Big_Stat

Big Statistical Data for a Mobile Society

Des données statistiques massives pour observer une société mobile

Total support requested from ANR: 291 585 €. Project duration 48 months

Table of contents

… / …

Big Statistical Data for a Mobile Society

Summary

New family and demographic behaviours are leading to greater individual and social mobility, making it more difficult to define and observe actual family and housing situations. Simultaneously, big statistical data, i.e. data from administrative files covering the whole population, are now becoming available to the research community. The project aims to extend our knowledge of complex and hard-to-measure situations, using several data sources including big data, and to assess the strengths and weaknesses of several data sources that will be disseminated in 2016 by the French National Institute of Statistics and Economic Studies (INSEE).

Data necessary for complex demographic studies, such as the French Demographic Panel based on censuses and civil registration, tax data and family allowance data, are now becoming widely available. So far, they have been rarely used for research purposes in demography. Therefore, we propose, in a first and crucial stage, to assess and document for general use the big data sources recently made available for research, in collaboration with INSEE. This collaboration is key to the constitution of reliable and well documented data sources. The knowledge from experts from several backgrounds and institutions will be essential to fully validate and test these data sources for various uses. In this step, we will check the consistency of population estimates based on censuses, surveys and administrative data, in terms of omissions and double counts, and the impact of discrepancies on the estimation of family situations and behaviours. Two research questions, which are normally difficult to evaluate with standard surveys, will then be addressed, making use of diverse methods and sources: administrative data, censuses, population surveys and qualitative data from semi-structured in-depth interviews. . First, intimate relationships at young adult ages are known for their volatility, and are therefore hard to study with standard survey data. The new data sources will make it possible to look at forms of partnership and union stability in relation to income, education, occupation and labour market integration. This will vastly increase our knowledge of the dynamics of early adulthood and will further our understanding of new forms of partnership. Second, we will look more closely at the situation of children whose parents are separated, and who are a major source of double counting in surveys and censuses. New data sources will provide a more accurate picture of the family situation of children, including those in complex living arrangements, in relation to their standards of living and poverty risk. Administrative data are very useful for studying transitions, while retrospective surveys are often complicated by recall bias and panel studies are weakened by attrition.

This project will be placed in a national and international perspective. It will provide an opportunity to combine the strengths of national institutions, while creating links with institutions abroad involved in the analysis of big administrative and census data. We will benefit from their experience and interact with big data networks to improve the quality and efficiency of our assessments and studies.

By making information, documentation and code for data use available on a website, this project will have a significant impact on the scientific community. It will also contribute to the enhancement of data quality and access. The publication of methodological and applied articles in internationally reputed journals will promote the dissemination of the project's progress and findings.

Members of the project come from the French Institute for Demographic Studies (INED), INSEE, and the universities of Paris 1 Panthéon-Sorbonne, Paris Descartes, Lyons, Nancy and Strasbourg. A major output of the project will be to encourage and facilitate the use of big statistical data amongst scholars working in the humanities and social sciences.

Des données statistiques massives pour observer une société mobile

Résumé

Les nouveaux comportements conjugaux induisent une augmentation des mobilités individuelles familiales, ce qui rend plus difficile une description simple des situations familiales et résidentielles qui tienne compte de leur complexité. Dans le même temps, les données statistiques massives issues des fichiers administratifs exhaustifs deviennent accessibles en France pour la recherche. Le projet a pour ambition de renouveler la connaissance sur des situations familiales particulières, difficiles à observer, en tirant bénéfice de sources statistiques diverses, incluant des fichiers issus des données administratives massives, mais aussi d'évaluer scientifiquement les forces et les faiblesses des différentes sources démographiques qui ont été récemment mises à disposition de la communauté scientifique, ou le seront prochainement, par l'Institut national de la statistique et des études économiques (Insee).

Les données nécessaires pour les analyses démographiques complexes, comme l'Échantillon démographique permanent qui regroupe des données issues des recensements et de l'État civil, et a été récemment enrichi avec des données sociales et fiscales, sont maintenant disponibles. Ces données n'ont pas encore été beaucoup utilisées pour des études démographiques. Notre ambition est, dans un premier temps, d'évaluer la qualité des fichiers massifs mis à disposition et de les documenter pour les utilisateurs, en collaboration avec les personnes en charge de ces données à l'Insee, à partir d'un diagnostic partagé sur les fichiers. Nos efforts se concentreront d'abord sur les estimations de population au recensement, à partir d'une estimation inédite des doubles comptes et des omissions, ainsi que sur les représentations des structures familiales qui en découlent. Deux situations familiales particulières seront ensuite analysées à partir d'analyses diverses fondées sur des sources très hétérogènes : données administratives, recensement, enquêtes en population générale, entretiens non directifs. Tout d'abord, les relations conjugales des jeunes adultes ne sont pas toujours clairement définies et stabilisées. Les différentes définitions de la vie en couple seront mises en regard des conditions sociales et professionnelles des jeunes adultes. Ensuite, nous examinerons la situation familiale des enfants de parents séparés, particulièrement au risque de doubles comptes dans les enquêtes et le recensement. Les données administratives permettent de décrire la situation familiale des enfants et des jeunes adultes telle qu'elle est déclarée à l'État et d'analyser les conditions de vie en termes économiques, y compris pour les enfants partageant leur temps entre deux logements parentaux.

Le projet se place dans des perspectives nationale et internationale. Des contacts existent déjà avec des institutions étrangères qui utilisent et documentent des données administratives massives. Nous tirerons bénéfice de leur expérience et serons en mesure de disposer de notre propre expertise nationale, fondée sur une collaboration étroite entre nos instituts. En rendant accessibles sur un site Internet les informations sur l'accès et la qualité des données massives, et en proposant des solutions concrètes aux difficultés identifiées, le projet rendra un grand service à la communauté des chercheurs, tout en contribuant à l'amélioration de ces données. La publication d'articles méthodologiques dans des revues internationales de premier rang garantira la diffusion de nos résultats.

Les participants au projet travaillent à l'Institut national d'études démographiques (Ined), à l'Insee et dans les universités de Paris 1 Panthéon Sorbonne, Paris Descartes, Lyon, Nancy et Strasbourg. Le projet permettra de développer et de faciliter l'utilisation par les chercheurs en sciences humaines et sociales des données massives bientôt disponibles.

Summary table of participants

Institution	First Name	Second name	Current position	Mths	Role and responsibility
French Institute for Demographic Studies (INED)	Laurent	Toulemon	Research supervisor	24	Scientific coordinator Coordinator, tasks 1, 2.1, 3.2 Member, tasks 2.2, 2.3, 3.1, 4
INED	Marie	Bergström	Researcher	12	Member, Task 2.2
INED	Arnaud	Bringé	Research engineer	12	Coordinator, task 4.2 Member, Task 1
INED	Carole	Bonnet	Researcher	10	Member, Tasks 2.3, 3.2
INED	Arianna	Caporali	Research engineer	6	Member, Task 4.2
INED	Marion	Leturcq	Researcher	6	Member, Task 2.3
INED	Elisabeth	Morand	Research engineer	12	Coordinator, task 3.1 Member, Task 3.2
INED	Lidia	Panico	Researcher	12	Coordinator, Task 2.3 Member, Tasks 2.1, 3.1
INED	Ariane	Pailhé	Researcher	12	Coordinator, Task 4.1 Member, Tasks 2.1, 3.1
INED	Arnaud	Régnier-Loilier	Research supervisor	6	Member, Tasks 2.2, 3.1
INED	Anne	Solaz	Researcher	6	Member, Task 2.3
INED	Xavier	Thierry	Researcher	6	Member, Task 2.3
INED		To be hired	Post-doc 1	24	Coordinator, task 3.2 Member, Tasks 1, 2, and 4
INED		To be hired	Post-doc 2	24	Member, Tasks 1, 2, 3, and 4
INED		To be hired	4 Interns	12	Members, Tasks 1, 2, 3, and 4
National Institute of Statistics and Economic Studies (INSEE)	Sébastien	Durier	Statistician, EED division	6	Member, Tasks 2.1, 3.2. In charge of the Permanent Demographic Sample
INSEE	Céline	Leroy	Statistician, RTI division	6	Member, Tasks 2.1, 3.2
INSEE	Stéfan	Lollivier	Senior adm-inistrator	3	Member, Tasks 2.1, 3.1, 3.2 In charge of the RSL
INSEE	Gaël	De Peretti	Chief, RTI division	3	Member, Tasks 2.1, 3.1
INSEE	Isabelle	Robert-Bobée	Chief, EED division	3	Member, Task 2.1, 3.1
University of Lorraine	Cécile	Bourreau-Dubois	Professor	3	Member, Task 2.3
University of Strasbourg	Didier	Breton	Professor	12	Coordinator, Task 4.3 Member, Task 2.1
University Paris Descartes / INED	Christophe	Giraud	Lecturer	12	Coordinator, Task 2.2 Member, Task 3.1
University Paris 1 Panthéon Sorbonne / INED	Angela	Greulich	Lecturer	12	Member, Tasks 2.2, 4.3
CNRS/University Lumière Lyon II / INED	Emmanuelle	Santelli	Researcher	5	Member, Task 2.2

RTI division: division du Recueil et traitement de l'information (Data Collection and Processing)

EED Division: division des Enquêtes et études démographiques (Demographic Studies and Surveys)

Progress since the pre-proposal

Since the previous phase of the proposal, several changes have been made. They do not modify the project with respect to the pre-proposal, but they increase its ambition and relevance.

Legal framework in France and in Europe

The law on health (*loi Santé*) was voted in January, 2016, and a new institute, (*Institut national des données de santé*, INDS) will soon be created, with a specific aim to centralize demands and facilitate the use of the national set of health data (*Système national des données de santé*). The discussions on the *Loi pour une république numérique* are ongoing. These changes, together with European rules on data availability for research, to be effective this year, will likely increase dramatically the possibilities for accessing large administrative datasets. Our aims of creating an open group of users and disseminating knowledge on these datasets and on how to use them are therefore becoming even more important in this new context.

At INED, a new competence centre

With the creation of the INDS, INED is in close contact with the National Institute of Health and Medical Research (INSERM), as well as with the minister of Health to get access to health data, and we are currently discussing the practical aspects of health data availability. The decision was made to create a competence centre on administrative data use and dissemination; the person in charge of this centre and three members have joined the project.

At INSEE, a decision to broaden collaboration

The acceptance of the pre-proposal has encouraged the French National Institute of Statistics and Economic Studies (INSEE) to join the project on a wider basis. The decision to broaden collaboration has been taken at the highest level at INSEE. Statisticians from INSEE will actively work on census statistical assessment, and we will collaborate on the cross-validation of census and survey data. Task 2.1 has thus been broadened to a more global approach on family structures in France.

Involvement of the coordinator in three international projects

The project coordinator has made contact with researchers involved in statistical data dissemination abroad. As a result he has joined two new projects and one that is ongoing. The Canadian network project *PopCan* submitted by McGill University to SSHRC in February 2016, the European Starting community project *Poplife* submitted by the National demographic institute of the Netherlands (NIDI) in March 2016, and the IPUMS-International programme (a project launched in 1998 by the University of Minnesota) as a member of the International Advisory Board launched in March 2016.

