The *Répertoire Statistique des Individus et des Logements* (Statistical Register of Individuals and Dwellings, RÉSIL)

A new reference universe for demographic and social statistics

Olivier Lefebvre *

INSEE is currently setting up a Statistical Register of Individuals and Dwellings (Résil) in order to modernise its system for producing demographic and social statistics, in particular by making greater use of administrative data.
This project is already well advanced. It should be completed by the end of 2025 and be operational from the beginning of 2026.
Like its counterpart for enterprises and their local units, this register will make it possible to set up sampling bases or to check the coverage of administrative data, as well as to create files enriched by matching different sources in a simpler and more reliable way. The various data production processes will thus be able to respond to the growing demands for coverage, processing speed and responsiveness.
There are a number of prerequisites for setting up such a system: the need to use different sources in order to aim for exhaustiveness; the ability to apply statistical processing, which is often innovative, in order to ensure quality; the need for a legal framework to protect the data processed; and finally, the need for a "social mandate", i.e. a legitimacy that goes beyond technical or legal capacity.

* Head of the RÉSIL Program, INSEE.
olivier.lefebvre@insee.fr

In order to be established, a statistic requires data to be gathered (Dupont, 2023), either directly (via a survey) or indirectly (using administrative data or even privately held data). The quality of coverage of the data gathered also needs to be assessed. Do they cover all the statistical units in our field of interest? Without any duplicates or omissions? Does the information collected relate to the "correct" statistical units?

To that end, it is extremely useful to have a list of all the statistical units of the field of observation, with no units being included incorrectly. This is because a survey sample can then be drawn from this sampling frame, and it also becomes possible to verify that the administrative data used cover the field exhaustively, to classify the representativeness of the data and to correct data if necessary, so as to avoid any bias linked to a lack of coverage. Such lists are referred to as "reference universes".

▶ Under Certain Conditions, Registers May Constitute a Reference Universe

A register is an exhaustive list of units, with very few variables (Rivière, 2022). If one were to draw up a register, it would be both very long, since it potentially contains all the observations of a field, but also very narrow because it handles few variables.

The variables included in a register must make it possible to unambiguously identify the units it contains, in particular to avoid duplicates, to allow for simpler updates and to enable them to be linked with other elements of the information system. A register forms the backbone of the information system. Registers are a living thing, as they are generally updated continuously; however, it is nevertheless possible to take a snapshot of one, reflecting the situation on a given day (often on 1 January), which will be the frame for the reference universe.

INSEE has a long history of and long-standing experience in registers. It has managed the *Répertoire national d'immatriculation des personnes physiques* (National Register for the Identification of Individuals, RNIPP) since 1946 (Espinasse and Roux, 2022). More recently, in 2019, INSEE built and took charge of the *Répertoire électoral unique* (Single Electoral Register) for the management of the electoral rolls (Desmotes-Mainard, 2019).

In the area of businesses, INSEE was entrusted in 1973 with the management of the *Système Informatique pour le Répertoire des ENtreprises et des Établissements* (Computer system for the business and establishment register, SIRENE)[1]).

Could these registers, particularly the SIRENE and the RNIPP, be used as "reference universes"? Not directly, as the information concerning units that leave the field of observation (departure from the national territory for individuals or cessation of activity for a company), are either unknown or ascertained with a delay. The RNIPP does not include individuals' addresses and there is no obligation to report departures from France. Formal cessation of activity for companies often occurs well after the actual cessation of economic activity.

1 *https://www.insee.fr/fr/information/6675111.*

> ***INSEE felt that it was necessary to go further, by creating downstream administrative registers, so-called "statistical" registers.***

INSEE therefore felt that it was necessary to go further, by creating downstream administrative registers, so-called "statistical" registers. Such registers make it possible to carry out processing for statistical purposes, without any impact on the persons or companies whose data are being processed: for example, statistical processing may conclude that a person no longer lives in the national territory, but this will not affect their inclusion in a pension plan.

Business statistics are ahead of demographic and social statistics in this regard. Indeed, in 2012, INSEE created the *Système d'immatriculation au répertoire des unités statistiques* (Statistical Business Register, SIRUS), a statistical register of companies and establishments, enriched by information collected or built by Official Statistics – profiles of groups, level of activity of companies, etc. (Hachid and Leclair, 2022).

The situation was more complex for demographic and social statistics. While we do have a sampling frame constructed from tax files used for household surveys, covering the entire field of ordinary housing[2], the absence of a shared identifier and the lack of exhaustiveness of the sources made it impossible to consider it to be a "reference universe" with which the entire set of administrative sources can be compared.

▶ Facing New Challenges and Opportunities: Creating the Statistical Register of Individuals and Dwellings

Several factors have converged to advance progress in demographic and social statistics.

On the one hand, there are increasing needs for administrative data use, along with new opportunities. In terms of needs, it was necessary to answer new questions, to more quickly provide answers to recurring questions or to better reflect the diversity of situations, particularly territorial ones, by producing more granular data than those obtained using samples; for example, FILOSOFI[3] provides information on living standards with a granular level of localisation, based on tax declarations and data on social security benefits. In terms of opportunities, there are more sources, which are more accessible, more effectively structured and documented, and of better quality, as well as IT processing capabilities that make it possible to process very high volumes of data securely and quickly (e.g. PASRAU[4]). Like most national statistical institutes, INSEE is therefore investing heavily in this area (***Box 1***).

2 Ordinary housing is defined as being in contrast to housing in a facility offering specific services (homes for the elderly, student halls of residence, accommodation for tourism, accommodation provided for social purposes or for persons with disabilities, etc.).
3 French acronym based on *FIchiers LOcalisés SOciaux et FIscaux*, meaning localised tax and benefits files.
4 The PASRAU (*Prélèvement À la Source pour les Revenus AUtres*, Withholding Tax for Other Incomes) resource is the result of work to simplify and streamline social security declarations and the need to send the information necessary for the withholding of taxes at source to the DGFiP (*Direction générale des Finances publiques*, French Directorate-General of Public Finances).

▶ Box 1. An Overview of National Statistical Institutes (NSIs) Outside France

In national statistical institutes outside France, while there is a shared need for richer and more responsive production of statistical data that relies more on administrative data, a range of responses can be seen, depending on the technical, organisational, cultural or legal context. Several major models have been identified:

– a highly integrated statistical system, which has long used administrative population registers and a shared identifier

Countries with administrative population registers based on an obligation to report changes in residence and those based on a personal identifier shared by all government bodies have been able to build a highly integrated statistical system. The common use of the phrase "population register" can cause confusion because it does not specify its purpose (administrative or statistical).

