

Présentation générale de l'EDP

L'échantillon démographique permanent : Un panel de données multi-sources

Natacha Gualbert

Un panel d'individus

L'échantillon démographique permanent (EDP) est un panel d'individus, sélectionnés (échantillonnés) par leur jour de naissance dans l'année

=> appartiennent potentiellement à l'EDP tous les individus nés un des 16 « jours EDP » (2-5 janvier et 1-4 des mois d'avril, juillet et octobre)

=> quels que soient le lieu et l'année de naissance

Dans la population générale, environ 4,4 % (16/365) des individus sont nés un jour EDP

Un panel avec un suivi sur longue période

- Lancement du dispositif en 1968

- Enrichissement chaque année :

 - => d'informations nouvelles sur les individus déjà suivis

 - => de nouveaux individus (naissances, migrations)

- Les trajectoires sont interrompues définitivement ou provisoirement suite à des décès ou à des migrations

Un panel avec un suivi sur longue période

•L'EDP propose actuellement des trajectoires pour environ 3,7 millions d'individus (vivants ou décédés)

•Pour plus de 200 000 individus, le suivi s'étend sur 50 ans

•Dès sa création, l'EDP mobilise des données administratives et des données du recensement

=> dispositif économe car sans coût supplémentaire de collecte ni de surcroît de charge de réponse pour les enquêtés, ce qui assure sa pérennité

=> pas d'attrition ni de biais de mémoire

Un panel multi-sources

• Actuellement, l'EDP compile pour les individus retenus des informations issues de 5 sources de données statistiques préexistantes :

• 1) les recensements généraux de la population (RP) de 1968 à 1999 puis les Enquêtes Annuelles de Recensement (EAR) depuis 2004

• 2) les bulletins d'état-civil (naissance, mariage, décès) depuis 1968

• 3) le fichier électoral de 1990 à 2018 puis (intégration en 2020) le répertoire électoral unique (REU) qui l'a remplacé