New demand for use of administrative data on children

A group of experts, including members of this project, is currently working on a report to the National council of information and statistics (CNIS) in order to gain better knowledge on family disruptions and their consequences. The report (forthcoming) emphasizes, among several points, the growing need to facilitate and to use current administrative data.

Larger, more interdisciplinary team

We are grateful to the reviewers of the pre-proposal and we have slightly reoriented the project in line with their recommendations. The main change is related to the complementarity between administrative data, large-scale population surveys, and a qualitative approach to new family situations. A specific effort has been made to expand the team working on young adults' family situations, allowing the project to embrace a large spectrum of methodological approaches.

No change has been made in the structure of the project, but stronger ties will be established with existing projects in France and elsewhere in Europe and in Canada. The team has been enlarged and now includes 28 members, so that the total support requested has increased to 291,585 euros.

1. Background, positioning and objective of the detailed proposal

New family and demographic behaviours are leading to more mobile individuals and society, making it more difficult to define and observe actual family and housing situations. Simultaneously, big statistical data, *i.e.* data from several administrative files covering the whole population and elaborated for processing and monitoring individual situations, are now becoming available to the research community.

By comparison with other big data which circulate via the Internet and other information channels, known for their "low information density", big statistical data usually contain a rich set of information on individuals, but they share two important common features with other big data: they are exhaustive on their field, and they were not constructed for scientific use based on statistical inference (Groves, 2011). Administrative data include records of actions, not answers to questions; this implies that the data are more closely related to actual behaviour, and that their use for statistical purposes requires specific data management and preparation prior to analysis. In addition, INSEE is now offering access to individual-level census-based data, which are no longer exhaustive but which are weighted using exhaustive administrative data. This set of merged and connected data thus fits the definition of big data proposed by Ruggles (2014).

The project aims to extend our knowledge of some specific, hard-to-measure situations, using several data sources including big data, and to assess the strengths and weaknesses of several data sources that will be disseminated in 2016 by the French National Institute of Statistics and Economic Studies (INSEE). New collaboration between statisticians from INSEE, researchers from INED and from universities will allow multi-source analyses and multi-disciplinary approach to data quality assessment. A major output of the project will be to encourage and facilitate the use of big statistical data by students and scholars working in the humanities and social sciences (HSS).

The project comes under Axis 6 of Challenge 8 (innovating, integrating and adaptative societies: digital revolution and social changes). It aims at organising and developing the use of big statistical data for both training and research in the social sciences. Our ambition is to promote the use of these big data, which remain underexploited. The aim is also to bring new answers to three classic questions in demography and to assess the scientific quality of available data, by cross-checking the results across different statistical sources. It corresponds to the ANR orientation n° 32: "*Disponibilité des données et extraction de connaissances*".

1.1. Background

Family situations have become more ambiguous and complex

The increasing complexity of life histories is making it more difficult to observe patterns of demographic behaviour and family situations, and the definition of the events that mark the life course is changing. For instance, couples are no longer defined by their marital status, since unmarried cohabitation has become common, first as a temporary state at the beginning of a couple's life, and now as a lasting form of union. The recognition of same-sex unions, with the introduction of the civil partnership (*Pacte civil de solidarité*, PACS) in 1999, and the legalisation of same-sex marriage in 2013, have contributed to changing the definition of the couple. Non cohabiting Living Apart Together (LAT) relationships have also been described as a new type of union. This diversification of intimate relationships is associated with a change in family transitions, which are increasingly gradual and less ritualised.

Complex or ambiguous family situations are often linked to multiple places of residence. For example, some people in a couple keep their own place of residence and spend only a part of their time with their partner (Levin 2004; Caradec 1997; Régnier-Loilier, Beaujouan, Villeneuve-Gokalp 2007; Duncan *et al*. 2013; Giraud 2014); The spread of new forms of partnership including partial coresidence or the use of two different dwellings has become a topic of interest in many countries. Couples Living Apart Together (LAT) correspond to new models of residency and union definitions that have become increasingly common among young adults (Roseneil, Budgeon 2004) as well as among elderlies (De Jong Gierveld 2004) as a result of not only increasing individualism in family

behaviour (Bauman 2003; Duncan 2011), but also new forms of coresidency and flat sharing (Bruun 2011). LAT relationships are now identified as a specific family form in many countries like for example England and Wales (Haskey 2005), Spain (Castro-Martín, Dominguez-Folgueras, Martin-Garcia 2008); United States (Strohm *et al*. 2009), Canada (Milan, Peters 2003) and Australia (Reimondos, Evans, Gray, 2011).

Children may also share their residence between different households. After parental separation, court decisions increasingly encourage ongoing relationships between children and both parents through shared custody and alternating residency. This trend is increasing in many developed countries (Kitterød, Lyngstad, 2012; Cashmore, Parkinson, Taylor 2008; Cancian *et al*. 2014). In France, shared residence of children is the solution chosen for almost one divorce out of five when minor children are involved (Carrasco, Dufour 2015; Bonnet, Garbinti, Solaz 2015). These situations make it difficult to accurately observe family situations based on the "main place of residence", a definition used in the census and in almost all statistics produced on family situations. In general population surveys, some 7% of inhabitants declare more than one "usual residence", the situation being most common around age 18 (Toulemon, Denoyelle 2012; Desplanques 2008). Individuals may commute between their different residences on a weekly, monthly or even yearly basis (Imbert *et al*. 2015).

Surveys on family situations complement basic counts based on census data. Most studies based on survey data do not consider the family situation outside the household, and the sampling frame does not take account of the fact that inhabitants with more than one usual residence are more likely to be included in surveys when they can be reached in two dwellings (Toulemon, Pennec 2010). In recent INSEE publications on families, efforts have been made to use different datasets (Bodier *et al*. 2015), and analyses on consistencies between survey and census data are now performed. Furthermore, "inconsistent" results are becoming increasingly acceptable. For instance, the census includes two definitions of "living as a couple" for respondents aged 14 or more. An explicit question asks whether people "live as a couple" and thus provides an indicator for couple life. However, the overall analysis on families and households introduce two specific constraints: both partners of a couple must declare that they live as a couple (or are married), but they must also be included in the same household in the census, and be of different sexes (see www.insee.fr, "definitions"). These definitions are subject to update: for instance, same sex couples will be included in the definition of "couples" in census results based on the 2015 annual survey onwards. New family forms constitute a challenge for census definition, which must evolve in order to identify new emerging forms of households, with strong constraints of time consistency and international comparability (Freguja, Valente, 2010). Household surveys use a different definition of "household membership" (including people living "usually elsewhere") and "life as a couple" (explicitly distinguishing partners living permanently in the household, partially or not at all). A file merging all household surveys using the same core questions on household membership and members' family situation is being constructed, but no results based on this source have been published by INSEE, and very few methodological studies have been published (Chardon, Vivas 2009; Toulemon, Denoyelle 2012; Trabut *et al*. 2015).

Almost all surveys on family situations and behaviours in France are conducted by INSEE and INED (often together). INED has a long tradition of scientific publications on family structures and fertility behaviour in France and in Europe, and the Family surveys conducted in conjunction with the 1999 and 2011 censuses were organised by INSEE and INED together (Rault *et al*., 2010). Recently, INED and INSEE conducted a joint survey on families, the *Enquête sur les relations familiales et intergénérationelles* (ERFI), the French component of the European Generation and Gender Programme (GGP, see Vikat et al. 2007), based on a first wave of interviews in 2005 (Régnier-Loilier 2014) and two three-year follow-ups in 2008 and 2011 (Régnier-Loilier 2016). INED organized the Study on individual and conjugal trajectories (*Étude des Parcours Individuels et Conjugaux*, EPIC) in 2014 (Bergström 2016). INSEE was involved in the survey team. These surveys are complemented by qualitative analyses based on in-depth interviews: team members include Christophe Giraud, Emmanuelle Santelli, Marie Bergström and Arnaud Régnier-Loilier, who all have conducted and analysed such interviews for their research. In-depth interviews allow understanding how people make sense with their situations and biographies, with all their complexities. They complement administrative data, census and surveys and such interviews have linked to ERFI and EPIC surveys, on a group of respondents chosen for their situation as observed in surveys. In-depth interviews may

help understanding how peculiar situations are coded in quantitative data, and how to adapt survey questionnaires (Manning, Smock 2005).

Increasing availability of administrative data

These new patterns of behaviour oblige us to adapt and build upon our usual data collection processes, currently based on population censuses and civil registration data. The growing availability of administrative and fiscal big data provides opportunities for innovation (in methods as well as analyses) in social and demographic research. The law on health (*loi Santé*) adopted in January, 2016 and discussions on the law for a digital republic in an information-based society (*Loi pour une république numérique*) are ongoing. Together with European rules facilitating data availability for research, that comes into effect this year, these laws will likely significantly extend the opportunities for accessing large administrative datasets.

This trend towards access to administrative data is not confined to France. Some countries already have a long tradition of population registers and data merging for research, especially in Scandinavia, but each set of data offers its own opportunities for research. Canada and Belgium are now on a process of opening more of their administrative data for research purpose (Citro 2014). In France the set of administrative data currently looks like "big statistical data", as no central population register exists: administrative files are seldom linked with one another, and most studies conducted up to now did not focus on an assessment of the statistical quality of each dataset.

Big data produced at INSEE are still under-used

Within INSEE, the process of integrating big data from different administrative and fiscal sources offers the opportunity for a dynamic, longitudinal observation of individual family situations and their changes over the life course. This will bring to light certain situations that, while rare, may be typical of new demographic trends.

Responding to strong incentives to use big data (principles 9 and 10 of the European Statistics Code of Practice), INSEE has recently developed considerable expertise in this area. Administrative data replaced specific surveys for annual business statistics in the 2000s (project RESANE).

INSEE has a long tradition of using big data and merging administrative and statistical files. The Permanent Demographic Sample (*Échantillon démographique permanent*, EDP) brings together, for a one-percent sample of the population, data from population censuses, the civil registration system, the electoral roll and administrative data on employees and their employers (DADS). With the move to annual census surveys in 2004, the EDP has changed dramatically, as individuals are no longer all included at the time of the census but randomly each year. Since 2008, the sample has quadrupled in size (one in 25 individuals are now included). The inclusion of annual social and fiscal data in 2016 has transformed the EDP into a very efficient instrument for research. It is now available via a secure remote access system (Secure Data Access Centre, CASD). Further enrichments with health data (from hospitals stays, medication use, and causes of deaths) are foreseen.

INSEE is also creating a file that merges the core questions on household members systematically included in most household surveys: list of household members, family ties, frequency of presence in the household, and existence of another "usual dwelling". This file, Households surveys core questionnaire, (*Tronc commun des enquêtes ménages*, TCM), homogenised since 2004, is designed to allow merging of samples and creation of a big dataset. It is currently being documented and should be released for research purposes in 2016, through the French Data Archive for Social Sciences (*Réseau Quetelet*).

INSEE has also set up a specific project for the creation and assessment of a Statistical Register of Dwellings (*Répertoire statistique des logements*, RSL) providing an annual file of fiscal data (mainly income tax and housing tax) of inhabitants and their demographic events over a year (birth of a child, death, marriage, divorce, PACS, union disruption), enriched with data on the dwellings, as well as a file identifying all the dwellings to which an individual is linked for fiscal purposes. Based on the same complete fiscal data, merged with data from family allowances (CNAF), the Social and Fiscal Localised File (*Fichier localisé social et fiscal*, FiLoSoFi) includes information on the total household

income and standard of living but can currently be used only at the household level. The release of an individual-level dataset is planned.

