 3 ou 4 octobre (quelle que soit leur année de naissance) ;
 - 9 200 personnes nées les 1^{er}, 2, 3 ou 4 octobre 1968 pour lesquelles on disposait du bulletin de naissance dans les fichiers de l'état civil. Elles n'ont pas été recensées en 1968, le recensement ayant eu lieu en mars 1968², donc avant leur naissance ;
 - 900 personnes ajoutées grâce à l'état civil sur les mariages (6 900 personnes nées un 1^{er}, 2, 3 ou 4 octobre se sont mariées en 1968, et parmi elles, 6 000 ont été recensées en 1968) ;
-

1. Arrêté du 6 août 2014 portant création d'un traitement automatisé de données à caractère personnel relatif à l'échantillon démographique permanent de l'Insee (voir les références réglementaires en fin d'article).
2. Les recensements généraux de population (jusqu'en 2004) se déroulaient généralement au mois de mars.

- ① et 500 personnes nées en 1967 mais non recensées en 1968 (sur les 8 300 bulletins de naissance d'enfants nés un 1^{er}, 2, 3 ou 4 octobre 1967, 7 800 concernaient des personnes déjà incluses dans l'EDP grâce au recensement de 1968).

Si le panel avait été limité aux seules personnes nées en 1968, il aurait fallu attendre de nombreuses années avant d'étudier des trajectoires. Pour éviter cet inconvénient, l'échantillon a dès le départ été constitué de personnes de tous âges. Le panel s'est enrichi ensuite chaque année des personnes nées l'un des quatre jours de référence au cours de l'année en France à partir d'informations recueillies dans des actes d'état civil et d'informations consignées dans les bulletins de recensement successifs. Ainsi, la base était proche d'un échantillon représentatif au 1/100^e (4 jours / 365 jours) de la population résidant en France (Couet, 2006).

« L'échantillon a dès le départ été constitué de personnes de tous âges. »

Depuis sa création, le renouvellement de l'échantillon est assuré par les naissances ou par la venue en France de nouvelles personnes. Si elles sont nées un jour EDP, ces dernières rejoignent l'EDP à l'occasion d'un recensement ou d'un événement enregistré dans un acte de l'état civil (Couet, 2006). À l'inverse, le suivi d'un individu cesse, de fait, en cas de décès ou de départ à l'étranger. Cependant, la trace de cet individu persiste dans l'échantillon avec

le détail des événements démographiques qui ont jalonné son parcours sur le territoire, et l'on peut ainsi comparer à tout moment les trajectoires de différentes cohortes.

① PASSER À 16 JOURS POUR S'ADAPTER AU RECENSEMENT CONTINU

Le recensement de la population sert avant tout à établir le nombre d'habitants de chacune des circonscriptions administratives (établissement des populations dites légales), mais c'est aussi une source privilégiée pour décrire la population (sexe et âge, mais aussi diplôme, catégorie sociale, configurations familiales, etc.) : à ce titre, cette source est fondamentale pour l'EDP.

Le recensement a changé de méthode : exhaustif tous les sept à dix ans jusqu'en 1999, il est devenu annuel au début des années deux-mille, et sa collecte repose désormais sur un sondage. Depuis 2004, des enquêtes annuelles de recensement (EAR) sont ainsi menées sur un échantillon d'adresses³. Conséquence pour les individus inclus dans l'EDP : ils ne sont donc plus recensés en même temps ; pour une année donnée, seul un individu sur sept⁴ est présent dans l'échantillon de l'enquête du recensement.

Pour compenser la dégradation de la qualité des estimations du fait de la réduction des données disponibles une même année, la taille de l'échantillon a été considérablement

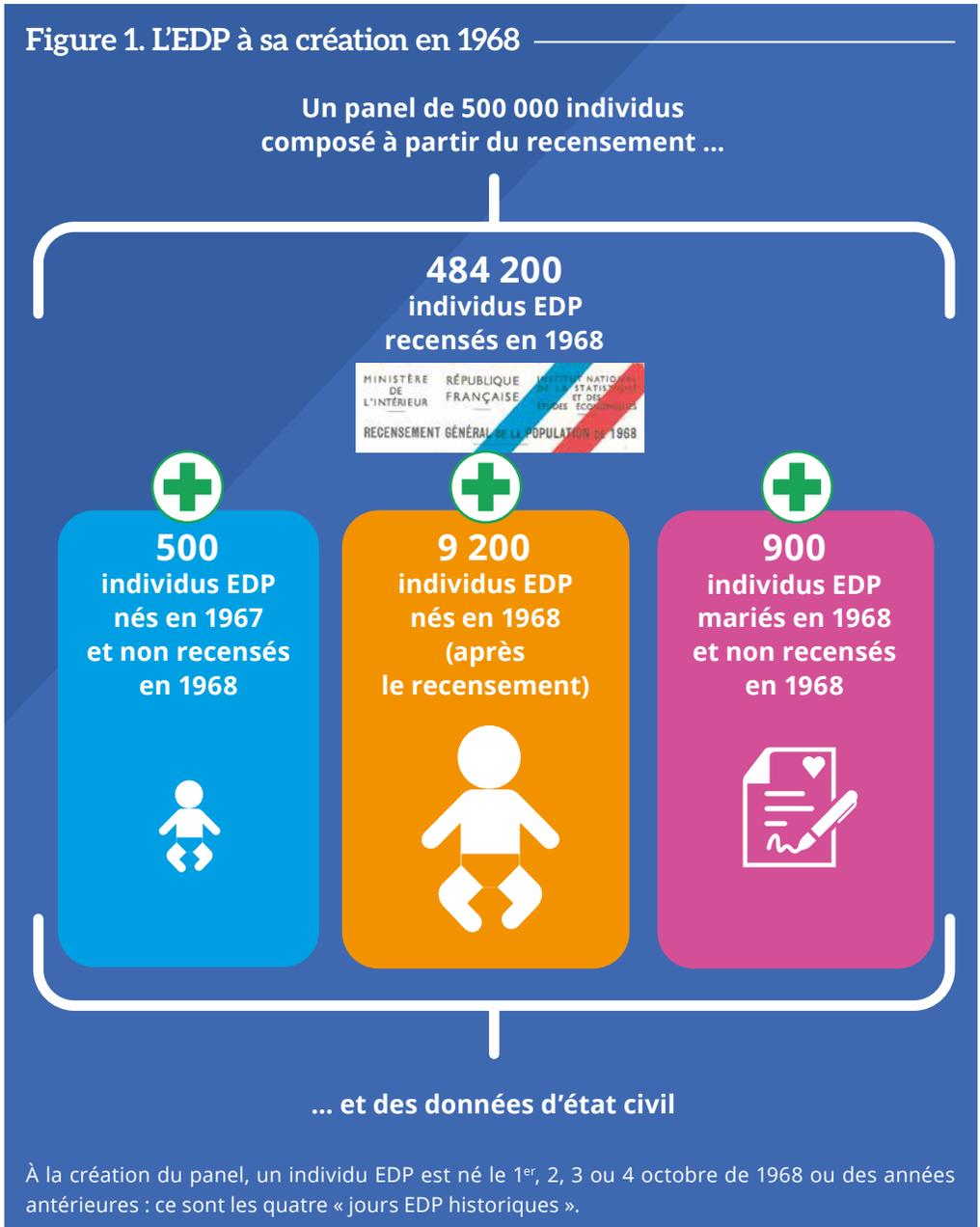
3. Sur une période de cinq années successives, l'ensemble du territoire est couvert par une collecte du recensement. Les premiers résultats avec la nouvelle méthode sont ceux millésimés 2006, qui combinent les enquêtes de 2004 à 2008 (Godinot, 2005).