In Finland, for example, the production of statistical data is 95% based on data from administrative sources or registers. The Finnish population census has been based entirely on this type of data since 1980. The population information system collects a lot of data that describe people while also allowing interconnection with other files (see the Finnish statistical institute website*). It exists in a State and a society in which the interconnection of files on individuals does not raise major technical or organisational problems, nor does it raise any problems in terms of acceptance by the population. It is neither conceivable nor envisaged for France to move towards such a system.

– in the Netherlands: a highly integrated system with registers, administrative sources and survey data

The Dutch statistical institute (CBS**) forms part of a general approach to using the data sources available, the "System of Social Statistical Datasets (SSD)" (Bakker et al., 2014). It is a system of interconnected and standardised registers and surveys. It is a trove of information on people, households, jobs and benefits, pensions, education, hospitalisations, crime reports, housing, vehicles, etc.

In the Netherlands, it is the largest source of official social statistics, in the form of aggregated results that protect the confidentiality of the data used. The highly detailed data on individuals accessed by researchers remain in the secure environment managed by the CBS, which systematically checks that the exported data or results do not pose any risk of a breach of confidentiality***.

- in New Zealand: a resource for the statistical identification of individuals and housing in order to facilitate matching, in particular for research purposes

There is no shared individual identifier or usable administrative register (Bycroft *et al.*, 2022). An initial statistical register of individuals was compiled using civil status data (births and deaths), border data (immigration-emigration) and tax data. This makes it possible to interlink several sources of administrative data, on request, for statistical or research purposes. This resource was very useful for the 2018 census, which was hindered by data collection problems that could lead to biases in the data produced. It has made it possible to produce statistical results by combining (administrative or survey) data on education, the labour market, benefits, justice, health and safety, migration and business data.

Since 2021, an overhaul has been under way to establish a fully fledged statistical register of individuals and housing that is more integrated and easier to manage.

The New Zealand NSI pays a great deal of attention to communications regarding these techniques and the use of data: information is released to reassure users about quality and to get the entire population to buy into this new type of data collection, in particular the indigenous Māori population.

* *https://dvv.fi/en/personal-data.*

** *https://www.cbs.nl/en-gb/.*

*** *https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research.*

On the other hand, the abolition of housing tax (HT) on primary residences accelerated the process and imposed the need to find new solutions for the statistics that were created using HT data.

INSEE has for many years relied on a file generated through the management of HT, which has an ever-increasing range of ways in which it can be used. This file first enabled INSEE to establish a list of dwellings in each municipality, in order to prepare and monitor the annual census surveys. It also made it possible to calculate the population of the municipalities by updating the results of the census surveys. It was then used as a sampling frame for household surveys and served as a framework for the *Fichier Démographique sur les Logements et les Individus* (Demographic File on Dwellings and

> ***The abolition of housing tax (HT) on primary residences accelerated the process and imposed the need to find new solutions for the statistics that were created using HT data.***

Individuals, FIDÉLI) and as a reference for the establishment of household profiles, an essential prerequisite for calculating standards of living, including at granular geographical levels (Lamarche and Lollivier, 2021).

The objective with RÉSIL is therefore to build a system that provides at least the same services as the file generated from HT, that is to say, performing the census, sampling household surveys, reconstructing standards of living and creating composite study files. INSEE took the opportunity to go further, in three areas: bolstering the sustainability of the system (so as to avoid a repetition of the issues that arose with the abolition of HT), making progress in terms of monitoring the coverage it provides, and providing the "reference universe service" for administrative data collection and use.

The goal is therefore to build a more robust resource, in multiple respects:

- creating a register with unambiguously identified units that are stable over time, resulting in better quality;
- extending the coverage to dwellings not covered by housing tax, in particular communal homes (EHPAD[5], boarding schools, etc.);
- using multiple sources to populate and update this register, for more comprehensive coverage of the field of observation, ensuring continuity in the event of a change in the data. The abolition of housing tax caused INSEE to realise that this was a genuine risk.

The centralised nature of the register means that it is also possible to develop data matching (and therefore to produce works based on multiple sources) and to monitor the coverage of administrative sources, with a view to using them for statistical purposes.

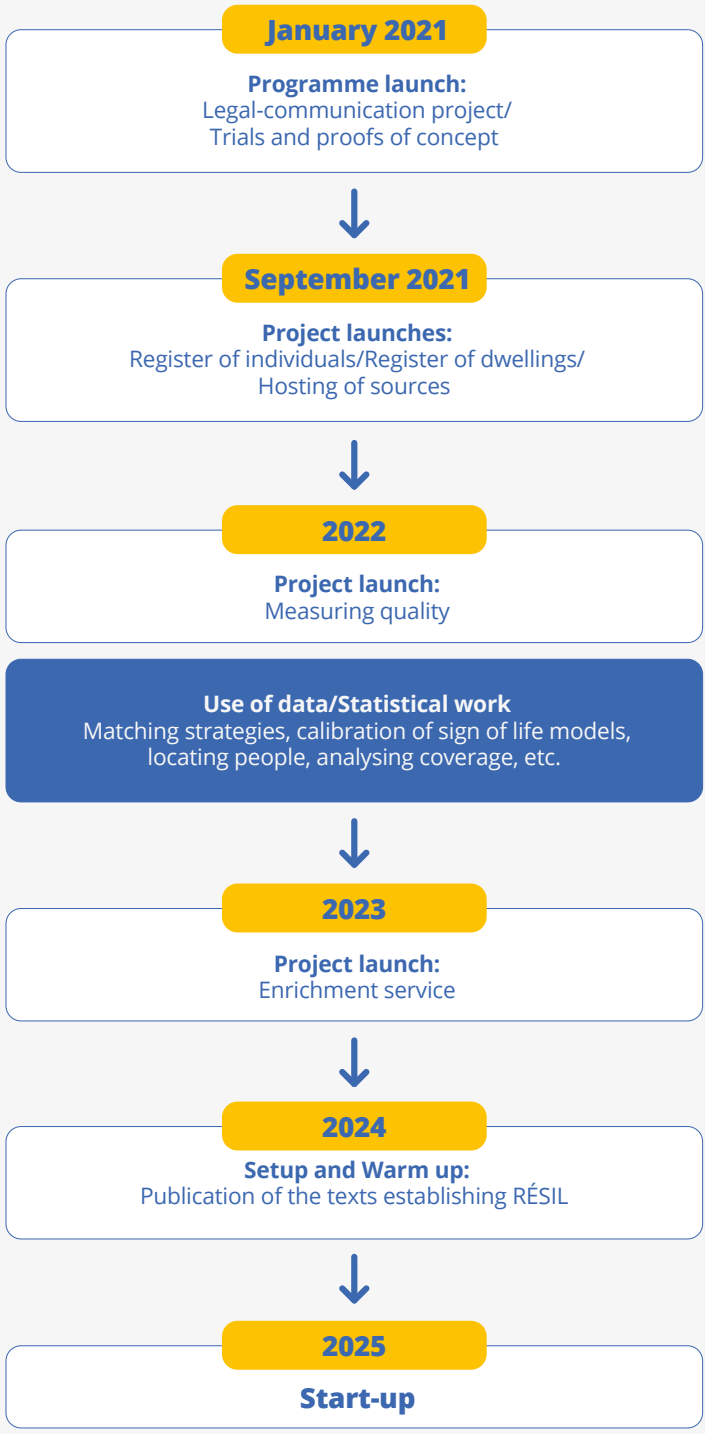
> ***The goal is to create the* Répertoire statistique des individus et des logements *(Statistical Register of Individuals and Dwellings, RÉSIL), not only making it possible to create the reference universes required for demographic and social statistics, but also facilitating the matching of administrative data.***

