AGENCE NATIONALE DE LA RECHERCHE ANR

# Compte-rendu intermédiaire

## Projet ANR-16-CE41-0007

# Big_Stat

Programme DS0806 2016


A IDENTIFICATION .......................................................... 2
B LIVRABLES ET JALONS .................................................... 2
C RAPPORT D'AVANCEMENT .................................................... 3
C.1 Objectifs initiaux du projet ............................................3
C.2 Travaux effectués et résultats atteints sur la période concernée...3
C.3 Difficultés rencontrées et solutions ....................................4
C.4 Faits et résultats marquants ............................................4
C.5 Travaux spécifiques aux entreprises (le cas échéant).................5
C.6 Réunions du consortium (projets collaboratifs) .........................5
C.7 Commentaires libres......................................................5
D VALORISATION ET IMPACT DU PROJET DEPUIS LE DEBUT ..................... 6
D.1 Publications et communications ..........................................6
D.2 Autres éléments de valorisation .........................................7
D.3 Pôles de compétitivité (projet labellisés) ..............................8
D.4 Personnels recrutés en CDD (hors stagiaires) ...........................8
D.5 État financier............................................................8
E ANNEXES EVENTUELLES ..................................................... 9
E.1 Annexe 1. Présentation très synthétique des projets de recherche 9
E.2 Annexe 2. Reproduction d'une note de l'Insee ......................... 10
E.3 Annexe 3. Résumé public mis à jour ..................................... 13

# A IDENTIFICATION

| | |
|---|---|
| Acronyme du projet | Big_Stat |
| Titre du projet | Big Statistical Data for a Mobile Society |
| Coordinateur du projet (société/organisme) | Laurent Toulemon |
| Date de début du projet<br>Date de fin du projet | Mars 2017<br>Mars 2021 |
| Labels et correspondants des pôles de compétitivité<br>(pôle, nom et courriel du corresp.) | |
| Site web du projet, le cas échéant | https://big-stat.site.ined.fr/ |

| | |
|---|---|
| Rédacteur de ce rapport | |
| Civilité, prénom, nom | M. Laurent Toulemon et Mme Giulia Ferrari |
| Téléphone | 01 56 06 21 16 |
| Courriel | Toulemon@ined.fr |
| Date de rédaction | 16/08/2019 |
| Période faisant l'objet du rapport d'activité | 1 Mars 2017 - 30 Septembre 2019 |

# B LIVRABLES ET JALONS

| N° | Intitulé | Nature* | Date de fourniture | | | Partenaires (souligner le responsable) |
|---|---|---|---|---|---|---|
| | | | Prévue initialement | Replanifiée | Livrée | |
| 1 | Accès aux données par le CASD | Jalon | 3-2017 | | 3-2017 | |
| 2 | Ajouts de membres et de sources (ajouts acceptés par le Comité du secret statistique les 9-6-2017, 13-10-2017, 2-2-2018, 29-6-2018 | Jalon | | | Divers | |
| 3 | Réunion de lancement | Séminaire | 10-2016 | | 10-2016 | |
| 4 | Réunion du WP « Structures familiales » | Séminaire | 10-2017 | | 10-2017 | |
| 5 | Réunion du WP « jeunes adultes » | Séminaire | 10-2017 | | 02-2018 | |
| 6 | Formation sur l'apprentissage supervisé | Séminaire | 01-2018 | | 01-2018 | |
| 7 | Réunion de l'advisory Board | Jalon | 6-2017 | 11-2019 | | |
| 8 | Réunion du WP « Estimations de population » | Séminaire | 11-2018 | | 11-2018 | |
| 9 | École d'été sur les Big Data | Formation | 07-2018 | | 07-2020 | |
| 10 | Présentation à des conférences, articles de recherche (voir liste ci-dessous) | Jalon | | | | |

# C Rapport d'avancement

## C.1 Objectifs initiaux du projet

Centré sur l'utilisation des données administratives françaises en sciences sociales, le projet comporte trois aspects.
Tout d'abord, des recherches sur des sujets importants en sociodémographie, pour lesquels les données administratives viennent compléter les données d'enquête. Trois sujets forment le cœur des projets de recherche. 1) évaluation des doubles comptes dans les enquêtes et le recensement, description de la situation familiale des habitants en tenant compte des personnes recensées ou enquêtées deux fois car elles ont deux logements habituels. 2) formation et rupture des couples par les jeunes adultes. 3) analyse des situations familiales et socioéconomiques des enfants de parents séparés qui partagent leur temps entre les deux domiciles parentaux.
Ensuite, une validation des données, en collaboration avec les producteurs, fondée d'une part sur une comparaison entre sources et une confrontation avec des données d'enquêtes sociologiques (y compris des entretiens qualitatifs) pour analyser les situations concrètes qui se cachent derrière des situations familiales mal recensées ou mal identifiées par la statistique publique et, d'autre part, par un retour aux données initiales en cas de résultat étrange. Ces validations peuvent conduire, comme c'est déjà le cas pour l'Échantillon démographique permanent (EDP) de l'Institut de la statistique et des études économiques (Insee), à des enrichissements ou des corrections des fichiers.
Enfin, un effort de diffusion et de mise à disposition des données administratives qui se traduit par la réalisation d'une liste de diffusion du projet, de sites internet dédiés à chaque source, et d'actions de formation des utilisateurs de ces données.

## C.2 Travaux effectues et resultats atteints sur la periode concernee

Le projet a rassemblé dans un premier temps des travaux portant sur l'Échantillon démographique permanent de l'Insee, qui regroupe les données du recensement et de l'état civil depuis 1968, et dont l'enrichissement récent aux données sociales et fiscales en fait un fichier de données extrêmement riche, mais dont la documentation complète et la validation gagnent à tirer bénéfice de retours des utilisateurs. Benjamin Marteau, Laurent Toulemon et Sébastien Durier ont estimé la proportion et le nombre d'individus en double compte au recensement et ils décrivent les situations familiales et sociales de ces individus. Julien Boelaert a utilisé des méthodes de *machine learning* pour distinguer les vrais couples de même sexe au recensement à partir de l'enquête Famille. Giulia Ferrari et Laurent Toulemon, en utilisant les données socio-fiscales appariées avec les EAR, ont analysé l'évolution des situations conjugales et les transitions entre les différents états conjugaux en France en 2010 et 2015. Giulia Ferrari, Carole Bonnet et Anne Solaz ont étudié la mobilité résidentielle suivant un divorce ou une rupture de PACS, avec un focus particulier sur les parents et sur le rôle du type de garde des enfants. Marie-Caroline Compans a exploré l'état civil pour étudier la fécondité.
De très nombreux projets de recherche ont été lancés (voir annexe 1).