With the exception of the Permanent Demographic Sample, these different datasets are still largely under-used in France. The current project will be the first comprehensive study of these datasets, and aims at disseminating their use among researchers in the social sciences.

1.2. Origins and foundation of the project, new directions

Previous collaboration between INSEE and INED

INED and INSEE have a long tradition of collaboration. INED researchers have been seconded at INSEE (including the principal investigator for this project), most surveys on family behaviour have been conducted by consortia including INSEE and INED members. The 2011 Family survey, conducted in conjunction with the census, was an opportunity to study multiple residency in the census (Breuil *et al*. 2016; Imbert et al., 2014; Trabut *et al*., 2015). Joint construction of the data file was partially funded by ANR: (from spatial to social ties, *des Lieux aux Liens*, LiLi, 2011-2014) (see http://lili-efl2011.site.ined.fr/en/). A book is currently under review.

INED and INSEE are in close contact on the publication of the INED annual report on the Demographic Situation in France (Mazuy *et al*. 2015), mostly based on the analysis of INSEE data, while the annual report on the Demographic balance (Bellamy, Beaumel 2015) is reviewed at INED. Two Ined researchers have worked on a joint project with INSEE using the fiscal data, and participated to the working group for a better knowledge of family disruptions and consequences.

New challenges, new collaboration

This project will be the first comprehensive study of these different sets of data, and aims at disseminating their use among researchers in the social sciences. The current period is a time of great opportunity: new laws are being adopted in France, in compliance with European rules. INSEE is currently willing to play a central role in dissemination of data, INSEE statisticians are working at INED (including the current director) and our previous director is now, back at INSEE, director of Demographic and Social Statistics (DSDS). This marks the start of a period where a spectacular increase in data availability will meet a common ambition to collaborate for greater efficiency in the use of this new opportunity for basic statistics as well as for research. The project aims at producing tools and resources which may be useful to a large audience of scholars, in France and abroad.

1.3. Objectives of the project

Focusing on three research questions related to residence

In this context of change in individual and family behaviour, the project addresses three questions that are key to describing current demographic trends in France, and that are well known to the project members. These three topics call for the use of big data, since they concern small subgroups within the population that are poorly identified when only one data source is used. They will be presented in detail in the next section. They have been chosen because they all correspond to a specific challenge in terms of definition and analysis, and their importance is likely to grow in the future. The issue of residence is at the heart of these three questions.

The first question deals with accurate counting of population size, the omissions and double counts in the census, and the consequences of the "main residence rule" (each inhabitant must be included in the census in one and only one household) on the definition of households and families. The project will give the team the opportunity to assess the omissions and double counts in the census, for the first time since the adoption of annual census surveys, and to compare the description of family structures in the census and other data sources. The second question relates to young adults, the group in which multiple residency is most common, in relation to the definition of living as a couple: when do young adults consider themselves as living as a couple, and when are they considered as such? The recent decrease in age at union formation among young adults (Rault, Régnier-Loilier 2015) could be an

artefact due the change in the definition of "being in a couple" in the recent cohorts. The third question deals with the family situation of children whose parents are separated. Statistics from the ministry of Justice show a sharp increase in court decisions leading to shared residence (Kesteman, 2007; Guillonneau, Moreau 2013), but no increase is visible from survey data (Toulemon, Denoyelle 2012). Double counting of children with two parental homes is very likely in household surveys, which strongly bias the analyses.

Other topics could be of interest, and other situations, such as people living alone (Toulemon, Pennec, 2011) or lone parenthood (Buisson, Costemalle, Daguet 2015) have recently been studied. These questions will be partially covered by our first topic which will include an overview of family structures. Our second and third topics are considered in this project because a large group of colleagues are planning to work on them in the coming years. The project will encourage team members to exchange their expertise and their knowledge of the datasets they use.

Complementarity of administrative data and surveys

The team will benefit from the use of very different data sources. Administrative data contain very precise information on some behaviours, transitions and situations, but produced for non-scientific purposes. Census data include answers to self-completed forms; survey data come from personal face face-to to-face or telephone interviews. Qualitative analyses are based on in-depth personal interviews. The complementarity of these approaches is twofold. First, they relate to different definitions of residence and of family situations; second, they focus on different questions, from the analysis of "actual" behaviour" based on official declarations, to more nuanced individual testimonies and thoughts in in-depth interviews.

Assessing the quality of administrative data

The collaboration between statisticians from INSEE, demographers and sociologists on these research questions will build bridges between these complementary approaches. On the one hand, administrative data offer unbiased data on specific items, and provide a means to check survey results which may be affected by several biases. On the other hand, administrative data have present two potential weaknesses. First, individuals usually "play with the rules" and the "truth" coming from administrative data may be very different from the actual situations of individuals. Whether children are declared in income tax declarations as present in one or the other parental home, and whether young adults are declared in the parental home or in another household, for instance, may depend on the way they affect the parents' tax liability. Second, administrative data may be limited or inconsistent, but merging different sources often makes it possible to overcome these limitations. In France, for example, couples who are married or in a civil union (PACS) declare the couple's income on a single form, while other couples fill in two separate forms. Unmarried couples are thus not identified in income tax data but can be detected by merging income tax and housing tax data. Despite some differences in the timing of declarations, such data merging is crucial.

The use of big statistical data comes with several challenges. The first one is the lack of accurate information and documentation on all variables and the need to provide other users with detailed information. The second challenge when using such huge datasets is to prioritise the numerous data problems to be corrected (missing or inaccurate data) according to their potential impact on the analyses and the results. The third one is to deal with internal inconsistencies of the database, and to propose solutions in order to get a coherent set of information. The fourth challenge is to see to what extent the data is nationally representative in its cross-sectional version or panel dimension, and on which group of population. Weighting procedures can be needed, as in surveys, in order to make the data representative. On all these points, the members of the research team built thanks to this project will be able to cooperate and efficiently profit from the experience of other members. The project is thus focusing on datasets produced and already well known by team members, in order to be efficient in data assessment, documentation and dissemination, and to share experiences learned from new research using and comparing these data sets.

Sharing, disseminating and training

In some cases inconsistent results from different types of data and analyses may be interpreted as the consequence of different definitions; in other cases, the inconsistency comes from a bias in one data source. The project aims at reaching a shared diagnosis, focusing on some situations studied for many years by the team members.

A prominent aim of the project is to disseminate knowledge, tools and resources for a wide audience. A specific website will be created, where easy-to-use tools will be made available to access administrative data available through the French secure data access centre (*Centre d'accès sécurisé aux données*, CASD) or other similar platforms (a project has been launched at INSERM on health data). It will provide links to the platforms, users' guides, and documentation of the files. Methodological papers on the quality of the datasets will be available, as well as source codes for the creation of standard variables from different data sets, and a statistical evaluation of their coverage and quality. INED has a extensive experience in this area. See, for instance, users' website of the Elfe child cohort (*Étude Longitudinale Francaise depuis l'Enfance*) that includes documentation for understanding the survey, codes to recreate constructed variables, as well as explanations of the data structure, how to request and access data, and Elfe news (http://util_elfe.site.ined.fr/en/), partially funded by the ANR project Veniromonde (2011-14). These resources may be useful to a large audience of scholars, in France and abroad. They will include information provided by the users' groups which already exist for some datasets (like the Permanent Demographic Sample), with the support of the Competence centre currently being launched at INED. This centre includes INED expertise in documenting and disseminating surveys in compliance with international standards like the Data Documentation Initiative, DDI (Caporali, Morisset, Legleye 2015). See, for example, the online codebook of the Generations and Gender Programme (GGP) surveys (http://www.ggp-i.org/online-data-analysis.html).

In parallel, training for PhD students, postdoctoral researchers and other users will be offered in association with INSEE and CASD. Didier Breton is setting up an academic data platform (*Plateforme universitaire de données*, PUD) in Strasbourg. A summer school will be devoted to the use of big data.

International collaborations

The writing of the proposal was an occasion to make contact with several teams working on administrative data abroad (see progress since the pre-proposal above). Similar projects are ongoing in several European countries. The current project provides opportunities for collaboration with European consortia and national teams, and will enable us to participate effectively in international research projects based on big data in the future.

1.4. Three research questions for which big statistical data are needed

Double counts, omissions and family structures in the new census

The first issue concerns total population size and structure. Since 2004, the complete population enumeration has been replaced by annual census surveys. The new census method, based on a regularly updated list of residential buildings and conducted by more "permanent" teams of census agents than in the past, has probably reduced the number of omissions. Conversely, there are probably more double counts of individuals who habitually live in more than one dwelling, because these dwellings are unlikely to be included in the same annual sample (Desplanques 2008; Toulemon 2012). So far, no estimate of double counts and omissions in the new census has been published. Using big data and data merging, it is becoming possible to produce estimates of this kind, the aim being to assess the statistical accuracy of key operations for public statistics. Three datasets will be used for this purpose.

First, the Permanent Demographic Sample may include some individuals twice from the same annual survey, if a form was collected for the same person in two different households. The probability that both households will be selected in the same year depends on the selection probability of each household (information present in the dataset) and the correlation between both probabilities (which has to be calculated from the sampling frame of the census). The probability of double selection ranges

broadly from 1in 5 (two households in the same small town) to 1 in 156 (two independent households in different large cities), setting aside specific situations. Preliminary studies (internal memos) have been produced at INSEE (Mambetov, 2014), showing that 0.25% of inhabitants were included twice each year, but with no estimate of the double count probabilities. A preliminary estimate has been proposed, based on several assumptions (Toulemon 2016), but the team will produce a more accurate estimate, based on complete information on selection probabilities of households, in close collaboration with and under the responsibility of INSEE team members. For the years 2004-2011, more than 8,000 individuals were included in the census twice in one annual survey. This sample is large enough to analyse the two household situations of these individuals, as the Permanent Demographic Sample contains census information for all household members in each household.

Second, the statistical register of dwellings includes more tax payers (including their beneficiaries living in the same household) than the number of inhabitants given by the census (Lollivier, personal communication). Some tax payers may be considered as not being "permanent inhabitants" according to the census definitions, and an estimate of omissions in the census will be produced, after a precise examination of the discrepancies and census weighting rules (annual and five-year weights will be provided by INSEE, under a specific agreement). Statistical matching between the two files will be performed to identify the characteristics of people present in one dataset, and not in the other. The Statistical register of dwellings will be updated and renamed as the Demographic Dataset on Households and Individuals Based on Tax Information (*Fichier démographique d'origine fiscale sur les logements et les personnes*, FLP), and income data will be added, including the type of income, and monetary exchanges with other households. These will not be panel data like the EDP, but information is present for two successive years in each annual dataset, making it possible to study current situation and flows from one year to the next. The file will be used inside and outside INSEE, through remote access at CASD planned for 2017.

