



Conseil national
de l'information statistique

LE PROJET DE RÉPERTOIRE STATISTIQUE D'INDIVIDUS ET DE LOGEMENTS « RÉSIL »

Rapport du groupe de concertation du Cnis

Président : Jean-Marie DELARUE

Rapporteuse : Françoise DUPONT

Novembre 2022

Sommaire

AVANT-PROPOS.....	5
SYNTHÈSE.....	7
RAPPORT.....	12
1. La démarche de concertation.....	12
1.1. Contexte.....	12
1.2. Démarche.....	14
2. Etat des travaux statistiques.....	16
2.1. Recours aux données administratives et appariements, bref historique et pratiques actuelles.....	16
2.1.1. La statistique publique.....	16
2.1.2. Les appariements pour la recherche.....	18
2.2. La situation à l'étranger : pratiques, projets et recommandations internationales en matière de répertoires de personnes et d'utilisation de sources administratives.....	19
2.2.1. Pratiques et projets.....	19
2.2.2. Recommandations, bonnes pratiques et prescriptions internationales.....	23
3. Etat du cadre juridique et déontologique.....	25
3.1. Le cadre juridique général de la statistique publique en France.....	25
3.2. Le cadre juridique général de la statistique publique européenne.....	28
3.3. L'accès de la statistique publique à des données administratives.....	29
3.4. L'accès de l'Insee à des données produites par des personnes morales de droit privé.....	30
3.5. La réalisation des appariements de données.....	31
3.6. Principes éthiques et déontologiques de la statistique publique.....	32
3.6.1. Au niveau mondial.....	32
3.6.2. au niveau européen.....	34
3.6.3. Deux exemples de pratiques : Royaume-Uni et Canada.....	34
4. Le projet de répertoires statistiques d'individus et de logements (Résil).....	36
4.1. L'objectif de Résil.....	36
4.2. Le contenu de Résil.....	37
4.3. Durée de conservation des données de Résil.....	38
4.4. La mise à jour de Résil.....	40
4.5. Le cadre juridique prévu pour Résil.....	43
4.6. La sécurité du système d'information Résil.....	43
4.7. Les services que rendra le répertoire Résil.....	44
4.7.1. Appariements de données réalisés grâce à Résil.....	44
4.7.2. Base de sondage pour tirer les échantillons des enquêtes réalisées auprès des personnes.....	46
4.7.3. Préparation et contrôle des enquêtes de recensement, calcul de la population des communes.....	46
5. Le point de vue du groupe de concertation.....	46
5.1. Considérations générales.....	46
5.2. Les principes au cœur de Résil.....	48
5.2.1. Sur la nécessité du répertoire et ses modalités d'alimentation.....	48
5.2.2. Sur l'utilisation de Résil pour réaliser des appariements pour d'autres utilisateurs.....	49
5.3. L'architecture du projet.....	50
5.3.1. les dérogations accordées à la statistique publique.....	50
5.3.2. les craintes concernant la vie privée.....	50
5.3.3. les craintes relatives au risque de stigmatisation d'un groupe social.....	51
5.3.4. Risque de trop grande rigidité du système.....	51
5.3.5. La gestion des identifiants et de l'identifiant pérenne.....	51

5.3.6 Le groupe de concertation s'est aussi interrogé sur le délai de conservation des données et sur l'exercice du droit à l'effacement.....	52
5.3.7 Concernant les données contenues dans les répertoires et les informations utilisées pour les alimenter.....	52
5.3.8 La sécurité des données.....	54
5.3.9 Concernant l'accès aux données par les chercheurs	55
5.3.10 Impact de Résil sur le recensement de la population.....	55
5.4 Les procédures par lesquelles Résil doit être approuvé et suivi.....	56
5.4.1 Le dispositif juridique.....	56
5.4.2 Un examen d'opportunité par le Cnis pour les appariements réalisés via Résil ?.....	56
5.5 Concertation et communication autour du projet Résil.....	57
6. Les préconisations du groupe de concertation.....	58
RÉFÉRENCES BIBLIOGRAPHIQUES.....	61
Gouvernance de la statistique publique.....	61
Aspects éthiques.....	61
Appariements de données et utilisation de données administratives.....	62
Expériences étrangères sur les registres de population.....	62
Communication et concertation sur les accès aux données, les appariements et le projet Résil.....	63
ANNEXES.....	64
Annexe 1 : Mandat du groupe de concertation.....	64
Annexe 1 : Mandat du groupe de concertation.....	65
Annexe 2 : Liste des membres du groupe de concertation.....	67
Annexe 3 : Liste des personnes auditionnées.....	68
Annexe 4 : Liste des sigles et termes utilisés.....	69
Annexe 5 : Situation dans les pays étrangers en matière de registres administratifs de population.....	70
Annexe 6 : La méthode des signes de présence pour statuer sur la présence d'une personne sur le territoire national.....	71
Annexe 7 : le cadre de nécessité et de proportionnalité de Statistique Canada.....	72
 LISTE DES ENCADRÉS	
Encadré 1 : le projet Safari et ses suites.....	13
Encadré 2 : Point de terminologie concernant les appariements.....	16
Encadré 3 : Registre de population/ Répertoire administratif/ Répertoire statistique.....	20
Encadré 4 : Un registre de population administratif semble à ce jour exclu en France en raison de la très grande méfiance attachée au NIR.....	23
Encadré 5 : Les dérogations accordées pour des finalités de statistique.....	27
Encadré 6 : Article 7 bis de la loi du 7 juin 1951.....	29
Encadré 7 : Article 3 bis de la loi du 7 juin 1951.....	30
Encadré 8 : Quelques précisions techniques sur la réalisation des appariements.....	31
Encadré 9 : Le Code Statistique Non Signifiant (CSNS).....	32
Encadré 10 : Le référentiel (répertoire) géographique de l'Insee en cours de construction.....	45
Encadré 11 : Risques d'atteinte au respect de chacun à sa vie privée.....	46

Avant-propos

Le projet Résil soumis au groupe de concertation, dont on va lire le rapport dans les pages qui suivent, consiste à utiliser deux répertoires (exhaustifs par conséquent) de la population résidente en France, d'une part, et du parc immobilier, d'autre part, pour apparier (ou interconnecter) des fichiers distincts de nature diverse, administratifs, d'enquête (par exemple fichier des revenus et fichier des étudiants de l'enseignement supérieur).

Ce projet s'inscrit dans un cadre désormais sévèrement réglementé au niveau national (loi de 1951 sur la statistique, loi de 1978 relative à l'informatique et aux libertés) comme à l'échelle européenne (le Règlement général de protection des données – appelé RGPD – entré en vigueur en 2018) et dans un contexte de multiples documents internationaux relatifs à l'éthique et à la rigueur du travail des statisticiens. Cette réglementation très précise sépare de l'ensemble des traitements de données ceux élaborés par les institutions de la statistique publique (Insee et services statistiques ministériels relevant de la compétence de l'Autorité de la statistique publique) et réalisés à des fins exclusivement statistiques pour un but d'intérêt général. Elle en allège les contraintes. Toutefois, ces traitements doivent présenter des « garanties appropriées pour les droits et libertés des personnes concernées » (RGPD, article 89). Ils doivent surtout être **nécessaires** à l'intérêt général et leurs moyens **proportionnés** à ce qui est recherché.

Résil sera, si le projet aboutit, mis en œuvre afin d'établir des nombres, par conséquent demeurera sans conséquence ni administrative, ni d'aucune sorte, pour chaque personne dont les données sont collectées. Son intérêt pour la connaissance de notre société ne fait pas de doute. Il élargit les possibilités d'investigation dans le temps (suivi de cohortes) comme dans l'espace (approche d'aires géographiques de faible dimension). Il s'intègre dans des travaux dont l'Insee a démontré depuis longtemps qu'il respectait scrupuleusement de strictes conditions de confidentialité et de méthode. En l'espèce, de surcroît, des précautions particulières sont prises (pas d'utilisation du numéro national d'identification des personnes physiques, par exemple).

Pourtant, dès lors que sont acceptés les principes des répertoires, d'une part, des appariements (qui ne sont pas chose nouvelle), d'autre part, une grande vigilance s'impose doublement : dans la manière dont Résil est utilisé et dans la manière dont il est perçu, donc accepté ; ce dernier élément est souvent insuffisamment apprécié.

Ainsi, la législation en vigueur appelle à une particulière attention les traitements comportant des données dites « sensibles ». Celles qui révèlent « l'origine prétendument raciale ou ethnique », « les opinions politiques, philosophiques ou religieuses », « l'appartenance syndicale », « la santé ou l'orientation sexuelle », celles qui sont « génétiques ou biométriques ». Il n'y a rien de tel ni dans les répertoires, ni dans les données tirées des appariements pour constituer Résil. Cependant, dans les fichiers appariés, peuvent figurer des données sensibles. L'Insee doit apporter toute assurance qu'aucune d'entre elles, même de manière accidentelle, ne sera en usage dans Résil et doit prévoir à cette fin des garanties particulières de mise en œuvre.

La réglementation ne fait pas de mention particulière des traitements exhaustifs comme le sont les deux répertoires de Résil. Pourtant l'exhaustivité est exigeante. Elle ne supporte pas l'à-peu-près. Ou du moins, doit-elle minimiser la marge d'erreur le plus possible. La recherche de cet objectif peut conduire à un recours à des données de plus en plus massives pour faire en sorte que l'objectif de complétude soit atteint. Il faut naturellement borner cette exigence. La perfection statistique trouve sa limite dans le respect de la protection des personnes. Autrement dit, Résil doit respecter de manière concrète et continue le délicat équilibre entre les fins et les moyens. La nécessité des investigations statistiques est aisée à établir d'emblée ; la proportionnalité des moyens employés est un combat de longue haleine.

Plus les travaux sur des données personnelles sont étendus, plus s'impose le respect visible des normes en vigueur. Résil doit, par conséquent, offrir l'occasion à l'Insee et à l'ensemble de la statistique publique de renforcer son indépendance et sa transparence.

Si le projet a rencontré dans le groupe de concertation un accord de principe, il a été maintes fois soulevé la question de savoir ce que risquait de devenir un tel instrument entre des mains non plus statistiques mais administratives voire policières. Il est des moments de péril dans une nation où tous les moyens pour conjurer les dangers sont recherchés, y compris ceux attentatoires aux libertés des personnes. Résil ne peut fonctionner sans bénéficier de la garantie qu'il demeurera toujours entre les mains des seuls statisticiens publics, pour des seules finalités de décompte, à l'exclusion de toute autre fin. En d'autres termes, Résil doit se rattacher comme une ombre à la réalité de l'indépendance de la statistique publique. Or, cette indépendance est fragile. L'Insee et les services statistiques ministériels sont des directions composées d'agents publics soumis à l'obéissance hiérarchique aux ministres. La manière dont ces derniers peuvent ou non interférer dans la production statistique peut être subtile... ou brutale. Résil et les études qui en seront issues doivent rester loin de toute interférence de cette nature. L'Autorité de la statistique publique, créée en 2008, est chargée de faire respecter les principes d'un travail réalisé selon les règles de l'art dans la production statistique. Mais elle est aussi responsable du « respect du principe d'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques. » L'Insee doit apprécier si cette garantie d'indépendance est suffisante pour faire obstacle à toute intrusion fâcheuse dans Résil (et ailleurs). Peut-être les dimensions du projet impliquent-elles des garanties supplémentaires par un tiers de l'indépendance nécessaire.

Une question identique se pose pour la transparence. Depuis plus de soixante-dix ans, la production statistique en France est soumise à un regard extérieur, celui du Conseil national de l'information statistique (Cnis), qui examine tous les projets d'enquêtes et formule des propositions d'études et d'enquêtes, ce qui constitue sans nul doute une grande originalité dans le système administratif français. Mais cette institution rassemble des experts et des utilisateurs de statistiques, a priori favorables à l'obtention de résultats (la nécessité), autrement dit moins sensibles aux questions de proportionnalité des moyens (principe de minimisation des données par exemple).

L'institution par l'Insee du présent groupe de concertation, après une réunion très ouverte de présentation du projet tenue le 28 janvier 2022, est un pas significatif dans la transparence requise. Mais celle-ci doit s'ouvrir davantage au grand public, dans la permanence, pour Résil et aussi les autres travaux. Ceux-ci sont exonérés par les textes en vigueur du droit d'accès et du droit de rectification des personnes dont les données sont conservées. La contrepartie de cette exonération réside dans la clarté qui doit être faite sur l'utilisation des données. L'exigence de transparence requiert donc une réflexion, qui devrait conduire ou bien à ouvrir le Cnis à de nouvelles préoccupations, en l'occurrence la sauvegarde des droits et libertés (par conséquent à une nouvelle composition), ou bien à de nouvelles formes de participation citoyenne, dont le groupe a eu, par exemple, l'écho inventif au Canada ou au Royaume-Uni.

Si Résil ouvre à bon droit de nouvelles et stimulantes perspectives de connaissance, il doit ouvrir aussi de nouvelles garanties des droits et libertés.

Ces réflexions n'auraient pas abouti sans l'investissement de chacun des membres du groupe de concertation, qui doit en être chaleureusement remercié ; sans le travail considérable et talentueux de la rapporteure, Françoise Dupont ; sans l'aide attentive et permanente du secrétariat général du Cnis (en l'occurrence Monsieur François Guillaumat-Tailliet) ; et sans la science discrète mais aiguë et sûre de Monsieur Olivier Lefebvre. A tous, je dois ma reconnaissante gratitude.

Jean-Marie Delarue

Synthèse

Un projet important, structurant, qui suscite des interrogations

Les appariements de données sont pratiqués de longue date depuis les années 60 par le service statistique public en France. Depuis quelques années, ils connaissent un développement important, portés par une disponibilité croissante de données administratives, des moyens technologiques performants et une demande accrue. Ils permettent en effet d'affiner les connaissances, dans l'espace (meilleure approche des territoires) et dans le temps (compréhension des parcours individuels, évaluation de politiques publiques) en rendant possibles des approches croisant plusieurs thématiques (par exemple emploi et formation). Ce développement n'est pas seulement français, mais international (dans les pays de développement comparable) et il est encouragé par les organismes internationaux et la communauté des praticiens de la statistique.