The goal is to create the *Répertoire statistique des individus et des logements* (Statistical Register of Individuals and Dwellings, RÉSIL), not only making it possible to create the reference universes required for demographic and social statistics, but also facilitating the matching of administrative data with other administrative data or data from other sources, mainly from surveys. The objective is to achieve this in 2025 (***Figure 1***).

In accordance with the provisions of the decree establishing RÉSIL[6], the register

5 EHPAD: *Établissement d'hébergement pour personnes âgées dépendantes* (Residential Facility for Dependent Elderly Persons).
6 See the legal references at the end of the paper.

▶ **Figure 1 - RÉSIL: a Project Built Step by Step**



"has as its purpose, with a view to contributing to the public debate as well as to the development and assessment of public policies, bolstering the ability of the National Institute for Statistics and Economic Studies (INSEE) and the Ministerial Statistical Offices (MSOs) to produce statistical data and studies, by enabling the establishment of a national register of individuals and dwellings and facilitating the matching of administrative data with other data sources".

Various Uses, which are Essential for Compiling Demographic and Social Statistics

RÉSIL is therefore a production infrastructure that makes it possible to meet several exclusively statistical objectives.

It will offer, to the Official Statistical Service only, a data matching service that has multiple applications[7] and can be used in particular (Dupont, 2023) to:

- reduce the collection of information through surveys by asking questions only on aspects not covered by administrative data;
- match administrative data to other administrative data, in order to produce statistics on a granular level, which cannot be produced using a sample;
- enrich a file with additional variables to deepen the analyses (for example, the addition of income information to the *enquête sur les ressources et les conditions de vie* (survey on income and living conditions (SRCV));
- provide clarity regarding specific methodological aspects: for example, performing matching between the files from the *enquête Emploi* (Labour Force Survey) and the historical job seekers' file in order to measure the difference between the concepts of unemployed persons as defined by the ILO[8] and job seekers registered with *France Travail*;
- assess public policies (monitor the trajectory of beneficiaries of specific forms of aid).

These matching operations will be better and less costly thanks to the inclusion of common identifiers; representativeness and quality will also be more easily measurable.

> ***RÉSIL will also make it possible to measure the quality of administrative sources – one of the key official statistics resources.***

RÉSIL will also make it possible to measure the quality of administrative sources – one of the key official statistics resources. It will allow a comparison between the actual coverage area of a statistical source and the list of individuals or dwellings in RÉSIL, and thus enable the detection of any potential lack of coverage.

By extending existing systems, the sampling frame used to draw samples for household surveys conducted by the Official Statistical Service will be produced by RÉSIL. The coverage provided will be better guaranteed than previously, not only due to the diversity of sources used, but also the addition of people living in communal

7 See the paper on matching by Koumarianos, Lefebvre and Malherbe in this issue.
8 ILO: International Labour Office.

homes and the taking of deaths into account more quickly; it will also be possible to add variables from sources other than tax sources to the sampling frame, which will make sampling more accurate.

RÉSIL will provide information that will help in preparing and carrying out annual census surveys and extrapolating the results, like the current use of files resulting from the housing tax system (***Box 2***). RÉSIL can thus be used to facilitate the production of more timely demographic indicators.

Lastly, through the use of shared and effective tools, RÉSIL will allow for data to be produced more consistently and for processing to be more consistent yet also more efficient. By facilitating the reconciliation of data, as well as their comparison, by unifying the reference data and the calibration margins[9], RÉSIL makes it possible to "break the silos" of the current information system organised by source.

▶ Box 2. RÉSIL and the Population Census, a Win-Win Partnership

RÉSIL will, in the place of housing tax, produce the data that the census needs to prepare, control and extrapolate results from census surveys. Specifically, it could make it possible to provide earlier estimates, as requested by many users and by Eurostat.

Conversely, the census makes it possible to assess the quality of the sources used as input for RÉSIL, whether in relation to individuals or dwellings.

Moreover, statistical methods known as dual system estimation (Zhang and Dunne, 2017) should ultimately make it possible to measure the coverage of RÉSIL and also the coverage of the census by comparing these two sources and thus allow the identification of coverage biases.

As in other countries, RÉSIL should allow for the modernisation of the census system. However, it is too early to indicate either the form of this modernisation or its date of implementation.

▶ What Will RÉSIL Contain?

In concrete terms, RÉSIL will be composed of two separate statistical sub-registers, linked to each other: a register of individuals and a register of dwellings. They will be updated regularly for births and deaths and with tax and social sources (data on social and family benefits from the CNAF[10], DSN[11] and PASRAU) or other sources covering a specific audience, such as files on enrolment in higher education.

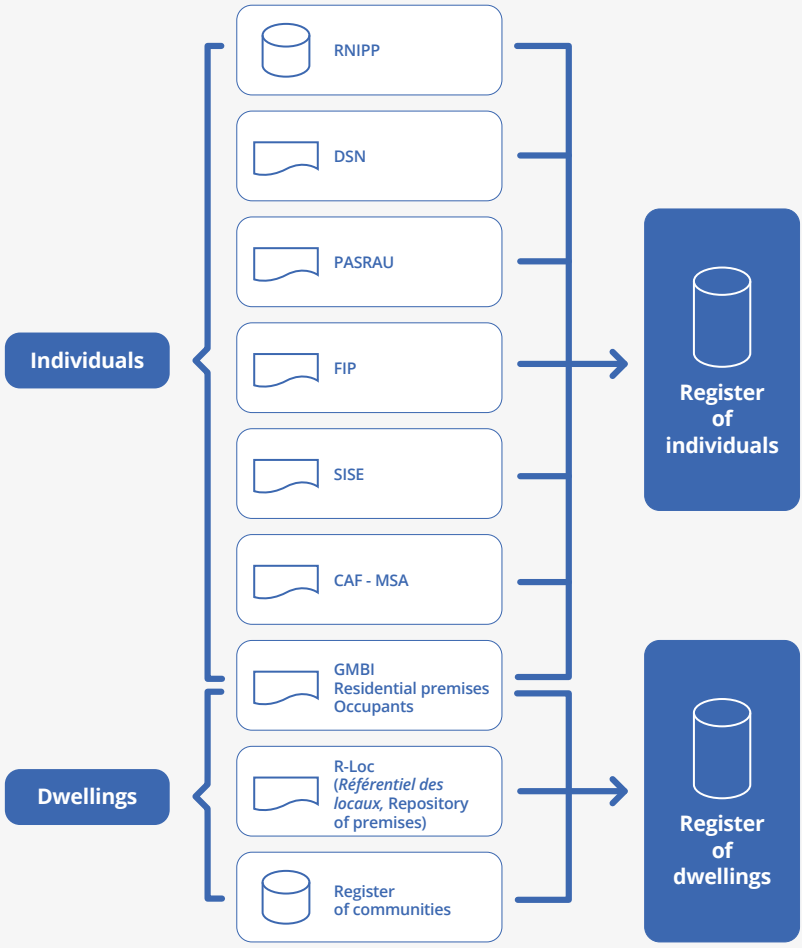
A key point is that, in these sources, only data identifying individuals and dwellings, address data and data on the links between individuals and dwellings are retained (***Figure 2***).