Third, the Household surveys core questionnaire data file (TCM), which contains information on all household members (time spent in the dwelling, family ties with other members, sociodemographic information), will be compared against census results on family structures. This comparative work has already started at INSEE and INED, with a focus on double count estimates (Lapinte 2013, Leroy 2015, Toulemon 2016) and data management. A project is ongoing at INSEE on a new version of the census housing form (three members of the present project are involved in the Scientific Committee), and a precise analysis of family ties and complex households from these household surveys and from the census is currently under way. A broad analysis of family situations in France, based on current situations drawn from censuses, tax data and survey data, will be conducted as part of the project, as an addition since the pre-proposal. The colleagues at INSEE who joined the team already have this work on their agenda. The current project will formalise collaboration on this topic, and allow a large dissemination of the documentation of the TCM dataset and methodological studies, in addition to data release at CASD (and possibly at the *Réseau Quetelet*, for an anonymised version of the file).

Couple status of young adults

The second research question concerns young adults living as a couple. Over the last ten years, the trend towards ever later union formation has stalled and even reversed (Brückner, Mayer 2005; Daguet, Niel 2010; Rault, Régnier-Loilier 2015), with large differences across socio-economic groups (Pailhé 2015; Sironi, Barban, Impicciatore 2015). The process of couple formation is complex and the intimate life of young adults no longer implies living with a partner. The recent increase in early unions could be related to a new form of romantic relationship which involves sharing the same residence "as partners or friends" but without any plans to enter a civil partnership or marry, let alone start a family in the short term. In such relationships, the issue of union duration is not explicitly raised. Unstable and precarious employment statuses, as well as individual mobility constraints (e.g. related to ongoing education or finding a job) make separation a possible scenario. If relationships of this kind are growing in number, unions among young people will be particularly unstable, and their disruptions may often be related to a geographical move. This challenges the definition of these relationships as "unions", that usually imply a lasting commitment. Administrative data linked to the census provides better information than surveys on changes in couple status and the spatial mobility of young adults.

The underlying hypothesis of this study is that of a growing heterogeneity of young couples, and an increasing mean risk of union disruption at early stages of adult life. The volatility of relationships may be linked to precarious or temporary situations such as being a student or looking for a job. Like for short term job contracts, the existence of short-term couples could be linked either to a phase of exploration considered as part of "being young", or the result of difficulties in committing and settling down due to economic hardship and job insecurity. We will explore this hypothesis using two complementary methods. First, administrative data on salaries, census data on student status, transition data from the EDP and tax data providing information on couple status for two successive years may all be very efficient tools for identifying the economic conditions of couple formation and dissolution. We will use this data and compare the results to those obtained from two household surveys recently conducted at INED on couple situations and trajectories (Rault, Régnier-Loilier 2015; Bergström 2016). Previous analyses were based on the comparison of the first sexual partner and the first cohabiting partner (Toulemon 2008; Bozon, Rault 2012), but the new survey will consider other types of partnerships. The triangulation of these different sources will shed new light on the terms in which young people enter romantic relationships.

Second, in-depth interviews made or planned for by team members will be used to identify subjective definitions of what a "couple" is and what "being in a relationship" means. Sociologists who work on romantic relationships before cohabitation, on couple commitment at the beginning of the first cohabiting union, and on couples "living-apart-together" (LAT) have joined the team and will participate in the evaluation of current trends: is a new form of cohabiting union becoming more common among young adults? After decades of increase in unmarried cohabitation and pre-conjugal sex, are we now confronted with a new form of pre-institutional relationships that precede "first unions"? The study of "complex households" including flatmates (people living together without any romantic relationship), and the analysis of the different homes of young adults, identified as having more than one "usual residence" in the census and in surveys, will complement this qualitative study of family situations of young women and men and their understanding of what a couple really is.

These different types of data may provide different types of evidence since they do not capture the same thing. The census registers couples following a twofold definition (individuals living as a couple; couples within households). Household surveys identify couples who live together all the time, part of the time or not at all. Tax data identifies flatmates. A typology of couples could then be made, distinguishing settled and "starting" couples, couples with a long-term commitment and temporary couples. The combination of large datasets, which include information on residential transitions and on economic and social situations, and in-depth interviews, will provide important pointers for interpreting current trends among young adults with different social backgrounds.

Children with shared residence

The third question relates to the increasing prevalence of shared custody of children between separated parents, as shown in statistics published by the Ministry of Justice and in fiscal data. This trend is not apparent in demographic surveys, suggesting that children may be declared by both parents as living with them, leading to double counts, irrespective of their actual residential situation. Fiscal data offer a complementary image of these situations, as the "family quotient" system explicitly offers shared custody as an option (Bonnet, Garbinti, Solaz 2015). In order to determine the post-divorce family situation of children, it is important to accurately describe multiple residency. Income tax data can be used to estimate the proportion of children declared by one or two parents, part-time or full-time. Surveys identify children with shared residence. The Permanent Demographic Sample gives information on both the parental homes of children included twice in the census. According to surveys, around 15% of children are living in multiple residence after a parental disruption, an estimate consistent with court decisions (Carrasco, Dufour 2015). Not only are the crude estimates from surveys much higher, due to double counting, but the corrected estimates vary largely between surveys, from 11 to 21% (Toulemon, Denoyelle 2012). The use of administrative data will shed light on these inconsistencies.

After a report published in 2014 by the High Council on Families (Haut conseil de la famille) on family disruptions (HCF, 2014), The French national council on statistical information (*Conseil national de l'information statistique*, CNIS) launched in March 2015 a working group on "Improving

statistical observation of family disruptions and their consequences on living condition of families". The group is lobbying for a better description of broken families and their living conditions (Jeandidier, Sayn, Bourreau-Dubois 2012; Fontaine, Stehlé 2014), based on their current situation as well as on the changes occurring after a disruption, using follow-up studies (Bonnet, Garbinti, Solaz 2015), and for international comparisons. One rapporteur and two members of the group have joined the team and the project aims at publishing scientific papers on this topic, with special attention on how these situations are identified in tax data (Bonnet, Garbinti, Solaz 2015), in the census (Toulemon 2011; 2012; Albouy, Breuil-Genier 2012; Bodier *et al*. 2015), and in surveys (Amossé, de Peretti 2011; Lapinte 2013; Trabut, Lelièvre, Bailly 2013). The issue of double counting is very important, as children living part-time with one parent and part-time with another are very likely to be counted twice (Toulemon, Denoyelle 2012). Members of the group will work on changes in families' standards of living after a union disruption, and will probably participate in a specific survey on these families in the future. This survey will probably make use of administrative data in several ways: to construct the sample, to include information on monetary exchanges registered in tax data, and to ensure follow-up after interviews. The survey is beyond the scope of this project, but it is an example of what can be done with administrative data, and how synergies between different data sources are useful for research.

2. Scientific and technical programme, project organisation

2.1. Scientific programme and project structure

The programme involves 16 researchers from INED and universities (including 2 post-docs to be hired), 3 research engineers from INED, and 5 key statisticians working on population datasets at INSEE. Four interns will also be hired, for a grand total of 243 person-months.

INED is responding to this call as a single institutional partner. Team members from INED all have a good experience using survey data; some have already worked on census or administrative data. Project members also include five university scholars having strong ties with INED, (research collaborations, members of our Labex iPOPs, secondments at INED). They have all worked on the research topics presented above, using diverse quantitative databases and their own qualitative data. They are already strongly involved in one or another research topic, and they are all committed to sharing and transmitting their know-how.

INSEE, which cannot apply as it is a part of the Ministry of Economy, Finance and Industry, was in the pre-proposal represented by a senior statistician: a former Director of Social and Demographic Statistics at INSEE and currently in charge of the creation and assessment of the RSL. After the pre-proposal acceptance, the INSEE teams in charge of the Permanent Demographic Sample (Demographic Studies and Surveys division) and of the Household Core Questionnaire Database (Data Collection and Management division) have joined the project. In both cases, the chief of the division as well as the person in charge of the dataset joined the group, to stabilize the institutional involvement of INSEE. The team in charge of the census could not join the project, but the evaluation of population counts based on census is currently being done primarily by team members of this project. The Demographic Studies and Surveys division is in charge of the annual demographic balance based on estimates coming from census data for population counts, from civil registration data for births and deaths, and administrative data from the ministry of Home affairs for immigration counts. The RSL is currently at INSEE the main database challenging the census in terms of population counts: it includes all people paying taxes on housing or incomes, and the count based on this dataset exceeds population at census. Works on population counts consistencies based on these two datasets are complemented by comparisons of census and household surveys data on family structures, under the supervision of team members. Team members from INSEE are thus statisticians in charge of three of the four main datasets used in this project. This will allow for the first external evaluation of the census based on empirical statistical methods, since it was renewed as annual census surveys.

The close collaboration between team members from INED and INSEE will ensure that the shared diagnosis on data assessment will be published in main scientific journals, and disseminated in our website. Specific summer courses will be devoted to efficient access to and analyses from the datasets used in this project.

Besides the Project management (task 1), the project is divided in three tasks. Task 2 is devoted to scientific analyses based on numerous and diverse data sets. These analyses are already ongoing, and each team member will provide, in addition to her/his own research papers, information on the strength and weaknesses of the data set used.

The project will start by launch meeting where all members will present their ongoing works and participation in the project. This will lead to a survey of potential databases which could be useful for the topics under study, after screening of their availability, usefulness and potential quality, as the availability of new datasets after 2016 may already be envisioned. Four internal annual one-day seminars will be organised, in order to exchange results and analyses, and to see what topics would need a specific focus in term of new results, as well as on data comparison and checking. A common culture of data accountability will thus be organized within the group members.

The advisory Board will meet at least once a year, in close relation to the competence centre on administrative data use and dissemination, including a INSEE senior manager, as well as other big data providers (CNAF, CASD), in order to share and update our knowledge on administrative data which are or could be made available in the near future.

In addition, open seminars will be organized in years 2 to 4, in order to exchange experience with scholars using the same or other administrative datasets. A users' group meeting will take place at the same time, to ensure that the website is useful and to encourage users to send feedback and additional materials based on their own experience. The final seminar will take the form of an international closing conference.

Support from the ANR is needed mainly for the recruitment of two postdoctoral researchers for 2 years each, working simultaneously on analysing data, assessing their strengths and weaknesses and creating a complementary body of documentation, alongside that offered by the data producers, in order to facilitate data access and use. The cost of these post-doctoral fellows and interns is 184,000 euros for 60 months. Total personal cost is 1,228,000 euros, including 243 person-months related to all team members. Other funds will be allocated to cover the cost of remote data access for the project members (26,000 euros). Finally, a budget is requested for missions and project meetings, as well as participation in international conferences for some project members, and some equipment (50,000 euros). The support requested from the French National Research Agency (ANR) is 291,585 euros, including INED overheads. INED will manage the budget, through its International affairs and partnerships department (DRIP).

2.2. Project management

The annual meetings of the team members, the advisory Board and the open seminars will be organized by Laurent Toulemon, project coordinator, and Arnaud Bringé, leader of the competence centre on administrative data use and dissemination which is currently being launched at INED.

Project coordinator

Laurent Toulemon (INED), scientific coordinator of the project, is a research supervisor at INED and head of its "Fertility, Family and Sexuality" research unit. He worked for three years at INSEE, organising a one-percent family survey alongside the 1999 census, and preparing an updated and standardised core questionnaire for household surveys (Herpin, Toulemon, Verger, 2001). He collaborated with colleagues at INSEE on this database and prepared its dissemination. His own analyses of this dataset explored the topic of multiple residency, cross-checking different surveys and statistical data. He was involved in the INSEE survey on families conducted alongside the 2011 annual census survey. He joined the INED competence centre on administrative data use and dissemination. His role will be to coordinate the project, in relation to other initiatives regarding big statistical data at INED and elsewhere, in France and abroad.