4. Une petite commune (moins de 10 000 habitants) sur cinq est recensée exhaustivement chaque année, ainsi que 8 % des adresses en grandes communes. Au total, environ un habitant sur sept est recensé une année donnée.

augmentée : le nombre de jours de référence EDP est passé de 4 à 16 (figure 2). Cet élargissement n'a pas été rétroactif, les données nominatives n'étant pas conservées dans les fichiers du recensement.

L'extension des « jours EDP » a volontairement été répartie sur l'année, pour améliorer les analyses de trajectoires pouvant être affectées par la saisonnalité des naissances (Cnis, 2006).

Figure 1. L'EDP à sa création en 1968



Depuis 2008⁵, le suivi du panel EDP porte donc sur les personnes nées les 4 premiers jours de chaque trimestre, avec une subtilité pour éviter le 1^{er} janvier⁶. Les « individus EDP » sont ceux nés un des 16 jours suivants : du 2 et au 5 janvier, du 1^{er} au 4 avril, du 1^{er} au 4 juillet, ou du 1^{er} au 4 octobre (ces 4 derniers jours étant les « jours historiques de l'EDP »).

L'élargissement du panel à 16 jours n'est pas la seule innovation mise en place pour pallier la fin de l'exhaustivité du recensement. L'EDP s'est tournée vers une autre source exhaustive : les données socio-fiscales⁷.

📍 AVEC LES DONNÉES FISCALES, L'EDP RETROUVE SON EXHAUSTIVITÉ...

À l'origine, l'EDP compile des données d'état civil et de recensements de la population. Il s'enrichit en 2008 de données du fichier électoral (dates d'inscription sur les listes électorales, dates de radiation, communes d'inscription), puis du panel « tous salariés » (qui décrit le parcours d'emploi salarié et les rémunérations, depuis 1968) et de données socio-fiscales avec Fidéli (Fichiers démographiques sur les logements et les individus) et Filosofi (Fichier localisé social et fiscal) (**encadré 1**).

L'intégration des données fiscales a redonné à l'EDP l'exhaustivité qu'il avait perdue lors du changement de méthode du recensement de la population. Les données fiscales apportent annuellement pour tous les individus nés un des jours EDP notamment des informations sur :

- 📍 le logement (localisation, caractéristiques du logement) ;
- 📍 et sur la situation familiale (car celle-ci influe sur le taux marginal de l'imposition).

« L'intégration de données socio-fiscales est sans doute la plus grande avancée de l'EDP ces dernières années. »

Cela compense donc le fait que ce type d'informations n'est désormais plus récupérable *via* le recensement une année donnée sur la totalité des individus.

L'intégration de données socio-fiscales est sans doute la plus grande avancée de l'EDP ces dernières années. Ces données administratives exhaustives sont issues des déclarations servant

à établir l'impôt sur le revenu et la taxe d'habitation, complétées par des données sur les prestations sociales. Elles sont utilisées à des fins statistiques et permettent à l'EDP d'enrichir les domaines couverts, grâce à l'introduction du niveau de vie, variable essentielle à de nombreuses analyses socio-économiques.

5. En pratique la date de passage varie selon les sources : 2004 pour l'état civil, 2008 pour les EAR, 2002 pour le panel « tous salariés », 2011 pour les données fiscales (revenus perçus en 2010), rétrospectif depuis le début des années 1990 pour les inscriptions électorales.

6. Le 1^{er} janvier a été exclu car c'est trop fréquemment la date retenue lorsque le jour de naissance est inconnu (Insee, 2019).

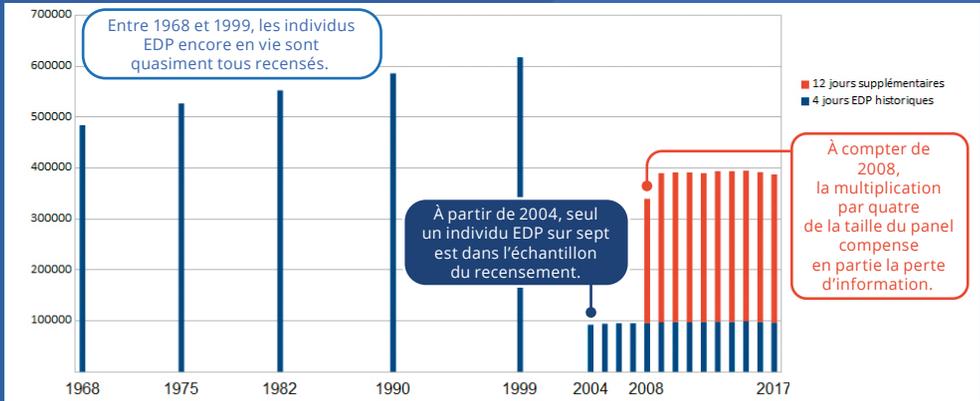
7. L'intégration de données fiscales à l'EDP répondait aussi à des recommandations du Cnis (Chaleix et Lollivier, 2004).