Ces travaux ont donné lieu à des retours vers l'Insee, producteur de l'EDP (voir annexe 2). Les travaux de validation des sources se poursuivent, notamment sur le fichier de données fiscales Fidéli, dont la structure rend les analyses longitudinales complexes.

Nous avons créé un site web du projet https://big-stat.site.ined.fr/ où nous présentons les sources potentiellement disponibles pour la recherche et les moyens d'y accéder, les travaux effectués à partir de ces données et financés par ce projet. Le site regroupe également des références techniques et théoriques sur l'utilisation des données administratives et des liens vers les expériences similaires, en France et à l'étranger.

Ensuite, nous avons créé un site participatif pour les utilisateurs de l'EDP (https://utiledp.site.ined.fr), où ils peuvent consulter la documentation des données (qui n'est pas disponible par ailleurs) et contribuer à son enrichissement en proposant des codes de variables construites par eux-mêmes et exploitables par d'autres utilisateurs, sur le modèle du site des utilisateurs de la cohorte d'enfants Elfe. Le site contient ainsi un ensemble cohérent de métadonnées : documentation officielle produite par les producteurs, variables ajoutées par les utilisateurs, notes sur les fichiers. Cet ensemble est mis à jour et corrigé en permanence.

Des sites similaires ont été réalisés pour les enquêtes annuelles de recensement (https://recensement.site.ined.fr), le fichier regroupant les « troncs communs » des enquêtes de l'Insee auprès des ménages (https://util-tcm.site.ined.fr), les données de la Caisse national des allocations familiales (Cnaf) qui ont été mises à disposition en février 2019 au CASD (https://utilcnaf.site.ined.fr/). D'autres sont envisagés pour les enquêtes européennes EU-Silc et le fichier des données fiscales Fidéli, maintenant disponibles sur le site du CASD.

Nous avons organisé au printemps 2018 une formation aux méthodes d'analyse des données massives et aux routines utilisables en langage R, et participé au financement d'une formation de l'Ined sur les données EU-Silc. Une école d'été sur l'Échantillon démographique permanent sera organisée à l'été 2020.

Le tout est conforme au plan initial.

## C.3 Difficultes rencontrees et solutions

La mise à disposition des données de l'Insee (Tronc commun des ménages TCM) et de la Cnaf ont été retardées par rapport au projet initial. Le projet est associé à la rénovation du TCM, dont les données devraient être mises à disposition fin 2019. La Cnaf a mis au point une procédure spécifique pour l'accès aux données pour la recherche, en cours de finalisation pour les réponses aux premières demandes. Les données sont d'ores et déjà déposées au CASD, mais les premières demandes arrivées fin 2018 n'étaient pas encore satisfaites fin juillet 2019.

Ces retards ne remettent pas en cause l'avancement du projet ; à l'inverse l'existence du projet facilite les progrès dans la mise à disposition en explicitant la demande et en proposant des solutions concrètes pour ce qui concerne les relations entre les producteurs et les chercheurs.

## C.4 Faits et resultats marquants

L'analyse de l'Échantillon démographique permanent, en collaboration entre l'Insee et l'Ined, a permis de mesurer (pour la première fois depuis l'adoption des enquêtes annuelles de recensement en 2004) la fréquence des doubles comptes au recensement à 2%, et de discuter de la précision et de la portée de ce résultat avec les responsables du recensement, notamment en termes d'estimation des situations familiales complexes souvent associées à des doubles comptes (enfants de parents séparés, jeunes adultes plus ou moins partis de chez leurs parents). Après de longues discussions avec le département de la démographie de l'Insee, les publications du projet seront intégrées dans un volume de présentation de la qualité du recensement à paraître.

La rénovation des formulaires du recensement a donné lieu début 2019 à un ajustement des estimations de population par l'Insee, pour lequel l'expérience acquise dans le cadre du projet a été déterminante.

Le site du projet https://big-stat.site.ined.fr , en français et en anglais, regroupe un ensemble d'informations utiles pour les chercheurs souhaitant utiliser des données administratives ou des données ouvertes ou des données contextuelles. Un ensemble cohérent d'articles théoriques sur les données administratives et les données massives en sciences humaines et sociales est présenté, et complété par de nombreux exemples de travaux utilisant de telles données, en France et ailleurs. Le lien vers www.data.gouv est complété par des liens vers les principales bases de données dans les domaines de la démographie, de la santé, de l'équipement et des services territoriaux, de l'économie et des transports. De même les principales bases de données contextuelles sont référencées.

L'évaluation de l'Ined par l'HCERES au printemps 2019 a salué la vision de l'Ined concernant les données administratives et leur utilité pour la recherche. Cette priorité, initiée entre autres par le projet Big_Stat, se traduit maintenant par le recrutement d'un ingénieur de recherche spécialisé en données administratives, prévu à l'automne 2019. Dans cette perspective, le contrat de Giulia Ferrari a été prolongé d'un an, pour lui permettre de candidater à ce poste dans de bonnes conditions.

## C.5 TRAVAUX SPECIFIQUES AUX ENTREPRISES (LE CAS ECHEANT)

## C.6 REUNIONS DU CONSORTIUM (PROJETS COLLABORATIFS)

| Date | Lieu | Partenaires présents | Thème de la réunion |
|---|---|---|---|
| 19/10/2016 | INED | Tous les membres du projet (18 participants) | Réunion de lancement du projet |
| 04/10/2017 | INED | 14 participants | Réunion des participants à l'axe 2, jeunes adultes |
| 09/02/2018 | INSEE | 13 participants | Réunion des participants à l'axe 1, comptage de la population et situations familiale |

## C.7 COMMENTAIRES LIBRES

### *Commentaires du coordinateur*

La collaboration efficace entre les membres du projet a permis de corriger ou de compléter les données de l'Échantillon démographique permanent (EDP). Les allers-retours entre l'Insee, producteur de l'EDP, et les utilisateurs sont tout à fait conformes à l'ambition initiale du projet, fondée sur le constat que les données administratives ou les données complexes comme l'EDP ne peuvent être parfaitement validées et documentées par les producteurs. Le projet a permis des recherches collaboratives et a conduit les utilisateurs à valider les éléments du fichier nécessaires à leur recherche, puis à faire des retours vers l'Insee, qui a pu ainsi améliorer le fichier et sa documentation. Une note de l'Insee, reproduite en annexe, en atteste.