Les pays du nord de l'Europe disposent de longue date de registres de population mis en place pour des usages administratifs, avec des identifiants uniques partagés au sein des administrations et facilitant les appariements ; c'est sur cette infrastructure que s'appuient les instituts nationaux de statistique pour réaliser leurs statistiques démographiques et sociales. Pour des raisons détaillées dans le rapport, un tel système n'est ni envisageable, ni souhaitable, ni envisagé en France. Les pays qui ne disposent pas de registre de population, comme la Nouvelle-Zélande et le Canada, se sont engagés dans des projets de répertoires statistiques, destinés exclusivement à des usages statistiques, afin d'améliorer la qualité de leurs productions dans le domaine démographique et social. C'est dans cette voie que s'engage l'Insee avec le Projet de **Répertoire Statistique d'Individus et de Logements** (Résil) dont la mise en service est prévue en 2025.

Résil sera constitué de deux **répertoires statistiques** exhaustifs portant l'un sur les individus résidant sur le territoire national et l'autre sur les logements, et des liens qui relient les individus et les logements. Il répond à plusieurs finalités exclusivement statistiques :

- Il permettra d'offrir au service statistique public uniquement un service d'appariement de données¹ qui permet de gagner en qualité, en harmonisation et en sécurité par rapport à ce qui est pratiqué actuellement.
- Il permettra de mesurer la qualité au regard d'une utilisation statistique des sources administratives qui constituent une des ressources principales de la statistique publique.
- Il permettra, dans le prolongement des dispositifs existants, de construire la base de sondage dans laquelle tirer des échantillons pour les enquêtes réalisées par le service statistique public auprès des ménages, avec une couverture encore mieux assurée.
- Résil fournira l'information permettant de préparer et réaliser les enquêtes annuelles de recensement, et d'en extrapoler les résultats, à l'instar de l'utilisation actuelle des fichiers issus de la taxe d'habitation.
- Enfin, Résil permettra de produire des indicateurs démographiques plus robustes et précoces.

Concrètement, il s'agit d'une liste exhaustive d'entités (des individus, des logements) dont le contenu est réduit aux seules informations permettant de faire la jonction entre deux fichiers, c'est-à-dire des clés d'identification et des liens entre individus et logements.

Ses caractéristiques essentielles en sont d'une part l'absence du Numéro d'Immatriculation au Répertoire national d'identification des personnes physiques (le NIR, également connu sous le terme « numéro de sécurité sociale »), d'autre part l'absence de données telles que le revenu, la catégorie sociale, l'état matrimonial.

¹ Ces appariements peuvent porter sur des données administratives entre elles, ou sur des données administratives et des enquêtes statistiques.

Mis à jour régulièrement grâce à plusieurs sources de données administratives de périodicités annuelle ou infra-annuelle dont une première liste est présentée et commentée dans le rapport, Résil sera un référentiel statistique sur les individus et les logements, structurant pour les statistiques démographiques et sociales. Une « photographie » des individus et des logements au premier janvier sera réalisée chaque année.

Compte tenu du caractère exhaustif des deux répertoires et du croisement de plusieurs sources pour le réaliser, une étude d'impact est nécessaire au sens du Règlement Général sur la Protection des Données (RGPD). L'Insee souhaite cependant disposer d'un cadre juridique solide et pour cela encadrer le projet Résil par un décret en Conseil d'État dès 2023.

Conscient que le cadre juridique fort ne suffit pas, et qu'il est important de disposer d'un mandat social explicite, c'est-à-dire une reconnaissance explicite par la société de l'intérêt du projet et de la légitimité de l'Insee à le faire vivre, l'Insee a souhaité associer des représentants de la société très en amont du projet afin que des points de vue divers sur le projet Résil s'expriment et que l'Insee puisse en tenir compte dans la construction du répertoire statistique et la délimitation de ses usages.

Une concertation indispensable, dont le groupe a assuré une première étape, avec un apport conséquent

Le groupe de concertation sur le projet de **Répertoire Statistique d'Individus et de Logements** (Résil), placé sous l'égide du Cnis, s'inscrit dans le prolongement de la rencontre du Cnis du 28 janvier 2022 sur les appariements. Celle-ci avait permis de présenter pour la première fois le projet Résil et ses finalités ainsi que le contexte statistique national et international en matière d'appariements de données personnelles afin de recueillir les points de vue des représentants de la société dans sa diversité (administration publique, chercheurs, associations, syndicats, collectivités locales) en amont du projet. Des intervenants de la table ronde sur la transparence et l'information du public, qui découvraient le projet Résil et le cadre juridique et déontologique des travaux de la statistique publique, avaient souligné que cette réflexion sur Résil et les appariements devait à leurs yeux être approfondie sur une plus longue période avec un panel d'expertise adapté aux enjeux.

L'Insee a donc proposé de constituer un groupe de concertation pour dresser la liste des interrogations que suscite le projet Résil, échanger sur les réponses apportées par l'Insee à ces interrogations, proposer des modalités concrètes pour poursuivre les réflexions tout au long de la construction du projet jusqu'en 2025 et, le cas échéant, au-delà. Le Bureau du Cnis du 12 mars 2022 en a approuvé la création et le mandat.

Le Cnis a demandé à Jean-Marie DELARUE, Conseiller d'État honoraire, de présider ce groupe de concertation compte tenu de son expertise et son intérêt bien connu pour le droit des libertés publiques et le respect de la vie privée. Outre son président, le groupe était constitué du maître d'ouvrage de Résil et de 14 personnes (voir annexe 2) ayant une expertise susceptible d'exprimer des interrogations de la société sur le projet Résil (protection des libertés fondamentales, protection des données sur le plan juridique et informatique, travaux statistiques, travaux de la recherche économiques et sociales, transformation numérique, communication sur les données, éthique).

Conséquence du parti pris de s'adresser avant tout à des personnes situées en dehors des interlocuteurs habituels de la statistique publique et donc des philosophes, juristes, membres d'associations ou de *think tanks*..., un grand effort de présentation synthétique des travaux d'appariement de la statistique publique, de son cadre juridique et déontologique national et européen a été nécessaire, ainsi que des efforts de pédagogie pour présenter le projet Résil et ses usages dans le temps imparti. De leur côté, les participants ont fourni un travail d'appropriation considérable sur la période mai-octobre 2022, pour comprendre le contenu des travaux réalisés par la statistique publique, les conditions dans lesquelles ils sont menés ainsi que la plus-value apportée par Résil.

Le groupe a rempli son mandat, qui était de comprendre le projet, de mettre en commun et de confronter les points de vue, pour finalement lister un certain nombre de questions et de craintes sur le projet, de mobiliser et d'analyser de premiers éléments de réponse, et de

proposer des garde-fous pour la suite des travaux. Son apport aux réflexions du Cnis et de l'Insee quant à la construction du projet et son accompagnement est très net, mais il est important de souligner qu'il n'a pas pu épuiser le sujet. D'abord parce que le temps était limité (7 réunions en 5 mois dont 2 mois d'été) et le sujet vaste, technique et peu débattu jusqu'à présent ; ensuite parce que si l'on a cherché à mobiliser des experts en dehors de la statistique publique, ils ne se considéraient pas comme investis de la mission de représenter toute la société dans ce groupe, parce que leur expertise rend leur approche un peu différente de celle du grand public, même s'ils contribuent à l'appréhender ; enfin parce qu'il s'agit d'un projet en construction à horizon 2025, dont l'évolution et les usages ne peuvent pas être figés une fois pour toutes.

La poursuite de la démarche est donc cruciale, elle occupe une grande part des recommandations.

Les interrogations du groupe et ses recommandations (parties 5 et 6 du rapport)

Le groupe a fait part en préalable du risque qu'il perçoit que l'accroissement des appariements n'encourage une tendance du politique à la « gouvernance par le nombre » en d'autres termes que la mise à disposition de données plus nombreuses n'encourage les gouvernants à ne décider qu'en fonction de données chiffrées. Ce risque existe, mais le groupe a considéré qu'il tenait davantage à l'usage des données qu'à leur production.

Le groupe a exprimé des craintes sur les risques de mésusages des données de Résil ou des appariements de données qui en résultent en cas de fuite de données consécutive à une cyber-attaque, ou lors d'un détournement de finalité à la suite d'une pression politique, dans le cadre d'un régime autoritaire ou non.

Le groupe a centré ses préoccupations sur trois ordres d'idées :

- Le principe d'un répertoire de cette nature ;
- Les atteintes à la vie privée qui pouvaient en résulter ;
- La sécurité "interne" et "externe" du dispositif.

Les réponses qu'il a données sont relatives à quatre domaines.

A- Le principe du répertoire

Le groupe a admis que le critère de la **nécessité** était satisfait dans la mesure où les besoins de connaissance avec un besoin de précision accrue sont réels et qu'il n'y a pas d'alternative convaincante pour permettre des améliorations dans ce domaine.

Il a estimé que le fondement juridique des appariements pouvait être mieux assuré. Les appariements ne sont en effet pas explicitement définis par la loi ou les règlements. Pour les statisticiens, il s'agit d'un cas particulier d'interconnexion. Le groupe a donc recommandé que le fondement juridique des appariements soit clarifié.

B- L'architecture du dispositif

Le groupe s'est interrogé pour savoir si les modalités de Résil répondent au principe de **proportionnalité au regard au regard des usages et des risques possibles sur les libertés**.

Il a noté que les services proposés par Résil ne seront accessibles que pour des agents du Service statistique public, pour des traitements à finalité statistique exclusivement, et que la maîtrise de l'outil demeurerait du seul ressort de l'Insee. Il a demandé que ces éléments soient bien précisés dans les textes.

Le groupe a relevé positivement l'absence d'usage du NIR et un usage approprié du Code Statistique Non Signifiant (CSNS). La gestion des identifiants apparaît satisfaisante et restent strictement internes. Quelques-uns seront pérennes : le groupe a demandé qu'ils soient modifiés, comme le CSNS, en cas de problème de sécurité intervenu sur le dispositif.

Il a relevé aussi l'engagement de l'Insee que ne figurerait dans le répertoire aucune donnée "sensible"² et a souhaité que toute modification ultérieure sur ce point fasse l'objet d'un avis de la Cnil et du Cnis.

Le groupe a souhaité que la vigilance sur le risque de stigmatisation de groupes sociaux (par caractéristiques de ménage...), déjà présente dans les travaux statistiques actuels, soit renforcée du fait de l'accroissement du nombre des appariements, donc des données disponibles diverses rassemblées pour un même individu, même si elles sont pseudonymisées.

Il est proposé de conserver les données dix ans en base active, c'est-à-dire dans le cadre de la gestion en production du répertoire Résil. Il s'agit du délai au terme duquel on supprime du répertoire les informations pour les personnes décédées ou sorties du territoire, les logements détruits ou dont l'usage n'est plus l'habitation. Un résident stable en France reste, lui, toute son existence dans le répertoire, avec un historique des mouvements de ses logements sur dix ans. Cette possibilité a été admise, pour les seuls besoins de la statistique publique.

Le répertoire proprement dit sera alimenté par différents fichiers administratifs existants, destinés notamment à compenser la disparition du fichier fiscal de la taxe d'habitation. Certains sont présentés comme d'utilisation certaine, d'autres comme étant ou devant être soumis à une étude de faisabilité technique, de pertinence statistique (gains en qualité du répertoire) et à une évaluation de la proportionnalité du traitement.

Le groupe³ a rendu un avis défavorable sur l'utilisation du fichier des titres de séjour des étrangers (AGDREF), en raison de la sensibilité de la question de la possession des titres de séjour par les étrangers. Il a également émis un avis défavorable en raison d'un risque d'image et d'une non-proportionnalité pour le fichier de la carte VITALE et pour le fichier RNCPS (répertoire national commun de la protection sociale, lui-même conglomérat de fichiers de diverses caisses de protection sociale).

De manière générale et au-delà des sources qui lui ont été présentées, le groupe met l'Insee en garde quant à l'utilisation de trois types de données : données de consommation, a fortiori celles issues de fichiers de sécurité (type Faed, Fichier automatisé des empreintes digitales), données de santé.

Sur le plan de la sécurité informatique, elle doit être garantie par l'Insee qui porte la responsabilité du dispositif et des données qu'il traite. Les précautions que l'institut prendra à cet égard, qui s'inscrivent dans une politique exigeante de sécurité informatique de l'Insee et plus largement de l'Etat, ont paru appropriées au groupe même si en matière de sécurité il ne peut exister de garantie absolue. Le groupe recommande de faire auditer la sécurité régulièrement par un intervenant externe.

C - Procédures d'adoption et de suivi du dispositif

L'encadrement par un décret en Conseil d'État pris après avis de la Cnil paraît souhaitable au groupe.

Le groupe recommande que les sources de données constitutives des répertoires d'individus et de logements figurent dans un arrêté ou un décret simple pris après avis du Cnis et de la Cnil, afin de s'assurer qu'un examen de la nécessité et la proportionnalité est réalisé pour chaque source utilisée. Ces textes devront être mis à jour, après avis de la Cnil et du Cnis, à chaque modification de la liste des sources.

2 Les **données sensibles** forment une catégorie particulière des données personnelles. Ce sont des informations qui révèlent la prétendue origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

Le règlement européen interdit de recueillir ou d'utiliser ces données, sauf, notamment, dans les cas suivants :

- si la personne concernée a donné son consentement exprès (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée) ;
- si les informations sont manifestement rendues publiques par la personne concernée ;
- si elles sont nécessaires à la sauvegarde de la vie humaine ;
- si leur utilisation est justifiée par l'intérêt public et autorisée par la Cnil ;
- si elles concernent les membres ou adhérents d'une association ou d'une organisation politique, religieuse, philosophique, politique ou syndicale.

3 Le détail est donné en partie 5.3.7

Le groupe recommande que le bureau du Cnis et les commissions compétentes soient régulièrement informés des avancées du projet Résil.

Le groupe recommande de poursuivre la concertation sur le projet Résil tout au long de la construction du projet et de l'élargir aux appariements. En particulier, il propose de tenir une rencontre du Cnis sur les appariements et la mise en œuvre de Résil au premier trimestre 2025 à l'image de la rencontre du Cnis sur les appariements du 28 janvier 2022. Concernant Résil et les appariements de données, une consultation directe d'un panel de citoyens, comme cela se pratique outre-Manche sur certains projets liés au numérique, pourrait être utile et pourrait même être étendue à d'autres projets de la statistique publique.

Le groupe recommande que le Cnis donne un avis d'opportunité sur les appariements réalisés par la statistique publique en prenant en compte les dimensions éthiques et celles liées aux droits et aux libertés des personnes. Le groupe recommande pour cela d'élargir les missions du Cnis et de modifier sa composition, en y incorporant quatre ou cinq membres compétents en matière de libertés et de protection des données.