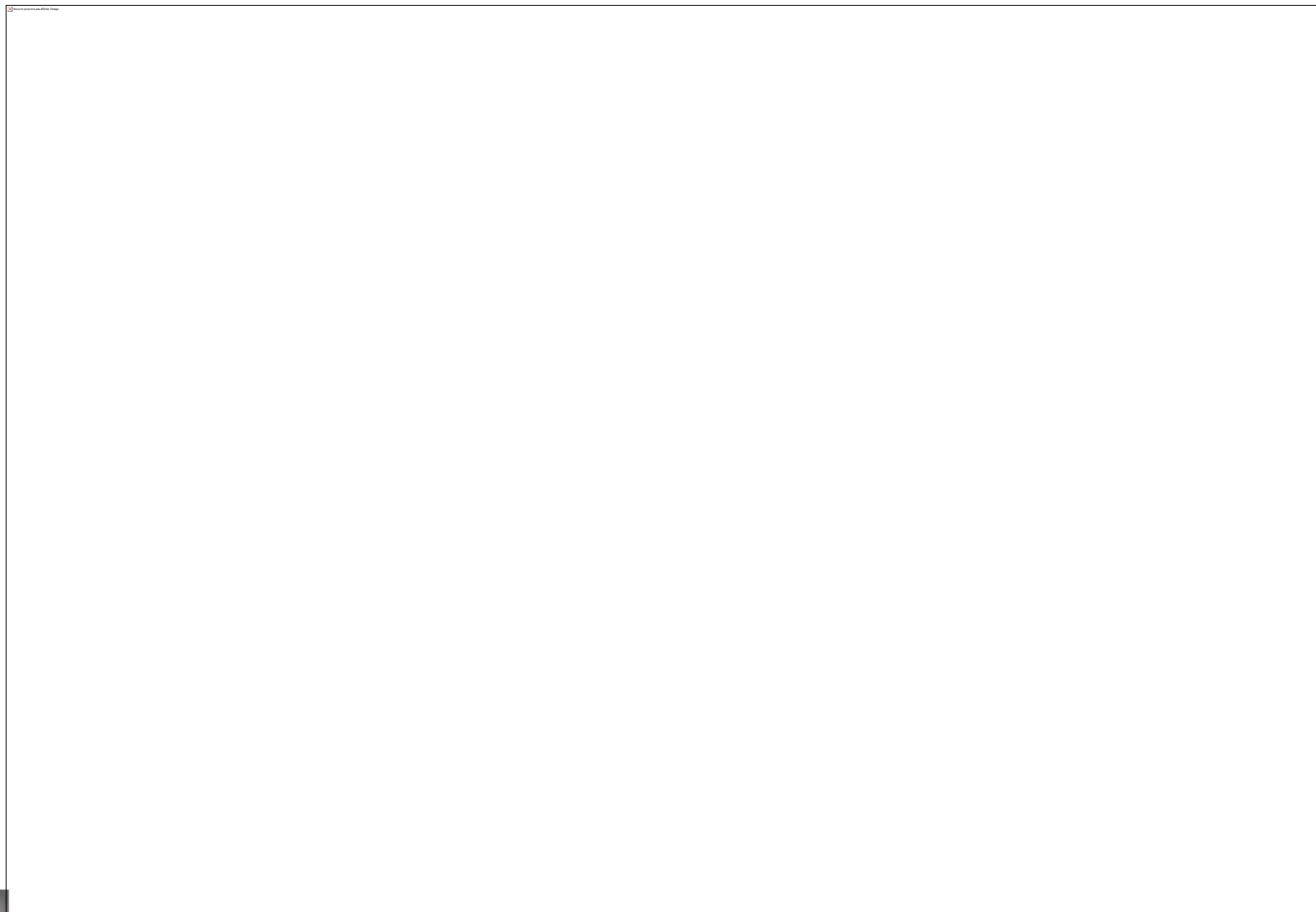
Un panel multi-sources

4) le panel « tous salariés » issu des DADS/FP/PE depuis 1968

5) les données socio-fiscales (déclarations de revenus et prestations sociales) FIDELI + FILOSOFI depuis 2011 (revenus 2010)

L'EDP est une source riche et précieuse 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 (séparation, **passage à la retraite**)