Le projet regroupe au Centre d'accès sécurisé aux données (CASD) pas moins de 46 chercheurs. Alors que le modèle du Comité du secret est plutôt fondé sur des petits projets déconnectés les uns des autres (pour garantir au mieux le respect de la confidentialité), le projet Big_Stat valorise la mise en commun des expériences et les retours vers les producteurs, offrant ainsi un autre modèle de recherche, plus collaboratif et associant pleinement l'Insee, producteur des données. L'accès aux données confidentielles a été prolongé pour 3 ans en juin 2019 par le Comité du secret statistique.

Outre l'Ined, les membres du projet appartiennent au CNRS, à l'Inserm ou à diverses universités en France (Bordeaux, Strasbourg, Nanterre, Paris Descartes, Paris Sorbonne) et à l'étranger (Penn University, université d'Anvers). Centré à l'origine sur l'EDP, le projet s'élargit à de nombreuses autres sources, soit déjà disponibles (Fichiers démographiques sur les logements et les individus (Fidéli), Enquête emploi en continu et Enquêtes Formation qualification professionnelle de l'Insee, enquêtes européenne EU-SILC) soit qui le seront bientôt (Tronc commun des enquêtes auprès des ménages de l'Insee, enquêtes annuelles de recensement avec leur pondération spécifique, fichiers de la Caisse nationale des allocations familiales, causes de décès enrichissant l'EDP).

De nouveaux membres sont venus se joindre au projet. Aux vingt-deux membres initiaux se sont ajoutés dix doctorants, cinq post-doctorants, six chercheurs et trois enseignants-chercheurs en France. Des contacts ont été pris avec des responsables de projets similaires en Belgique (université d'Anvers), en Allemagne (MPIDR de Rostock), au Canada (McGill à Montréal) et aux États-Unis (Penn University ; Madison, Wisconsin). Un chercheur et un post-doctorant de l'université d'Anvers ont également rejoint le groupe. Des synergies se mettent également en place avec d'autres chercheurs travaillant sur d'autres projets mais utilisant l'environnement créé par le projet pour faciliter leurs relations avec les producteurs.
En rajoutant trois administratifs et neuf inscriptions pour information, la liste de diffusion comprend actuellement 67 membres actifs ou proches du projet.

### *Commentaires des autres partenaires*

### *Question(s) posée(s) à l'ANR*

# D VALORISATION ET IMPACT DU PROJET DEPUIS LE DEBUT

## D.1 PUBLICATIONS ET COMMUNICATIONS

| Liste des publications monopartenaires (impliquant un seul partenaire) | | |
|---|---|---|
| **International** | **Revues à comité de lecture** | 1. Toulemon Laurent. 2017. Undercount of young children and young adults in the new French census, Statistical Journal of the IAOS, Vol 33, p. 311–316. https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji1054<br>2. Ferrari, G., Bonnet, C., Solaz, A. 2019. Will the one who keeps the children keep the house? Residential mobility after divorce by parenthood status and custody arrangements in France. *Demographic research*, 40, p. 359-394. https://www.demographic-research.org/volumes/vol40/14<br>3. Tomkinson, J. 2019. « Age at first birth and subsequent fertility : The case of adolescent mothers in France and England and Wales ». *Demographic Research*, 40(27), p. 761-798. https://www.demographic-research.org/volumes/vol40/27 |

| | | |
|---|---|---|
| | **Communications (conférence)** | 1. Giulia Ferrari et Laurent Toulemon « Is the French PACS similar to cohabitation or marriage? New data on family situations and transitions ». Présenté à la conférence annuelle de la PAA (Population Association of America) à Denver (USA) en avril 2018.<br>2. Giulia Ferrari et Laurent Toulemon « Is the French PACS similar to cohabitation or marriage? New data on family situations and transitions ». Présenté à la conférence Européenne de Population (EPC) à Bruxelles en juin 2018.<br>3. Giulia Ferrari, Carole Bonnet et Anne Solaz « Who keeps the children will keep the house? Residential mobility after divorce and civil partnerships' dissolution ». Présenté à un workshop dédié aux différences internationales dans la mobilité résidentielle après le divorce à Saint Andrews (Écosse) en mai 2017.<br>4. Giulia Ferrari, Carole Bonnet et Anne Solaz « Who keeps the children will keep the house? Residential mobility after divorce and civil partnerships' dissolution ». Présenté à Divorce Conference à Anverse en octobre 2017.<br>5. Laurent Toulemon, Sébastien Durier et Benjamin Marteau "Two homes, two families? People counted twice in the French rotating census". Présenté à XXVIII IUSSP International Population Conference, Session 257: Improving collection and quality of demographic data in censuses and surveys, Cape Town (South Africa), 29 October - 4 November 2017.<br>6. Laurent Toulemon, Sébastien Durier et Benjamin Marteau, "Towards a Precise and Accurate Analysis of the Situation of Two-Home Children in France from New Demographic Panel Data". Présenté à la European Population Conference 2018, European Association for Population Studies (EAPS), Brussels (Belgium) en juin 2018 .<br>7. Bonnet, C., Ferrari, G., Solaz, A., Toulemon, L. 2019, Marital shocks and mortality in France: recent evidence from panel tax and civil registration data. Présentation à la ECSR 2019 Conference, 12-14 Sep 2019 Lausanne (Switzerland)<br>8. Bonnet, C., Ferrari, G., Solaz, A., Toulemon, L. 2019, Marital shocks and mortality in France: recent evidence from panel tax and civil registration data. Présentation à la SLLS Annual Conference, Potsdam, Germany 25 - 27 September 2019 |
| **France** | **Ouvrages ou chapitres d'ouvrage** | 1. Toulemon Laurent, 2018, Combien de personnes ont plusieurs résidences habituelles en France ? in Imbert Christophe, Lelièvre Éva, Lessault David, La famille à distance : mobilités, territoires et liens familiaux, Paris : Ined, 2018, Chap. 1, p. 27-49.<br>2. Toulemon L., Durier S., Marteau B., 2019, Estimation des doubles comptes statistiques au recensement à partir de l'Échantillon démographique permanent, chapitre à paraître dans un document de l'Insee sur le recensement de la population, 12 pages. |
| | **Communications (conférence)** | 1. Laurent Toulemon, Sébastien Durier et Benjamin Marteau, « Au recensement, 2,3 % de doubles comptes, d'après l'échantillon démographique permanent ». Présenté aux 13es Journées de Méthodologie Statistique de l'Insee (JMS 2018), Paris (France), 12-14 juin.<br>2. Pennec S., Flammant C., 2018, Évolution de l'orphelinage précoce au XXIe siècle, Colloque « Le vécu de jeunes après le décès d'un (des) parent(s). Expérience sociale, soutiens et acteurs à l'épreuve de la recherche sur les orphelins en France », Fondation OCIRP – CADIS-EHESS, 3 octobre 2018, p. 8-11, https://www.ocirp.fr/sites/default/files/fondation_ocirp_actes_03102018.pdf. |
| | **Notes techniques** | 1. Toulemon L., 2018, Faut-il ajuster les estimations de population de 2016 ? Note pour Chantal Cases, directrice des statistiques démographies et sociales de l'Insee, 2 novembre 2018, 9 pages. |