... ET S'ENRICHIT ANNUELLEMENT DE DONNÉES SUR LES REVENUS ET LE NIVEAU DE VIE

Alors que l'exploitation statistique des données fiscales en dehors de l'EDP ne permet un suivi individuel des niveaux de vie que sur deux années successives⁸, l'intégration dans l'EDP de données issues des dispositifs Fidéli et Filosofi offre désormais un suivi sur plus longue période.

Cette panélisation des données fiscales depuis la déclaration fiscale de 2011 (revenus 2010), pour l'échantillon des personnes nées un « jour EDP », ouvre de larges perspectives d'études. Citons à ce titre l'étude récente sur l'évolution du niveau de vie des retraités (Abbas, 2020) : au moment des débats sur la réforme des retraites, cette analyse éclaire non seulement la situation financière des retraités l'année de leur départ à la retraite, mais aussi l'évolution de leur niveau de vie au cours des trois années qui le précèdent et des trois années qui le suivent. Elle met ainsi en lumière la dégradation des conditions

Figure 2. Le passage à 16 jours permet de compenser en partie l'impact du recensement rénové



Nombre d'individus EDP recensés, selon la date de recensement

Lecture : Parmi les 3,7 millions d'individus présents dans l'EDP, 484 200 sont présents au recensement de 1968 et 617 000 sont présents au recensement de 1999. 92 000 sont présents à l'EAR de 2004, effectifs plus réduits du fait du passage de recensement exhaustif à des enquêtes annuelles par sondage. L'extension de l'EDP à 16 jours de naissance au lieu de 4 a permis d'augmenter les effectifs suivis à compter de 2008. Une année donnée, près de 400 000 individus EDP sont recensés dans une EAR. L'effectif est plus réduit en 2008 que pour les années suivantes, la procédure pour identifier les personnes EDP ayant été plus limitée cette année-là. Seules les personnes dont on avait pu retrouver le numéro d'identification au répertoire (NIR) automatiquement au répertoire national d'identification des personnes physiques ont été intégrées, sans traitement complémentaire (Jugnot, 2014).

Source : (Insee, 2019)

8. En dehors d'une étude ponctuelle réalisée avec un suivi de 5 ans (Bonnet, Garbinti et Solaz, 2015).

de vie pour certains en fin de carrière et l'amélioration de leur situation financière après leur départ à la retraite, notamment pour les retraités les moins diplômés. Cette étude inédite n'aurait pas été possible sans l'EDP : elle combine des caractérisations sociales (données du recensement) et la connaissance des ressources au fil des années (données fiscales en panel), disponibles uniquement dans l'EDP.

Les trajectoires individuelles suivies dans l'EDP se sont ainsi étoffées (**encadré 2**). L'introduction des données socio-fiscales exhaustives a aussi permis des études inédites sur la situation des familles après une rupture d'union, rares étant les données permettant de telles approches, comme le souligne le Cnis (Thélot *et alii*, 2016). Il en est ainsi de l'étude sur le logement des familles l'année de la rupture d'une union et dans les années qui la suivent (Durier, 2017). Ou des analyses de l'évolution du niveau de vie après une rupture d'union, qui révèlent une forte baisse en moyenne l'année de la rupture pour les femmes et une « récupération » ensuite dans les années qui suivent (Costemalle, 2017), amélioration qui s'observe surtout pour les parents formant rapidement une nouvelle union (Abbas et Garbinti, 2019).

La force de l'EDP réside aussi dans la taille de son échantillon : l'étude sur le niveau de vie après une séparation a ainsi été déclinée pour la région en Auvergne-Rhône-Alpes (Balouzat et Labosse, 2020), ce qui n'aurait pas été possible avec un échantillon national moins important.

À la fois riche par la diversité des sources qui l'alimentent, la finesse géographique potentielle des analyses du fait de la taille de l'échantillon et la profondeur historique d'un suivi sur plus de 50 ans (**encadré 3**), l'EDP est une source incontournable pour les études en panel : près de 60 équipes de recherche travaillent actuellement à partir de ces données (CASD, 2021). Mais revers de la médaille, cette richesse s'est accompagnée d'une plus grande complexité pour pouvoir exploiter ces données.

CONSTRUIRE SA POPULATION D'INTÉRÊT..

L'EDP permet de croiser les informations individuelles de plusieurs sources : on peut par exemple étudier la mortalité (données d'état civil) selon à la fois le diplôme (recensement), la catégorie sociale (recensement ou panel tous salariés) et le niveau de vie (données fiscales). La richesse des informations va toutefois de pair avec la complexité de l'usage des données : réfléchir à la manière de construire sa population d'intérêt et à la nécessité ou non de pondérer les données est un préalable incontournable avant de mener une étude à partir de l'EDP. Ces pré-requis font que l'EDP s'adresse à des chargés d'études ou chercheurs à l'aise avec l'exploitation des données et ayant de solides compétences statistiques, pour ne pas introduire de biais dans les résultats et analyses.

C'est en combinant des informations des différentes sources intégrées à l'EDP que chaque chargé d'études construit sa population d'intérêt et les données nécessaires à son étude (**encadré 1**). Il doit alors faire appel à différentes tables statistiques de la base études de l'EDP (**figure 3**), reliées entre elles par un identifiant commun (**encadré 4**).

Encadré 1. Quelles informations sont disponibles dans l'EDP? —

1967

2000

2011

PAR ANNÉE, DEPUIS 1968 ET 1967 POUR LA NAISSANCE DES INDIVIDUS EDP



- Bulletin de naissance de l'individu EDP : sexe, date de naissance, lieu de naissance (en France), date de naissance des parents, lieu de naissance des parents
- Bulletin de naissance de ses enfants

- Bulletin de mariage : date et lieu, informations sur les conjoints (date de naissance, lieu de naissance, état matrimonial légal antérieur)
- Bulletin de décès : date, lieu, état matrimonial légal

RECENSEMENTS 1968, 1975, 1982, 1990, 1999 ET ENQUÊTES ANNUELLES DE RECENSEMENT DEPUIS 2004



- Individu : informations sur l'individu EDP et les personnes résidant dans le logement (sexe, date de naissance, vie en couple, situation conjugale légale ou de fait, diplôme, lieu d'étude, CS, activité professionnelle, situation sur le marché du travail, lieu de travail, moyen de transport etc.), lieu de résidence antérieur