Le groupe recommande également de s'appuyer sur l'Autorité de la statistique publique (ASP) pour prévenir les risques de mésusage qui seraient en contradiction avec les règles fixées : l'ASP doit pouvoir intervenir à titre préventif en cas de pression sur les services, dénoncer tout mésusage ; il est important qu'elle soit régulièrement informée de l'avancement du projet, puis de ses usages.

D- présentation du projet et de ses usages au public

La transparence doit rester la règle dans la mise en œuvre de Résil :

- Le groupe recommande d'ouvrir une rubrique grand public très complète pour présenter le projet Résil, ses finalités, les apports attendus, les modalités de protection des données, les textes juridiques et la délibération de la Cnil.
- Le groupe recommande de publier la liste des appariements réalisés grâce à Résil avec le responsable de traitement, les sources utilisées, les finalités des traitements, les populations concernées de manière à rendre compte du respect des principes de nécessité et de proportionnalité.

Plus largement le groupe suggère de développer une communication à destination du grand public sur l'utilisation des sources administratives et les appariements de données en s'inspirant des bonnes pratiques des instituts nationaux statistiques étrangers, en particulier celui du Canada.

Le groupe considère que son rapport, même s'il résulte d'efforts importants de pédagogie et d'explicitation, ne peut constituer un document « grand public » sur Résil, l'utilisation des sources administratives et les appariements. Il est important pour cela d'approfondir la démarche et de préparer un dispositif approprié.

Le groupe recommande de recourir à un focus groupe dans la préparation d'une communication grand public.

Rapport

1. La démarche de concertation

1.1. Contexte

L'Insee s'engage dans un projet de construction d'un système de répertoires statistiques d'individus et de logements (Résil) avec une mise en service prévue pour 2025 pour des usages statistiques exclusivement. Il s'agit de bâtir un référentiel statistique sur les individus et les logements, sur la base de plusieurs sources de données présentées, puis discutées, dans ce rapport. Ce projet structurant pour les statistiques démographiques et sociales répond aux finalités suivantes :

- Résil offrira un service d'appariement de données permettant de gagner en efficacité, en qualité et en harmonisation des méthodes par rapport à ce qui est pratiqué actuellement. Il conduira à l'élargissement des possibilités d'appariements
- Résil permettra, en constituant un point de référence pour la comparaison, de mesurer la qualité des sources administratives qui constituent une des ressources principales de la statistique publique.
- Résil permettra, dans le prolongement des dispositifs existants, de construire la base de sondage dans laquelle tirer des échantillons pour les enquêtes réalisées par le Service Statistique Public auprès des ménages, avec une couverture encore mieux assurée.
- Résil fournira l'information permettant de préparer et réaliser les enquêtes annuelles de recensement, et d'en extrapoler les résultats, à l'instar de l'utilisation actuelle des fichiers issus de la taxe d'habitation.
- Résil permettra de produire des indicateurs démographiques plus robustes et précoces

Ce projet s'inscrit dans un contexte national et international marqué à la fois par :

- une forte attente au regard de l'accroissement des données et exploitations statistiques, à des fins d'analyse, de préparation et d'évaluation de politiques publiques, d'alimentation du débat public, une demande de plus de réactivité dans la fourniture des statistiques⁴, attentes dont le Conseil national de l'information statistique (Cnis) se fait régulièrement l'écho ;
- et une attention croissante portée par les citoyens et leurs représentants aux enjeux de protection des données individuelles, face à l'accroissement à la fois du volume de ces données et de leurs capacités de traitement ; cela se traduit également par le renforcement des règles relatives à la transparence sur le traitement des données, permettant aux citoyens d'avoir des garanties sur leur bon usage.

Il vise à concilier ces deux dimensions.

Au lancement du projet Résil (pour Répertoires Statistiques d'Individus et de Logements), l'Insee a considéré que les enjeux juridiques, déontologiques et de concertation étaient à la hauteur des enjeux techniques et nécessitaient donc une attention particulière.

En effet, avec la transformation numérique de la société, la masse de données potentiellement mobilisables s'accroît continûment. La puissance des outils d'analyse augmente également en parallèle sous l'effet des progrès techniques de l'intelligence artificielle et des capacités de stockage et de calcul. De nouvelles opportunités sont offertes et une forte demande de régulation des nouveaux usages se fait jour dans la société pour profiter des bénéfices en limitant les mésusages.

⁴ Il s'agit d'une demande récurrente d'Eurostat. La pandémie a également souligné l'intérêt de fournir des remontées rapides d'information.

Depuis une dizaine d'années, l'encadrement des nouveaux outils et des nouveaux usages fait l'objet de très nombreuses discussions sous l'angle de la protection des données et celui de la protection des libertés fondamentales aux niveaux national, européen et international dans un champ très vaste, beaucoup plus large que celui qui s'applique aux travaux du service de la statistique publique.

Ainsi, par exemple, l'entrée en vigueur en 2018 du Règlement général sur la protection des données (RGPD) a constitué un pas important vers de nouveaux droits pour les personnes, une responsabilisation des pilotes de traitements de données assortie de nouvelles obligations, la mise en place d'instances de coordination européenne : le comité européen de la protection des données est institué par le règlement général sur la protection des données.

Les réflexions sur l'intelligence artificielle font également depuis quelques années l'objet de réflexions nombreuses de recommandations ou de chartes éthiques de différents organismes ([Cnil 2017](#), [Sénat 2019](#), [OCDE 2019](#), [Commission Européenne 2019](#), [Unesco 2021...](#)).

Même si elle a pris récemment une nouvelle dimension, l'attention de la société sur les conditions de l'utilisation des données individuelles ne date pas d'hier. En France, on peut considérer que la première manifestation de cette attention est la polémique surgie en 1974 autour du projet Safari, projet d'automatisation de la gestion du répertoire national d'identification des personnes physiques, permettant un partage de cet identifiant au sein des administrations et services publics. Ce projet, porté par l'Insee, s'inscrivait dans un contexte général d'informatisation et de modernisation des administrations. La modernisation des administrations nécessitait la création d'identifiants individuels. Le partage du même identifiant, celui créé par l'Insee, apparaissait comme la solution la plus simple techniquement et la plus efficiente. La contestation a surgi en raison d'une extension possible de l'utilisation de cet identifiant à des finalités non statistiques par les ministères chargés de la police et de la justice (voir encadré), dans un contexte de défiance à l'égard du rôle de l'Etat.

Non seulement le projet a été abandonné mais les discussions qui ont suivi ont permis également de définir des garanties très significatives en matière de protection des données avec la mise en place de la loi Informatique et libertés en 1978 et la création de la Cnil.

L'évocation récente ([13] et [14]) des fichiers mécanographiques du Service National de la Statistique mis en place durant l'Occupation a également fait ressurgir ces questionnements et montré que la préoccupation de la société face à ces grands répertoires et aux risques de détournement de finalité qui les accompagnent restait forte.

Encadré 1 : le projet Safari et ses suites

L'Insee avait créé en 1947 un répertoire **administratif** d'identification des individus (actuellement le RNIPP) dont il assurait la gestion. Dans ce répertoire à vocation administrative, chaque individu portait un numéro qui l'identifiait de manière unique : le NIR, créé par le Service National de Statistique durant l'Occupation. En 1970, un projet de gestion informatisée du RNIPP, dont la gestion était auparavant manuelle, est conçu au niveau de l'Insee. Ce projet avait pour objectif secondaire d'étendre progressivement l'utilisation du NIR à l'ensemble des administrations par souci de modernisation, de simplification et d'efficacité ».

Un article du journal Le Monde en 1974 « Safari ou la chasse aux Français » lance une controverse en s'interrogeant sur la concomitance de la création de deux unités informatiques au ministère de l'intérieur et au ministère de la justice et du projet Safari où, selon l'auteur, le ministère de l'Intérieur cherche à jouer un rôle important. Précédemment, plusieurs personnes avaient fait une analogie avec les travaux menés par le Service National de la Statistique durant l'Occupation.

La suite est décrite par S.Jugnot [15] « *le Premier ministre interdit aux départements ministériels de procéder sans son autorisation à de nouvelles interconnexions de fichiers. Le gouvernement crée une première « Commission informatique et libertés ». Composée de personnalités qualifiées, elle est placée auprès du ministère de la justice pour « proposer au Gouvernement, dans un délai de six mois, des mesures tendant à garantir que le développement de l'informatique dans les secteurs public, semi-public et privé se réalisera dans le respect de la vie privée, des libertés individuelles et des libertés publiques ».* En juin 1975, le rapport de la commission, dit « rapport Tricot », est finalisé. En juillet 1976, un projet de loi est présenté et, le 6 janvier 1978, la loi n° 78-17, dite « Informatique et Libertés », est promulguée donnant ainsi aux traitements de données personnelles un cadre protecteur pour les personnes concernées par ces données ». La Cnil est ainsi créée en 1978 et porte depuis sa création une vigilance particulière aux appariements de données.

Une première rencontre sur les appariements de données sociodémographiques a eu lieu le 28 janvier 2022 sous l'égide du Conseil National de l'Information Statistique, instance de concertation entre les producteurs de la statistique publique, les utilisateurs des statistiques et les représentants des différentes composantes de la société française.

Près de 360 personnes comprenant des représentants de l'administration publique, des chercheurs, des associations, des syndicats et des collectivités locales, ont participé à cette manifestation qui proposait un large tour d'horizon des pratiques d'appariements et des besoins des utilisateurs pour alimenter la préparation et l'évaluation des politiques publiques ainsi qu'une table ronde sur le besoin de transparence pour le grand public. Elle associait au niveau des intervenants des utilisateurs venus du monde de la recherche, des études et du monde politique, des personnes qualifiées en termes de protection des libertés fondamentales et d'éthique, des collègues des instituts nationaux étrangers. Une première présentation du projet Résil a été faite à cette occasion.

L'effort de pédagogie et la vue d'ensemble de la diversité des pratiques ont été appréciés [32]. La dernière table ronde a soulevé des interrogations sur les conditions dans lesquelles les appariements sont réalisés ainsi qu'une demande de poursuivre les discussions, en croisant des points de vue variés sur le sujet pour mieux l'expliquer, approfondir la réflexion sur les bénéfices et les risques du projet Résil et accompagner sa mise en œuvre.

En réponse à cette attente, et parce que l'Insee et le Cnis étaient conscients que la rencontre du 28 janvier n'avait pas épuisé le sujet, le Cnis a lancé un groupe de concertation qui poursuit la démarche entreprise avec la rencontre du 28 janvier dernier.

1.2. Démarche

Le mandat du groupe (annexe 1) proposé a été validé par le bureau du Cnis le 17 mars 2022. Un mois a été consacré à la recherche de personnalités qualifiées, permettant d'éclairer les différents aspects des interrogations autour du projet en allant au-delà du regard que peuvent porter les statisticiens et les utilisateurs habituels de données statistiques. (Quels risques perçus ? Quelles opportunités ? Quelles questions éthiques ? Quel encadrement juridique ? Quelles garanties données par la statistique publique ? Quelle transparence pour le grand public et plus largement quelle communication autour du projet ?).

Une cinquantaine de personnalités, universitaires ou membres d'associations ou d'institutions ont été approchées sur la base de recherches bibliographiques et de contacts du Conseil économique social et environnemental, du Cnis et de la Ligue des droits de l'homme. La Quadrature du net et le Mouton Numérique ont décliné la proposition de participation pour des raisons de moyens. Le Défenseur des droits a décliné l'offre pour des raisons de positionnement incompatible avec sa participation. La Cnil n'a pas été sollicitée, ne pouvant être juge et partie. La composition finale du groupe a bien permis grâce à ses quinze membres (annexe 2) de croiser les points de vue de personnes apportant un point de vue citoyen et de personnes qualifiées apportant les éclairages souhaités.

Le groupe s'est fixé comme objectifs de :

- dresser la liste des interrogations que suscite ce projet ;
- présenter à la discussion les réponses apportées par l'Insee ;
- proposer des modalités concrètes pour poursuivre les réflexions tout au long de la construction du projet et, le cas échéant, au-delà.

La majorité des membres du groupe ne connaissaient pas, ou très peu, la nature et les modalités des travaux de production menés par les instituts nationaux de statistique français et étrangers pour établir les statistiques diffusées. De même, ils ne connaissaient pas le cadre juridique dans lequel évolue le service statistique public français. Les auditions et les exposés

ont permis de présenter le cadre général français et international et de resituer le projet Résil dans la logique de l'évolution des pratiques françaises et internationales sur longue période afin de bien expliciter les enjeux du projet actuel. Un travail important d'appropriation du cadre général et du projet a été effectué par les membres du groupe dans un calendrier serré ; les représentants de la statistique publique ont essayé de faciliter au maximum cette appropriation sur des sujets assez techniques par des réponses en séances aussi simples et concrètes que possible et quelques compléments bibliographiques.

Sept réunions se sont tenues en présentiel (éventuellement en distanciel pour faciliter la participation de certains), les 4 mai, 25 mai, 28 juin, 5 juillet 14 septembre, 21 septembre, 28 septembre avec des séquences d'auditions et des séquences de questions/réponses. Sept auditions se sont tenues en séance plénière et trois en séance plus restreinte (annexe 3).

Ces auditions ont permis de brosser le cadre juridique général français et européen, en particulier sur les accès aux données, sur les modalités du secret statistique et les exceptions qu'il comporte (notamment l'accès des chercheurs aux données).

L'exposé d'Eurostat a permis de faire le point sur les pratiques des autres pays et d'évoquer les projets de répertoires statistiques en cours dans de nombreux pays. L'exposé d'un chercheur de l'Institut des Politiques Publiques, Antoine Bozio, a permis d'éclairer l'usage des appariements pour la préparation et l'évaluation des politiques publiques et leur intérêt pour la recherche..

L'audition de Statistique Canada a illustré les questionnements déontologiques d'un institut statistique sur l'utilisation des données administratives et sur les appariements de données, une pratique ancienne qui, comme en France, devient prégnante pour la production statistique et qui n'est pas connue du grand public.

L'audition de la secrétaire générale du Conseil National de l'Information Statistique a permis d'exposer les pratiques de concertation du service statistique public qui portent surtout sur les besoins en matière d'information statistique à travers la discussion sur les programmes de travail annuels et les programmes à moyen terme des services producteurs, la charge de réponse des personnes et des entreprises aux enquêtes, l'accès aux sources administratives.