## D.2 AUTRES ELEMENTS DE VALORISATION

| **Liste des éléments. Préciser les titres, années et commentaires** | |
|---|---|
| **Autres (sites web)** | 1. Site du projet : https://big-stat.site.ined.fr/<br>2. Site des utilisateurs de l'EDP : https://utiledp.site.ined.fr<br>3. Site des utilisateurs du recensement : https://recensement.site.ined.fr/<br>4. Site des utilisateurs du Tronc commun des enquêtes auprès des ménages: https://util-tcm.site.ined.fr/<br>5. Site des utilisateurs des données de la Cnaf : https://utilcnaf.site.ined.fr/ |

| | |
|---|---|
| | 6. Site des utilisateurs des données des enquêtes européennes EU-SILC : https://eu-silc.site.ined.fr/en/ |

## D.3 POLES DE COMPETITIVITE (PROJET LABELLISES)

### *Collaboration du projet avec le(s) pôle(s) ayant labellisé*

### *Activités financées par le complément de pôle (laboratoires publics uniquement)*

## D.4 PERSONNELS RECRUTES EN CDD (HORS STAGIAIRES)

| **Identification** | | | | **Avant le recrutement sur le projet** | | | **Recrutement sur le projet** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Nom et prénom | Sexe H/F | Adresse email (1) | Date des dernières nouvelles | Dernier diplôme obtenu au moment du recrutement | Lieu d'études (France, UE, hors UE) | Expérience prof. antérieure (ans) | Partenaire ayant embauché la personne | Poste dans le projet (2) | Date de recrutement | Durée missions (mois) (3) |
| Ferrari Giulia | F | giulia.ferrari@ined.fr | | Doctorat | UE | 6 | Ined | Post-Doc | 01/03/2017 | 24 |
| Ferrari Giulia | F | giulia.ferrari@ined.fr | | Doctorat | UE | 8 | Ined | Post-Doc | 01/03/2019 | 12 |

## D.5 ÉTAT FINANCIER

| **Nom du partenaire** | **Crédits consommés (en %)** | **Commentaire éventuel** |
|---|---|---|
| INED | 53% | Recrutement d'un post-doc et abonnements au CASD conformes aux prévisions. Post-doc prolongé d'un an.<br>Dépenses de personnel : 128 208 €<br>Dépenses de fonctionnement : 26 614 €<br>Total : 154 822 € (sur 269 986 € attribués au total) |

# E Annexes eventuelles

## E.1 Annexe 1. Presentation tres synthetique des projets de recherche

Le travail est collaboratif, et de nombreux projets de recherche sont inclus dans le programme. Cette annexe les présente très brièvement sous forme de liste.

- Benjamin Marteau et Marie Bergström s'intéressent à étudier le lien entre précarité professionnelle et instabilité conjugale des jeunes adultes avec l'enquête SRCV et les données fiscales.
- À partir des données des EAR de 2004 à 2016, Antoine Robin examine les conditions de logements et l'insertion professionnelle des jeunes natifs des Dom en métropole.
- Yajna Govind étudiera les inégalités des distributions de revenus dans les DOM et la comparera à la situation en Metropole.
- Louise Caron, Myriam Khlat et Lidia Panico utiliseront l'EDP pour décrire l'évolution des conditions de vie des enfants nés de parents immigrés, et leur impact sur les profils sociodémographiques et la mortalité à l'âge adulte.
- Julien Boelaert, Giulia Ferrari et Benjamin Marteau proposent de retravailler la question des inégalités de trajectoires dans le passage à la vie adulte à partir de l'EDP et des données sociales et fiscales en employant des réseaux de neurones convolutifs.
- Baptiste Coulmont a repéré des couples mariés du même sexe dans le fichier détail du recensement et dans l'EDP. Avec Gaëlle Meslay ils étudieront leurs caractéristiques sociodémographiques et la présence d'enfants.
- Baptiste Coulmont a travaillé sur le fichier électoral de l'EDP, pour décrire la population des non-inscrits et de personnes "inscrites ailleurs". Il compte poursuivre cette étude en lien avec l'exploitation de l'enquête Participation électorale 2017.
- Marie-Caroline Compans compte étudier les déterminants du report et du rattrapage de la fécondité à des âges tardifs.
- John Tomkinson travaillera sur les enfants de moins de 5 ans non recensés ainsi que sur l'agrandissement de famille à partir des données fiscales et panel DADS.
- Alessandra Trimarchi utilisera l'EDP pour étudier la formation des couples, l'homogamie éducative, socio-économique et d'âge en France.
- Cyril Jayet travaillera sur les variables mesurant l'origine sociale de l'enquêté.
- Cécile Flammant a soutenu en mai 2019 une thèse sur les orphelins en France. Elle a utilisé les données de l'EDP et de la Cnaf et a mis en ligne un site Internet présentant ses données et facilitant l'accès aux fichiers statistiques auxquels elle a eu accès.
- Marion Leturcq et Yajna Govind analyseront les effets des changements de législation sur l'acquisition de nationalité française par mariage et leur impact sur l'évolution des mariages mixtes
- Claire Kersuzan et Matthieu Solignac analysent les disparités spatiales d'accès à l'autonomie précoce (départ de chez les parents, soutien parental et position professionnelle)
- Karel Neels et Jonas Wood comparent l'impact des cycles économiques en France et en Belgique sur la formation des couples et la fécondité des jeunes, à partir de données administratives belges et de l'EDP
- Isabelle Attané utilise le recensement pour préparer une enquête auprès des habitants d'Île de France nés en Chine ou de nationalité chinoise
- Guillaume Le Roux s'intéresse à la ségrégation urbaine à Paris et le long de l'Axe Seine, à partir d'une approche longitudinale pour les individus et les quartiers dans l'EDP
- Magali Mazuy et Didier Breton tirent profit des données non diffusées de l'état civil et des recensements pour compléter les données disponibles (ruptures de Pacs, pondérations annuelles du recensement, naissances et enfants sans vie, etc.)
- Lucas Sage analysera la dynamique des inégalités salariales au cours du temps à partir de différentes sources, dont le « British Household Panel Survey » anglais (BHPS) et l'EDP
- Marta Veljkovic revisitera les enquêtes FQP pour analyser l'évolution de la mobilité sociale en cours de carrière, de 1964 à 2015, sus la direction de Delphine Remillon