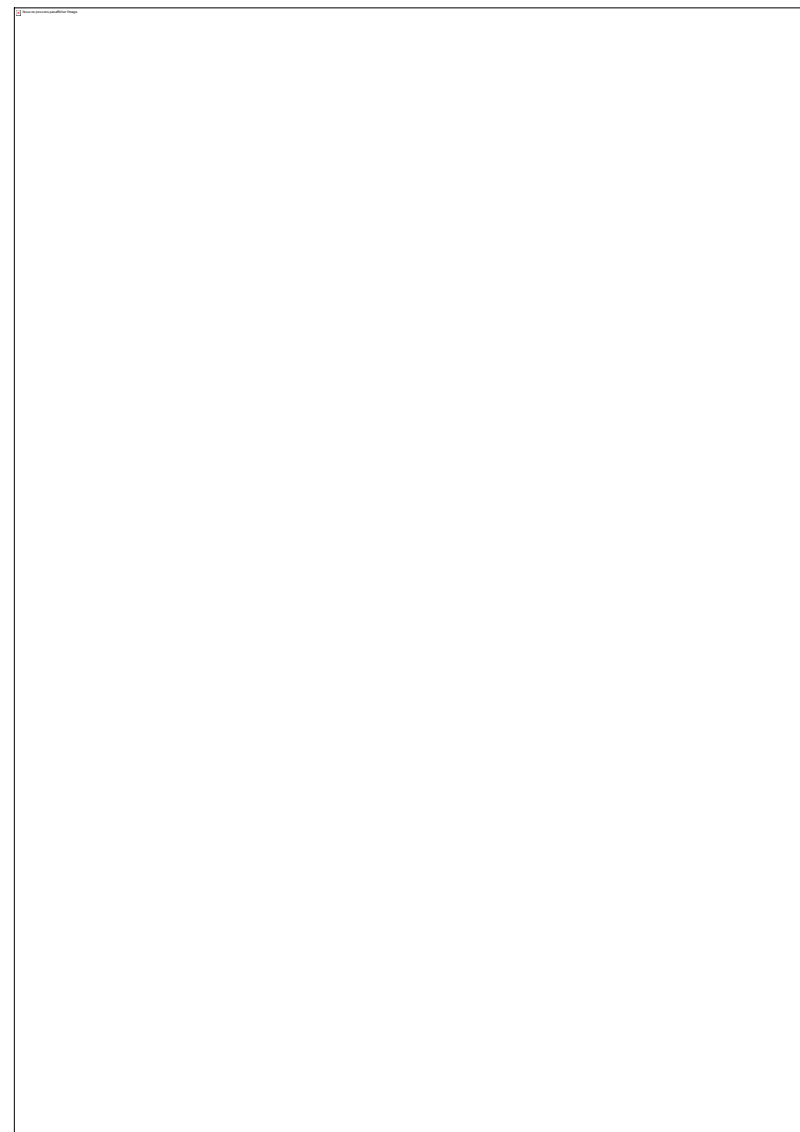
Les sources de l'EDP : l'état civil



Les sources de l'EDP : l'état civil

- Cette source permet un suivi de la fécondité, de la nuptialité et de la mortalité aux travers des bulletins de naissance, décès et mariage.
- Le champ est complet (yc les Doms) et sans “trous” depuis 2004 pour les naissances et mariages, et 2008 pour les décès

Les sources de l'EDP : le fichier électoral



Les sources de l'EDP : le fichier électoral

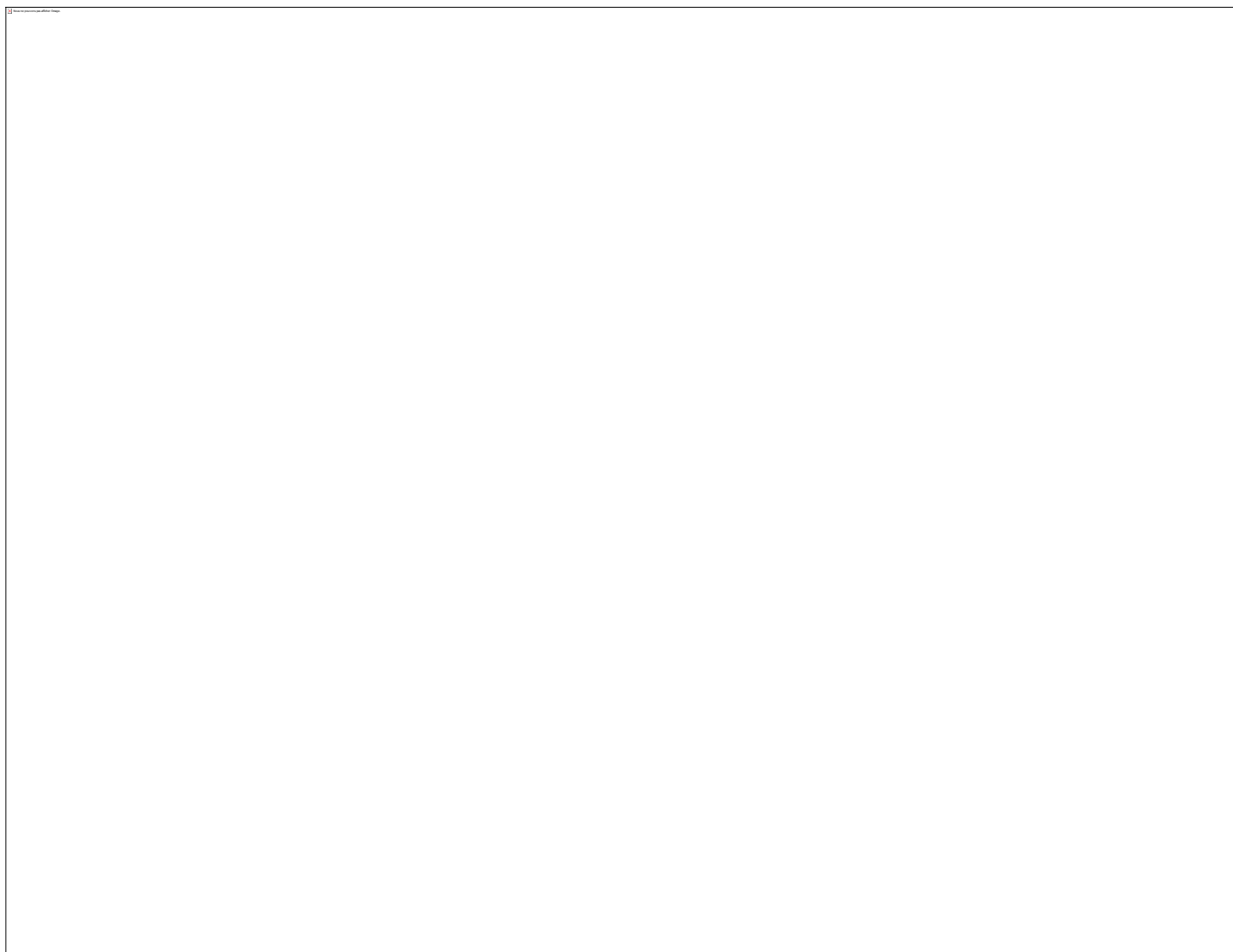
Pour chaque inscription au fichier électoral, on dispose :