L'audition de Claude Poulain a permis d'analyser les raisons de la contestation de Safari et de replacer le projet dans le contexte de l'époque (au tournant des années 70, le partage des identifiants était fortement encouragé comme une condition de modernisation de l'administration).

L'audition du représentant de l'ANSSI a permis aux participants de mieux comprendre les enjeux de sécurité du système d'information, le rôle que l'ANSSI joue en ce domaine, notamment pour accompagner les administrations dans leurs projets, l'importance de caractériser les risques contre lesquels on souhaite se prémunir.

L'audition de Serge Slama a permis de passer en revue le cadre juridique de Résil et de soulever des questionnements sur les mésusages possibles des informations contenues dans Résil.

L'audition de Daniel Agacinski a permis d'évoquer les différentes modalités d'échanges sur un projet, de souligner la difficulté de l'exercice de concertation, les écueils à éviter et les attitudes à privilégier.

L'audition de l'Institut Ada Lovelace, *think tank* britannique sur l'éthique du numérique, a permis de présenter une pratique de concertation sur des sujets liés à l'éthique du numérique, reposant en partie sur des panels de citoyens.

2. Etat des travaux statistiques

2.1. Recours aux données administratives et appariements, bref historique et pratiques actuelles

2.1.1. La statistique publique

L'Insee, comme l'ensemble du service statistique public, utilise depuis longtemps des données administratives pour ses missions afin de réduire la charge d'enquêtes auprès des ménages et des entreprises tout en orientant le questionnement vers des dimensions non prises en compte dans des fichiers administratifs (sur le ressenti, les pratiques, les difficultés...). Cette pratique permet également d'alléger les coûts des opérations statistiques, de produire des informations plus fines, notamment au niveau géographique, et de renforcer la performance du service statistique public. Les données les plus couramment utilisées sont issues des Déclarations Sociales Nominatives (DSN, qui sont remplies par les employeurs des secteurs privé et public), des déclarations fiscales des entreprises et ménages (pour l'impôt sur le revenu, les impôts fonciers, l'impôt sur les sociétés, la TVA, etc.), et des caisses d'allocations familiales.

La pratique d'enrichissement d'une enquête par une source administrative ou d'enrichissement d'une source administrative par une autre est également assez ancienne dans la statistique publique. La première enquête sur les revenus fiscaux, qui rapprochait des données du recensement et des données fiscales, date de 1956. La constitution de l'échantillon démographique permanent (EDP), panel⁵ qui rapproche des données du recensement et des données d'état civil date de 1968, les panels sur l'ensemble des actifs (avec des trajectoires sur dix ans) datent de 1976.

Encadré 2 : Point de terminologie concernant les appariements

Un **enrichissement** de données d'un fichier par un autre fichier consiste à trouver, pour un individu donné, les informations qui le concernent dans deux fichiers différents, puis à constituer un troisième fichier avec les données ainsi rassemblées. Les statisticiens désignent cette technique souvent sous le nom d'« **appariement** » de fichiers (ou de données) qui figure dans la loi de 1951 et le décret d'application n° 2016-1930 de la loi pour une République numérique, là où les rédacteurs de la loi relative à l'Informatique, aux fichiers et aux libertés de 1978 ont employé le terme « **interconnexion** » de fichiers⁶ qu'on retrouve également dans la loi de 2016 pour une République numérique. D'autres termes peuvent être utilisés par les statisticiens pour la même opération : **combinaison ou enrichissement ou couplage** de fichiers (le dernier terme étant celui utilisé au Canada à la fois par Statistique Canada et dans la loi qui régit ces opérations).

La question de l'équivalence juridique des termes appariement et interconnexion a été soulevée ; le fait que le terme « appariement » n'apparaisse quasiment pas dans les textes de niveau législatif peut imposer une clarification à cet égard : soit pour définir précisément ce qu'est un appariement par rapport à une interconnexion, soit pour indiquer que ces deux termes désignent le même type de traitement. Quelle que soit l'option retenue, elle doit être explicitée, soit dans le texte fondant Résil, soit dans son exposé des motifs.

Viennent ensuite, à partir de 2012, le Fichier Localisé Social et Fiscal (FiLoSoFi), issu du rapprochement des données fiscales exhaustives en provenance de la Direction générale des finances publiques (déclaration de revenus des personnes physiques, taxe d'habitation et fichier d'imposition des personnes physiques) et des données sur les prestations sociales émanant des organismes gestionnaires de ces prestations (CNAF, CNAV, MSA). Enfin dès les années 1990, les enquêtes auprès des ménages récupèrent les données de revenus - pour lesquelles le questionnement serait trop lourd - grâce à des appariements avec les données fiscales et sociales.

Certaines bases de données, qui jouent un rôle de pivot dans le système d'information en raison de leur panélisation⁷, sont ensuite continuellement enrichies au fil du temps pour mieux

5 Juxtaposition d'information portant sur les mêmes individus relatifs à des années différentes

6 <https://www.cnil.fr/fr/comment-determiner-la-notion-dinterconnexion>

7 Suivi dans le temps sur plusieurs années des mêmes individus

répondre à la demande de statistiques. Ainsi l'échantillon démographique permanent⁸ (EDP) qui rassemblait au départ des informations issues des recensements successifs s'est étendu aux départements d'outre-mer et a vu son échantillon augmenter en 2008 pour tenir compte de l'arrivée du recensement par sondage, et a été complété par l'ajout de nouvelles sources : les inscriptions électorales pour réaliser des études sur le vote, des données sur les salaires et les périodes d'emploi pour effectuer des analyses de parcours de vie professionnelle, enfin des données socio-fiscales sur les revenus. On peut ainsi étudier la mortalité ou la fécondité selon le niveau de diplôme, la catégorie sociale, le niveau de vie. Plus récemment, il a été apparié avec des données de santé pour donner une nouvelle source, EDP-santé⁹, pour le service statistique compétent en matière de santé (la Drees) en 2019, ce qui permet d'étudier les inégalités sociales de santé [20].

Le panel de salariés créé en 1976 sur la base de la source DADS puis, plus récemment, de la source DSN, qui contenait à sa création les salaires et périodes d'emploi déclarés par l'employeur pour un échantillon de salariés du secteur privé, a été progressivement étendu à tous les salariés puis à tous les actifs (avec les non-salariés) pour donner une vision complète des situations d'emploi des personnes sur une année : salariés du secteur public ou privé, de l'agriculture, de particuliers employeurs, non-salariés, et pour mesurer notamment la multi-activité ; la dimension « panel » permettant d'étudier les trajectoires des individus, par exemple les transitions entre salariat et non-salariat¹⁰.

Depuis quelques années, on assiste en France comme à l'étranger à une intensification du recours aux sources administratives et à la combinaison de sources entre elles (enrichissement d'enquêtes à partir de données administratives, données administratives entre elles) qui permet de répondre à de nouvelles questions. On peut par exemple citer le développement lors de l'élaboration des politiques publiques des questionnements prenant en compte la dimension territoriale - dont la prise en compte dans les enquêtes, ne serait-ce qu'au niveau du département, imposerait des échantillons très importants - ainsi qu'à une demande croissante d'analyse approfondie des effets croisés à laquelle il n'est pas possible de répondre avec une seule enquête. Cette analyse des effets croisés permet de disposer d'éléments pour mieux orienter les politiques publiques vers des groupes sociaux particuliers.

Enfin, la préparation des politiques publiques et l'évaluation de leurs effets accroissent le besoin d'un suivi dans le temps des mêmes individus à partir de données annuelles ou celui de croiser plusieurs fichiers.

Ainsi, des appariements sont réalisés pour permettre de décrire des trajectoires individuelles, notamment les trajectoires d'insertion sur le marché du travail, les trajectoires professionnelles d'emploi et de salaire, ou encore celles de certaines populations telles que les bénéficiaires de minima sociaux. Ils peuvent permettre de décrire des trajectoires relatives à différents domaines, par exemple l'évolution du niveau de vie ou la mobilité résidentielle lors du passage à la retraite, en mobilisant l'EDP.

L'appariement est aussi employé dans le cadre d'évaluation de réformes ou de pratiques de l'éducation nationale, avec la mobilisation des panels d'élèves qui existent depuis 1973. Basés sur l'utilisation de données administratives issues du système scolaire, ils ont permis d'abord d'observer les entrants en 6e, puis les élèves entrant en CP à partir de 1978 et se sont enrichis depuis les années 90 des évaluations des acquis et enquêtes auprès des élèves ou des familles.

Les services statistiques publics compétents en matière d'emploi (la Dares) et d'éducation (la DEPP) recourent également à cette pratique pour évaluer les effets des mesures d'aide à l'emploi sur l'insertion des jeunes passés par l'apprentissage ou la voie professionnelle à l'aide du dispositif **InserJeunes**.

Sur le plan méthodologique, les appariements permettent également de mieux comprendre certains phénomènes, notamment en aidant à analyser les écarts entre des sources concernant les mêmes concepts ou des concepts voisins. Ainsi, l'appariement de l'enquête Emploi et des

⁸ <https://www.insee.fr/fr/information/5349625> et <https://www.insee.fr/fr/information/5398685>

⁹ <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/ledp-sante-enrichissement-de-lechantillon-demographique-permanent-par>

¹⁰ <https://www.insee.fr/fr/statistiques/4470790?sommaire=4470890#onglet-3>

déclarations des employeurs – déclarations annuelles des données sociales (DADS) puis déclarations sociales nominatives (DSN) – vise à mieux comprendre le concept d'emploi et sa mesure. Le rapprochement de l'enquête Emploi, qui mesure le chômage au sens du Bureau international du travail (BIT), et du fichier historique des demandeurs d'emploi de Pôle emploi permet de mieux comprendre le concept du chômage et les écarts qui existent entre ses différentes définitions.

Les opérations statistiques recourant aux sources administratives sont facilitées par la mobilisation plus rapide des sources, liée au développement de la culture de la donnée (en particulier à la prise de conscience plus forte de la valeur contenue dans les données) ainsi qu'au progrès constant des techniques sur le plan statistique comme sur le plan des performances des outils informatiques. L'objectif de limiter la charge des enquêtés et de maîtriser les coûts de collecte reste, bien sûr, un argument crucial en faveur de cette tendance de fond.

Ainsi, encouragés explicitement par le Conseil national de l'information statistique dans son programme de moyen terme 2019-2023 afin d'enrichir l'analyse des liens entre différents domaines dans le strict respect de la protection des données à caractère personnel, les appariements entre sources de données (enquêtes et/ou sources administratives) sont de plus en plus fréquemment mis en œuvre au sein du service statistique public. La constitution d'une source de données en appariant plusieurs sources administratives devient une méthode de collecte de données à part entière au même titre que la collecte par enquête qui conserve tout son intérêt.

Cette tendance de fond peut cependant se heurter à des difficultés de réalisation lorsque les données d'identification nécessaires pour l'appariement sont incomplètes ou comportent des erreurs (voir encadré 8 de la partie 3.5 sur les appariements de données).

2.1.2 Les appariements pour la recherche

Les chercheurs sont également demandeurs de données plus riches permettant d'intégrer la dimension locale dans leurs recherches ainsi que le croisement des thématiques et le suivi dans le temps pour contribuer utilement au débat public.

En effet, comme l'a indiqué Antoine Bozio lors de son audition, les questions de recherche ne sont pas « monothème » et donc pas circonscrites à une seule source de données ; elles nécessitent d'adopter une vision transverse, tout comme pour la production demandée à la statistique publique. Par exemple, les problèmes environnementaux auxquels sont exposés les ménages et leur capacité à y faire face peuvent dépendre de leur situation sociale ou de leur lieu de résidence. Il y a donc un enjeu scientifique (progresser dans la connaissance des phénomènes économiques et sociaux) dans les appariements. Il y a aussi un enjeu démocratique, car l'évaluation des politiques publiques, qui touchent souvent plusieurs domaines, ne peut être correctement menée que si on dispose de données sur les personnes, décrivant leur situation dans ces différents domaines.

Plusieurs exemples illustrent le déficit de connaissance imputable à des besoins d'appariement qui ne sont pas satisfaits. Ainsi, bien que la France dispose de très bonnes données sur l'éducation d'une part, et sur le marché du travail d'autre part, il est impossible, à tout le moins très difficile, pour un chercheur de suivre les élèves à leur entrée sur le marché du travail et dans leurs trajectoires d'activité ultérieures, parce que ces bases ne contiennent pas le même identifiant individuel. En conséquence, il n'y a pas de travaux de recherche sur l'impact des différentes politiques éducatives sur l'insertion à long terme dans l'emploi.

Pour répondre à la question d'actualité : quelle politique de rénovation des logements dans le contexte de rationnement ou du renchérissement de l'énergie ? il est nécessaire de rapprocher des informations sur la « qualité » du logement, sur les revenus du ménage et sur sa consommation d'énergie ; ces informations sont pour le moment réparties dans plusieurs bases, gérées par des entités différentes.

Les chercheurs ont un accès possible aux données qui résultent d'appariements réalisés par le service statistique public en obtenant un avis favorable du [comité du secret statistique](#)¹¹. Ils peuvent par ailleurs réaliser un appariement de leurs données avec celles de la statistique publique avec des formalités simplifiées depuis la loi de 2016 pour une République numérique, en ayant recours à deux tiers de confiance et en utilisant un code non signifiant spécifique à chaque projet de recherche [18].

Antoine Bozio indique que dans les pays européens, dont la France, qui n'ont pas de registre de population avec un identifiant partagé, il faut multiplier les appariements entre bases de données pour mener des analyses croisées, ce qui joue sur les délais d'obtention et la qualité du fichier obtenu.

2.2 La situation à l'étranger : pratiques, projets et recommandations internationales en matière de répertoires de personnes et d'utilisation de sources administratives

2.2.1. Pratiques et projets

Les instituts nationaux de statistique font régulièrement évoluer leurs méthodes de collecte et de traitement des données, en cherchant à mobiliser et intégrer davantage de données externes, d'origine administrative, voire privée, pour les statistiques sur les personnes comme sur les entreprises.