## E.2 Annexe 2. Reproduction d'une note de l'Insee

Note n° 2018_12063_DG75-F170 du 20 juillet 2018

Objet : Apports des échanges avec les utilisateurs à la production et à la documentation de l'Échantillon démographique permanent

Isabelle Robert-Bobée ; personne chargée du dossier : Sébastien Durier

La division Enquêtes et études démographiques (EED) de l'Insee prépare chaque année une nouvelle base étude de l'échantillon démographique permanent (EDP), qui est un panel d'individus, en y ajoutant des données plus récentes. Au départ le panel EDP était centré sur la compilation de données extraites de trois sources : des données des enquêtes annuelles de recensement, des données d'état civil et des données du fichier électoral. L'EDP s'est ensuite enrichi de deux nouvelles sources. Depuis l'enrichissement de l'échantillon démographique permanent par les données du panel « tous salariés » dans la base Études 2013 (diffusion début 2015) et les données fiscales (Fidéli et FiLoSoFi) dans la base Études 2014 (diffusion début 2016), la communauté des utilisateurs de l'EDP s'est grandement élargie. La possibilité d'accéder à l'EDP via le CASD a aussi fortement contribué à cette expansion. Pour accompagner les utilisateurs, la division EED a décidé de mettre en place depuis novembre 2015 un groupe des utilisateurs. Elle a aussi ouvert la possibilité aux utilisateurs d'échanger par mail avec les producteurs. Les échanges avec les utilisateurs ont aussi lieu grâce à la participation de la division EED aux réunions sur les comparaisons de sources utilisant l'EDP, pilotées par l'Ined dans le cadre de l'ANR « des données massives pour une société mobile ». Cela contribue aussi à des contacts plus proches entre producteurs et utilisateurs, et une meilleure diffusion de l'EDP, la documentation de la base étant disponible sur un site internet accessible à tous.

Si l'objectif initial de ces échanges est évidemment d'aider les utilisateurs dans la réalisation de leur projet d'études ou de recherche, ceux-ci en retour peuvent contribuer à l'amélioration de l'EDP. On propose dans cette note de présenter les apports des utilisateurs dans trois domaines : les enrichissements par de nouvelles données ou variables, les corrections ou améliorations de la documentation et enfin les corrections d'erreurs dans les données de l'EDP.

### 1- Enrichissements de l'EDP suite aux demandes ou aux questions des utilisateurs

Un des objectifs du groupe des utilisateurs est de pouvoir recueillir les besoins des utilisateurs. En particulier, l'EDP récupérant des données de cinq sources différentes, des choix ont été opérés par les producteurs parmi la masse des informations disponibles. Ces choix peuvent ne pas s'avérer finalement les meilleurs au regard des besoins des utilisateurs.

#### a- Ajout de variables brutes

Les sources que l'EDP mobilisent sont des données initialement transversales, qui proposent souvent des variables redressées (notamment par imputation de la non-réponse). Pour l'EDP, les variables redressées peuvent être utiles, mais dans le cas très fréquent d'une mobilisation en panel, disposer des données brutes s'avère essentiel pour les utilisateurs, pour reconstituer eux-mêmes des données manquantes en tenant compte des différentes réponses déjà apportées par le passé ou dans d'autres sources intégrées à l'EDP. La stratégie des producteurs consiste donc à fournir autant que possible les variables à la fois dans leurs versions brutes et dans leurs versions redressées. Cette stratégie générale n'avait en pratique pas toujours été adoptée pour chacune des variables. Deux oublis de variables brutes ont ainsi été corrigés dans la base études 2016 (BE2016) suite à des demandes d'utilisateurs :

- ajout de la variable IRAN_X, qui donne la réponse à la question du recensement « Où étiez-vous un an auparavant ? », et qui permet de compléter la variable IRAND dans laquelle les non-réponses à la variable IRAN_X sont imputées (étude sur les double-comptes au recensement, INED)
- rang de naissance : remplacement des valeurs redressées des variables CTX_NAV_PREC_DATE, donnant la date de la naissance précédente, et CTX_MERE_VIVANT_ENF_PREC_NBR, donnant le nombre de naissances précédentes, par les valeurs brutes issues du bulletin de naissance. (étude sur les naissances tardives, INED)

### b- Variables apparues dans les sources

Les sources mobilisées par l'EDP ont évolué au cours du temps. En particulier des variables nouvelles peuvent faire leur apparition dans ces sources. La veille opérée par les producteurs peut s'avérer insuffisante. Par exemple, une demande d'une chargée d'étude de la Drees sur la disponibilité de deux « cases » de la déclaration fiscale (sommes versées pour la garde d'enfant à domicile et sommes versées pour l'emploi de personnes à domicile) a permis aux producteurs de constater l'apparition des deux variables correspondantes dans les fichiers sources (GARDEMM et SERVDOMM) et de les ajouter à la base études 2016.

### c- Récupération de données anciennes

L'EDP compile des données depuis le recensement de 1968. Ces données anciennes sont en théorie complètes et bien documentées. Il peut cependant arriver que des données anciennes puissent être récupérées et ajoutées à la base études. C'est le cas pour les recensements 1990 et 1999. Dans ces deux recensements les identifiants de l'îlot du recensement sont disponibles dans l'EDP, mais sans possibilité de faire le lien entre eux. Des questions d'un utilisateur ont permis aux producteurs de connaître l'existence d'une table de passage entre les îlots 1990 et les îlots 1999 qui sera ajoutée à la base études 2017 (étude sur la mobilité des immigrés, Paris 1).

## 2- Améliorations de la documentation

La mise au point d'une documentation exhaustive et efficace est une des tâches les plus difficiles, notamment pour l'EDP qui compile plusieurs sources sur une longue période. Les erreurs, manques ou insuffisances soulignées par les utilisateurs permettent ainsi une amélioration continue de la documentation.