Participants in the project

Marie Bergström (INED) is a researcher at INED since 2015 and member of the research units "Family, Fertility and Sexuality" and "Demography, Gender and Society". Her research focuses on social and gender disparities in couple formation. Recent projects revolve around experiences of singlehood and young people's entry into stable romantic relationships. A part of her research addresses methodological questions regarding the use of "big data" and is based on a study using data from online dating sites to assess patterns of partner selection.

Arnaud Bringé (INED) is the chief of the Statistical Methods service at INED. He has been appointed as the head of the INED competence centre on administrative data use and dissemination; the competence centre will centralize current and future use of administrative data at INED. It will help researchers intending to use administrative data, and will be the main interlocutor for researchers from other institutions providing administrative data or participating it their dissemination He joins the project to organize synergies between the project and the activities of the centre. He will take prior responsibility of the project website.

Carole Bonnet (INED) is an economist and a researcher at INED, specialised in retirement and ageing, and a researcher affiliated with France's Public Policy Institute. She currently heads the economic demography unit at INED. Her recent work has been concerned with gender inequality (in retirement pensions and wealth), the family component of pension policy, and the impact of family events (widowhood and divorce) on standard of living.

Arianna Caporali (INED) is a Research Engineer in the Surveys support service, in charge of providing access to survey data and of developing contextual databases, in relation with the TGIR PROGEDO. She is a member of the INED competence centre on administrative data use and dissemination. She will bring to the team her experience in metadata preparation and creation of online datasets. For example, she maintains the online codebook of the Generations and Gender Programme (GGP) surveys and participates to its Contextual Database.

Marion Leturcq (INED) is a researcher at INED since 2013. She is specialized in the economic impact of the legal treatment of marriage and civil unions, and in gender inequalities. She is currently working on the impact of marriage and marriage contracts on wealth accumulation, but also on the gender wage gap, and on the links between childbearing and labour force participation of women. She defended her thesis in 2011 at the Paris School of Economics and CREST.

Elisabeth Morand (INED) is a data scientist at INED, member of the Statistical Methods service. As a statistician, she participates in the dissemination and use of open and big data. She is a member of the INED competence centre on administrative data use and dissemination. During 2014-2016 she co-organised seminars on techniques to use big data and open data (http://methodes-et-logiciels.sfds.asso.fr/archives/). As a research engineer at INED, she is involved in data management and data mining from different sources used at INED like the ELFE children cohort.

Lidia Panico (INED) is a researcher at INED since 2012 and member of the research units "Family, Fertility and Sexuality" and "Economic demography". She is responsible for the INED key research project "Children and Their Families". Previously, she was an ESRC Post-Doctoral Research Fellow at the London School of Economics, and a Research Fellow at University College London (UCL). She obtained her PhD from UCL, where she was based in the International Centre for Lifecourse Studies. She is actively involved in the new French birth cohort, ELFE, and in the creation and dissemination of constructed socio-economic variables for this new data source.

Ariane Pailhé (INED) is a research supervisor at INED and member of the research units "Economic demography" and "International migrations and minorities". Her current research fields are gender relations within households, patterns of transition to adulthood and living conditions of children. She analyses the effect of employment status on family formation. She also leads the thematic group Economic Status & Insecurity of the ELFE cohort survey (INED-INSERM-EFS) and is editor in chief of the INED website.

Arnaud Régnier-Loilier (INED) is a researcher at INED since 2003. He was in charge of the implementation of the French INED-INSEE Generation and Gender Survey, *Étude des relations*

familiales et intergénérationelles, ERFI), a panel study based on three waves of data collection in 2005, 2008 and 2011, and coordinated, with Wilfried Rault, a new INED-INSEE survey on couple formation and dissolution (*Étude des Parcours Individuels et Conjugaux* 2013-2014, EPIC), a survey with an important retrospective part on couple biographies. Both surveys were completed with qualitative works based on in-depth interviews. He is currently working on couple behaviour, especially on Living-apart-together couples, among young and older adults, using different data sources.

Anne Solaz (INED) is an economist and a research supervisor at INED at the "Economic demography" and "Family, fertility and sexuality" research units. She is co-editor in chief of the journal *Population*. Her recent publications in economics and demography concern the evaluation of parental leave policies, fertility and remarriage, time allocation in couples over the life cycle, and gender inequality on the labour market and in the private sphere. She also currently works on the economic consequences of divorce.

Xavier Thierry (INED) is a demographer and a coordinator of the ELFE children cohort. His research focuses on family dynamics using quantitative data (Charles *et al*. 2011). He used to work on international migration statistics based on statistical surveys as well as administrative data. He is familiar with linkage procedure at the individual level and also builds contextual local database for ELFE surveys. His recent publications concern breastfeeding. He also currently works on child care arrangement among migrant population.

Sébastien Durier (INSEE) is in charge of the Permanent Demographic Sample (EDP) within the Demographic Studies and Surveys division. He organizes the EDP users' group. His participation in the team will ensure close collaboration for the assessment of doubles counts in the census based on the EDP, as well as an assessment of its quality, after the recent enrichment of the EDP with administrative data.

Céline Leroy (INSEE) is a statistician in charge of the Household Core Questionnaire Database (TCM) within the Data Collection and Processing division since 2015. She worked on the adaptation of this core questionnaire to longitudinal surveys and on the weighting procedures that could be improved in order to take into account double counts in household surveys. She also collaborates on the assessment of family structure estimates based on a comparison of household surveys and census data. Her participation in the team will facilitate the dissemination and documentation of the TCM database within the CASD and the *Réseau Quetelet* in 2017. The project website will be the tool for sharing experiences and statistical assessment of the data.

Stéfan Lollivier (INSEE) is a former Director of Social and Demographic Statistics at INSEE and currently in charge of the creation and assessment of the He wrote in 2004 a seminal report on panel data at INSEE based on census, surveys and administrative data (Chaleix, Lollivier 2004; 2005), which led to the broadening of the Permanent demographic sample, in size as well as in content, and to the creation of the Demographic dataset on households and individuals based on tax information and the Social and Fiscal Localised File. He is currently in charge of the evaluation of census and administrative data. He is the best connoisseur of the statistical quality of the new census and of the opportuinities related to several administrative databases. His participation in the project will facilitate the dissemination of his work outside INSEE, as well as further collaborations with INED.

Gaël De Peretti (INSEE) is the chief of the division Data Collection Processing division in the Department of Statistical Methods at INSEE. The division is responsible, among other subjects, of survey methodology issues, and as such, is responsible for the Household Core Questionnaire Database (TCM). Statistician, specializing in survey methodology issues, he participated in the design of many official statistics household surveys and has occurred or occurs as such in different scientific committees (ELFE, ELIPSS, etc.). His participation will facilitate collaborations between household survey producers within INSEE and the project team.

Isabelle Robert-Bobée (INSEE) is the Chief of the Demographic Studies and Surveys division, in charge of the EDP. The division is also publishing the annual demographic balance, and of population projections. She worked on fertility and family behaviour in France, as well as on mortality

differentials. Within the team, she will work on young adults couple situations assessment and ensure an efficient availability of the EDP.

Cécile Bourreau-Dubois (Université de Loraine) is professor in economics at the Université de Loraine (within the Bureau for Economic Theory and Applications, *Bureau d'économie théorique et appliquée*, BETA). She is specialized in family law (especially in case of divorce) and social policies (on poverty, indebtedness and old-age dependency). She is rapporteur of the CNIS group on family disruptions, which pleas for a new survey on divorced couples and parents, based on links with administrative data.

Didier Breton (University of Strasbourg) is professor at the University of Strasbourg, among the research unit Societies, Actors and Government in Europe (SAGE). He coordinates the research axis "*Politique sociales, dynamiques familiales et professionnelle*" in this unit, with Jay Rowell. During the last years, he has contributed to set up two national surveys: ELFE and "*Migrations, Famille et Vieillissement*" (MFV), a survey in French overseas departments. His recent publications concern Family and Fertility. He manages the Master program in demography in Strasbourg University. From this year he leads the Strasbourg's academic data platform for Social Sciences (*Plateforme universitaire de données*, PUD-S).

Christophe Giraud (University Paris Descartes / INED) is lecturer at Paris Descartes University since 2002, statistician and sociologist at CERLIS (*Centre de recherches sur les liens sociaux*) and seconded at INED (UR3). He realized a qualitative panel research on the intimate life of young adults in France (2006-2014) and will publish in 2016 a book based on this research entitled *Realistic Love*.

Angela Greulich is Lecturer at the University Paris of Paris 1 (Economics Department). She has been working for 10 years on the economics of gender and demography and is a consultant for the OECD and the World Bank. As former post-doc at INED, she was a member of the international research project REPRO (reproductive decision making in a micro-macro perspective) for the European Commission (Seventh Framework Programme under the Socio-economic Sciences and Humanities theme. She has extensively worked with international survey and census data on the topics of budget-, labour market- and fertility- decisions of young adults.

Emmanuelle Santelli (CNRS/ University Lumière Lyon II / Centre Max Weber) is a sociologist, research supervisor at CNRS and seconded researcher at INED. Specialised in the study of second-generation immigrants from North Africa, her recent publications concern the way these children of immigrants enter into adulthood and their choices with respect to union formation. She is currently working on the sociology of young adults. Her current qualitative survey addresses the question of love and the stages of marital life.

Two post-doctoral fellows will be hired within the project. Both will primarily publish their own work on the research topics, in collaboration with other team members. The first one will focus on data access and documentation, and will establish the users' website. The second one will collaborate with similar projects in Europe or in North America. They will have close contacts with the team members at INED as well as at INSEE.

Four interns will join the project for a three-month period. They will focus on one research topic, and on two or more datasets, the aim being to teach them how to combine two different data sources and to cross-validate their results.

2.3. Tasks

The project is organised into four tasks: management and organisation, analyses, data assessment, and dissemination.

Tasks 2 and 3 will be strongly interconnected. On the one hand, the analyses have their own scientific aims, and each task group will produce its own academic publications. On the other hand, the examination of (in)consistencies between data sources will be based on an accurate comparison of the definitions used in and the field covered by each data source, in order to propose comparable estimates, as well as on an evaluation of potential bias specific to each source. This will allow identifying the strengths and weaknesses of each source, and offering definitions and concrete tools to

use them for further analyses based on common explicit definitions. Feed backs from data assessment to research will be organized on a permanent basis. In addition, annual team meetings will offer the occasion to share experiences and to reach shared diagnosis on results as well as on data assessments.

Task 4 will complement the project with dissemination of results and data documentation targeted to scholars in France and abroad, as well as to doctoral students. It will be performed through presentations at conferences, users' group meetings, the settlement of a website devoted to Administrative data access and documentation, and training sessions.

Task 1: Management and organisation

This task will involve traditional project management activities such as organising and overseeing the participation of project members in all tasks. It will be devoted to preparing data access for participants, organizing each year the team meetings and Advisory Board meetings in June, open seminars, users' group meetings, and the closing conference in December.