9 Margin calibration is a statistical technique used to improve the accuracy of sample surveys. It involves modifying the sampling weights of the individuals contained in the sample so that the weighted totals of certain variables in the sample correspond to the known totals for these variables over the entire field of observation (population, housing stock or companies) (Deville *et al.*, 1992).

10 CNAF: *Caisse nationale d'allocations familiales* (National Family Allowances Office).

11 DSN: *Déclaration sociale nominative* (Nominative Social Declaration).

▶ Figure 2 - RÉSIL and its Input Sources*



RNIPP: *Répertoire National d'Immatriculation des Personnes Physiques* (National Register for the Identification of Individuals)
DSN: *Déclaration sociale nominative* (Nominative Social Declaration)
PASRAU resource: *Prélèvement À la Source pour les Revenus AUtres* (Withholding Tax for Other Incomes)
FIP: *Fichier d'imposition des personnes* (Tax File for Individuals)
SISE: *Système d'information sur le suivi de l'étudiant* (Student Monitoring Information System)
CAF: *Caisses d'allocations familiales* (Family Allowances Office)
MSA: *Mutualité sociale agricole* (Agricultural Mutual Benefit Fund)
GMBI: *Gérer mes biens immobiliers* (Manage My Property)

** This list of sources may change over time, subject to the receipt of a favourable opinion from the CNIL and the CNIS.*

These registers will be used to produce annual "snapshots", which will make up the reference universe:

- the list of individuals present within the national territory on 1 January;
- the list of dwellings located within the national territory on 1 January, and their status (primary residence, secondary residence or vacant dwelling);
- the list of households[12] and their composition on 1 January.

The lists of individuals and dwellings will be used as a reference for demographic and social statistics. Lists of households are essential for establishing data, such as on standards of living, or for conducting surveys.

The information in the RÉSIL registers will essentially be identification keys to fulfil the role of a reference list (avoiding omissions and duplicates) and to perform the record linkage:

> ***The registration number for the RNIPP (the NIR or social security number) will not be stored in RÉSIL.***

- identifiers of individuals: the *code statistique non signifiant* (Non-Significant Statistical Code, CSNS) (Espinasse *et al.*, 2023) created by the French Law for a Digital Republic in 2016[13] to facilitate record linkage between sources within the Official Statistical Service, an identifier that is specific and strictly internal to RÉSIL, which is permanent to allow history management, the identifiers of the sources used in RÉSIL (for individuals and households). The registration number for the RNIPP (the NIR or social security number) will not be stored in RÉSIL;
- dwelling identifiers: a permanent identifier specific to RÉSIL, the identifiers of the sources used for RÉSIL;
- Each dwelling will have an address identifier from INSEE's address repository in the format of "BZ140JD" (and not the plain-language address "*8 rue Zéphyrin Brioché à Gleux-lès-Lure, département de la Haute-Saône*"), which makes it unusable outside INSEE;
- civil status data: surname, forename, date and place of birth and, where applicable, date of death;
- links between individuals and their dwelling or their home address[14], with, where appropriate, several possible dwellings for the same individual depending on the sources; a primary residence will ultimately be determined for each individual.

Some other variables to allow register management and quality measurement:

- dates of data updates;
- effective dates (start date and end date) for certain variables in RÉSIL for which a history is ideally kept;

12 A household includes all people sharing the same dwelling.
13 See the legal references at the end of the paper.
14 In some cases, it will not be possible to distinguish between multiple dwellings located at the same address.

- indicator of presence within French territory;
- presence of the person or dwelling in each administrative source (yes/no).

RÉSIL will not contain any other information. Data in connection with aspects such as income, marital status, occupation, housing area, etc. will be included in specific databases separate from RÉSIL and will only be used for separate processing on request. RÉSIL will not be a "mega-database" containing everything we know about each individual or dwelling.

The RÉSIL register is based on four major pillars: data of diverse origins, statistical processing allowing these data to be transformed into a quality statistical register, a solid legal basis and a "social mandate".

▶ Four Pillars for RÉSIL

First Pillar: Diverse Sources of Information for a Robust Result

RÉSIL will use several data sources to ensure:

- the best possible coverage of the population (no administrative source is exhaustive and no administrative source is fully in line with the statistical concepts of the resident population[15]);
- more precise location of individuals and better understanding of multiple residences;
- the sustainability of the system in the absence of a source, or even if the source changes or disappears, so as to avoid breaks in data collection (such as in the case of the abolition of housing tax).

The results of the initial trials confirm the benefit of using each of these sources, in addition to the tax source alone, in terms of population coverage. Using the *Enquête Annuelle de Recensement* (Annual Census Survey, EAR) as a reference point, the overall gain in terms of coverage[16] is around 2 percentage points for individuals over 18 years of age, but it is 10 points for those aged 21–25. Coverage by age is thus more homogeneous than when using the tax source alone.

For people living in communities[17] (approximately 1.3 million people in retirement homes, student halls of residence, boarding schools, workers' homes, penitentiary institutions, religious communities, etc.), the coverage rate increases by 10 points, from 80% to 90%.

15 The resident population includes all persons residing in France, regardless of their nationality and situation, from the point at which they have been in France for at least one year or, if they have just arrived, when they intend to stay for at least one year. However, people passing through (tourists, seasonal workers or foreign students coming for a 9-month academic year) are not included. This definition corresponds to international rules and thus allows comparisons between countries.
Persons without a usual residence in another country are counted among the resident population of France if they are in the country on the reference date for the calculation of this population.

16 The rate of coverage is estimated by the proportion of persons identified who are listed in the tax source or other data sources.