- de la date de l'inscription sur les listes électorales
- de la commune d'inscription
- de son type (volontaire ou d'office)
- du type de liste électorale (liste principale, complémentaire municipale, complémentaire européenne)
- de la date de radiation éventuelle (déménagement, incapacité). Cette variable est bien remplie à partir de 2010 et correctement depuis les années 2000

A partir du passage au REU qui remplace le fichier électoral (intégré en 2022 pour BE 2020) :

- le fichier électoral s'arrête (l'historique est conservé) et l'actualisation de la situation électorale se fera par le répertoire électoral unique

Les sources de l'EDP : le recensement



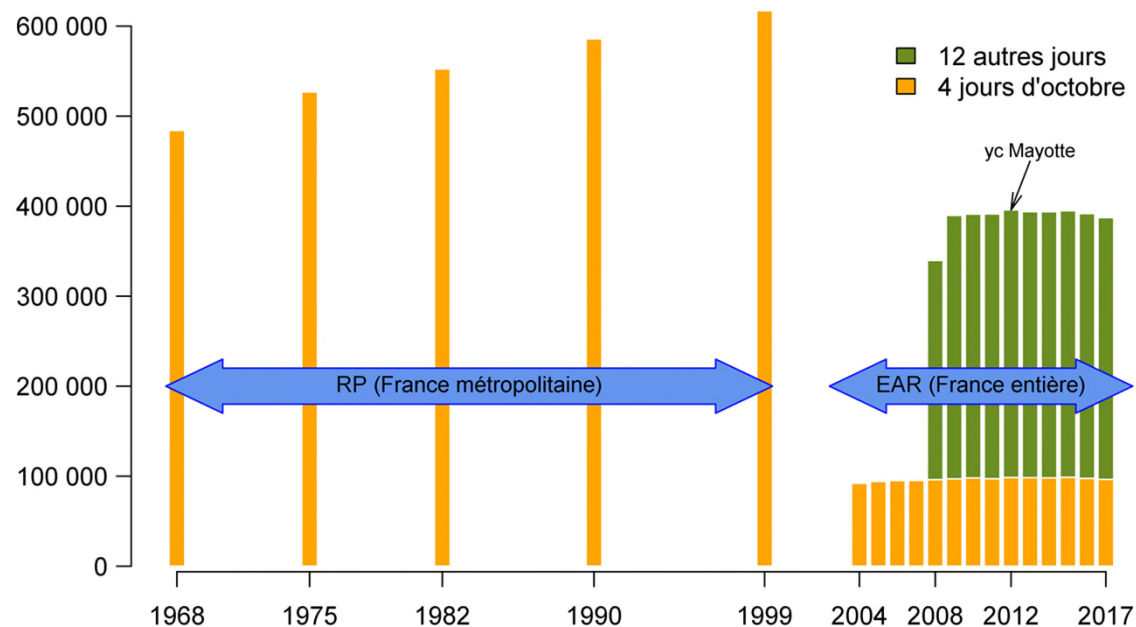
Les sources de l'EDP: le recensement

- En 2004, changement de méthode : on est passé d'un recensement exhaustif (RP) à des enquêtes annuelles de recensement par sondage (EAR) => à partir de 2004, on a donc un *échantillon* annuel des individus EDP (1/7 de l'effectif total)

- 16 jours EDP à partir de 2008/2009 : environ 400 000 individus EDP recensés chaque année, pour compenser la dégradation de la qualité des estimations liée à la réduction des données disponibles une même année

- Extension aux Doms à partir de 2004 (yc Mayotte en 2012)