Person in charge: Laurent Toulemon (INED). Members, Arnaud Bringé (INED), and postdoctoral researchers (INED).

Task 2: Analyses

The second project task is to pursue analyses using one or more of the above-mentioned sources, with a view to cross-validate the various available sources.

Task 2.1 Census population counts and family structure distribution

This task aims to validate the INSEE census results by comparing them with other sources, especially the Permanent demographic sample (EDP) to identify double counting, and the tax-based dataset (FLP) which provides information on omissions. The involvement of INSEE in this validation guarantees its utility for data collection and production processes at INSEE. An additional part of this task will be devoted to family structures as observed in the census and in household surveys. Three programmes have been launched at INSEE, in which the coordinator of this project is involved. 1. A new version of the census housing form, in relation to the increasing participation through Internet, with two specific aims: first, a more efficient allocation of inhabitants to one and only one household, and eventually a better identification of potential double counts and omissions within the census housing form; second, a better identification of family ties between household members. 2. The documentation and dissemination of the Households surveys core questionnaire (TCM) which is foreseen for this year. 3. The use of this file to compare family structures in the census and in surveys. The current project will allow these two last projects to be conducted in an open way, with an external expertise, and the results will be published in scientific journals and disseminated outside INSEE.

Three papers will be published during the project: one on census counts assessment (census, EDP, FLP), one on family structures (census, FLP, TCM), and one on inhabitants includes twice in the census (EDP).

Person in charge: Laurent Toulemon (INED). Participants: Didier Breton (University of Strasbourg), Sébastien Durier (INSEE), Céline Leroy (INSEE), Gaël de Peretti (INSEE), Isabelle Robert-Bobée (INSEE), Lidia Panico (INED) and a postdoctoral researcher (INED).

Task 2.2. Young adult couples, their formation and dissolution

This task brings together several researchers who work on transitions to adulthood, intimate relationships and couple formation among young adults in France. Analysis of differential stability of young adult unions by school enrolment and labour market integration will make it possible to test the hypothesis of the emergence of a new form of co-residence for students and young people in an unstable or uncertain situation, weakened by constraints imposed by individual geographic mobility.

The first aim of the task members is to publish analyses on new forms of union among young adults, based on their own data (qualitative in-depth interview and household surveys). The second aim is to compare results and write common articles based on different datasets, and to analyse how changes in

union behaviour among young adults may lead to discrepancies in the levels and trends based on different data sources.

At least two specific papers will be published during the project: one on the new forms of unions and their sequence among youth, one on social differentials and the relations between student status and job situations and union behaviour of men and women, one on the assumptions needed to reconcile evidence based on in-depth interview and surveys and census data.

Person in charge: Christophe Giraud (Paris Descartes). Participants: Marie Bergström (INED), Arnaud Régnier-Loilier (INED), Angela Greulich (Paris 1 Panthéon Sorbonne), Emmanuelle Santelli (Max Weber Centre) and a postdoctoral researcher.

Task 2.3. Children in joint custody

The participating researchers have worked on the situation of children using fiscal data and the French Longitudinal Study of Children (*Étude longitudinale française depuis l'enfance*, ELFE), coordinated by INED. The precise description of the situation of children in their two homes and their evolution from one year to the next provides information on the frequency of shared residence, the consequences in terms of living standards of families and rearrangements in the years following parental break-up.

Three papers will be published during the project: one on the family situations of children with separated parents and its variation according to socio-economic characteristics of parents, according to different data sources: one on the short-term economic consequences of union disruptions of married and unmarried couples; one on the long term consequences for children.

Person in charge: Lidia Panico (INED). Participants: Cécile Bourreau-Dubois (University of Lorraine), Carole Bonnet (INED), Marion Leturcq (INED), Ariane Pailhé (INED), Anne Solaz (INED), Xavier Thierry (INED), and a postdoctoral researcher.

Task 3: Data Assessment

Task 3.1. Cross-validation of the sources

In close collaboration with the supervisors and the authors of the studies conducted in task 2.1, 2.2, and 2.3, expertise on data will be based on critical analysis of the plausibility of the proposed results and on assessment of the strengths and weaknesses of each of the data sources used. This expertise will be based on both external data (published data, international comparisons) and on an assessment of the consistency of the results (on global and local geographical scales).

A precise documentation on the variables used in each of the works performed in task 2, and the differences found between the results, will be collected.

Three papers will be produced from this documentation. A paper on the future use of big statistical data for demographic purposes, based on the analyses of works conducted on population counts and family structures, one on a critical assessment of the definitions of a union and a child in statistical administrative data, and proposals to improve data collection on these topics, one on the specific population sub-groups which are badly observed in each data source, using statistical matching between census and administrative data.

Person in charge: Elisabeth Morand (INED). Participants: Christophe Giraud (Paris Descartes), Lidia Panico (INED), Arnaud Régnier-Loilier (INED), Laurent Toulemon (INED), Stéfan Lollivier (INSEE), Gaël de Peretti (INSEE), Isabelle Robert-Bobée (INSEE).

Task 3.2. Assessment of the statistical quality of each source

A critical evaluation will be produced for each source to pinpoint the origin of differences with respect to other sources as the consequence of the process of data collection or construction. This scientific and documented assessment will ensure that collection biases specific to each source, including administrative sources, can be taken into account. It may lead to changes in the methods of statistical data collection or extraction from administrative files.

Specific documentation and improvement of the datasets are the first aim of this task. The first aim is to take into account works done in tasks 2 and 3.1 in the documentation of the files. The second is to improve the quality of each file, when possible, to propose codes creating corrected or constructed variables, and to discuss with other team members on the sensibility of their results to the improvement of the files. Finally, a global view on potential issues related to each file will lead to propose changes in data collection and management, eventually considering new available datasets and possibilities in terms of file merging. Results from the EDP users' group will also be useful for this task.

This task will produce technical documentation and new version of the files released for research at the CASD (EDP, RLP, TCM) or at the *Réseau Quetelet* for anonymous files (an anonymized version of the TCM).

Person in charge: Laurent Toulemon (INED). Participants: Carole Bonnet (INED), Elisabeth Morand (INED), Postdoctoral researcher (INED), Sébastien Durier (INSEE), Céline Leroy (INSEE), and Stéfan Lollivier (INSEE).

Task 4: Dissemination

Besides the traditional tasks of article publications, the project aims at participating in the dissemination and exploitation of newly available files.

Task 4.1. Academic dissemination

Scientific publications will be the main output of the project. The scientific evaluation of big databases in France will give rise to collaboration with teams using similar data in other countries, as well as with administrations producing and teams using other big data sets. The project will include an annual one-day seminar and a project closing conference.

Person in charge: Ariane Pailhé (INED). Participant: Laurent Toulemon (INED).

Task 4.2. Documentation on a website

A website will be set up to prepare the opening of big data to users in the social sciences, offering computer programs that reduce the entry cost to these databases.

Technical support will be organized within INED in collaboration with the Statistical Methods and Surveys Support services, which are both present in the project. The bilingual (English and French) website will be open at month six, and will serve for the dissemination effort envisioned by the competence centre launched at INED. The website will benefit from our previous experiences at INED: websites produced for INED surveys (http://epic.site.ined.fr/en/), INSEE surveys (http://lili-efl2011.site.ined.fr/en/), and users' group (http://util_elfe.site.ined.fr/en/); the latter was settled by the current project team members, and team members participated to the formers. It will also rely on INED expertise in creating online databases and codebooks of survey data like, for the GGP, http://www.ggp-i.org/online-data-analysis.html.

Collaboration with other groups of data providers, notably of Health data (Institute of Health and medical research, INSERM, INDS) and economic and social data (CASD, CNAF) will be pursued during the project.

Person in charge: Arnaud Bringé (INED). Participants: Postdoctoral researcher (INED), Arianna Caporali (INED).

Task 4.3. Users training

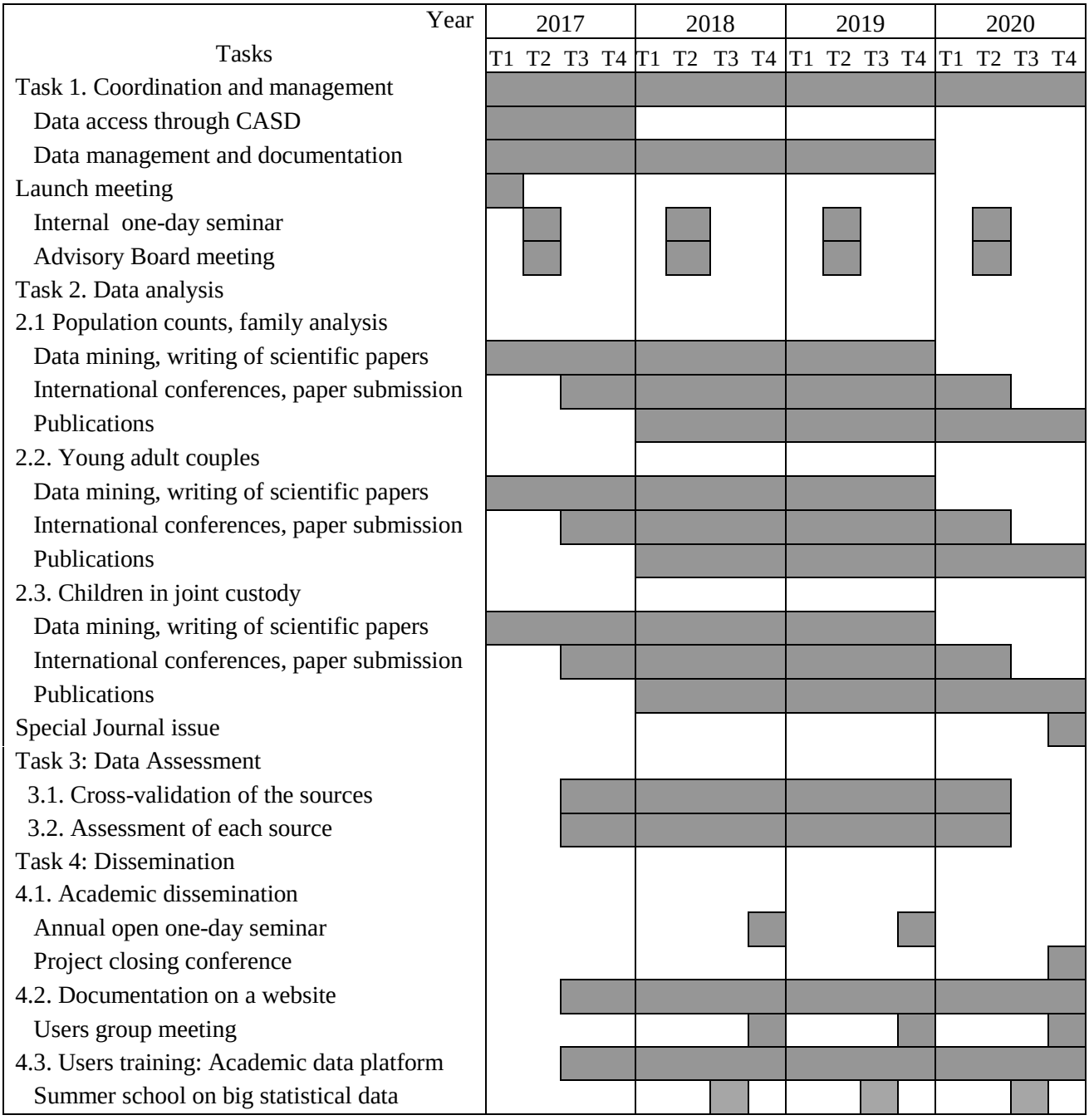
In parallel, training for PhD students, postdoctoral researchers and other users will be offered in association with INSEE and CASD. Didier Breton will set up an academic data platform (*Plateforme universitaire de données*, PUD) in Strasbourg. Three users' group meetings will be organized within the open seminars, allowing for exchanges with users of our datasets and their documentation, and exchanges with other data sets users and providers. A summer school will be devoted to the use of big data, starting in the second year of the project. This task will be coordinated with several training

projects within university academic data platforms and our Laboratory of Excellence Individuals, Populations, Societies (Labex iPOPs, http://www.ipops.fr/en/). We thus do not ask for funding from the ANR, as our aim is to organize the summer school through iPOPS and to encourage universities to organize the summer school with their own partners.