17 See definition (as per the population census): *https://www.insee.fr/en/metadonnees/definition/c1134*.

▶ Table of Sources Used in the RÉSIL Population Process

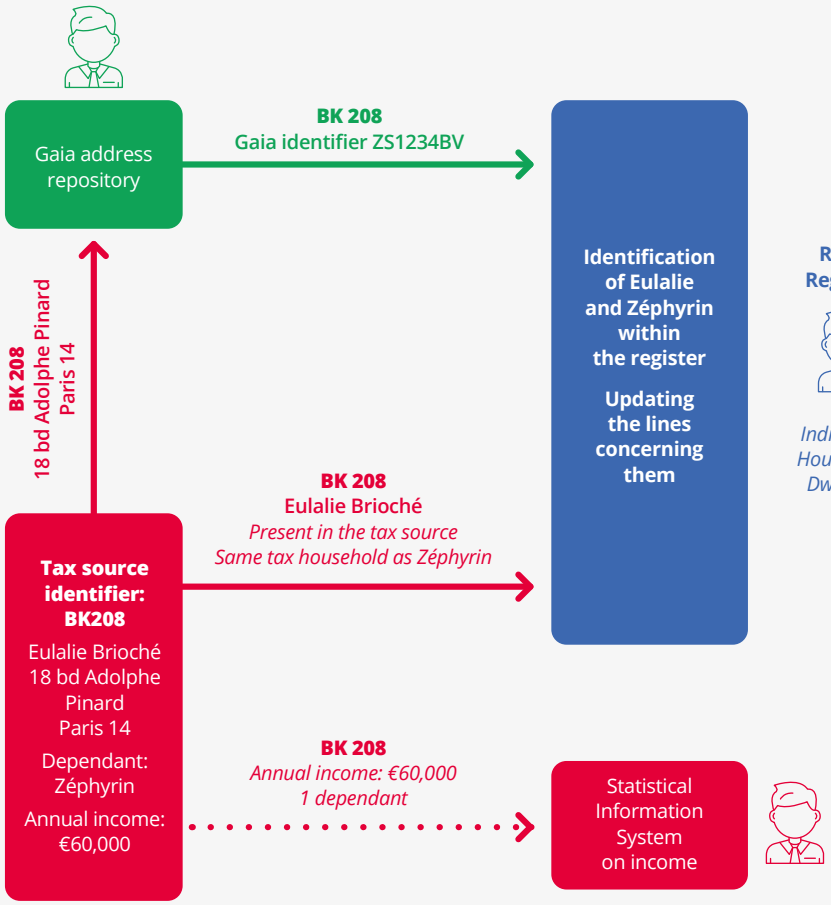
Source	Justification
Répertoire national d'identification des personnes physiques **(National Register for the Identification of Individuals, RNIPP)**	• To update the list of individuals with births and deaths. • To update civil status data.
Tax files: • ***Fichier d'imposition des personnes*** **(Tax File for Individuals, FIP)** • ***Fichier permanent des occurrences de traitement des émissions*** **(Income tax declarations for tax households, POTE)** • ***Fichier de mise à jour des informations cadastrales*** **(Cadastral Information Update File, Majic)** • **R-Loc (*Référentiel des Locaux*, Repository of Premises)** • **"*Gérer mes biens immobiliers*" ("Manage My Property", GMBI) file**	• Knowledge of the housing stock (Repository of Premises, gradual relay of Land Register data). • Knowledge of housing use (primary or secondary residence) and occupants (via GMBI data for the declared occupant, FIP and POTE for the other occupants of the dwelling). • Identifying and locating data are reprocessed, excluding any information on tax bases, tax revenues, taxes due or taxes paid.
Déclaration Sociale Nominative **(Nominative Social Declaration, DSN)**	The monthly frequency of this source allows for a reliable update of the register of individuals (duration of presence within the territory, taking into account young active persons in addition to tax data, taking changes in address into account earlier).
Prélèvement à la source pour les revenus autres **(Withholding Tax for Other Incomes, PASRAU)**	The monthly frequency of this source allows for a reliable update of the register of individuals (duration of presence within the territory, taking into account recipients of income other than salary, in addition to tax data).
Caisses d'allocations familiales **(Family Allowances Office, CAF) beneficiary reference file**	This source provides details on the composition of beneficiary households, which are essential for reliable information on households (particularly their composition).
Mutualité sociale agricole **(Agricultural Mutual Benefit Fund, MSA) beneficiary file**	• This source provides details on the composition of beneficiary households, which are essential for reliable information on households (particularly their composition). • The coverage of the MSA supplements that of the CAF.
Annual file on enrolment in higher education	Ce fichier permet d'améliorer la couverture du répertoire sur les jeunes de 17 à 25 ans, parfois absents des sources fiscales ou localisés de manière ambiguë.
Register of communities prepared for conducting the population census	Communities are the primary residence of approximately 1.3 million residents in France. It is therefore essential to supplement the ordinary housing stock with that of communal homes. The register of communal homes feeds into the RÉSIL register of dwellings.
Statistical surveys to monitor the quality of the register	INSEE may need to conduct statistical surveys to monitor the coverage of the register, or even completeness surveys for areas poorly covered by administrative sources.

> ***For each of the sources, decisions made regarding the data used are selective.***

For each of the sources (***see Table of Sources***), decisions made regarding the data used are selective. The data will be selected and directed as input into INSEE's information system to feed only into the processes that need them. The hosting resource functions as a sorting station. During this step, the NIRs included in the sources will be replaced by the corresponding non-significant statistical codes (CSNSs): the identity data will be directed towards RÉSIL, either for updating or to ensure the quality of the identification. Address data will also be processed at source to replace them with a non-significant identifier from INSEE's register of addresses.

"Profession" data, along with non-significant identifiers for people and addresses, will be integrated into the information systems to produce statistical data (***Figure 3***).

▶ Figure 3 - Populating RÉSIL: Data Selection and Processing



RÉSIL adheres to a principle of minimisation of the data processed: it will contain very few variables and will make it possible to delete directly identifying data from other INSEE information systems and replace them with pseudonyms. In addition, RÉSIL will not contain any statistical data that make it possible to describe individuals and dwellings, as such data will only be processed by applications intended to produce statistical data.

To populate the register, a modernised tool has been developed to make the administrative data used for statistical purposes available.