Individus EDP recensés depuis 1968



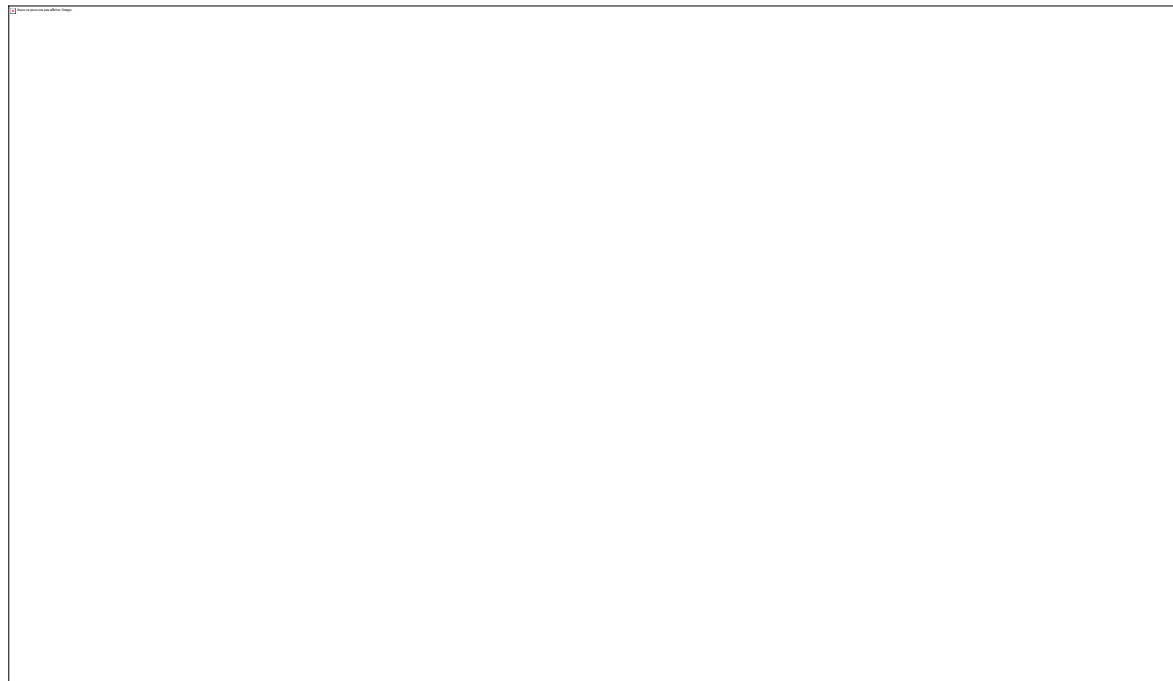
Source : Échantillon démographique permanent BE2017, Insee

Les sources de l'EDP: le recensement

- Les recensements apportent diverses informations : sur le logement, les diplômes, etc.
- Les recensements jusqu'en 1999 étaient exhaustifs, ce qui garantissait une couverture totale de la population.
 - => **Avantages** : effectifs importants pour les études, caractéristiques à une date précise.
 - => **Inconvénient** : pas d'informations sociales entre deux recensements, espacement de 10 ans.
- Les EAR obligent à mettre en œuvre des pondérations plus complexes, le tirage au sort est plus complexe.