L'utilisation des données administratives par les instituts nationaux de statistique pour leur production statistique régulière est ancienne et date de l'informatisation de ces fichiers dans les années 60. Elle a connu une évolution régulière permettant d'alléger ou de supprimer des enquêtes, de créer de nouvelles statistiques. Avec les améliorations de l'efficacité des traitements, et l'évolution de la demande de statistiques (plus fréquentes, plus détaillées sur différentes dimensions dont la dimension territoriale et l'analyse des trajectoires des personnes), cette utilisation de données administratives est devenue la source principale de données pour les instituts nationaux de statistique. Dans ce contexte, les appariements de données pratiqués de longue date¹² par les instituts de statistique se sont régulièrement développés.

Devant cette réalité, la réflexion s'est logiquement portée sur la façon de rationaliser l'organisation de la production statistique en mettant les données administratives au centre du dispositif du point de vue du système d'information, en facilitant les articulations et en mutualisant les traitements le plus possible au sein de chaque institut.

Sur ce sujet, des échanges sur les plans méthodologique et stratégique se sont intensifiés dans les dernières années au niveau international (cf. récemment le projet Admin 2020¹³ d'Eurostat et les échanges au sein de l'UNECE¹⁴ dans le cadre de la préparation des recensements de la population).

Chaque institut national mène ces évolutions de fond selon les moyens dont il dispose, les sources qui lui sont accessibles et leur niveau de qualité, mais aussi selon son contexte

11 Le rôle, le fonctionnement du comité ainsi que sa composition sont fixés par la [loi n° 51-711 du 7 juin 1951](#) sur l'obligation, la coordination et le secret en matière de statistiques (articles 6bis, 7bis et 7 ter). Il émet un avis sur les demandes de communication de données couvertes par le secret statistique (données collectées par voie d'enquête statistique ou transmises au service statistique public à des fins d'établissement des statistiques) ou par le secret fiscal.

Le comité est présidé par un conseiller d'État. Il comprend, outre son Président et les représentants de l'Assemblée nationale et du Sénat, un membre du Conseil économique, social et environnemental, 4 représentants de l'administration, 2 représentants de la Cnil, 2 représentants des chercheurs, 2 représentants des organisations syndicales de salariés, 4 représentants des entreprises, 1 représentant de l'Institut national d'études démographiques, 1 représentant d'élus locaux.

12 Le premier article de référence du statisticien de Statistique Canada, Yvan Fellegi, date de 1969.

13 https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-admin-administrative-data-sources_en

14 Commission économique pour l'Europe des Nations unies

administratif ou culturel. Les transitions sont souvent longues et se prêtent à des consolidations progressives. Chaque pays a ses spécificités, mais plusieurs modèles se dégagent.

Parmi ces différents modèles, Résil se rapprochera plutôt des exemples néo-zélandais et canadien (pas de registre de population administratif, pas d'identifiant inter-administratif partagé, mais la construction de répertoires statistiques).

Un système statistique bien installé, très intégré, qui utilise de longue date des registres de population administratifs mis en place dans le cadre de la gestion administrative et qui utilise un identifiant partagé

Michel Poulain et Anne Herm [23] retracent l'histoire des registres administratifs de population¹⁵ et montrent que l'idée en est ancienne et remonte à 1665 pour l'Europe (à l'origine il s'agissait de registres paroissiaux) ; elle s'est développée au XIX^e siècle, et son apport pour la statistique a été relevé en 1906 lors du congrès de l'Institut International de la Statistique. Puis dans la deuxième moitié du XX^e siècle, on voit apparaître des registres de population centralisés : en Islande (1953), en Suède (1966), Norvège (1968), Danemark (1968) et Finlande (1970), en Belgique (1968, légalisé en 1985), plus récemment aux Pays-Bas (1994), en Espagne (1996) et Autriche (2000), enfin dans les Etats baltes.

Les pays dotés de registres de population administratifs reposant sur une obligation de déclaration de changement de domicile et ceux qui peuvent s'appuyer sur un identifiant personnel partagé par toutes les administrations ont pu construire un système statistique très intégré reposant essentiellement sur la mobilisation de données administratives. **L'utilisation courante du terme « registre de population » (register en anglais) induit de la confusion dans la mesure où il ne précise pas sa finalité (administrative ou statistique).**

En Finlande, par exemple, la production de données statistiques s'appuie à 95 % sur des données tirées de registres ou de données administratives. Le recensement de la population finlandais est, depuis 1980, entièrement fondé sur ce type de données. Le Registre Central de Population a été créé à la fin des années 60 (mais la pratique de registres est bien plus ancienne) ; il est désormais désigné sous le terme de Système d'Information sur la population et rassemble de nombreux fichiers administratifs issus d'origines très diverses (cf. schéma).

Il s'inscrit dans un Etat et une société où l'interconnexion de fichiers individuels de toute nature ne soulève pas de problèmes techniques ou organisationnels majeurs, ni de problème d'acceptation par la population. Il n'est ni envisageable ni envisagé que la France s'oriente vers un tel dispositif.

Encadré 3 : Registre de population/ Répertoire administratif/ Répertoire statistique

Un répertoire est constitué d'une liste exhaustive d'objets assortie des informations permettant d'identifier ces objets sans ambiguïté, et de les relier à d'autres données du système d'information, en particulier un identifiant.

Un répertoire administratif est destiné à un usage de **gestion administrative** qui a des répercussions pour les entités qui le composent. Il contient pour la gestion des pièces justificatives attestant l'identité des objets qui le composent. Le répertoire Sirene (système informatique pour le répertoire des entreprises et des établissements), le Répertoire électoral unique (REU) le Répertoire national d'identification des personnes physiques (RNIPP) sont des répertoires administratifs gérés par l'Insee.

¹⁵ « Pour un registre de population, les unités individuelles sont les personnes physiques et la mise à jour s'applique à toutes les informations à caractère administratif les concernant. Généralement, les personnes physiques considérées sont celles résidant administrativement sur un territoire donné, une commune, une région ou un pays. La mise à jour des données se fait en tenant compte des naissances, décès, mariages et divorces, mais également de tous les changements de résidence ou migrations. Le caractère continu de la mise à jour des données, y compris celle de l'adresse de résidence habituelle, constitue la spécificité du registre de population et représente son intérêt majeur pour la production de statistiques démographiques ».

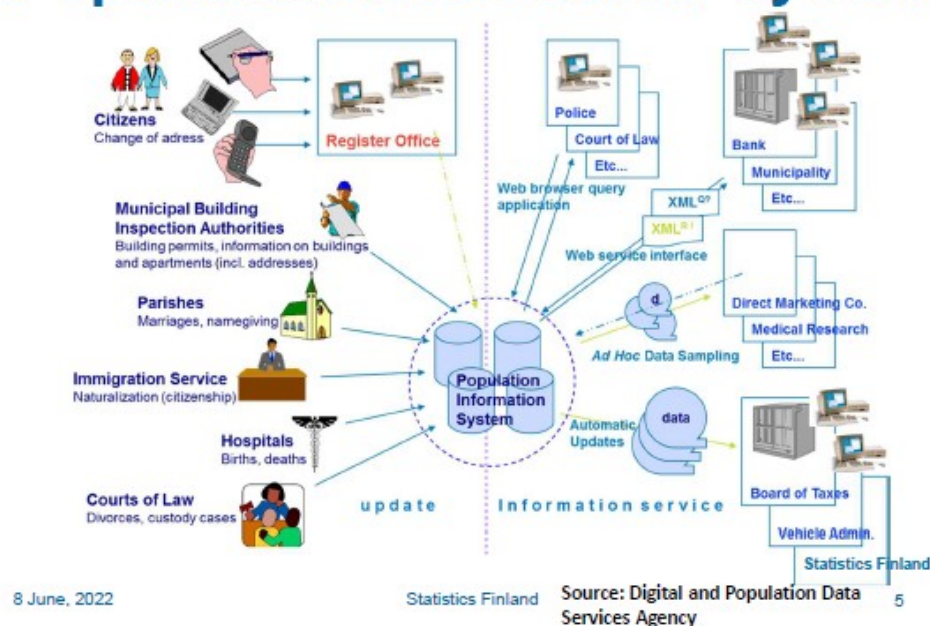
Un **répertoire statistique** est **exclusivement** destiné à un **usage de production de statistique**. Il n'entraîne aucune répercussion administrative directe pour les unités qui le composent.

Un **registre de population** a une **vocation administrative**. Il repose sur une obligation de déclaration de changement de domicile des citoyens et le plus souvent sur un identifiant des personnes partagé par toutes les administrations.

L'utilisation courante du terme « **registre de population** » (*register* en anglais) induit de la confusion dans la mesure où il ne précise pas sa finalité administrative ou statistique. En général, elle est administrative.

Le Système d'Information sur la Population en Finlande

Population Information System



Aux Pays-Bas : un système très intégré avec registres, sources administratives et données d'enquêtes

L'institut de statistique hollandais (CBS) s'inscrit dans une démarche générale d'utilisation des sources de données disponibles, le *System of social statistical datasets* (SSD). C'est un système de registres et d'enquêtes interconnectés et normalisés. Tous les processus de production au sein de CBS qui ont trait aux statistiques sociales ou spatiales convergent vers ce SSD, qui contient une mine d'informations sur les personnes, les ménages, les emplois et les prestations, pensions, éducation, hospitalisations, rapports de criminalité, logements, véhicules par exemple. Aux Pays-Bas, c'est la plus importante source de statistiques sociales officielles et, parce que les données sont disponibles sur demande au moyen d'un accès à distance, également très populaires en sciences sociales. Les statistiques diffusées sont bien sûr des résultats agrégés qui préservent la confidentialité des données utilisées pour les produire. Comme en France pour le CASD, les données individuelles très détaillées auxquelles accèdent les chercheurs ne sortent pas de l'environnement sécurisé géré par CBS, qui vérifie systématiquement que les données ou résultats qui seront exportés pour publication ne présentent pas de risque de rupture de confidentialité¹⁶.

¹⁶ <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>

Un des débouchés de ces registres est le recensement. Ainsi, aux Pays-Bas, le dernier recensement « traditionnel » remonte à 1971, celui de 1981 et les suivants, qualifiés de « *virtual censuses* » sont produits à base de registres et de données administratives. La Norvège et la Suède font un recensement en partie basé sur des registres depuis les années 80.

En Nouvelle-Zélande et au Canada, où il n'y a pas d'identifiant unique permettant le croisement de données, se sont mises en place progressivement des infrastructures pour faciliter le croisement de données individuelles à des fins de recherche et de production statistique (en particulier pour conduire le recensement). Des projets de répertoires statistiques de personnes sont en cours de développement pour renforcer l'utilisation et la valorisation des données administratives.

La Nouvelle-Zélande a mis en place en 2011 un système d'identification statistique des individus et des logements afin de faciliter les appariements, en particulier pour un objectif de recherche dans un contexte où il n'existe pas d'identifiant individuel partagé et pas de répertoire administratif mobilisable [26]. Un premier répertoire statistique d'individus a été constitué à partir des données d'état civil (naissances et décès), des données aux frontières (immigration-émigration) et des données fiscales.

Ce dispositif a été très utile pour la finalisation du recensement de 2018, perturbé par des difficultés de collecte plus importantes et plus « concentrées » qu'à l'accoutumée. Il a permis d'obtenir des résultats statistiques en combinant des données (administratives ou enquêtes) sur l'éducation, le marché du travail, les prestations, la justice, la santé et la sécurité, les migrations et les données commerciales. Depuis 2021, une rénovation est en cours pour disposer d'un véritable répertoire statistique des individus et des logements plus intégré et plus facile à gérer. L'institut national de statistique de Nouvelle-Zélande est très soucieux de la communication sur ces techniques et sur la mobilisation de ces données, à la fois pour rassurer les utilisateurs sur leur qualité, mais aussi pour obtenir l'adhésion de toute la population face à ce nouveau type de collecte, et en particulier la population autochtone maori.

Statistique Canada travaille sur un projet de répertoire statistique de personnes depuis 2016 [27]. En parallèle, le répertoire d'adresses qui a émergé dans les années 90 fait l'objet d'un projet de registre de bâtiments utilisant plus de sources administratives et couvrant les logements mais aussi les bâtiments non résidentiels. Le répertoire statistique des personnes utilise de nombreuses sources administratives : les informations sur les naissances, les décès, des données sur les résidents permanents et temporaires issues de la gestion de l'immigration, des informations concernant la population indigène, les données fiscales, le numéro d'assurance sociale (SIN) dont l'utilisation au Canada n'est pas obligatoire et pas uniforme. L'objectif de Statistique Canada est de mettre en place une architecture très sécurisée, bien articulée pour faciliter l'intégration et l'exploitation de données administratives. Les données sont « désidentifiées » dès que possible en se basant sur un identifiant statistique. La sécurité repose également sur l'accès très contingenté au sein de Statistique Canada aux données identifiantes.

Statistique Canada a fortement développé depuis 2018 le questionnement éthique et la communication sur l'ensemble des opérations de collecte de données, notamment la mobilisation et l'appariement de données administratives ou privées, suite à une polémique d'importance sur un projet d'utilisation de données bancaires. Afin de conserver la confiance et l'adhésion de la population canadienne à ses travaux, Statistique Canada a mis en place un cadre d'examen systématique de la nécessité et de la proportionnalité pour fonder le mandat social de collecte de données et de traitement, indispensable complément du mandat juridique. Statistique Canada porte également une grande attention à la prise en compte des points de vue de la population autochtone du Canada.

Certains pays ont un système intermédiaire récent fondé sur la consolidation de registres communaux, dont la couverture peut être contrôlée via des sources administratives, voire des enquêtes de terrain.

C'est le cas, par exemple, en Italie, où les statistiques de population sont établies en rassemblant les données des registres administratifs municipaux de population au niveau de l'institut national de statistique italien. La couverture de cette compilation est contrôlée d'une part par voie d'enquêtes de terrain, d'autre part via une méthode dite de « signes de vie » fondée sur la fréquence d'apparition dans des fichiers administratifs. Une personne qui figure dans un registre mais dans aucun fichier administratif pourra ainsi être considérée comme non résidente.

En définitive, partant de situations bien différentes, les instituts nationaux de statistiques s'organisent de plus en plus avec des **répertoires statistiques de population** et des répertoires d'adresses, dans le respect de la confidentialité des données et d'autres principes éthiques, afin de mieux organiser leur production majoritairement basée sur des sources administratives.

Encadré 4 : Un registre de population administratif semble à ce jour exclu en France en raison de la très grande méfiance attachée au NIR

Nous avons déjà évoqué le projet Safari qui a donné lieu à une polémique et qui a été à l'origine de la création de la Cnil et de la loi Informatique et liberté en 1978. Dès sa création la Cnil a donc cherché à encadrer les interconnexions de fichiers.