Person in charge: Didier Breton (University of Strasbourg). Participant: Angela Greulich (University of Paris 1 Panthéon Sorbonne)

2.4 Organisation and calendar

The Gantt chart below presents the time schedule of the different tasks. Despite the large number of team members, the structure of the project is simple: each year, one internal and one open meeting will be organized. The tasks closely intertwined, as many research projects being held simultaneously.



2.5 Budget

The budget is divided in three parts. The biggest item is devoted to two post-doctoral contracts. The second item is related to scientific networking and research. The meetings will be hosted at INED or INSEE, and their cost is very limited. Most networking will take place at national and international conferences, where the team members will advertise the project and its website. The closing conference will be more expensive, as scholars from abroad will be invited. The third item is devoted to data access through the CASD and other operating expenses

	Unit cost	No.	Total
Short-term staff			
Post-docs (Just after PhD, 24 months plus travel)	3 500	24	84 000
Post-docs (3-5 years after PhD, 24 months plus travel)	3 900	24	93 600
Interns 3 months (4)	555	12	6 660
Total, short-term staff			184 260
Task 1. Coordination and management			
Launch meeting (missions)	300	2	600
Launch meeting (lunch)	30	21	630
Internal one-day seminar (missions * 4)	300	8	2 400
Internal one-day seminar (lunch * 4)	30	84	2 520
Task 2. Data analysis			
2.1 Population counts, family analysis			
Conference in Europe	2 000	3	6 000
Conference outside Europe	1 000	3	3 000
2.2. Young adult couples			
Conference in Europe	2 000	3	6 000
Conference outside Europe	1 000	3	3 000
2.3. Children in joint custody			
Conference in Europe	2 000	3	6 000
Conference outside Europe	1 000	3	3 000
Task 4: Dissemination			
Annual open one-day seminar (missions * 2)	300	6	1 800
Annual open one-day seminar lunch * 2)	30	80	2 400
Project closing conference			
Two speakers from the US or Canada	2 000	2	4 000
Roundtable with users and providers	1 000	1	1 000
Missions	300	6	1 800
Catering expenses (tea and coffee)	3	80	240
Catering expenses (final conference dinner)	35	40	1 400
Total, Task related "missions"			38 600
Total, Other task related expenses			7 190
Operating expenses			
Journal submission fees	100	2	200
Publications (translation, open access)	1 000	10	10 000
Data access through CASD	25 736	1	25 736
Total, operating and other expenses			43 126
Computer equipment (2 laptops for the post-docs)	2 000	2	4 000
Total			269 986
INED, 8% expenses charged (operating plus structure)			21 599
GRAND TOTAL			291 585

3. Strategy to promote, protect and use the results, overall impact

This project will provide an important panorama for several stakeholders, especially given our focus on complex families, and the variation of these family structures across socio-economic groups. Potential stakeholders include civil servants, governmental and intra-governmental organizations such as national and international statistical services, non-governmental organisations, and academic researchers. Research, policy, governmental and non-governmental users will be invited to a final closing conference, where key project results will be presented by project researchers, and discussed by invited non-academic users to interpret findings within a policy perspective. We will reach academic audiences through peer-reviewed publications, presentations at conferences, and a project website. Both methodological and empirical papers will be submitted.

The coding and documentation of harmonized variables will be made freely available to the research community initially on an INED website. This will foster further use of these new data sources, especially within the social and economic sciences communities. We will explore further dissemination practices by exchanging with different teams at INSEE.

Big statistical data are already used by economists and epidemiologists, in France and abroad. Other big statistical data, including Health data currently being made available, will be used at INED and INSEE, including by team members. These data sets are out of the scope of the current project, but INED is aiming at playing a major role in their dissemination. The collaboration between INSEE and INED, as well as the documentation and dissemination efforts performed within the project, will be crucial for the development of relevant skills at INED, and will have a major impact in expanding the use of such big statistical data among scholars working in the humanities and social sciences (HSS), as well as in epidemiology.

3.1. An applied research project

The competence centre launched at INED in March 2016 will have three main tasks: information on current databases used by INED researchers, dissemination of knowledge on the administrative data used at INED, centralization of the ongoing projects at INED and collaboration with other research institutes to provide a streamlined access to administrative data. The competence centre's focus will also include other big statistical data coming from health, economic and social sources. The current project will help deliver these three aims quickly and efficiently , with concrete outputs for three major data files, central for population studies: the Permanent Demographic Sample (EDP), the Statistical Register of Dwellings (RSL), which will be made available in its new form as the Demographic Dataset on Households and Individuals Based on Tax Information (FLP), and the Household surveys core questionnaire data file (TCM).

These administrative files require data processing and cleansing. Cross-checks with other data sources, at a global as well as at an individual level based on data merging, will dramatically strengthen the accuracy and consistency of basic estimates. The project will allow unifying efforts from INED and INSEE to broaden the use of administrative data through sharing experience, documentation and training. Some of these databases are already available; others will become accessible in 2016 to a limited audience that will expand in the future. While data collection costs for these bases are small compared to traditional surveys, the cleaning and matching tasks demand considerable efforts and specialist skills. The project intends presenting a model for further dissemination and use of administrative data for research, based on synergies between producers and users of these big datasets.

3.2. Engaging in the dissemination and use of big data in social sciences

Beyond these direct goals, the project aims at mobilising the scientific community to ensure maximum use of these data, whose complexity entails a significant entry cost. This is why the project brings together researchers with an expertise on one or more of these sources, who are eager to compare different sources within the same research object, and who wish to share their experience in using such data. These joint operations will allow testing the data quality while disseminating these sources.

Dissemination to scholars and teaching

Thanks to its central position in population studies in France, INED is in the best position to undertake the task of increasing data accessibility to the scientific community and training in the use of these sources. INED has a long tradition of disseminating and providing access to its surveys through the *Réseau Quetelet* (the French Data Archive for Social Sciences) of which it is a founding member (Caporali, Morisset, Legleye, 2015). INED will mobilise the networks necessary for this dissemination: academics and students via the iPOPs labex and the RESODEMO training network; data users via CASD; the TGIR PROGEDO and the *Réseau Quetelet*; academic data platforms and training networks in applied statistics. The project will also allow French researchers to join international research communities which have already addressed the issue of big data in social science.

The team will produce a series of articles and documents on the methodological issues relating to the use of the complex data sources.. The exploration and assessment of these datasets will be of high methodological utility, and will pave the way for new data collection based on the limits and weaknesses identified in these administrative data. These activities will be disseminated through papers submitted to peer reviewed journals with a statistical orientation. Such dissemination strategy will provide scientific validation by peers of our proposed methods for data dissemination. Documents dedicated to data dissemination will be available to all users directly on the project website. At the same time, the activity of data validation, cleaning and processing will undoubtedly result in high quality papers on innovative topics. These papers will be submitted in priority to the best journals in the field of study, including Demography, Population Studies or Demographic Research.

Dissemination for non-academic audiences

With the current changes in French laws on administrative data availability, the need for open access documentation on the statistical data constructed from administrative big data is crucial for the extension of their use. The project will combine documentation on the creation of the files, accurate and explicit statistical assessment of their quality, and results based on these data accessible to a number of non-academic audiences.

3.3. Collaborations with international institutions involved in big data dissemination

The Nordic countries have benefitted of register data linked to various administrative data for a long time, which has allowed very precise studies of family trajectories and living arrangements. Though there are technical limitations in most European countries, given the absence of data registers with individual identification, the example of the Nordic countries has been seminal and attempts to create big statistical databases are increasing. Many European countries have moved to survey- and register-based estimates for population counts (Valente, 2010). INSEE has been very innovative with the new French census based on annual surveys and a limited use of administrative data, but there is room for a broader integration of administrative data sources in population estimates.

At first, relationships with foreign institutions will be based on communication about data management and technical aspects of setting up such data schemes. This will take place through seminars and conferences. Exchanges with scholars working on similar questions will be extremely useful to determine quality assessment, data management, and data use strategies. Furthermore, the project will allow increasing possibilities for collaborations with institutions across the world, by bringing together researchers and sharing results. Creating a pool of users for new data is always an issue, notably due to access limitations. We will support broad access to the proposed data sources through depositing data on the *Réseau Quetelet* and the CASD, supplemented by the project website offering data assessment, practical tools for users, training opportunities and scientific references. This will extend the pool of researchers able to use the French data at the end of this project. Some contacts have already been taken in the preparatory phase of this project, but our current aim is to put France on the map in the ongoing discussion of big data through the use, documentation and dissemination of major population-based French datasets.

Bibliographical references

Note: the names of team members are in bold in the reference list

Albouy V., Breuil-Genier P., 2012, Démographie et famille : les différences sociales se réduisent-elles? *France, portrait social - Insee Références*.

Amossé T., **de Peretti** G, 2011, Men and Women in Household Statistics: A Piece in Three Acts, *Travail, Genre et Sociétés*, 26, p. 23-46.

Bauman, K, 2003, *Liquid love: On the frailty of human bonds*, Cambridge: Polity Press.

Bawin-Legros B, Gauthier A, 2001, Regulation of intimacy and love semantics in couples living apart together, *International Review of Sociology*, 11, 1, p. 39-46

Bellamy V., Beaumel C., 2015, Bilan démographique 2015. Le nombre de décès au plus haut depuis l'après-guerre. *Insee Première*, 1581, 4 p.

Bergström 2016, Who uses online dating sites in France? Who finds their partner this way?, *Population and Societies*, 530, 4 p.

Bodier M., Buisson G., Lapinte A., Robert-Bobée I., 2015, *Couples et familles. Insee Références*, 190 p.

Bonnet C., Garbinti B., **Solaz** A., 2015, Les variations de niveau de vie des hommes et des femmes à la suite d'un divorce ou d'une rupture de Pacs, *Couples et familles. Insee références*, p. 51-61.

Bozon, M., Rault W., 2012, From Sexual Debut to First Union. Where do Young People in France Meet their First Partners? *Population*, 67, 3, p. 377-410.

Breuil P., Buisson G., **Robert-Bobée** I., Trabut L., 2016, Enquête famille et logements adossée au recensement de 2011 : comment s'adapter à la nouvelle méthodologie du recensement et quels apports au recensement ? *Économie et Statistiques*, in print.