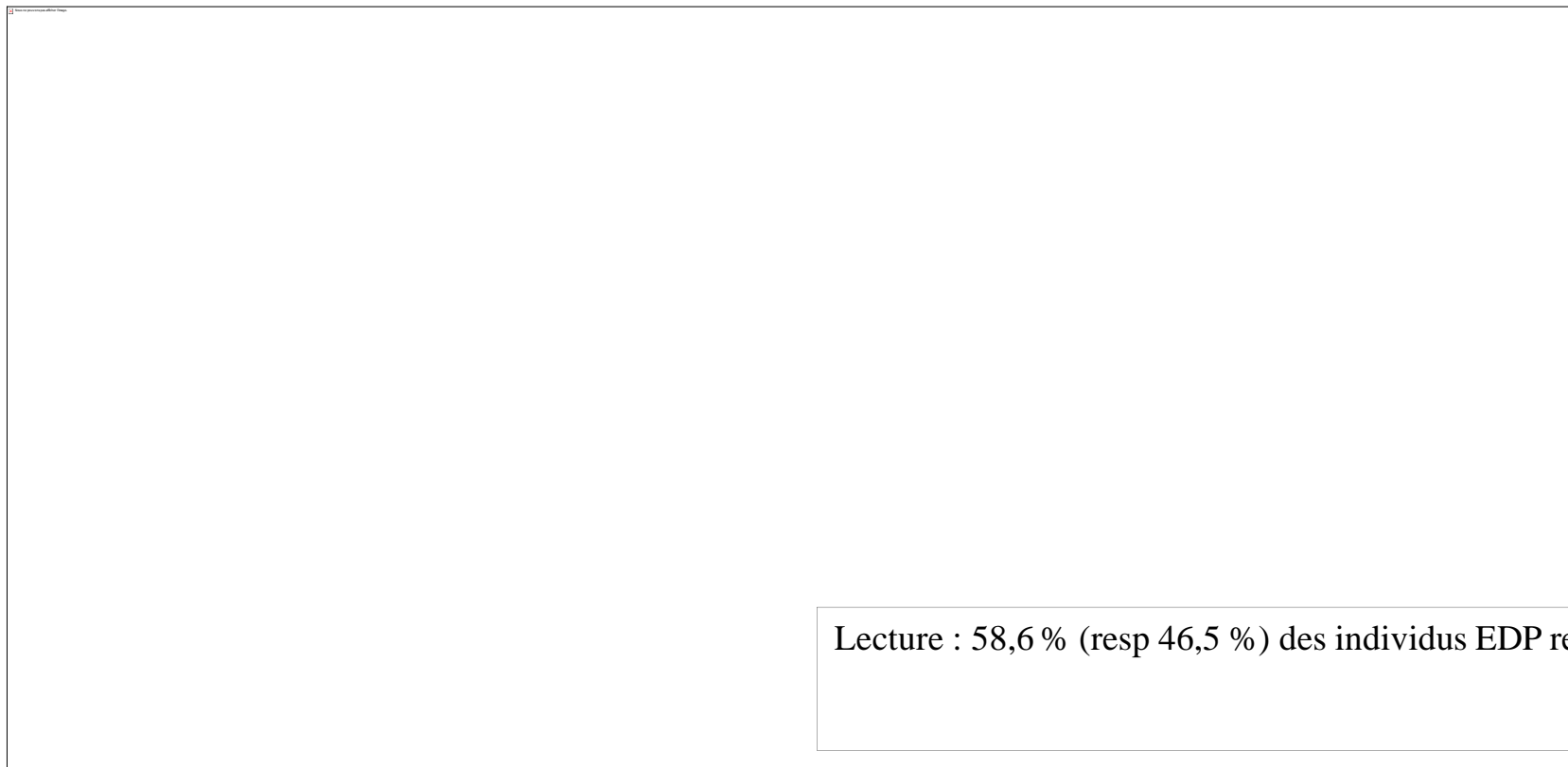
Les sources de l'EDP: le recensement

Entre deux RP successifs, environ 85% des individus EDP étai



Les sources de l'EDP: le recensement

En raison du plan de sondage des EAR, le panel EAR a un pas de 5 ans



Lecture : 58,6 % (resp 46,5 %) des individus EDP recensés en 200

A conduit à intégrer 2 nouvelles sources : panel tous salariés et données fiscales

Le panel tous salariés et les données fiscales, outre leur richesse propre (notamment sur l'ajout des niveaux de vie), apportent :

- Des données **exhaustives** : pour compenser la non exhaustivité des données du recensement depuis 2004 (localisation, caractéristique du logement situation familiale par exemple)
- Des données **annuelles** : pas de « trous » d'analyse de 10 ans, ou de 5 ans contrairement au RP ou EAR

Les sources de l'EDP : le panel tous salariés



Les sources de l'EDP: le panel tous salariés

- Le panel tous salariés permet d'avoir des éléments notamment sur : salaires nets, durée de paie, nombre d'heures travaillées, communes de travail et de résidence, PCS, type de contrat de travail, condition d'emploi, secteur d'activité.
- Le panel « tous salariés » de l'EDP compile des données issues des déclarations de données sociales (DADS puis DSN), des fichiers de paie de l'état et des déclarations des particuliers-employeurs jusqu'à l'année N-2 (2017 pour BE2019).
- Les postes des individus EDP sont agrégés au niveau annuel et en conservant les caractéristiques du poste principal.
- Le champ des individus EDP est complet à partir de 2002
- Dans un futur proche, seront intégrés les non-salariés, ce qui fournira une couverture exhaustive des actifs en emploi