### a- Suivi des modifications du contenu des variables dans les sources

Un défaut possible de la documentation de l'EDP consiste en la non prise en compte de modification dans le contenu des variables. Par exemple, à partir de 2015, le questionnaire individuel des EAR a été modifié notamment sur la question de la vie en couple ainsi que sur le plus haut niveau de diplôme atteint. Ces changements de modalités ont bien été intégrés à la documentation. Cette refonte de l'enquête de recensement a aussi été accompagnée de changements de traitement de variables sur la famille (intégration des couples de même sexe dans le traitement dits de l'analyse-ménage-famille du recensement), qui n'avaient pas été mentionnés dans la documentation de l'EDP et ont conduit des utilisateurs à poser des questions au producteur, qui a complété la documentation. Deux ajouts ont été faits :

- les variables type de ménage détaillé (TYPMD) et type de famille détaillé (TYPFD) prennent en compte à partir de 2015 les couples de même sexe dans leurs modalités concernant les couples (étude sur la mobilité sociale, Insee Île-de-France)
- pour faire le lien entre les niveaux de diplôme sur longue période (depuis le recensement de 1968) des programmes pour convertir les variables de diplôme dans une nomenclature harmonisée (SAPHIR) étaient fournis dans la documentation de l'EDP. Le changement de questionnaire de 2015 n'avait cependant pas été pris en compte dans le programme fourni. Il l'est depuis la BE2016 suite à une question d'un utilisateur.

### b- Amélioration de la documentation des données anciennes

Lorsque les utilisateurs mobilisent l'EDP en panel sur longue période, ils sont amenés à demander des précisions sur des variables anciennes. Si la documentation n'a pas été suffisamment complète au moment de la récupération des données, les informations ne sont pas ou plus facilement disponibles. Par exemple, la variable catégorie socio-professionnelle pour les DADS avant 1982 (CS2_ANC) n'était pas documentée car non-présente dans la documentation fournie par la source. Un utilisateur souhaitant mobiliser la variable a repéré le manque et une recherche dans les archives a permis de compléter la documentation pour la BE2016 (étude sur la mobilité sociale et le niveau de vie, France Stratégie).

## 3- Corrections des erreurs dans des données de l'EDP

Malgré les efforts déployés par les producteurs pour contrôler le contenu des variables dans l'EDP, des erreurs peuvent toujours passer inaperçues, d'autant que les producteurs n'utilisent pas en pratique l'intégralité des données, mais seulement une partie pour leurs propres études. L'utilisation des données conduit à mieux les connaître et facilite le repérage d'erreurs.

### a- Pondération en panel

Depuis la BE2014, une variable de pondération pour utiliser les EAR en panel dans l'EDP (POIDS_PANEL_5) est disponible. Cependant, une erreur dans le programme de calcul entraînait une légère surpondération pour les départements et régions d'outre-mer qui a été détectée par un utilisateur (étude sur la mobilité résidentielle, Université de Bordeaux et INED) : dans un premier temps, une solution à mettre en oeuvre dans la BE2016 par les utilisateurs eux-mêmes a été proposée ; dans un second temps la variable de pondération corrigée sera livrée avec la BE2017.

### b- Identifiant famille dans l'EAR 2004

Un des intérêts de l'EDP est de disposer d'informations sur les individus habitant le même logement que les individus EDP. Par exemple, dans les EAR l'identifiant famille (ID_FAM_DIFF) permet de connaître les caractéristiques des parents, du conjoint ou des enfants de l'individu EDP. Dans ses travaux préparatoires, un utilisateur a pu détecter l'absence d'un grand nombre d'identifiants famille dans l'EAR 2004 ce qui a été corrigé pour la BE2016 (étude sur les unions libres, Insee DG)

### c- Code ilot1999

Les tests sur la table de passage îlot 1900-îlot 1999 qui a été mentionnée dans la partie 1c ont permis de détecter une erreur dans la variable îlot 1999, due à une mauvaise gestion des espaces dans l'identifiant, probablement lors de la rénovation de l'EDP aux débuts des années 2010 (étude sur la mobilité des immigrés, Paris 1).

## E.3 ANNEXE 3. RESUME PUBLIC MIS A JOUR

### *Des données statistiques massives pour observer une société mobile*

http://www.agence-nationale-recherche.fr/Projet-ANR-16-CE41-0007
(DS0806) 2016
Projet Big_Stat

***Centré sur l'utilisation des données administratives françaises en sciences sociales, le projet comporte trois aspects : recherche, retour vers les producteurs de données, formation des utilisateurs.***

Tout d'abord, des recherches sur des sujets importants en sociodémographie, pour lesquels les données administratives viennent compléter les données d'enquête. Trois sujets forment le cœur des projets de recherche. 1) évaluation des doubles comptes dans les enquêtes et le recensement, description de la situation familiale des habitants en tenant compte des personnes recensées ou enquêtées deux fois car elles ont deux logements habituels. 2) formation et rupture des couples par les jeunes adultes. 3) analyse des situations familiales et socioéconomiques des enfants de parents séparés qui partagent leur temps entre les deux domiciles parentaux.

Ensuite, une validation des données, en collaboration avec les producteurs, fondée d'une part sur une comparaison entre sources et une confrontation avec des données d'enquêtes sociologiques (y compris des entretiens qualitatifs) pour analyser les situations concrètes qui se cachent derrière des situations familiales mal recensées ou mal identifiées par la statistique publique et, d'autre part, par un retour aux données initiales en cas de résultat étrange. Ces validations peuvent conduire, comme c'est déjà le cas pour l'Échantillon démographique permanent (EDP) de l'Institut de la statistique et des études économiques (Insee), à des enrichissements ou des corrections des fichiers.

Enfin, un effort de diffusion et de mise à disposition des données administratives qui se traduit par la réalisation d'une liste de diffusion du projet, de sites internet dédiés à chaque source, et d'actions de formation des utilisateurs de ces données.

***Confronter les sources et les méthodes***

Sur les trois sujets de recherche, l'idée générale est de confronter les sources et les méthodes pour enrichir la compréhension des phénomènes étudiés. Un ensemble très riche de sources peut être mobilisé : les données sociofiscales et le recensement fournissent des estimations très précises des situations, mais fondées sur des définitions particulières. Par exemple le foyer familial est défini par le foyer fiscal (ou le logement avec la taxe d'habitation). Le recensement s'appuie sur la définition des ménages–logements, au sein desquels une ou deux familles peuvent être repérées. Les notions de couple sont également appréhendées à partir de questions individuelles et, depuis la vague 2018, d'informations sur les relations familiales entre habitants du logement, ainsi que sur un éventuel autre logement habituel de chacun des membres du ménage. La plupart des enquêtes auprès des ménages repèrent également les couples co-résidents ou partiellement co-résidents, tandis que des enquêtes plus spécifiques et des entretiens approfondis permettent d'une part de décrire l'ensemble des situations de couple (résidents ou non), ainsi que les situations de co-résidence sans vie de couple.