- Famille : situation familiale de individu EDP
- Ménage-logement : année d'emménagement, nombre de pièces, nombre d'habitants du logement, statut d'occupation, confort, région, département, commune, zone infra-communale

NB :

- Les informations intégrées peuvent être différentes selon les années.
- Dom depuis 2004



PANEL TOUS SALARIÉS DEPUIS 1967

Activité salariée, agrégée par année : salaire, nombre d'heures, CS, lieu de résidence, lieu de travail, condition d'emploi, activité économique

NB :

- Champ variable selon les années : introduction de la fonction publique depuis 1988, Dom et secteur agricole depuis 2002, particuliers employeurs depuis 2009
- Prochainement le panel non salariés viendra compléter les données



INSCRIPTIONS SUR LES LISTES ÉLECTORALES APPROXIMATIF AVANT LES ANNÉES 2000, FIABLE DEPUIS

Commune d'inscription, date d'inscription, date de radiation

Note : Des données peuvent ne pas être disponibles certaines années ou pour certains individus EDP, notamment pour les années antérieures à l'extension de l'EDP à 16 jours, ou pour d'autres raisons (restriction de traitements des données d'état civil des individus EDP à certains jours de naissance par exemple dans les années 1990 pour certaines données). Le schéma ci-dessus offre un rapide panorama des informations contenues dans l'EDP. La consultation de la documentation détaillée est indispensable pour tout projet d'exploitation.



Source : (Insee, 2019).



REVENUS ET COMPOSITION DES FOYERS FISCAUX ET MÉNAGES, PAR ANNÉE

Depuis 2011 (revenus 2010)

- Individu : informations sur l'individu EDP et ceux des logements où cet individu est connu au titre de l'impôt sur le revenu ou de la taxe d'habitation (date de naissance, situation conjugale, département commune de l'adresse fiscale, type de déclaration fiscale)
- Logement : informations sur le logement au sens de la taxe d'habitation principale (nombre de pièces, confort du logement, année d'emménagement, statut d'occupation)
- Revenus (Filosofi) : niveau de vie du ménage, composition du revenu, nombre de personnes
- Revenus individuels : revenu déclaré par l'individu EDP

Par exemple, pour estimer les espérances de vie par niveau de vie, catégorie sociale et, diplôme, étude réalisée pour la première fois en 2018 grâce à l'intégration des données fiscales dans l'EDP, (Blanpain, 2018a), il a fallu sélectionner des personnes présentes au recensement (diplôme, CS), pour lesquelles on a recherché des données de l'état civil (état vital) et des données fiscales (niveau de vie). La constitution de sa base d'études a nécessité des expertises préalables. L'auteur a comparé l'ampleur des différentiels sociaux de mortalité (Blanpain, 2016) selon que l'on retient la catégorie sociale d'après le recensement ou d'après le panel tous salariés (Costemalle, 2016). Il a imputé un niveau de vie à partir de variables de revenus lorsque l'information sur le niveau de vie n'était pas disponible. Ceci est en effet le cas pour des personnes ne résidant pas en logement ordinaire par exemple, qui sont souvent des personnes âgées, et pouvait donc avoir un impact sur la mesure de la mortalité par niveau de vie (Blanpain, 2018b). L'auteur a également comparé sa population cible à d'autres données pour vérifier que la population d'intérêt qu'elle avait sélectionnée dans l'EDP était bien représentative de l'ensemble de la population étudiée (comparaison des espérances de vie estimées avec l'EDP à celles issues des bilans démographiques (Blanpain, 2018b)) et pouvoir recalculer si besoin la population sélectionnée.

Une fois la population sélectionnée dans l'EDP, il faut se demander si elle représente bien la population générale, et se poser donc la question des pondérations.

... SAVOIR PONDÉRER

Le passage du recensement exhaustif aux enquêtes de recensement a introduit une nouvelle pratique pour les études menées à partir de l'EDP : les pondérations. La sélection sur les jours de naissance adoptée dans l'EDP ne biaise pas les analyses (aux limites mentionnées *supra*). Seul un facteur d'échelle était parfois utilisé pour donner des ordres de grandeur des effectifs concernés (par exemple en multipliant les effectifs concernés dans l'EDP par 365/4 ou 365/16 selon les années concernées). Mais estimer des répartitions ou des coefficients de modèles économétriques ne nécessitait pas toujours l'usage de poids : tous les individus avaient le même poids.

Or depuis 2004, le tirage de l'échantillon des enquêtes annuelles de recensement (EAR) marque une différence entre les petites et les grandes communes : ceci rend indispensable l'usage des pondérations associées aux EAR dans l'EDP dès lors qu'elles interviennent dans la définition de la population d'intérêt. Des variables de pondérations ont ainsi fait leur apparition dans l'EDP avec la fin de l'exhaustivité du recensement.

Mais on peut aussi utiliser les EAR sans utiliser ces pondérations, si par exemple cette source sert uniquement à compléter d'autres données en ajoutant des variables complémentaires notamment.

Ainsi, dans l'étude de l'évolution du niveau de vie des personnes parties à la retraite en 2013 (Abbas, 2020), une analyse est menée par diplôme. La population d'intérêt a été définie à partir des données fiscales et de la déclaration de revenus sous forme de pensions. Les informations pour ces personnes ont été complétées par le niveau de diplôme retrouvé dans une des EAR disponibles dans l'EDP. Il n'y a alors pas de raison de prendre en compte dans ce cas les pondérations des EAR, une fois vérifié que la population pour laquelle on a pu associer un niveau de diplôme ne diffère pas de la population cible totale (âge, niveau de vie, etc.).