Brückner, H., Mayer, K.U., 2005, De-Standardization of the life-course: What it might mean? And if it means anything, whether it actually took place? *Advances in Life Course Research*, 9, p. 27-54.

Bruun, M.H., 2011, Egalitarianism and Community in danish housing Cooperatives: Proper Forms of Sharing and Being Together, *Social Analysis*, 55, 2, p. 62-83.

Buisson G., Costemalle V., Daguet F., 2015, Depuis combien de temps est-on parent de famille monoparentale ?, *Insee Première*, 1539, 4 p.

Cashmore J., Parkinson P., Taylor, A., 2008, Overnight stays and children's relationships with resident and nonresident parents after divorce, *Journal of Family Issues*, 29, 6, p. 707-733.

Castro-Martín T., Dominguez-Folgueras M., and Martin-Garcia T., 2008, Not truly partnerless: Non-residential partnerships and retreat from marriage in Spain. *Demographic Research*, 18,16, p. 443-468.

Cancian M., Meyer D., Brown P. Cook S., 2014, Who Gets Custody Now? Dramatic Changes in Children's Living Arrangements After Divorce, *Demography*, 51, p. 1381–1396.

Caporali A., Morisset A., Legleye S., 2015, Providing Access to Quantitative Surveys for Social Research: The Example of INED, *Population*, 70, 3, p. 537-566.

Caradec, V., 1997, Forms of conjugal life among the "young elderly". *Population, an English Selection*, 9, p. 47-73.

Carrasco V., Dufour C., 2015, Les décisions des juges concernant les enfants de parents séparés ont fortement évolué dans les années 2000, *Infostat Justice*, 132, 6 p.

Chaleix M., **Lollivier** S., 2004, Outils de suivi des trajectoires des personnes en matière sociale et d'emploi. Rapport au Consil National de l'information statistique, N° 98/B010, 37 p. http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/publications/rapports/2005/RAP_2004_98_outils_trajectoires_personnes_emploi.pdf

Chaleix M., **Lollivier** S., 2005, Des panels pour les statistiques sociales, *Courrier des statistiques*, 113-114, p. 53-56.

Chardon O. Vivas E., 2009, Les familles recomposées : entre familles traditionnelles et familles monoparentales, Document de travail de l'INSEE, n°F2009/04.

Charles M.-A., Leridon H., Dargent P., Geay B, 2011, Tracking the lives of 20,000 children - Launch of the Elfe child cohort study, *Population and Societies*, 475, 4 p.

Citro, C.F., 2014, Des modes multiples pour les enquêtes à des sources de données multiples pour les estimations. *Techniques d'enquête*, 40, 2, p. 151-181.

Daguet F., Niel X., 2010, Vivre en couple. La proportion de jeunes en couple se stabilise, *Insee première*, 1281, 4 p.

De Jong Gierveld, J., 2004, Remarriage, unmarried cohabitation, living apart together: partner relationships following bereavement or divorce. *Journal of marriage and family*, 66, 1, p. 236-243.

Desplanques, G., 2008, Strenghts and Uncertainties of the French Annual Census Surveys, *Population*, 63, 3, p. 415-439

Duncan S, 2011, The world we have made? Individualisation and personal life in the 1950s, *Sociological Review*, 9, 2, 242-65

Duncan, S., Carter, J., Phillips, M., Roseneil, S., Stoilova, M., 2013, Why do people live apart together? *Families, Relationships and Societies*, 2, 3, p. 323-338.

Giraud, Christophe, 2014, *Les chemins du couple. Une sociologie de la vie personnelle des étudiants*, HDR, Université Paris Descartes.

Fontaine M., Stehlé J., 2014, Les parents séparés d'enfants mineurs : quel niveau de vie après une rupture conjugale ?, *Politiques sociales et familiales*, 117, p. 80-86.

Freguja C., Valente P. (eds.), 2010, Measurement of different emerging forms of households and families. Report by the Task Force on Families and Households, UN-ECE – Eurostat, Conference of European Statisticians.

Groves R., 2011, Three eras of survey research, *Public Opinion Quarterly*, 75, 5, p. 861–871.

Guillonneau M., Moreau C., 2013, La résidence des enfants de parents séparés. De la demande des parents à la décision du juge. Exploitation des décisions définitives rendues par les juges aux affaires familiales au cours de la période comprise entre le 4 juin et le 15 juin 2012. Report to the Minister of Justice, 63 p. www.justice.gouv.fr/art_pix/1_rapportresidence_11_2013.pdf

Haskey, J.,2005, Living arrangements in contemporary Britain: having a partner who usually lives elsewhere and living apart together (LAT), *Population Trends*, 122, p. 35-45.

Haut conseil de la famille, 2014. *Les ruptures familiales : état des lieux et propositions*, 218 p. http://www.ladocumentationfrancaise.fr/rapports-publics/144000594/index.shtml

Herpin N., **Toulemon** L., Verger D., 2001, Rénovation du tronc commun des enquêtes de l'Insee auprès des ménages. Questions sur la composition du ménage et les comportements familiaux, note interne de l'Insee, 25 p.

Imbert C., Deschamps G., Lelièvre É., Bonvalet C., 2014, Living in two residences: mainly before and after working life, *Population and Societies*, 507, 4 p.

Jeandidier B., Sayn I., **Bourreau-Dubois** C., 2012, Séparation des parents et contribution à l'entretien et l'éducation de l'enfant. Une évaluation du barème pour la fixation du montant de la pension alimentaire, *Politiques Sociales et Familiales*, n°107, mars 2012, pp. 23-39

Kesteman N., 2007, La résidence alternée : bref état des lieux des connaissances sociojuridiques, *Recherches et prévisions*, 89, p. 80-86.

Kitterød R.H., LyngstadJ., 2012, Untraditional caring arrangements among parents living apart: The case of Norway, *Demographic Research*, 27, p. 121-151.

Lapinte A., 2013, Un enfant sur dix vit dans une famille recomposée, *Insee première*, 470, 4 p.

Leroy C., 2015, Le Tronc Commun des Ménages : adaptation aux panels et empilement. Master thesis, ENSAI, Master mention Statistique Économétrie, spécialité statistique publique. Parcours Méthodologie de la Statistique Publique. 137 p.

Levin I., 2004, Living Apart Together: A New Family Form, *Current Sociology*, 52, 2, p. 223-240.

Mambetov D., 2014, Étude des BI multiples associés à une même personne dans les enquêtes annuelles de recensement, à partir de l'EDP (base étude 2011). Note interne à l'Insee n° 1253/DG75-F170/DM, 9 p.

Mazuy M., Barbieri M., Breton D., d'Albis H., 2015, The Demographic Situation in France: Recent Developments and Trends over the Last 70 Years, *Population*, 70, 3, 417-486.

Manning, W.D., Smock, P.J., 2005, Measuring and modeling cohabitation: new perspectives from qualitative data, *Journal of marriage and the family*, 21, 4, p. 75-108.

Milan, A., Peters, A., 2003, Couples living apart. *Canadian Social Trends*, 11-008, p. 2-6.

Pailhé A., 2015, Partnership Dynamics across Generations of Immigration in France: Structural vs. Cultural Factors, *Demographic Research*, 33,16, p. 451-498.

Reimondos A., Evans A., Gray E., 2011, Living-apart-Together (LAT) Relationships in Australia. *Family Matters*, 87, p. 43-55.

Rault W., Mazuy M., Rivière A., **Toulemon** L., 2011, L'enquête Famille et logements associée au recensement de 2011, in Tremblay Marie-Ève, Lavallée Pierre, El Haj Tirari Mohammed, *Pratiques et méthodes de sondage, Actes du colloque Sondages 2010*, p. 113-117.

Rault W., **Régnier-Loilier** A., 2015, First cohabiting relationships: recent trends in France, *Population and Societies*, 521, 4 p.

Régnier-Loilier A., Beaujouan É., Villeneuve-Gokalp C., 2009, Neither single, nor in a couple: a study of living apart together in France, *Demographic Research*, 21, 4, p. 75-108

Régnier-Loilier A. (ed.), 2014, The contemporary Family in France. *Partnership Trajectories and Domestic Organization*, Springer/Ined, 273 p.

Régnier-Loilier A. (ed.), 2016, Parcours de familles. L'enquête Étude des relations familiales et intergénérationnelles, Paris, Ined, 432 p. (Grandes Enquêtes)

Robert-Bobée I., 2006, Étudier la fécondité en France à l'aide de l'échantillon démographique permanent, *Courrier des statistiques*, 117-119, p. 15-20.

Roseneil S., Budgeon S., 2004, Culture of intimacy and care beyond the conventional family: personal life and social change in the 21st century, *Current Sociology*, 52, 2, p. 135-159

Ruggles, S., 2014, Big Microdata for Population Research. *Demography*, 51, 1, p.287-297

Sironi M., Barban N., Impicciatore R., 2015, Parental social class and the transition to adulthood in Italy and the United States, *Advances in Life Course Research*, 26, p. 89-104.

Strohm C, Seltzer J, Cochran S, Mays V, 2009, 'Living apart together' relationships in the United States, *Demographic Research*, 21, 7, p. 177-214

Toulemon, L., 2008, Between First Intercourse and First Union: The Early Trajectories of Men and Women Are Still Different. Chapter 9 In Bajos N, Bozon M. (eds), *Sexuality in France. Practices, gender and health*, Oxford, The Bardwell Press, 554 p.

Toulemon L., 2011, Counting and Describing Individuals, Families, Households, Homes, *Travail, Genre et Sociétés*, 26, p. 47-66.

Toulemon L., 2012, Changes in Family Situations as Reflected in the French Censuses, *Population*, 67, 4, p. 551-572.

Toulemon L., 2016, Combien de personnes ont plusieurs résidences habituelles en France ?, in Imbert, Lelièvre, Lessault (eds.), La famille à distance : configurations, pratiques et norms. Book submitted to INED collections.

Toulemon L., Denoyelle T., 2012, La définition des ménages dans les enquêtes françaises : comment tenir compte des multi-residences ?, *Actes des journées de méthodologie statistique*, 16 pages, http://jms.insee.fr/files/documents/2012/943_3-JMS2012_S26-1_TOULEMON-ACTE.PDF

Toulemon, L. Pennec, S., 2010, Multi-residence in France and Australia: Why count them? What is at stake? Double counting and actual family situations, *Demographic Research*, 23, 1, p. 1-40.

Toulemon, L. Pennec, S., 2011, How many people live alone in France?, *Population and Societies*, 484, 4 p.

Trabut L., Lelièvre É., Bailly E., Equipe LiLi. 2015, Does the Household-Based census capture the diversity of Family configurations in France?, *Population*, 70, 3, p. 637-665.

Valente P., 2010, Census taking in Europe: how are populations counted in 2010? *Population and Societies*, 467, 4 p.

Vikat, A., Spéder, Z., Beets, G., Billari, F.C., Bühler, C., Désesquelles, A., Fokkema, T., Hoem, J.M., MacDonald, A., Neyer, G., Pailhé, A., Pinnelli, A., and Solaz, A., 2007, Generation and Gender Survey (GGS): Towards a better understanding of relationships and processes in the life course. *Demographic Research*, 17, 14, p. 389-440.