Auparavant les travaux de répertoire sur 1940-1944 avaient également laissé une empreinte forte sur les esprits, notamment en raison des dérives potentielles d'un tel répertoire et de l'usage de l'identifiant qui lui était associé.

S. Jugnot [15] rappelle ainsi que « *Dès la Libération, le numéro d'inscription au répertoire (Nir) est retenu comme identifiant pour gérer les droits des bénéficiaires de la Sécurité sociale nouvellement créée mais son origine alimentera les suspicions sur les déviations possibles de son utilisation par l'État.* ».

La déclaration obligatoire de son dernier domicile par chaque citoyen renvoie également au régime de Vichy. En effet, en mai 1941, une loi rend obligatoire les déclarations de changement de domicile, applicable à partir de mars 1942 ; elle est abrogée à la Libération. Ces informations ont été consignées dans des dossiers personnels [15bis]. L'objectif était celui d'un système administratif relié par un identifiant unique.

Dans les deux cas, les projets n'ont été abordés que sur le seul plan technique.

Bien que le sujet des répertoires administratifs dépasse le périmètre de compétence de l'Insee, on peut relever les points suivants :

Actuellement, un registre administratif tel que pratiqué dans d'autres pays n'est pas possible dans l'état actuel du droit. Il n'y a pas d'identifiant partagé dans tous les fichiers administratifs depuis la loi Informatique et Libertés de 1978 : les usages du NIR sont très encadrés par le décret n° 2019-341 du 19 avril 2019 relatif à la mise en œuvre de traitements comportant l'usage du numéro d'inscription au répertoire national d'identification des personnes physiques ou nécessitant la consultation de ce répertoire qui en liste les usages possibles.

Il n'y a pas d'obligation de déclarer ses changements de domicile en France depuis la Libération. Il existe néanmoins un service, facultatif, de « guichet unique » pour la déclaration de changement d'adresse, géré par la Direction de l'information légale et administrative qui répercute l'information vers les organismes que l'utilisateur souhaite informer (il a le choix parmi une quinzaine d'administrations ou fournisseurs d'énergie).

La vigilance de la Cnil, des parlementaires et d'une certaine partie de l'opinion reste forte quant à l'interconnexion des informations.

L'Insee cherche donc plutôt à s'inspirer des exemples de la Nouvelle-Zélande et de Statistique Canada.

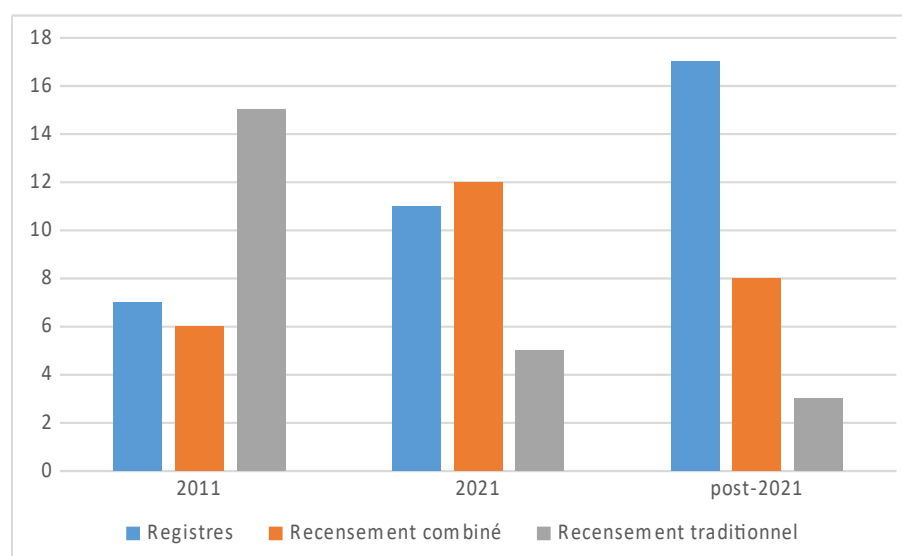
2.2.2. Recommandations, bonnes pratiques et prescriptions internationales

Les recommandations de l'ONU

Face aux contraintes qu'impose un recensement général (ressources humaines, financières, difficultés logistiques), la Division statistique de l'ONU recommande aux instituts statistiques de construire des répertoires statistiques d'individus et de logements qui pourraient faciliter la production de données de population. Un guide méthodologique « handbook » est en cours de finalisation¹⁷.

Alors que la collecte traditionnelle directe représentait en 2011 la majorité des pays de l'UNECE, les méthodes utilisant des registres¹⁸ ou des registres combinés avec une collecte représentent la majorité des recensements de l'année 2021¹⁹. Les méthodes par registres seuls devraient devenir la norme pour l'année 2031, comme le montre le graphique ci-dessous, avec une préparation qui commencera dès le cycle 2021 achevé.

Graphique : Les méthodes de recensement au niveau européen²⁰



La Commission économique pour l'Europe des Nations unies, l'UNECE (un peu plus large que l'Europe, elle regroupe également nombre de pays développés), a également produit un guide pour l'utilisation de données administratives dans le cadre des recensements et de la mise en œuvre de répertoires statistiques [29].

Le Code de bonnes pratiques de la statistique européenne

L'utilisation de données administratives pour la production de statistiques, qu'il s'agisse ou non de recensements, figure également dans le Code de bonnes pratiques de la statistique européenne adopté en 2005 [21]. Plusieurs principes l'encouragent explicitement, à la fois pour des raisons d'efficacité, de réduction de la charge de réponse aux enquêtes, de précision et de pertinence. D'autres portent sur les conditions à remplir pour que leur mobilisation soit le plus possible légitime, efficace, partenariale.

Principe 2 de mandat pour la collecte de données et l'accès aux données ; « *Les autorités statistiques sont autorisées par la législation à accéder rapidement et gratuitement aux données administratives et à les utiliser à des fins statistiques. Elles participent, dès le démarrage, à la* [17 *https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3e-Handbook-E.pdf*](https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3e-Handbook-E.pdf)

18 Selon cette approche, il n'y a pas de collecte directe de données auprès de la population, et le dénombrement traditionnel est remplacé par l'utilisation de données administratives détenues dans divers registres (registre de la population, registre d'immeubles/d'adresses, registre de sécurité sociale, etc.) qui sont appariées à partir de numéros d'identification uniques et partagés.

19 Pour faciliter les comparaisons l'ONU encourage les Etats à réaliser des recensements à des dates proches, au moins tous les 10 ans, les années terminant par 0 ou 1. On parle de « round 2021 » pour les recensements réalisés autour de l'année 2020.

20 Dans ce graphique, la France est considérée comme ayant un recensement traditionnel pour les trois dates.

conception, au développement et à l'abandon des fichiers administratifs, afin, le cas échéant, de rendre leurs données plus adaptées à des finalités statistiques »,

Principe 9 de charge non excessive pour les déclarants ; « Afin d'éviter la multiplication des demandes de données, les sources administratives ou autres sont mobilisées autant que possible »,

et **Principe 10 de rapport coût-efficacité** : « Les ressources sont utilisées de façon efficiente. Tout est mis en œuvre pour améliorer l'exploitation statistique des sources de données administratives ou autres et pour limiter le recours à des enquêtes directes ».

Avec l'adoption du Code de bonnes pratiques, le Système statistique européen s'est engagé à respecter les principes du Code et à travailler à sa mise en œuvre. Tous les instituts statistiques nationaux et l'office européen Eurostat sont soumis depuis 2006 à des évaluations régulières, tous les cinq ans environ, par des professionnels de la statistique publique européens (membres d'instituts étrangers, experts d'Eurostat et experts extérieurs) dites « revue par les pairs ». Les résultats sont publics et publiés sur le site d'Eurostat et sur ceux des Etats membres.

La récente « revue par les pairs » menée en 2021²¹ sur le système statistique français a noté les progrès réalisés par la France en ce domaine et a encouragé le système statistique public à se rapprocher encore davantage des producteurs de données pour en tirer le meilleur profit (et anticiper d'éventuels changements).

Le projet de règlement européen sur les statistiques de population

Enfin, le projet de règlement européen sur les statistiques de population (ESOP²²) visant à remplacer trois règlements actuellement en vigueur et dont l'adoption est prévue par Eurostat en 2023-2024, devrait imposer aux Etats membres des contraintes de temporalité et de finesse de l'information à fournir allant dans le sens d'une mobilisation accrue de données administratives, voire de la construction de répertoires statistiques.

En 2021 la situation en matière de répertoires statistiques de population connue d'Eurostat était la suivante :

- 12 pays ayant un répertoire statistique de population (Belgique, Pays-Bas, Danemark, Suède, Finlande, Italie, Espagne, Autriche, Bulgarie, Estonie, Lettonie, Slovaquie) ;
- 7 pays ayant prévu d'en construire un : France, Portugal, Croatie, Chypre, Hongrie, Lituanie, Malte ;
- 6 pays n'ayant pas de tels projets : République tchèque, Allemagne, Luxembourg, Pologne, Roumanie, Slovaquie.

La réponse de la France est basée sur l'existence du projet de répertoire statistique Résil.

Les négociations sont loin d'être achevées à la date de finalisation de ce rapport, mais on peut penser que les exigences d'Eurostat resteront fortes. Eurostat insiste sur le fait que les traitements nécessaires ainsi que les transmissions de données en découlant seraient réalisés en parfait accord avec le Règlement général pour la protection des données.

3. Etat du cadre juridique et déontologique

3.1 Le cadre juridique général de la statistique publique en France

Le cadre juridique de référence de la statistique publique est constitué essentiellement par deux textes :

[La loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques](#)

²¹ <https://www.insee.fr/fr/information/4137897>

²² <https://population-europe.eu/network/news-network/european-statistics-population-esop>

Cette loi, modifiée à plusieurs reprises pour tenir compte de l'évolution du contexte institutionnel et européen, des besoins et des opportunités, définit le **service statistique public** - composé de l'Institut national de la statistique et des études économiques et des [services statistiques ministériels](#) -- et les **statistiques publiques** qui regroupent l'ensemble des productions issues :

- des enquêtes statistiques dont la liste est arrêtée chaque année par un arrêté du ministre chargé de l'économie ;
- de l'exploitation, à des fins d'information générale, de données collectées par des administrations, des organismes publics ou des organismes privés chargés d'une mission de service public.

Elle définit le [secret statistique](#), qui interdit tout usage ou communication de données confidentielles, sauf pour des finalités statistiques ou de recherche scientifique ou historique.

Quel que soit leur mode de collecte, les informations individuelles traitées par l'Insee dans le but d'établir des statistiques publiques sont couvertes par le secret professionnel en vertu de la loi de 1951. Elles ne peuvent être utilisées qu'à des fins de statistique²³ par le service statistique public, à l'exclusion donc de finalités de contrôle ou de décisions individuelles. Elles n'ont aucune répercussion administrative directe pour un individu donné. Elles ne peuvent avoir d'effet qu'à travers une disposition légale de portée générale comme la détermination de seuils de minima sociaux faisant appel à des statistiques publiées par l'Insee.

La loi encadre les différentes modalités de collecte de l'information utilisée à des fins statistiques : enquêtes (dès l'origine), données recueillies dans le cadre de sa mission, par une administration, une personne morale de droit public, ou une personne morale de droit privé gérant un service public (depuis 1986), les informations présentes dans les bases de données détenues par des personnes morales de droit privé (depuis 2016), et la protection des données ainsi collectées. Cf. parties 3.3. et 3.4.

Elle crée différentes instances qui jouent un rôle essentiel dans le fonctionnement et la régulation des activités de la statistique publique parmi lesquelles :

- l'[Autorité de la statistique publique \(ASP\)](#), qui veille au respect du principe d'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques ainsi que des principes d'objectivité, d'impartialité, de pertinence et de qualité des données produites ;
- le [Conseil national de l'information statistique \(Cnis\)](#), organe de concertation entre producteurs et utilisateurs de la statistique publique, chargé de suivre les travaux statistiques ;
- le [Comité du secret statistique](#), qui organise l'accès aux données confidentielles pour des finalités statistiques ou de recherche scientifique ou historique.

Les agents de l'Insee sont tenus au respect du secret statistique. Comme tout agent public, ils sont également tenus au respect du secret professionnel : les fonctionnaires de l'Insee sont ainsi astreints au secret professionnel conformément aux dispositions de l'article 26 de la loi n° 83-634 du 13 juillet 1983 portant droits et obligations des fonctionnaires, sous les peines prévues par l'article 226-13 du code pénal (et pouvant aller jusqu'à un an d'emprisonnement et 15 000 euros d'amende). Les agents contractuels se sont engagés sur l'honneur à ne faire, pendant la durée du contrat, comme après son expiration, aucune communication écrite ou verbale concernant les documents ou les résultats de travaux qu'ils auront exécutés, dirigés ou contrôlés sans y avoir été autorisés préalablement par le chef de service lui-même, sous les peines prévues par l'article 226-13 du code pénal (rappelées ci-dessus).

[La loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés](#)

Elle vise à garantir à toute personne physique le contrôle de l'usage des données qui la concernent, et ce faisant, à concilier le développement de l'informatique des données personnelles avec le respect des droits fondamentaux. Elle s'inscrit, depuis 2018, dans un

²³ L'accès des chercheurs aux données produites par la statistique fait l'objet de dispositions particulières sous l'égide du comité du secret évoquées au 3.1.

cadre juridique européen harmonisé défini par le [Règlement général sur la protection des données \(RGPD\)](#). La loi impose à tout responsable de traitement des obligations qui s'articulent pour l'essentiel à deux grands principes : un principe de transparence – [informations des personnes concernées et réponse à leurs demandes, documentation des traitements](#) – et un principe de maîtrise et limitation des risques d'impact des traitements sur la vie privée des personnes concernées.

Compte tenu de la nécessité d'effectuer les traitements statistiques qui permettent de piloter la qualité et les analyses au niveau individuel, la majorité des informations sur les personnes traitées par la statistique publique ne peuvent pas être anonymisées, mais au mieux pseudonymisées²⁴. Elles relèvent donc dans la grande majorité des traitements de la catégorie des données personnelles. Les résultats statistiques qui sont diffusés une fois les traitements réalisés sont, eux, agrégés et ne posent pas de problème de confidentialité car ils ne permettent pas de retrouver des individus.