Concernant les situations familiales des enfants, les enfants partageant leur temps entre les deux résidences parentales après une rupture du couple sont identifiés dans les données fiscales pour les gardes alternées (partage des parts fiscales), dans le recensement depuis 2018, et dans les enquêtes auprès des ménages depuis 2004 à partir de la question sur la présence d'une autre résidence habituelle. La confrontation avec des enquêtes spécifiques, comme l'enquête Famille et logements de 2011, permet d'identifier les situations de ces enfants ; l'échantillon démographique permanent permet, pour les enfants présents la même année dans les deux logements parentaux, de décrire leur situation familiale à partir de la composition de leurs deux logements, où vivent leurs deux familles.

*Résultats*
Le projet a rassemblé dans un premier temps des travaux portant sur l'Échantillon démographique permanent (EDP) de l'Insee, qui regroupe les données du recensement et de l'état civil depuis 1968, et dont l'enrichissement récent aux données sociales et fiscales en fait un fichier de données extrêmement riche, mais dont la documentation complète et la validation gagnent à tirer bénéfice de retours des utilisateurs. L'analyse de l'Échantillon démographique permanent, en collaboration entre l'Insee et l'Ined, a permis de mesurer (pour la première fois depuis l'adoption des enquêtes annuelles de recensement en 2004) la fréquence des doubles comptes au recensement à 2,4%, et de discuter de la précision et de la portée de ce résultat avec les responsables du recensement, notamment en termes d'estimation des situations familiales complexes souvent associées à des doubles comptes (enfants de parents séparés, jeunes adultes plus ou moins partis de chez leurs parents).

De très nombreux projets de recherche ont été lancés : observation des couples de même sexe au recensement et à l'enquête Famille, évolution des situations conjugales et les transitions entre les différents états conjugaux en France entre 2010 et 2015, mobilité résidentielle suivant un divorce ou une rupture de PACS, avec un focus particulier sur les parents et sur le rôle du type de garde des enfants, mesure de la fécondité selon le rang de naissance dans l'EDP et à l'état civil. Ces travaux ont donné lieu à des retours vers l'Insee, producteur de l'EDP. Les travaux de validation des sources se poursuivent.

*Perspectives*
Le travail est collaboratif, et de nombreux projets de recherche sont inclus dans le programme. Nous avons créé un site web du projet https://big-stat.site.ined.fr/, en français et en anglais, où nous présentons les sources et les moyens d'y accéder, les travaux effectués à partir de ces données ainsi que l'ensemble des projets. Le site présente également un ensemble cohérent d'articles théoriques sur les données administratives et les données massives en sciences humaines et sociales, ainsi que de nombreux exemples de travaux utilisant de telles données, en France et ailleurs. Le lien vers www.data.gouv est complété par des liens vers les principales bases de données dans les domaines connexes à la démographie. De même les principales bases de données contextuelles sont référencées.

Ensuite, nous avons créé un site participatif pour les utilisateurs de l'Échantillon démographique permanent (https://utiledp.site.ined.fr), où ils peuvent consulter la documentation des données (qui n'est pas disponible par ailleurs) et contribuer à son enrichissement en proposant des codes de variables, sur le modèle du site des utilisateurs de la cohorte d'enfants Elfe (https://util-elfe.site.ined.fr). Le site est mis à jour et corrigé en permanence.

Des sites similaires ont été construits pour le recensement, le « tronc commun » des enquêtes de l'Insee auprès des ménages, les données de la Caisse nationale des allocations familiales (Cnaf) qui sont mises à disposition depuis novembre 2018. D'autres sont envisagés pour les enquêtes européennes EU-Silc et le fichier des données fiscales Fidéli.

Nous avons organisé en 2018 une formation aux méthodes d'analyse des données massives et aux routines utilisables en langage R, et participé au financement d'une formation de l'Ined sur les données EU-Silc (statistiques sur les revenus et les conditions de vie). Une école d'été sur l'Échantillon démographique permanent sera organisée à l'été 2020.

Le tout est conforme au plan initial.

***Productions scientifiques et brevets***

Toulemon Laurent. 2017. Undercount of young children and young adults in the new French census, Statistical Journal of the IAOS, Vol 33, p. 311–316. https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji1054

Ferrari, G., Bonnet, C., Solaz, A. 2019. Will the one who keeps the children keep the house? Residential mobility after divorce by parenthood status and custody arrangements in France. Demographic research, 40, p. 359-394. https://www.demographic-research.org/volumes/vol40/14

Tomkinson, J. 2019. « Age at first birth and subsequent fertility : The case of adolescent mothers in France and England and Wales ». Demographic Research, 40(27), p. 761-798. https://www.demographic-research.org/volumes/vol40/27

13es Journées de Méthodologie Statistique, Paris (12-14 juin 2018)

- Sébastien DURIER: L'échantillon démographique permanent a 50 ans : retours sur un dispositif statistique original
- Laurent TOULEMON, Sébastien DURIER, Benjamin MARTEAU: Au recensement, 2,3 % de doubles comptes, d'après l'échantillon démographique permanent

Journée Doctorale 2018 de l'Ined, Paris (24 mai 2018)

- Louise CARON, Les effets d'une mobilité internationale sur les trajectoires professionnelles : le cas de la France

Colloque annuel de la Population Association of America 2018 (26-28 avril 2018)

- Giulia FERRARI, Laurent TOULEMON: Is the French PACS similar to cohabitation or marriage? New data on family situations and transitions
- Matthieu SOLIGNAC: Within and beyond national borders: How administrative linked data can revive internal migration studies

28e Congrès international de la population, Cape Town, Afrique du Sud (30 octobre-3 novembre 2017)

- Sébastien DURIER, Vianney COSTEMALLE: Studying unmarried cohabitation with the French Demographic Panel
- Laurent TOULEMON; Sébastien DURIER; Benjamin MARTEAU: Two homes, two families? People counted twice in the French rotating census

15th Meeting of the European Network for the Sociological and Demographic Study of Divorce, Anvers (5-7 octobre 2017)

- Giulia FERRARI: Who stays in the house and who moves out after divorce in France?"

***Partenaires***

INED Institut National d'Etudes Démographiques

Aide de l'ANR **291 585 euros**

Début et durée du projet scientifique **mars 2017 - 48 mois**

***Big_Stat***
Big Statistical Data for a Mobile Society

***Centred on the scientific use of French administrative data in social sciences, the project has three main components: research, feedback to data producers, users training.***
First, the project aims at researching on important topics in sociodemography, for which administrative data complement survey data. Three topics form the core of the research projects. 1) Assessment of double counts in surveys and censuses, description of the family situation of the inhabitants taking into account the persons enumerated or surveyed twice because they have two usual dwellings. 2) Formation and dissolution of young adults' couples. 3) Analysis of the family and socio-economic situations of children of separated parents who share their time between the two parental homes.