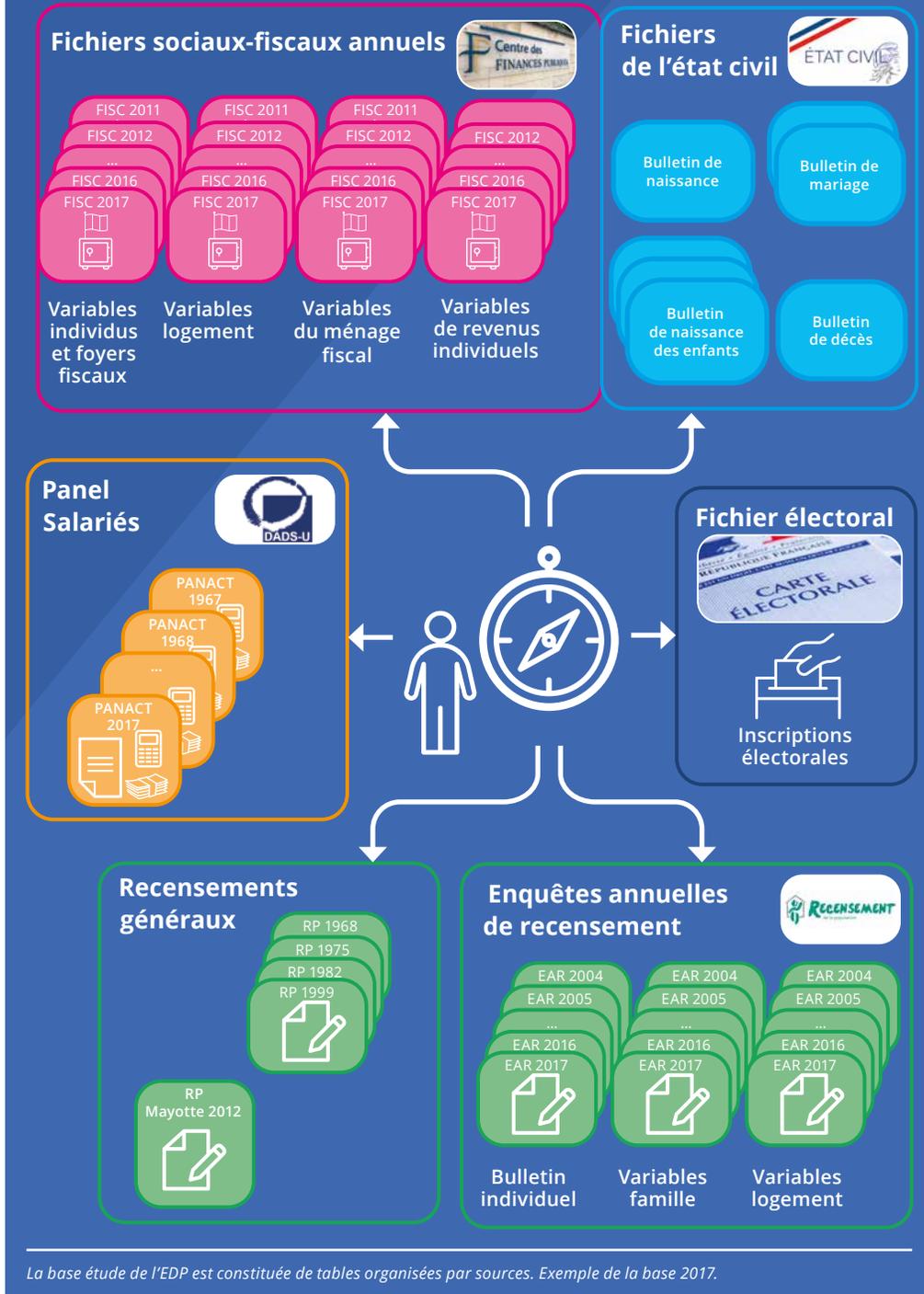
En contrepartie de la richesse des données contenues dans l'EDP, son usage s'est donc complexifié. C'est pourquoi un groupe d'exploitation a été créé en 2015 par l'Insee. Il réunit des chargés d'études et des chercheurs, afin d'échanger sur les nouveautés introduites dans l'EDP et sur les travaux réalisés à partir de ce panel.

Encadré 2. Un exemple fictif de trajectoire individuelle

Source	Informations obtenues
État civil	Un garçon naît un jour EDP en 1966 à Caen dans le Calvados, il est de parents résidant à Courseulles-sur-Mer (Calvados). On connaît leur année de naissance et leur date de mariage. 
Recensement général de 1968	En 1968, il vit toujours à Courseulles-sur-Mer. 
Recensement général de 1975	En 1975, toujours à Courseulles-sur-Mer, dans une maison de 5 pièces avec ses parents et trois autres enfants âgés de 7 à 16 ans (on ne connaît pas leur âge exact). 
Recensement général de 1982	En 1982, il réside toujours à Courseulles-sur-Mer. 
Fichier électoral	En 1984, à 18 ans, il s'inscrit sur les listes électorales à Courseulles-sur-Mer. 
Recensement général de 1990	En 1990, il vit dans une maison de 2 pièces en location à Hérouville-Saint-Clair (Calvados). On connaît son diplôme et sa profession. Il est célibataire (état matrimonial) et vit en couple avec une femme née en 1968 dans le Calvados. 
Fichier électoral	Il se ré-inscrit sur les listes électorales de Hérouville-Saint-Clair (on connaît la date). 
État civil	Il a deux jumelles (on connaît leurs date et commune de naissance ainsi que celles de leur mère et la commune de résidence). Les parents ne sont pas mariés. 
État civil	Il se marie (dates du mariage et de naissance de sa conjointe, commune de résidence). 
Recensement général de 1999	En 1999, il est propriétaire d'une maison de 7 pièces à Hérouville-Saint-Clair et réside dans ce logement depuis 1996. Il y vit avec sa femme et ses filles. On connaît son diplôme, la commune de son lieu de travail et sa profession. 
Panel salariés	En 2002 (et jusqu'en 2012), il est employé dans la fonction publique hospitalière de Caen. On connaît son salaire brut. 
Enquête annuelle de recensement	En 2008, il réside à Hérouville-Saint-Clair. On connaît sa date d'emménagement. Il vit avec sa femme et trois enfants. On connaît leur date et lieu de naissance, leur emploi, diplôme et commune du lieu de travail. 
Enquête annuelle de recensement	En 2013, divorcé, il vit à Hérouville-Saint-Clair dans une maison de 2 pièces en couple avec une femme divorcée et un enfant. On connaît leur date et lieu de naissance, diplôme, profession et commune du lieu de travail. 
Fichiers socio-fiscaux Fideli et Filsofi	Divorcé (date du divorce), il a sa résidence fiscale de 2011 à 2013 à Hérouville-Saint-Clair et réside avec une femme (date et le lieu de naissance). Ils ont un enfant à charge (année de naissance). Il verse une pension alimentaire. On connaît son niveau de vie. 
État Civil	Il décède le 12 décembre 2013 à Caen. 

Note : la finalité de l'EDP n'est pas de suivre un individu, mais d'établir des statistiques sur les trajectoires d'un groupe d'individus définis par des caractéristiques socio-démographiques pour les analyser au regard d'autres groupes par exemple.