Le secret statistique joue un rôle essentiel dans l'application, aux traitements statistiques mis en œuvre par la statistique publique, des obligations qui découlent de la loi Informatique et libertés.

La loi fixe également le cadre dans lequel doivent être réalisés les appariements mobilisant le NIR (Numéro d'Inscription au Répertoire) et donne la possibilité pour ceux qui sont réalisés à des fins de statistique publique d'utiliser le Code Statistique Non Signifiant (cf. 3.5).

La loi confère à la Cnil un pouvoir de contrôle de la conformité des traitements, assorti de possibilités d'action (mise en demeure d'interrompre le traitement) et de sanctions en cas de manquement.

En cas de divulgation de données personnelles, la loi prévoit des sanctions pouvant aller jusqu'à 300 000 € d'amende et 5 ans d'emprisonnement, donc bien plus lourdes que celles encourues en cas de violation du secret professionnel.

Encadré 5 : Les dérogations accordées pour des finalités de statistique ²⁵

Un des considérants du RGPD dit que « *Le traitement des données à caractère personnel devrait être conçu pour servir l'humanité. Le droit à la protection des données à caractère personnel n'est pas un droit absolu ; il doit être considéré par rapport à sa fonction dans la société et être mis en balance avec d'autres droits fondamentaux, conformément au principe de proportionnalité.* »²⁶. Les **finalités statistiques** bénéficient ainsi de certaines dérogations au niveau français comme au niveau européen dans le cadre des missions **d'intérêt public**.

L'article 89 du RGPD prévoit des garanties et dérogations applicables au traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques. Il impose d'abord l'existence de garanties particulières, liées à la mise en place "de mesures techniques et organisationnelles, en particulier pour assurer le respect du principe de minimisation des données. Ces mesures peuvent comprendre la pseudonymisation²⁷, dans la mesure où ces finalités peuvent être atteintes de cette manière. Chaque fois que ces finalités peuvent être atteintes par un traitement ultérieur ne permettant pas ou plus l'identification des personnes concernées, il convient de procéder de cette manière." En revanche, il ouvre la possibilité aux Etats membres de prévoir des dérogations à certains droits pour les traitements à des fins de recherche scientifique ou historique ou à

24 <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles> La pseudonymisation permet de traiter les données d'individus sans pouvoir identifier ceux-ci de façon directe. En pratique, il est toutefois bien souvent possible de retrouver l'identité de ceux-ci grâce à des données tierces : les données concernées conservent donc un caractère personnel. La pseudonymisation constitue une des mesures recommandées par le RGPD pour limiter les risques liés au traitement de données personnelles.

25 Le considérant 162 du RGPD précise « Par « fins statistiques », on entend toute opération de collecte et de traitement de données à caractère personnel nécessaires pour des enquêtes statistiques ou la production de résultats statistiques ».

26 Les considérants du RGPD mentionnent en outre que : « Le présent règlement respecte tous les droits fondamentaux et observe les libertés et les principes reconnus par la Charte, consacrés par les traités, en particulier le respect de la vie privée et familiale, du domicile et des communications, la protection des données à caractère personnel, la liberté de pensée, de conscience et de religion, la liberté d'expression et d'information, la liberté d'entreprise, le droit à un recours effectif et à accéder à un tribunal impartial, et la diversité culturelle, religieuse et linguistique. »

27 <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>

des fins statistiques (articles 15 (accès), 16 (rectification), 18 (limitation de traitement) et 19 (Obligation de notification pour la rectification, l'effacement ou la limitation du traitement) 20 (droit à la portabilité des données) 21 (opposition).

L'article 78 de la loi du 6 janvier 1978 prévoit que la France exerce cette faculté de dérogation « *dans la mesure où ces droits rendent impossible ou entravent sérieusement la réalisation de ces finalités* » dans des conditions fixées par un décret en CE. Celui-ci²⁸ dispose à son article 116 que ces dérogations sont admises aux conditions suivantes :

- (i) ces données ne sont accessibles ou modifiables que par des personnes autorisées soumises à des règles de déontologie
- (ii) elles ne peuvent être diffusées sans être anonymisées "*sauf si l'intérêt des tiers à cette diffusion prévaut sur les intérêts ou les libertés et droits fondamentaux de la personne concernée*",
- (iii) elles doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités
- (iv) un encadrement de l'ouverture de ces données en tant qu'archives publiques en application de [l'article L. 213-3 du code du patrimoine](#) « *après autorisation de l'administration des archives, après accord de l'autorité dont émanent les documents et avis du comité du secret statistique* »

Le RGPD établit, pour les traitements à finalité statistique, une dérogation au principe de finalité : la réutilisation à des fins statistiques d'une donnée traitée pour une autre finalité est possible, la finalité statistique venant se « surajouter » à la finalité initiale. Cette dérogation au principe de finalité est également conférée aux traitements à finalité archivistique et aux traitements à finalité de recherche scientifique ou historique.

D'autres dérogations sont accordées par le RGPD aux traitements à finalité d'établissement de statistiques publiques :

- La première, partagée avec tous les autres traitements d'intérêt général, concerne le droit à l'effacement, connu également sous le nom de « droit à l'oubli ». Ce droit n'est pas applicable aux données résultant d'un traitement à finalité de statistique publique, parce qu'il est susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs statistiques. La contrepartie est que les données individuelles ne sont jamais rediffusées.
- La seconde concerne le droit d'opposition, qui n'est pas applicable aux données traitées à des fins statistiques lorsque ce traitement est nécessaire à l'exécution d'une mission d'intérêt public.

Enfin, le RGPD donne la possibilité aux droits nationaux de définir d'autres dérogations. C'est par exemple le cas pour la France avec la possibilité d'utiliser le Code Statistique Non Signifiant (CSNS).

3.2 Le cadre juridique général de la statistique publique européenne

Les dispositions générales applicables aux contributions des États membres à la production des statistiques européennes font l'objet d'un règlement dédié, le [règlement \(UE\) n° 223/2009 révisé](#) du Parlement européen et du Conseil du 11 mars 2009 relatif à la statistique européenne. Ce règlement définit le système statistique européen, dont font partie l'Insee et les services statistiques ministériels, et la gouvernance de ce système, en particulier le rôle de coordination d'Eurostat. Il fait référence plus largement, pour ses principes, au [Code des bonnes pratiques de la statistique européenne](#), qui est le cadre déontologique de référence pour l'ensemble des agents du système statistique européen et donc du service statistique public français.

3.3 L'accès de la statistique publique à des données administratives

Le dispositif juridique français encadrant la mobilisation de ces données à des fins statistiques et leur protection s'est construit progressivement. Le recours aux sources administratives est possible grâce à l'introduction en 1986 de l'article 7 bis de la loi du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques. La transmission est rendue obligatoire depuis 2004 lorsqu'elle émane d'une demande du ministre en charge de l'économie établie après avis du Cnis. La loi de 2016 pour une République numérique a introduit la gratuité pour les échanges de données administratives au sein de l'administration. Enfin, ces données

²⁸ [Décret n° 2019-536 du 29 mai 2019 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés](#)

sont placées sous la protection du Comité du secret statistique grâce à l'article 7 ter de la loi du 7 juin 1951

Encadré 6 : Article 7 bis de la loi du 7 juin 1951

« Sur demande du ministre chargé de l'économie, après avis du Conseil national de l'information statistique, et sauf disposition législative contraire, les informations relatives aux personnes physiques, à l'exclusion des données relatives à la vie sexuelle, et celles relatives aux personnes morales, recueillies dans le cadre de sa mission, **par une administration, une personne morale de droit public, ou une personne morale de droit privé gérant un service public** sont cédées, à des fins exclusives d'établissement de statistiques, à l'Institut national de la statistique et des études économiques ou aux services statistiques ministériels.

Les données à caractère personnel relatives à la santé recueillies dans les conditions prévues à l'alinéa précédent ne peuvent être communiquées, sur demande du ministre chargé de la santé, à l'Institut national de la statistique et des études économiques ou aux services statistiques des ministères participant à la définition, à la conduite et à l'évaluation de la politique de santé publique que dans le cadre d'établissement de statistiques sur l'état de santé de la population, les politiques de santé publique ou les dispositifs de prise en charge par les systèmes de santé et de protection sociale en lien avec la morbidité des populations. Des enquêtes complémentaires, revêtues du visa préalable mentionné à l'article 2, peuvent être réalisées auprès d'échantillons de ces populations.

Les modalités de communication des données à caractère personnel relatives à la santé, recueillies dans les conditions prévues à l'alinéa précédent, ne doivent pas permettre l'identification des personnes. Il ne peut être dérogé à cette dernière obligation que lorsque les conditions d'élaboration des statistiques prévues au deuxième alinéa nécessitent de disposer d'éléments d'identification directe ou indirecte des personnes, notamment aux fins d'établissement d'échantillons de personnes et d'appariement de données provenant de diverses sources, dans le respect des dispositions de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

Seules les personnes responsables de l'opération, désignées à cet effet par la personne morale autorisée à mettre en œuvre le traitement, peuvent recevoir les données à caractère personnel relatives à la santé, transmises à l'Institut national de la statistique et des études économiques ou aux services statistiques des ministères participant à la définition, à la conduite et à l'évaluation de la politique de santé publique. Après utilisation de ces données, les éléments d'identification des personnes doivent être détruits.

Sous réserve de l'article 777-3 du Code de procédure pénale, les dispositions des alinéas précédents s'appliquent nonobstant toutes dispositions contraires relatives au secret professionnel.

Les cessions portant sur des données à caractère personnel, telles qu'elles sont définies à l'article 2 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés sont soumises aux dispositions de ladite loi ; l'acte réglementaire et, lorsque les cessions se font entre deux personnes morales distinctes, les conventions entre le cédant et le cessionnaire de ces informations prévoient les modalités de la transmission, la finalité du traitement envisagé et le sort des informations après leur utilisation aux fins de traitement statistique.

Les cessions portant sur des informations concernant des personnes morales sont autorisées par décision conjointe du ministre chargé de l'économie et des ministres intéressés.

Sous réserve des dispositions des articles 40, 56, 76, 97 et 99 du code de procédure pénale, les informations transmises en application du présent article et permettant l'identification des personnes physiques ou morales auxquelles elles s'appliquent ne peuvent faire l'objet d'aucune communication de la part du service bénéficiaire.

Les agents de l'Institut national de la statistique et des études économiques et ceux des services statistiques ministériels sont astreints, pour les données dont ils ont à connaître en application du présent article, au secret professionnel sous les sanctions prévues aux articles 226-13 du code pénal²⁹. »

Reprenant les préconisations du Code de bonnes pratiques de la Statistique Européenne, le **Règlement européen 223/2009 relatif aux statistiques européennes** renforce la tendance de fond du recours aux données administratives en indiquant que les informations demandées doivent, autant que possible, pouvoir être aisément extraites de fichiers ou de sources disponibles. L'article 17 bis, entièrement consacré à l'utilisation des fichiers administratifs, précise que les Instituts nationaux de statistique (INS) y ont « *un accès gratuit et immédiat de même qu'un droit d'utilisation et d'intégration de ces fichiers aux statistiques* ». Enfin le règlement stipule que « *Les INS³⁰ et la Commission (Eurostat) sont consultés sur la conception initiale, le développement ultérieur et la cessation de l'utilisation des fichiers administratifs créés et mis à jour par d'autres organes, et y sont associés, facilitant de ce fait l'utilisation ultérieure de ces fichiers aux fins de la production de statistiques européennes* »

3.4 L'accès de l'Insee à des données produites par des personnes morales de droit privé

Bien que cette question ne concerne pas le projet Résil, lequel ne s'applique qu'à des données détenues par une administration, une personne morale de droit public, ou une personne morale de droit privé gérant un service public et des données d'enquêtes, le rappel des dispositions applicables aux données produites par des personnes morales de droit privé a paru nécessaire, pour compléter le panorama de l'accès aux différents types de données.

Avec l'apparition du Big Data, les instituts de statistiques nationaux ont mené collectivement depuis 2013, sous l'égide de l'ONU et d'Eurostat, une réflexion sur les plans stratégique et technique pour examiner quelle utilisation régulière de données détenues par des personnes morales de droit privé était souhaitable et faisable dans leur production. Des discussions sont en cours au niveau européen pour prendre en compte ces besoins tant dans le *Data Act*, l'un des règlements de la stratégie européenne des données, que dans le Règlement 223/2009 relatif aux statistiques européennes. De son côté, l'Insee a mené des études techniques, puis une concertation avec les principaux détenteurs de sources dans le cadre du Cnis sur ce sujet [30], avant qu'une inscription dans la loi ne voie le jour.

L'**article 3 bis** de la loi du 7 juin 1951 permet à la statistique publique d'avoir accès depuis 2016 à des données détenues par des personnes morales de droit privé sous des conditions très strictes :

Encadré 7 : Article 3 bis de la loi du 7 juin 1951

« I. - Le ministre chargé de l'économie peut décider, après avis du Conseil national de l'information statistique, que **les personnes morales de droit privé sollicitées pour des enquêtes** transmettent par voie électronique sécurisée au service statistique public, à des fins exclusives d'établissement de statistiques, **les informations présentes dans les bases de données qu'elles détiennent**, lorsque ces informations sont recherchées pour les besoins d'enquêtes statistiques qui sont rendues obligatoires en application de l'article 1^{er} bis.

Cette décision est précédée d'une **concertation avec les personnes de droit privé sollicitées** pour ces enquêtes et d'une **étude de faisabilité et d'opportunité rendue publique**.

Les données transmises par ces personnes morales ne peuvent faire l'objet d'aucune communication de la part du service dépositaire. Seules sont soumises au livre II du Code du patrimoine les informations issues de ces données qui ont été agrégées et qui ne permettent pas l'identification de ces personnes morales.

Les conditions dans lesquelles sont réalisées ces enquêtes, notamment leur faisabilité, leur opportunité, les modalités de collecte des données de même que, le cas échéant, celles de leur enregistrement temporaire et celles de leur destruction sont fixées par voie réglementaire. »

29 Malgré le pluriel, il n'y a qu'un article qui indique que « La révélation d'une information à caractère secret par une personne qui en est dépositaire soit par état ou par profession, soit en raison d'une fonction ou d'une mission temporaire, est punie d'un an d'emprisonnement et de 15 000 euros d'amende. »

30 Instituts Nationaux de Statistiques.

Actuellement, seules des données de caisse des enseignes de la grande distribution sont utilisées dans le cadre de l'élaboration de l'indice des prix à la consommation depuis 2020.