Second, it aims at validating data, in collaboration with producers, based on a comparison between sources and sociological survey data (including qualitative interviews) to analyse the concrete situations behind family situations that are poorly identified or misidentified by official statistics and, on the other hand, by returning to the initial data in the event of a weird result. These validations can lead, as it is already the case for the Permanent Demographic Sample of the Institute of Statistics and Economic Studies (INSEE), to file enrichments or corrections.

Finally, it aims at disseminating and making available administrative data, which is reflected in the creation of a project mailing list, websites dedicated to each source, and training activities for users of this data.

***Compare data sources, using different methods***
For the three research topics, the main idea is to compare results from different data sources and using different methods, in order to enrich our understanding of the behaviours under study. We use a very rich set of data: tax and social administrative data, as well as the census, provide very precise estimates of family situations, based on their own specific definitions. For example, the family unit is defined as the taxable family (housing tax data allow identifying households and flatmates). Census data are based on the household definition as the group of people living in the same dwelling; family units are constructed within the households. These data come from the housing form, which has been renewed in 2018, in order to include more precise information on family links, as well as on any other usual residence for each member of the household. Living as a couple is also identified through a question in the individual census form. Most household surveys identify co-resident or partially co-resident couples, while more specific surveys and in-depth interviews describe all couple situations (co-resident or not), as well as situations of co-residence without a life as a couple, or life as a co-resident couple, but without long-term commitment.

Regarding children's family situations, children sharing their time between the two parental residences after a break-up of the couple are described in the tax data for alternate care (sharing of tax shares), in the census since 2018, and in household surveys since 2004 based on the question on the presence of another usual residence. Comparison with specific surveys, such as the 2011 Family and Housing Survey, makes it possible to identify the situations of these children; the Permanent Demographic Sample allows, for children present in the same year in both parental homes, to describe their family situation based on the composition of their two homes, where their two families live.

***Results***
The project initially gathered work on INSEE's Permanent Demographic Sample (PDS), which has been collecting census and vital statistics data since 1968, and whose recent addition to social and tax data makes it an extremely rich data file, but whose complete documentation and validation benefit from user feedback. The analysis of the Permanent Demographic Sample, in collaboration between INSEE and INED, made it possible to measure (for the first time since the adoption of the

annual census surveys in 2004) the frequency of double counting in the census at 2.4%, and to discuss the accuracy and scope of this result with the census authorities, particularly in terms of estimating the complex family situations often associated with double counting (children of separated parents, young adults more or less leaving their parents' homes).

A large number of research projects have been launched: observation of same-sex couples in the census and the Family survey, changes in couple situations and transitions between different conjugal states in France between 2010 and 2015, residential mobility following a divorce or a break-up of PACS, with a particular focus on parents and the role of the type of childcare, measurement of fertility by birth order in the PDS and in civil registration data. This work has led to returns to INSEE, the producer of the PDS. Work on data source validation is going on.
Outlook

The work is collaborative, and many research projects are included in the program. We have created a website for the project https://big-stat.site.ined.fr/, in English and French, where we present the sources potentially available for research and the means of accessing them, the work carried out on the basis of these data and funded by this project, as well as all the projects. The website also presents a coherent set of theoretical articles on administrative data and big data in the humanities and social sciences, as well as numerous examples of work using such data, in France and elsewhere. The link to www.data.gouv is complemented by links to the main databases in the fields of demography, health, territorial equipment and services, economy and transport. Similarly, the main contextual databases are referenced.

Then, we created a participatory site for Permanent Demographic Sample (PDS) users (https://utiledp.site.ined.fr), where they can consult the data documentation (which is not available elsewhere) and contribute to its enrichment by proposing variable codes built by themselves and usable by other users, on the model of the ELFE cohort of children users' website (https://util-elfe.site.ined.fr/en). The site is constantly updated and corrected.

Similar websites have been developed for the Annual Census Survey, the file containing the "common core questions" of INSEE household surveys, and data from the *Caisse nationale des allocations familiales* (Cnaf), which are available since November 2018. Others are being considered for the European Survey on Income and Living Conditions (EU-SILC) surveys and the Fidéli tax data file.
We organized in spring 2018 training in big data analysis methods and routines that can be used in R language, and helped finance an INED training workshop on EU-SILC data. A summer school on the INSEE Permanent Demographic Sample will be organized in the summer of 2020.

This is in line with the original plan.

***Scientific outputs and patents***
Toulemon Laurent. 2017. Undercount of young children and young adults in the new French census, Statistical Journal of the IAOS, Vol 33, p. 311–316. https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji1054
Ferrari, G., Bonnet, C., Solaz, A. 2019. Will the one who keeps the children keep the house? Residential mobility after divorce by parenthood status and custody arrangements in France. Demographic research, 40, p. 359-394. https://www.demographic-research.org/volumes/vol40/14
Tomkinson, J. 2019. « Age at first birth and subsequent fertility : The case of adolescent mothers in France and England and Wales ». Demographic Research, 40(27), p. 761-798. https://www.demographic-research.org/volumes/vol40/27

13es Journées de Méthodologie Statistique, Paris (12-14 juin 2018)

- Sébastien DURIER: «L'échantillon démographique permanent a 50 ans : retours sur un dispositif statistique original«
- Laurent TOULEMON, Sébastien DURIER, Benjamin MARTEAU: «Au recensement, 2,3 % de doubles comptes, d'après l'échantillon démographique permanent«

Journée Doctorale 2018 de l'Ined, Paris (24 mai 2018)

- Louise CARON, «Les effets d'une mobilité internationale sur les trajectoires professionnelles : le cas de la France«

Population Association of America annual conference 2018 (26-28 avril 2018)

- Giulia FERRARI, Laurent TOULEMON: «Is the French PACS similar to cohabitation or marriage? New data on family situations and transitions«
- Matthieu SOLIGNAC: «Within and beyond national borders: How administrative linked data can revive internal migration studies«

28th International Population Conference, Cape Town, south Africa (30 Octobre-3 November 2017)

- Sébastien DURIER, Vianney COSTEMALLE: «Studying unmarried cohabitation with the French Demographic Panel«
- Laurent TOULEMON; Sébastien DURIER; Benjamin MARTEAU: «Two homes, two families? People counted twice in the French rotating census«

15th Meeting of the European Network for the Sociological and Demographic Study of Divorce, Anvers (5-7 octobre 2017)

- Giulia FERRARI: «Who stays in the house and who moves out after divorce in France?«

***Partners***

INED Institut National d'Etudes Démographiques

ANR grant: **291 585 euros**

Beginning and duration: **mars 2017 - 48 mois**