Figure 3. En s'enrichissant, l'EDP rend son utilisation plus complexe



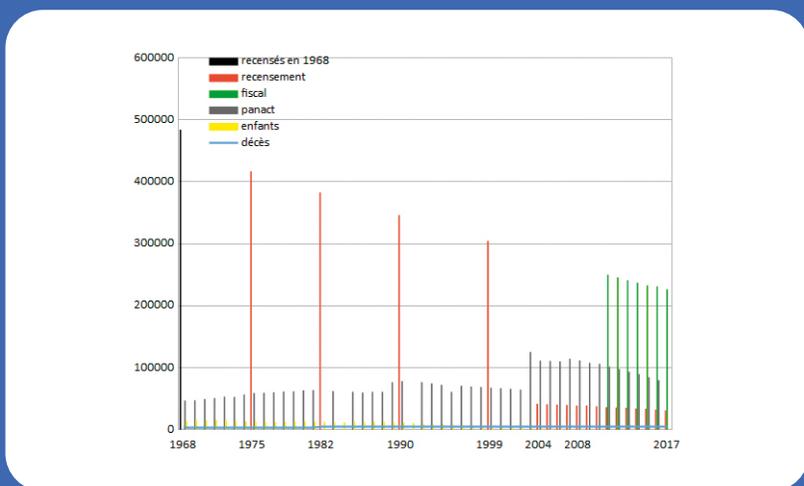
La base étude de l'EDP est constituée de tables organisées par sources. Exemple de la base 2017.

Encadré 3. Que sont devenues les personnes recensées en 1968? —

484 200 personnes nées un 1^{er}, 2, 3 ou 4 octobre, quelle que soit leur année de naissance, et recensées en 1968 sont suivies dans l'EDP. Au fil des années, elles sont présentes sur le marché du travail, ont des enfants, sont recensées, et décèdent pour une partie :

- pour 416 900 (soit 86 %) d'entre elles, on dispose d'informations statistiques les concernant au recensement de 1975 ;
- pour 346 200 personnes, on dispose de données du recensement de 1968 et de données du recensement de 1999, pour décrire leurs trajectoires sur 30 ans donc ;
- le nombre de personnes retrouvées aux recensements post-1968 décroît avec les années, car il y a des décès et de potentielles migrations ;
- il baisse fortement avec le passage des recensements exhaustifs aux enquêtes annuelles de recensement, du fait d'un recueil des données désormais sur un échantillon de la population : ainsi on dispose pour 41 600 individus du panel EDP recensés en 1968 d'informations les concernant à l'enquête annuelle de recensement (EAR) de 2004, et pour 31 400 de données de l'EAR de 2017, année la plus récente intégrée à l'EDP à ce jour. Mais on retrouve beaucoup plus de personnes avec des informations issues des données fiscales : **226 700 personnes recensées en 1968 ont aussi dans l'EDP des informations statistiques les concernant dans les données fiscales de 2017, avec un suivi statistique sur près de 50 ans.**

L'approche par cohorte peut aussi se faire par année de naissance. **On peut ainsi suivre de la même manière le devenir des 9 200 personnes nées les 1^{er}, 2, 3 ou 4 octobre 1968 à partir des événements retracés dans l'EDP.**



Champ : individus nés entre les 1^{er} et 4 octobre (quelle que soit l'année) et dont on dispose dans l'EDP d'informations statistiques dans le recensement de 1968 - Source : (Insee, 2019).

Note : l'effectif des naissances des enfants dont au moins un parent est né un jour EDP et a été recensé en 1968 est estimé en multipliant par 2 les effectifs pour les parents EDP nés un 1^{er} ou un 4 octobre, pour tenir compte de trous de collecte (Durier, 2018).

UN ACCÈS AUX DONNÉES TRÈS ENCADRÉ

On dispose dans l'EDP d'une information de plus en plus précise sur les personnes qui composent le panel (*encadré 1*), mais pas de données directement identifiantes. Cependant, la répétition d'informations dans le temps rend ces données plus sensibles au risque de non-respect des critères d'anonymisation. En effet, si on connaît une personne dont la date de naissance correspond à un jour EDP, le croisement de ses autres caractéristiques connues avec les informations contenues dans le panel pourrait conduire à l'identifier dans le panel, selon une probabilité plus forte lorsque les informations sont répétées dans le temps que si on dispose uniquement de caractéristiques à une date donnée. Il y a donc un risque

« Pour les chercheurs, l'accès se fait par l'entremise du CASD. »

d'apprendre plus d'informations sur cette personne que celles que l'on connaît déjà, et donc d'atteinte potentielle à la confidentialité. C'est pour cette raison que la constitution et l'accès aux données de l'EDP sont très réglementés (*encadré 5*).

À l'Insee, les chargés d'études peuvent accéder aux données dans un espace dédié après demande nominative, et ils exploitent les données dans un autre espace, dédié aux traitements.

Pour les chercheurs, l'accès se fait par l'entremise du CASD, Centre d'accès sécurisé aux données (Gadouche, 2019), après avis favorable du comité du secret sur leur projet. Ce mode d'accès, en vigueur depuis 2010, a permis de développer les exploitations de l'EDP, sur des thématiques aussi différentes que les inégalités territoriales, la mobilité géographique des immigrés en France, les parcours professionnels et les transitions de carrières entre secteurs public et privé.

Pour les agents des services statistiques ministériels, la situation varie selon les conditions de sécurité pour les accès aux données au sein du SSM. Au besoin, ils peuvent aussi recourir à l'intermédiaire du CASD.

Encadré 4. Un suivi en panel exigeant un identifiant unique et invariant

Depuis son origine, le suivi longitudinal des personnes présentes dans l'EDP s'est appuyé sur le NIR, numéro d'identification au Répertoire national d'identification des personnes physiques. Cet identifiant est unique et invariant*. Les personnes appartenant à l'échantillon suivi dans l'EDP sont dans un premier temps identifiées à partir de leurs traits d'identité (nom, prénom, sexe, date et lieu de naissance (Jugnot, 2014)) : il s'agit, à partir de ces informations, de retrouver leur NIR, pour ensuite enrichir leurs trajectoires dans le panel. Pour des sources, comme le panel « tous salariés », qui contiennent déjà le NIR, cette procédure d'identification n'est évidemment pas nécessaire*. **Le NIR n'est utilisé qu'à des fins de production de l'EDP, et les noms et prénoms ne sont pas conservés une fois l'identification réalisée.**

Les fichiers mis à disposition des chargés d'études et des chercheurs à des fins statistiques ne contiennent pas le NIR, mais uniquement un identifiant de diffusion non signifiant (qui n'apporte donc pas d'information sur la personne). Cet identifiant de diffusion leur permet ainsi de réaliser les appariements entre les différentes bases de l'EDP.