Les sources de l'EDP: les données socio-fiscales



Les sources de l'EDP: les données socio-fiscales

- des données administratives exhaustives issues des déclarations serv
- de deux dispositifs statistiques :
 - Le Fichier Démographique sur les Logements et les individus (**Fidéli**)
 - Le Fichier Localisé Social et Fiscal (**FiLoSoFi**)

Le champ retenu dans l'EDP est celui des individus « Fidéli » connus :

Les sources de l'EDP: les données socio-fiscales

Fideli a « pour finalité une meilleure connaissance du parc de logement et de

- Les revenus
- La localisation fine des individus (commune, section cadastrale,
=> L'Iris et la QPV nécessitent une expertise avant exploitation.
=> Il ne s'agit pas de faire des exploitations par zone fine, mais de
- Le logement (HLM, nombre de pièces, superficie, etc ...)
- Le statut matrimonial fiscal de l'individu, et la date de l'événement

Les sources de l'EDP: les données socio-fiscales

Le dispositif Filosofi permet de reconstituer *par ménage* :

- un revenu déclaré (avant redistribution et imputation des revenus financiers)
- et un revenu disponible (après redistribution et imputation des revenus financiers)

L'intérêt majeur est d'enrichir les trajectoires des individus avec des variables socio-fiscales

La suppression de la TH va faire évoluer Fideli => à partir de 2023, utilisation de Filosofi

Apport de l'EDP par rapport à ses sources

L'EDP enrichit les sources qui l'alimente :

- mutuellement entre elles, en cumulant leurs informations respectives s
 - en les « panelisant », pour permettre de suivre les individus sur longue
-
- Mais il est tributaire de la qualité des sources qu'il compile,
 - et il ajoute un risque supplémentaire à leur qualité :
 - C'est un échantillon (précision moins bonne que la source de départ)
 - Il peut y avoir des défauts d'identification (~ « taux de réponse » pour u

Accès aux données de l'EDP

• La base études mise à disposition ne contient aucune information nominative (nom, prénoms, NIR, identifiant fiscaux), mais la disponibilité des dates exactes et des lieux précis peut la rendre **indirectement nominative**

• La constitution et l'accès aux données de l'EDP sont donc **très réglementées**.

• Pour les chercheurs hors Insee, il est possible d'accéder à l'EDP sous conditions :

– Dossier de demande à adresser au comité du secret statistique, qui donnera un avis, puis demande d'autorisation aux archives nationales.

– En cas d'accord, accès aux données via le centre d'accès sécurisé aux données (CASD)

Exploiter l'EDP : de nombreuses tables de données à combiner

Une base de données avec de nombreuses tables organisées par sources, qui grossit tous les ans.

Types de données	Tables	Années de collecte des données	Nombre de tables	
			2018	2019
VARIABLES DE GESTION	INDIVIDU	1968-2019	1	1
État civil	NAISSANCE	1968-2019	1	1
	DESCENDANCE		1	1
	MARIAGE		1	1
	DECES		1	1
	Recensement		Exhaustifs	1968, 1975, 1982, 1990, 1999
	Mayotte	2012	1	1
Enquêtes annuelles de recensement	INDIVIDU	2004-2019	15	16
	FAMILLE		15	16
	LOGEMENT		15	16
	LIENS		2019	0
Données fiscales et sociales	FISC_INDIVIDU	2011-2019	8	9
	FISC_LOGEMENT		8	9
	FISC_REVENU		8	9
	FISC_REVDET		8	9
Panel tous salariés	PANACT	1967-2019	1	1
Fichier électoral	FE	1967-2017	1	1
			90	98

Tenir compte des spécificités des sources

Au moment de se lancer dans une étude, il faut :

- Estimer l'impact de ces spécificités sur mon projet d'études
- Tenir compte de l'évolution des variables récupérées
- Tenir compte de l'évolution des champs de collecte des sources
- Définir sa population d'intérêt en combinant des sources de données : se poser la question de la représentativité et donc des pondérations

Des données riches, qui nécessite un travail spécifique pour les études

- la liste des variables de l'EDP est disponible sur le site du CASD et u
- il est utile d'échanger : un groupe d'exploitation mis en place par l'Ins
 - => objectif : échanger sur les données, présenter ses travaux et
- Et pour une description rapide de l'EDP, voir par exemple :

Robert-Bobée, I., Gualbert, N. (2021). «L'échantillon démographique permanent : en 50 ans, l'EDP a bien

L'échantillon démographique permanent (EDP)

Merci de votre attention