> ***The emergence of RÉSIL is an opportunity to streamline the system for hosting administrative data.***

For most of the sources used, the RÉSIL register is a new "user", in addition to the producers of statistical data on employment, income or housing. However, the emergence of RÉSIL is an opportunity to streamline the system for hosting administrative data, which currently consists of several juxtaposed resources, linked to both sources and uses. In addition to the project to build the registers, the RÉSIL project is accompanied by a project to modernise and unify the system for hosting and structuring administrative data into raw statistical data that can be easily used to produce figures relating to statistical units (individuals, dwellings and households). This system is based on the *Accueil-Réception-Contrôle* (Receipt-Acceptance-Control) tool, known as ARC[18]. The principle is to use the same tool to host the various sources and to make them available to different users within the Official Statistical Service. This streamlining makes it possible to invest in the performance and functional enrichment of such a tool, without sacrificing its security or its ability to adapt quickly to changes in sources, or even to host new sources. Robustness and processing performance of such a tool are crucial, taking into account not only the number, volume and frequency of sources to be hosted (at least three of the largest sources are monthly), but also the need to quickly populate the systems that produce statistical indicators. The tool's ability to adapt rapidly to changes in administrative sources (dictated by the policies they implement and not by the statistics they help to produce) and to the emergence of new sources is also essential.

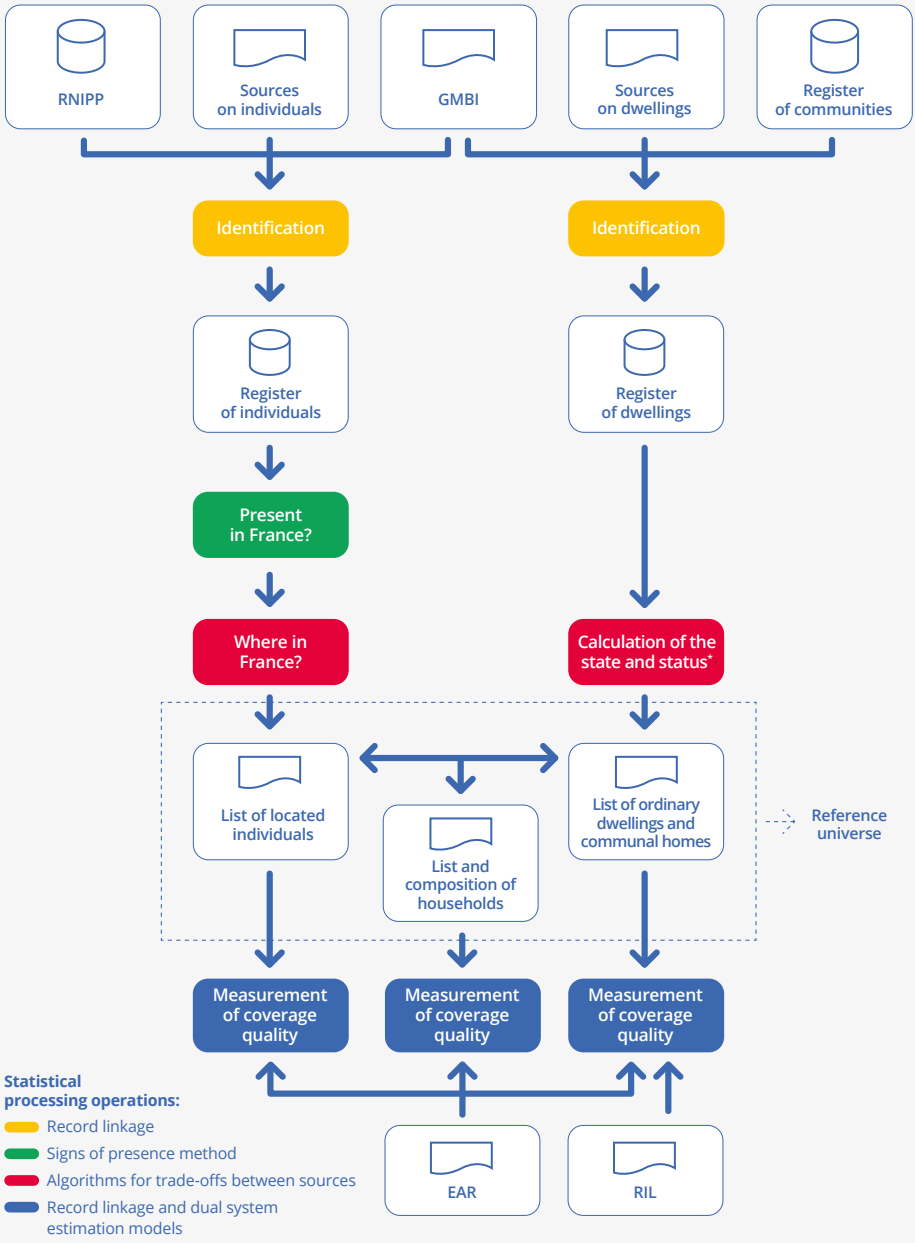
Second Pillar: to Ensure the Quality of the Register, Effective and Innovative Statistical Processing, Inspired by International Practices

Having the data is essential, but that is not enough. We also need effective tools to transform these data into a register and then a reference universe (list of individuals actually resident on 1 January, list of habitable dwellings, list of households), which is of a suitable standard for statistical use.

The first task is to identify individuals, to "update the correct lines of the file", on the basis of information that is sometimes incomplete or contains errors; this is the challenge faced by the identification engines (***Figure 4***). RÉSIL will use a more selective process than that used for the CSNS, insofar as it can, for the most uncertain cases, use additional information such as the composition of the household or the person's address.

18 See the paper by Lefebvre, Soulier and Tortosa on the hosting of administrative data in this issue.

▶ Figure 4 - From the Data Sources to the Reference Universe: a Succession of Statistical Processing Operations



RNIPP: *Répertoire national d'identification des personnes physiques* (National Register for the Identification of Natural Persons)
GMBI: "*Gérer mes biens immobiliers*" ("Manage My Property") file
EAR: *Enquêtes annuelles de recensement* (Annual census surveys)
RIL: *Répertoire d'immeubles localisés* (Register of Localized Buildings)

** State and status calculation: the purpose is to determine whether or not the dwelling is inhabited and whether it is a primary or secondary residence or a vacant dwelling.*

It is then necessary to specify which of the individuals included in the register actually reside within the national territory and to determine the primary residence for each of them. This makes it possible to build the reference universe.

Does the person still live within the national territory? The "signs of presence" method (or "signs of life" in the academic and professional literature) involves using, in addition to indisputable civil status data (births and deaths), information relating to the presence of individuals in a given administrative source, with an address located within the national territory. If an individual who is not dead in the RNIPP is absent in all administrative sources, there is a high probability that they no longer reside within the national territory. If the person appears in only some of the sources in which they should normally be found and has an address abroad in at least one source, there is a significant probability that they have left the national territory. Each of the signs of presence in the sources can be weighted by the quality of the identification of the individual[19] and its relevance to the individuals concerned. For example, using the student file will be relevant for those aged 18–25. This method is used in several countries, such as Estonia, Ireland, Italy and Australia. Its use is encouraged by Eurostat[20], in connection with the development of the use of administrative data.