* En toute rigueur, le NIR peut être modifié à de très rares exceptions (changement de genre par exemple, qui modifiera alors le premier chiffre du NIR).

DES PROJETS EN COURS ET À VENIR

L'intégration des données socio-fiscales a permis de développer des études sur un champ nouveau : l'évolution du niveau de vie suite à un événement, et dans les années qui précèdent ou suivent cet événement. Elle a suscité l'intérêt des chercheurs en économie ou sur les familles : citons par exemple le projet *Big_Stat*, des données statistiques massives pour observer une société mobile (Ined, 2021). Elle a également élargi la possibilité d'analyses régionales (Lacour, 2018, Bertaux *et alii*, 2019, Balouzat et Labosse, 2020, Dherbécourt et Kenedi, 2020).

L'élargissement des sources de l'EDP va se poursuivre prochainement, avec le panel non salariés, pour couvrir de plus larges trajectoires d'emploi : l'EDP comprendra des données annuelles sur l'activité salariée et sur l'activité non salariée, toujours sur longue période, permettant ainsi d'analyser les trajectoires entre différents types d'emploi au fil de la carrière, en lien avec les caractéristiques socio-démographiques des individus (diplôme, famille, etc.). Par le cumul d'informations depuis maintenant plus de 50 ans, l'EDP permet déjà de suivre des trajectoires longues, combinant notamment les parcours de vies familiale et professionnelle⁹.

L'EDP s'ouvre également à un nouveau domaine, dans un cadre juridique spécifique : celui de la santé. En 2019, la Drees a apporté à l'EDP les informations du système national des données de santé (SNDS) : cette source, gérée par la Caisse nationale de l'assurance maladie comprend notamment les consommations de soins et les causes médicales de décès, sur 10 ans mais sans informations ni sur les revenus, les milieux sociaux ou les situations professionnelles et familiales. Les données de santé faisant l'objet de procédures spécifiques, ce traitement, autorisé par la Cnil, est limité dans le temps (5 ans) (Drees, 2020). Il s'inscrit dans une finalité délimitée, pour évaluer la stratégie nationale de santé 2018-2022. Il permet de répondre à des questions sur l'évaluation des inégalités sociales de santé, et de compléter les analyses menées avec les données de santé en panel, mais ne comprenant que peu de descripteurs sociaux. La première publication de la Drees avec l'« EDP-santé », offre ainsi un complément d'analyse sur l'interruption volontaire de grossesse, sur la fréquence des IVG selon le niveau de vie (Vilain *et alii*, 2020).

Enfin, l'émergence de nouvelles sources sur les carrières professionnelles et les motifs d'interruption¹⁰ offrent des pistes qui ne demandent qu'à être explorées.

Encadré 5. Le cadre juridique de l'EDP en quelques mots

L'EDP est un traitement de données à caractère personnel mis en œuvre en conformité avec le Règlement général sur la protection des données et la loi Informatique et Liberté. À ce titre, il fait l'objet de règles et de mesures strictes garantissant la sécurité et la confidentialité des données.

Toute personne ayant accès aux données est astreinte au secret statistique. Les chargés d'études et les chercheurs peuvent obtenir communication des données, mais après avis du Comité du secret statistique, en application des dispositions de la loi n°51-711 du 7 juin 1951 sur l'Obligation, la coordination et le secret en matière de statistiques. Les données identifiantes, notamment le NIR, ne sont cependant pas communicables dans ce cadre.

9. Des études longitudinales sur les parcours professionnels peuvent être menées à partir de panels sur les salariés ou les non salariés. Mais l'EDP intègre en plus des données socio-démographiques, sur les situations familiales par exemple.
10. Voir l'article de Christian Sureau et Richard Merlen sur le Répertoire général des carrières unique, dans ce même numéro.

BIBLIOGRAPHIE

ABBAS, Hicham et GARBINTI, Bertrand, 2019. De la rupture conjugale à une éventuelle remise en couple : l'évolution des niveaux de vie des familles monoparentales entre 2010 et 2015. In : *France, portrait social, édition 2019*. [en ligne]. 19 novembre 2019. Insee. Pp. 99-113. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/4238447/FPS2019_D1.pdf.

ABBAS, Hicham, 2020. *Des évolutions du niveau de vie contrastées au moment du départ à la retraite*. [en ligne]. 12 février 2020. Insee Première, n°1792. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4308750>.

BALOUZAT, Bruno, LABOSSE, Aline, 2020. *Lors d'une séparation, les femmes basculent plus souvent dans la pauvreté que leur conjoint*. [en ligne]. Octobre 2020. Insee Analyses Auvergne-Rhône-Alpes, n° 103. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4774341>.

BERTAUX, Frédéric, BOUSSAD, Nadia et SAGOT, Mariette, 2019. *En quinze ans, la moitié des Franciliens résidant dans des espaces « pauvres » ont changé de commune*. [en ligne]. 30 septembre 2019. Insee Analyses Île-de-France, n° 104. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4218834>.

BLANPAIN, Nathalie, 2016. *L'espérance de vie par catégorie sociale et par diplôme – Méthode et principaux résultats*. [en ligne]. 18 février 2016. Insee. Documents de travail, Direction des Statistiques Démographiques et Sociales, n° F1602. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/2022138>.

BLANPAIN, Nathalie, 2018a. *L'espérance de vie par niveau de vie : chez les hommes, 13 ans d'écart entre les plus aisés et les plus modestes*. [en ligne]. 6 février 2018. Insee Première, n°1687. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/3319895>.

BLANPAIN, Nathalie, 2018b. *L'espérance de vie par niveau de vie – Méthode et principaux résultats*. [en ligne]. 6 février 2018. Documents de travail, Direction des Statistiques Démographiques et Sociales, n° F1801. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/3322051>.

BONNET, Carole, GARBINTI, Bertrand et SOLAZ, Anne, 2015. Les variations de niveau de vie des hommes et des femmes à la suite d'un divorce ou d'une rupture de Pacs. In : *Couples et familles*. [en ligne]. 16 décembre 2015. Insee Références, pp. 51-61. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/2017528/COUFAM15.pdf>.

CASD, 2021. EDP : Échantillon Démographique Permanent. In : *site du Centre d'accès sécurisé aux données*. [en ligne]. Les sources de données déjà disponibles au CASD. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.casd.eu/source/echantillon-demographique-permanent/>.

CHALEIX, Mylène et LOLLIVIER, Stéfan, 2004. *Outils de suivi des trajectoires des personnes en matière sociale et d'emploi*. [en ligne]. Juin 2004. Cnis, Mission Panels, note n° 98/B010, class. 1.5.91. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.cnis.fr/wp-content/uploads/2018/03/RAP_2004_98_outils_trajectoires_personnes_emploi.pdf.

CNIS, 2006. *Dynamique et trajectoires. Compte rendu de la séance du 3 avril 2006*. [en ligne]. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.cnis.fr/wp-content/uploads/2018/01/DP_R_2006_6e_reunion_GT_inegalites_dynamique_territoire.pdf.

COSTEMALLE, Vianney, 2016. *Catégorie sociale d'après les déclarations annuelles de données sociales et catégorie sociale d'après le recensement : quels effets sur les espérances de vie par catégorie sociale ?* [en ligne]. 18 février 2016. Insee. Documents de travail, Direction des Statistiques Démographiques et Sociales, n° F1603. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/2022136>.

COSTEMALLE, Vianney, 2017. Formations et ruptures d'unions : quelles sont les spécificités des unions libres ? In : *France, portrait social, édition 2017*. [en ligne]. 21 novembre 2017. Insee. Pp. 95-111. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/3197289/FPORSOC17.pdf>.

COUET, Christine, 2006. L'échantillon démographique permanent de l'Insee. In : *Courrier des statistiques*. [en ligne]. Insee. N° 117-119, pp. 5-14. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.insee.fr/fr/metadonnees/source/fichier/echantillon_demograp_permanent_courrier_stat117_119.pdf.

DHERBÉCOURT, Clément et KENEDI, Gustave, 2020. *Quelle influence du lieu d'origine sur le niveau de vie ?* [en ligne]. 12 juin 2020. France Stratégie. La note d'analyse, n°91. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/fs-2020-na91-niveau-territoire-juin.pdf>.

DREES, 2020. *L'EDP-Santé : enrichissement de l'échantillon démographique permanent par les données du système national des données de santé (SNDS)*. [en ligne]. 8 juillet 2020. Direction de la recherche, des études, de l'évaluation et des statistiques. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/ledp-sante-enrichissement-de-lechantillon-demographique-permanent-par>.

DURIER, Sébastien, 2017. *Après une rupture d'union, l'homme reste plus souvent dans le logement conjugal*. [en ligne]. 17 juillet 2017. Insee focus n°91. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/2896525>.

DURIER, Sébastien, 2018. Une nouvelle source de données sur la famille : l'EDP enrichi de données socio-fiscales. In : *Observer, décrire et analyser les structures familiales*. [en ligne]. Édité par Nicolas Cauchi-Duval. Association internationale des démographes de langue française. Pp. 5-15. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.erudit.org/fr/livres/actes-des-colloques-de-lassociation-internationale-des-demographes-de-langue-francaise/volume-5-observer-decrire-et-analyser-les-structures-familiales/000364li.pdf>.

GADOUCHE, Kamel, 2019. Le Centre d'accès sécurisé aux données (CASD), un service pour la data science et la recherche scientifique. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 76-92. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4254227/courstat-3-7.pdf>.

GODINOT, Alain, 2005. *Pour comprendre le recensement de la population*. [en ligne]. Insee Méthodes hors série mai 2005. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/information/2579979>.

INED, 2021. Big_Stat. In : *site de l'Ined*. [en ligne]. Institut national des études démographiques. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://big-stat.site.ined.fr/>.

INSEE, 2019. *Base études 2017 de l'Échantillon Démographique Permanent, manuel de l'utilisateur*. [en ligne]. 25 avril 2019. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://utiledp.site.ined.fr/fichier/s_rubrique/26440/manuel_edp_be2017.fr.pdf.

JUGNOT, Stéphane, 2014. *La constitution de l'échantillon démographique permanent de 1968 à 2012*. [en ligne]. 19 septembre 2014. Insee. Documents de travail, Direction des Statistiques Démographiques et Sociales, n° F1406. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/1381113/F1406.pdf>.

LACOUR, Cédric, 2018. *Les séparations : un choc financier, surtout pour les femmes*. [en ligne]. 16 octobre 2018. Insee Analyses Nouvelle-Aquitaine n° 64. Institut national des études démographiques. [Consulté le 30 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/3631116>.

SAUTORY, Olivier, 1988. Plus de la moitié de la population a changé au moins une fois de commune en vingt ans. In : *Économie et statistique*. [en ligne]. Avril 1988. Insee. N° 209, pp. 39-47. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.persee.fr/doc/estat_0336-1454_1988_num_209_1_5197.

THÉLOT, Claude, BOURREAU-DUBOIS, Cécile, CHAMBAZ, Christine, 2016. *Les ruptures familiales et leurs conséquences : 30 recommandations pour en améliorer la connaissance*. [en ligne]. Mars 2016. Cnis, rapport de groupe de travail. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.cnis.fr/wp-content/uploads/2017/10/RAPPORT-RUPTURES-FAMILIALES_-nouvelle-version-29mai2017.pdf.

VILAIN, Annick, ALLAIN, Samuel, DUBOST, Claire-Lise, FRESSON, Jeanne et REY, Sylvie, 2020. *Interruptions volontaires de grossesse : une hausse confirmée en 2019*. [en ligne]. Septembre 2020. Drees. Études et Résultats, n°1163. [Consulté le 30 mai 2021]. Disponible à l'adresse : https://www.repere.re/fileadmin/user_upload/DREES_ivg_2019.pdf.

FONDEMENTS JURIDIQUES

Arrêté du 6 août 2014 portant création d'un traitement automatisé de données à caractère personnel relatif à l'échantillon démographique permanent de l'INSEE. In : *site de Légifrance*. [en ligne]. [Consulté le 30 mai 2021]. Disponible à l'adresse : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000029556902>.