Decision-Making Rules in the Event of Multiple Addresses

When a person has different addresses in different files, it is also necessary to determine the address of their primary residence. The discrepancy may result from a delay in updating administrative files (the person has moved but the information has not yet been taken into account) or from multiple residences ("geographical singles", students staying at their place of study but still attached to the tax household of their parents, children in shared custody, etc.). The decision-making rules to be adopted must make it possible to locate persons at their primary residence, in accordance with the concepts established for international comparisons and implemented for the population census.

> ***In order to work with stable data, it is important to have a snapshot showing the situation on 1 January.***

In order to work with stable data, it is important to have a snapshot showing the situation on 1 January.

The reference universe generated from RÉSIL must correspond to a stable situation to allow its use to be shared. But the register is active, with regular updates. It is therefore necessary to "take a snapshot" of the register reflecting the situation on a given date, with 1 January being the date conventionally chosen.

RÉSIL will be updated using various sources, not all of which will be available at the same time. Rather than waiting for the last source to take the snapshot, the decision was made to produce three versions of each snapshot, as the data arrived.

19 If the identity traits of the source do not make it possible to reliably identify an individual in the register, less weight will be given to the sign of presence in that source, as it could be a matching error.
20 Eurostat is the Community statistical institute, associated with the European Commission.

For example, the provisional reference universe for 1 January 2025 will be produced in the summer of 2025, with the semi-final version arriving in January/February 2026 and the final version in early summer 2026.

Each user can therefore prioritise recency over completeness, or vice versa, based on how they will use the reference universe.

It is also necessary to measure the quality of coverage, which any register must be able to do. Coverage quality measurements use comparisons with the annual census surveys and rely on the dual system estimation method.

The principle behind this estimate is to compare the two sets of data collected – RÉSIL data and census survey data – to count persons present in both sources and persons present in only one of the sources in order to then determine the number of persons absent from both sources, therefore the size of the total population, under the assumptions, in particular, that the two sets of data are collected independently and that there are no persons incorrectly included in either of the datasets.

This method is used in several countries to estimate the coverage of comprehensive censuses. It is currently being developed to measure the coverage of directories, for example in Italy for the directory compiling municipal population registers.

Third Pillar: a Clear and Solid Legal Foundation that Authorises Processing and Protects the Resulting Data

RÉSIL forms part of the national and European legal framework on the production of official statistics and the protection of personal data (the French Act of 7 June 1951 on Legal Obligation, Coordination and Confidentiality in Statistical Matters, the General Data Protection Regulation (GDPR) and the French Data Protection Act[21]). No legislative provisions are needed as the above-mentioned legal environment guarantees access to the necessary data and the obligations relating to the protection thereof.

INSEE believed that the appropriate level of legal instrument to act as a basis for such processing is a decree issued by the Council of State. An order by the data controller, in accordance with the GDPR, could have been sufficient if it had been a purely technical text, but the questions of a more political or societal nature to which the resource may give rise justified an instrument of this legal level, providing a critical preliminary examination, by the CNIL[22] and by the Council of State, thereby strengthening the legitimacy of the resource.

The decree in the Council of State creating RÉSIL was published in the French Official State Gazette on 7 January 2024. It includes provisions that feature in most texts creating the processing of personal data: the creation of the processing and its purpose, the list of variables and their retention period, the users of the register, the recipients of the data it manages and the files it will produce, and provisions relating to the

21 See the legal references at the end of the paper.
22 CNIL: *Commission Nationale de l'Informatique et des Libertés* (National Commission on Information Technology and Liberties).

security of the information system. It also includes several provisions more specific to RÉSIL: the definition of the matching operations, the conditions for changing the list of sources and an explicit reference to the ethical requirements specific to the profession of statistician. An order[23] establishing the list of sources used to build and update RÉSIL, promulgated pursuant to this decree, was also published in the French Official State Gazette on 7 January 2024.

Fourth Pillar: a Social Mandate that Must Be Reinforced Continuously

The aim of all of the foregoing is to give INSEE the technical or legal capacity to build and manage the register. This capacity must be supported by legitimacy or a social mandate. The bodies constituted and the competent authorities confer part of this legitimacy, through the legal instruments governing the functioning of official statistics and the processing by RÉSIL in particular. However, this legitimacy would not be sufficient without a "social mandate", proof that this processing is recognised and accepted by the population, which trusts INSEE to carry it out.

> ***INSEE wanted to involve representatives of civil society ahead of the project in the form of a thorough consultation with various stakeholders.***

INSEE wanted to involve representatives of civil society ahead of the project in the form of a thorough consultation with various stakeholders in order to take into account the views expressed therein in the construction and uses of the statistical register.

The consultation[24] carried out in 2022 was based on two main actions: first, a meeting[25] of the *Conseil national de l'information statistique* (National Council for Statistical Information, CNIS)[26] on 28 January 2022 and, second, the creation of a consultation group, under the aegis of the CNIS, which ran from May to September and produced a report[27] that is published on its website.

The meeting afforded INSEE and various representatives of official statistics an opportunity to present the practices, uses and techniques of matching and to show how it contributes to knowledge and public action (measuring the professional integration of young people, studying outcomes for beneficiaries of statutory minimum income benefits, understanding gaps between two data sources, etc.). It was an opportunity for participants to ask questions or even express concerns about what a resource such as RÉSIL could enable if it were misused, but also revealed the participants' desire for the efforts made in terms of communication and transparency during the meeting to continue with a more in-depth consultation on the project.

23 See the legal references at the end of the paper.
24 See the paper by Dupont, Dussart and Guillaumat-Tailliet on the ethical challenges facing RÉSIL in this issue.
25 *https://www.cnis.fr/evenements/appariements-de-donnees-individuelles-entre-richesse-de-linformation-statistique-et-respect-de-la-vie-privee/.*
26 The *Conseil national de l'information statistique* (National Council for Statistical Information, CNIS) facilitates interactions between the producers and users of official statistics.
27 *https://www.cnis.fr/wp-content/uploads/2022/11/rapport-version-dfinitive.pdf.*

The consultation group was an opportunity to bring together a wide range of expertise (protection of fundamental freedoms, data protection from legal and IT points of view, digital transformation, ethics, research, etc.) to draw up a list of questions raised by the RÉSIL project and to discuss the responses provided by INSEE. This group concluded that the project was legitimate and in line with the principle of proportionality of the processing, provided that some of the sources initially envisaged by INSEE were not used. The group also felt that it was necessary to hear external views (Council of State, CNIL, CNIS, Official Statistics Authority and National Cybersecurity Agency) during the construction and use of RÉSIL. It also emphasised the need for transparency regarding the register and its uses, allowing another external point of view: that of the persons whose data would be processed.

INSEE is following this path, not only in terms of the content of the decree[28] and the methods of constructing the information system, but also in communications regarding the register and its applications. A page dedicated to RÉSIL[29] has thus been set up on INSEE's website. The CNIS will play an important role, in accordance with the arrangements to be established, in terms of the continued consultation on RÉSIL and on the services it will provide, in particular the construction of files enriched by record linkage.

Over the next two years, developments in the information system will be finalised. The register will be initialised and statistical processing will be carried out to ensure a high level of quality. The first reference universes (dated 1 January 2025) will be produced gradually between mid-2025 (provisional version) and mid-2026 (final version), with the first services of record linkage being provided in early 2026.

28 *https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000048866207.*
29 *https://www.insee.fr/fr/information/7748883.*

▶ Legal References

- Act N° 51-711 of 7 June 1951 on Legal Obligation, Coordination and Confidentiality in Statistical Matters. At: *Légifrance website*. [online]. Updated 25 March 2019. [Accessed 5 March 2024]. Available at: *https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573*.

- Act N° 78-17 of 6 January 1978 on Information Technology, Data Files and Civil Liberties. At: *Légifrance website*. [online]. Updated 21 February 2024. [Accessed 30 April 2024]. Available at: *https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460*.

- Act No 2016-1321 of 7 October 2016 for a Digital Republic. At: *Légifrance website*. [online]. [Accessed 5 March 2024]. Available at: *https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746*.

- Decree N° 2024-12 of 5 January 2024 creating an automated processing system for personal data called the "répertoire statistique des individus et des logements" (Statistical Register of Individuals and Dwellings, RÉSIL). At: *Légifrance website*. [online]. [Accessed 5 March 2024]. Available at: *https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000048866207*.

- Order of 5 January 2024 promulgated pursuant to Article 2 of Decree N° 2024-12 of 5 January 2024 creating an automated processing system for personal data called the "répertoire statistique des individus et des logements" (Statistical Register of Individuals and Dwellings, RÉSIL). At: *Légifrance website*. [online]. [Accessed 5 March 2024]. Available at: *https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000048866233*.

- Deliberation N° 2023-080 of 20 July 2023 providing an opinion on a draft decree in the Council of State relating to the creation of the automated processing system for personal data allowing the management of the statistical register of individuals and dwellings, and on the related order. At: *Légifrance website*. [online]. [Accessed 4 April 2024]. Available at: *https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000048934586?page=1&pageSize=100&searchField=ALL&searchType=ALL&sortValue=DATE_DECISION_DESC&tab_selection=cnil&timeInterval=01%2F07%2F2023+%3E+31%2F08%2F2023&typePagination=DEFAULT*.

▶ Bibliography

- BAKKER, Bart F.M., VAN ROOIJEN, Johan and VAN TOOR, Leo, 2014. The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. In: *Statistical Journal of the IAOS*. Volume 30, nº 4, pp. 411-424. [online]. [Accessed 5 March 2024]. Available at: *https://www.cbs.nl/-/media/imported/documents/2016/53/system-of-social-statistical-datasets.pdf*.

- BÉNICHOU, Yves-Laurent, ESPINASSE, Lionel and GILLES, Séverine, 2023. The Non-Significant Statistical Code (CSNS): a service to facilitate file matching. In: *Courrier des statistiques*. [online]. 30 June 2023. Insee. Nº N9, pp. 64-85. [Accessed 5 March 2024]. Available at: *https://www.insee.fr/en/information/8232634?sommaire=8232646*.

- BYCROFT, Christine, EATHERLEY, Clara, PAGE, Mathew and TA'ALA, Shane, 2022. A statistical person register in New Zealand: Progress and challenges. In: *Statistical Journal of the IAOS*. [online]. 21 March 2022. Volume 38, nº 1, pp. 225-230. [Accessed 5 March 2024]. Available at: *https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji210922*.

- DEVILLE, Jean-Claude and SÄRNDAL, Carl-Erik, 1192, Calibration Estimators in Survey Sampling. In: *Journal of the American Statistical Association*. [online]. June 1992. Vol. 87, Nº 418, pp. 376-382. [Accessed 23 May 2024]. Available at: *https://www.jstor.org/stable/2290268*.

- DEMOTES-MAINARD Magali, 2019. Elire, un projet ambitieux au service du Répertoire électoral unique. In: *Courrier des statistiques*. [online]. June 2019. Insee. Nº N2, pp. 58-71. [Accessed 23 May 2024]. Available at: *https://www.insee.fr/fr/information/4168399?sommaire=4168411*.

- DUPONT, Françoise, 2023. Quels types de sources l'Insee utilise-t-il pour construire ses statistiques ? In: *Blog Insee*. [online]. 16 May 2023. [Accessed 5 March 2024]. Available at: *https://blog.insee.fr/quels-types-de-sources-l-insee-utilise-t-il/*.

- DUPONT, Françoise, 2023. Les appariements de données de la statistique publique : des analyses enrichies, un cadre juridique protecteur. In: *Blog Insee*. [online]. 1st September 2023. [Accessed 5 March 2024]. Available at: *https://blog.insee.fr/appariements-de-donnees-de-la-statistique-publique/*.

- ESPINASSE, Lionel and ROUX, Valérie, 2022. The National Register for the Identification of Individuals (RNIPP) at the core of French administrative life. In: *Courrier des statistiques*. [online]. 29 November 2022. Insee. Nº N8, pp. 72-92. [Accessed 5 March 2024]. Available at: *https://www.insee.fr/en/information/6692672?sommaire=6692684*.

- HACHID, Ali and LECLAIR, Marie, 2022. Sirus, the business register for statisticians. In: *Courrier des statistiques*. [online]. 29 November 2022. Insee. Nº N8, pp. 115-130. [Accessed 5 March 2024]. Available at: *https://www.insee.fr/en/information/6692680?sommaire=6692684*.

- LAMARCHE, Pierre and LOLLIVIER, Stéfan, 2021. Fidéli, l'intégration des sources fiscales dans les données sociales. In: *Courrier des statistiques*. [online]. 8 July 2021. Insee. Nº N6, pp. 28-46. [Accessed 5 March 2024]. Available at: *https://www.insee.fr/fr/information/5398683?sommaire=5398695*.

- RIVIÈRE, Pascal, 2022. Qu'est-ce qu'un répertoire ? De multiples exigences pour un système complexe. In: *Courrier des statistiques*. [online]. 29 November 2022. Insee. Nº N8, pp. 52-71. [Accessed 5 March 2024]. Available at: *https://www.insee.fr/fr/information/6665186?sommaire=6665196*

- ZHANG, Li-Chun and DUNNE, John, 2017. Trimmed Dual System Estimation. In: *Capture-Recapture Methods for the Social and Medical Sciences*, pp. 239-259. Chapman and Hall/CRC editions. ISBN 978-1-49-874531-4.