3.5 La réalisation des appariements de données

Les appariements de données sont considérés comme des traitements au sens de la loi de 1978 et du RGPD et sont donc déclarés et tracés comme tels. Les appariements réalisés à des fins statistiques bénéficient des régimes dérogatoires accordés à tout traitement à finalité statistique. Pour les statisticiens, il s'agit d'un type particulier d'interconnexions de fichiers, débouchant sur la construction d'un nouveau fichier, appariement étant, comme il a été dit plus haut, le terme généralement utilisé par les statisticiens et chercheurs pour désigner ce type d'opérations.

Il y a plusieurs façons de réaliser un appariement de données personnelles sur le plan technique (voir encadré 8 ci-dessous). Du point de vue juridique,

1. Si on veut utiliser le NIR (Numéro d'inscription au répertoire RNIPP) pour un appariement, le traitement est décrit et encadré par un décret en Conseil d'État (décret n° 2019-341 du 19 avril 2019 relatif à la mise en œuvre de traitements comportant l'usage du numéro d'inscription au répertoire national d'identification des personnes physiques ou nécessitant la consultation de ce répertoire), en application de l'article 30 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ;
2. Pour simplifier juridiquement et sécuriser les processus, le service statistique public peut utiliser un même identifiant pour chaque individu, le Code statistique non signifiant (CSNS), et ainsi rapprocher différents fichiers pour des traitements à finalité statistique, mais à plusieurs conditions :
 - ces traitements doivent être mis en œuvre au sein du seul service statistique public ;
 - ils ne doivent mobiliser que des données produites au sein du service statistique public ;
 - ils ne doivent pas porter sur des données sensibles au sens de la loi du 6 janvier 1978.

Cette possibilité est encadrée par l'article 30 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et ses modalités de mises en œuvre sont précisées par un décret en Conseil d'État (décret n°2016-1930). L'article 30 a été créé par la loi pour une République numérique (article 34, titre 1 « la circulation des données et du savoir ») dans un but de faciliter certains appariements pour des finalités de statistique publique ou de recherche.

Encadré 8 : Quelques précisions techniques sur la réalisation des appariements

Pour rapprocher pour une même personne les données la concernant dans différentes sources, plusieurs cas de figures sont possibles pour le service statistique public :

Cas 1, le plus simple : les différentes sources ont un identifiant en commun qui permet d'identifier sans ambiguïté les individus

Certains identifiants sont certifiés : numéro d'inscription au répertoire (NIR) géré par l'Insee (« numéro de sécurité sociale ») dans le Répertoire national d'identification des personnes physiques (RNIPP), numéro d'identification national des étudiants (INE) qui sont des identifiants à usage administratif.

Pour diminuer la sensibilité des appariements, un identifiant spécifique à usage du service statistique public a été créé pour des finalités de production de statistiques publiques : Code statistique non signifiant (CSNS), par la loi pour une République numérique de 2016.

Cas 2 : des données d'état civil complètes (nom, prénom, sexe, date et lieu de naissance) sont disponibles :

On cherchera alors à rapprocher l'individu d'un enregistrement qui figure au RNIPP, de manière à

se ramener au cas précédent en s'appuyant sur le Code statistique non significiant.

L'appariement dépend alors beaucoup de la qualité des variables. Il est courant de rencontrer des orthographes différentes des noms (Lefebvre et Lefevre, par exemple, ou Dupond et Dupont) et prénoms (Carine ou Karine) voire des erreurs de frappe ou de transcription (Fabrice et Fbrice), ou encore des dates de naissance qui ne comportent que l'année ; tout ceci diminue grandement le taux d'appariement et la qualité du résultat.

Cas 3 : les données d'état civil sont incomplètes (pas de nom par exemple) mais on dispose d'autres informations comme une adresse ou une commune de résidence.

Le temps passé pour réaliser l'appariement et sa qualité dépend alors beaucoup de la qualité des variables d'état civil, comme on l'a vu plus haut, mais aussi de la qualité de l'information sur l'adresse qui peut générer de difficultés très variées et fréquentes. Il peut ainsi y avoir des erreurs orthographiques, des noms de rues anciens, des bâtiments connus à deux adresses différentes (qui donne sur deux rues différentes ou qui a deux dénominations usuelles différentes par exemple dans le cas d'une résidence avec plusieurs immeubles), pas de numérotation de la rue etc. (voir [21]) Il est courant d'avoir des taux d'appariements de 60 % avec des données « brutes » et de gagner 20 à 30 points sur le taux d'appariement en retravaillant finement les différents cas de figure.

Cas 3 bis: d'autres caractéristiques individuelles permettent de rapprocher les sources (adresse, année de naissance, sexe, autre variable commune aux deux sources)

Dans certains cas, l'adresse n'est pas disponible dans les sources et on utilise une autre information complémentaire qui permet d'identifier avec suffisamment de précision les individus pour réussir l'appariement des données.

Encadré 9 : Le Code Statistique Non Signifiant (CSNS)

Le Code statistique non significiant (CSNS) a été créé par l'article 34 de la loi de 2016 pour une République numérique afin de permettre au service statistique public de réaliser des appariements de données sans utiliser le NIR (et donc sans devoir passer par un décret en Conseil d'État). Les conditions concrètes de production et d'utilisation du CSNS sont précisées par le [décret n° 2016-1930](#) et un [arrêté du 28 septembre 2020](#).

Il est non significiant, identique pour un même individu pendant 10 ans, sauf si un problème de sécurité intervient qui amène à le changer sans délai. Son utilisation n'est possible que pour des données fournies par des entités du système statistique public, pour des traitements à finalité statistique et qui ne portent pas sur des données sensibles au sens de la loi.

Il est produit par un algorithme ; transformation irréversible (hachage) des éléments d'identification, obtention d'une signature, puis chiffrement de cette signature à l'aide d'une clé de chiffrement, ajout d'un identifiant d'environnement informant sur la période de validité, puis obtention du CSNS. La procédure s'appuie sur des algorithmes de chiffrement éprouvés, conformes à la politique de sécurité des systèmes d'information de l'Etat (PSSIE) et une organisation stricte au sein de l'Insee, visant à sécuriser la gestion des clés de chiffrement et confier les différentes phases du processus à des équipes séparées.

3.6 Principes éthiques et déontologiques de la statistique publique

3.6.1 Au niveau mondial

En 1992 la Conférence des statisticiens européens, puis en 1994 l'ONU au niveau mondial, ont adopté dix « [Principes fondamentaux de la statistique officielle](#) » qui sont les suivants :

Principe 1 : Pertinence, impartialité et égalité d'accès à l'information statistique

La statistique officielle constitue un élément indispensable du système d'information de toute société

démocratique, fournissant aux administrations publiques, au secteur économique et au public des données concernant la situation économique, démographique et sociale et la situation de l'environnement. À cette fin, des organismes responsables de la statistique officielle doivent établir les statistiques officielles selon un critère d'utilité pratique et les rendre disponibles, en toute impartialité, en vue de rendre effectif le droit d'accès des citoyens à l'information publique.

Principe 2 : Standards professionnels, principes scientifiques et comportement professionnel

Pour que se maintienne la confiance dans l'information statistique officielle, les organismes responsables de la statistique doivent déterminer, en fonction de considérations purement professionnelles, notamment de principes scientifiques et de règles déontologiques, les méthodes et les procédures de collecte, de traitement, de stockage et de présentation des données statistiques.

Principe 3 : Responsabilisation et transparence

Pour faciliter une interprétation correcte des données, les organismes responsables de la statistique doivent fournir, en fonction de normes scientifiques, des informations sur les sources, les méthodes et les procédures qu'ils utilisent.

Principe 4 : Prévention de l'usage abusif des statistiques

Les organismes responsables de la statistique ont le droit de faire des observations sur les interprétations erronées et les usages abusifs des statistiques.

Principe 5 : Sources de la statistique officielle

Les données utilisées à des fins statistiques peuvent être tirées de toutes sortes de sources, qu'il s'agisse d'enquêtes statistiques ou de fichiers administratifs. Les organismes responsables de la statistique doivent choisir leur source en tenant compte de la qualité des données qu'elle peut fournir, de leur actualité, des coûts et de la charge qui pèse sur les personnes sondées.

Principe 6 : Confidentialité

Les données individuelles recueillies pour l'établissement des statistiques par les organismes qui en ont la responsabilité, qu'elles concernent des personnes physiques ou des personnes morales, doivent être strictement confidentielles et ne doivent être utilisées qu'à des fins statistiques.

Principe 7 : Législation

Les textes législatifs et réglementaires et toutes dispositions régissant le fonctionnement des systèmes statistiques doivent être portés à la connaissance du public.

Principe 8 : Coordination nationale

À l'intérieur de chaque pays, il est essentiel que les activités des différents organismes responsables de la statistique soient coordonnées pour assurer la cohérence et l'efficacité du système statistique.

Principe 9 : Utilisation de standards internationaux

L'utilisation, par les organismes responsables de la statistique de chaque pays, des concepts, classifications et méthodes définis à l'échelon international favorise la cohérence et l'efficacité des systèmes statistiques à tous les niveaux officiels.

Principe 10 : Coopération internationale

La coopération bilatérale et multilatérale dans le domaine de la statistique contribue à l'amélioration des systèmes d'élaboration des statistiques officielles dans tous les pays.

La [Déclaration d'éthique professionnelle de l'Institut international de statistique](#), qui vaut pour les statisticiens privés comme les statisticiens publics, a été adoptée en 1985 et révisée en 2010. Elle préconise :

des valeurs professionnelles :

- Respect³¹
- Professionnalisme
- Honnêteté, intégrité

et des valeurs éthiques :

- S'attacher à respecter l'objectivité
- Expliciter les obligations et les rôles (des employeurs, des clients, des commanditaires et des

31 « Nous respectons la vie privée des personnes et l'assurance de confidentialité qui leur est donnée. Nous respectons les groupes auprès desquels nous collectons des données et prenons garde aux torts qui pourraient leur être faits par un usage inadéquat ou improprie des résultats. Nous ne devons pas dissimuler ou critiquer à tort le travail fait par d'autres. »

- statisticiens)
- Évaluer avec impartialité les diverses options
- Prendre en considération les intérêts contradictoires
- Éviter tout résultat prédéterminé
- Protéger les informations d'accès privilégié
- Faire preuve de compétence professionnelle
- Assurer la confiance dans la statistique
- Exposer et évaluer les méthodes et les résultats
- Faire connaître les principes éthiques
- Porter la responsabilité de l'intégrité de la discipline
- Protéger les intérêts des sujets (étudiés)³²

3.6.2 au niveau européen

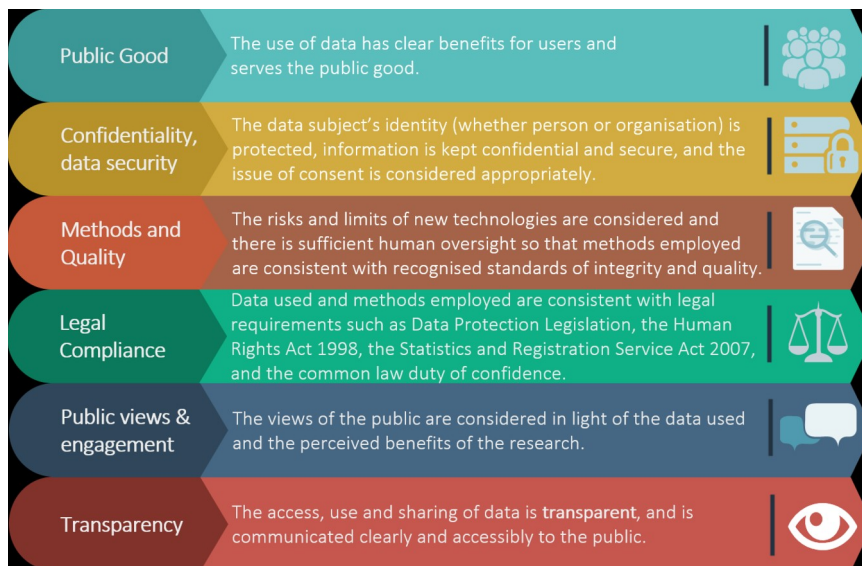
Enfin le premier [Code de bonnes pratiques de la statistique européenne](#) a été adopté en 2005. La version en vigueur contient seize principes portant sur différents aspects, dont trois ont été commentés précédemment :

Environnement institutionnel	Procédures statistiques	Produits statistiques
1 - Indépendance professionnelle	7 - Méthodologie solide	11 - Pertinence
1bis - Coordination et coopération	8 - Procédures statistiques adaptées	12 - Exactitude et fiabilité
2 - Mandat pour la collecte de données et l'accès aux données	9 - Charge non excessive pour les déclarants	13 - Actualité et ponctualité
3 - Adéquation des ressources	10 - Rapport coût-efficacité	14 - Cohérence et comparabilité
4 - Engagement sur la qualité		15 - Accessibilité et clarté
5 - Secret statistique et protection des données		
6 - Impartialité et objectivité		

3.6.3. Deux exemples de pratiques : Royaume-Uni et Canada

L'**institut national de statistique anglais** a mis en place un comité d'éthique en 2015 sous l'égide de l'Autorité de la statistique, puis les principes éthiques suivants :

³² « Les statisticiens se doivent de protéger autant que faire se peut les sujets de leurs travaux, comme individus et comme groupes, contre des effets potentiellement nuisibles de leur participation aux investigations menées. Cette responsabilité n'est en rien diminuée par l'accord explicite de participation donné par les sujets ou si cette participation résulte d'une obligation légale. La capacité intrusive potentielle de certaines enquêtes statistiques rend indispensable qu'elles soient entreprises avec les plus grandes précautions, qu'elles soient pleinement justifiées par des besoins réels et que les divers participants soient bien informés. Ces enquêtes doivent reposer, pour autant que cela soit réalisable, sur un consentement librement donné par des sujets dûment informés. L'identification et les enregistrements de tous les sujets de l'enquête, comme des seules personnes y répondant, doivent être gardés confidentiels. Des mesures appropriées doivent être prises pour empêcher la diffusion des données sous une forme qui permettrait de divulguer ou de déduire l'identité d'un répondant. »



L'Institut national **Statistique Canada** a adopté en 2019, à la suite d'une polémique survenue en 2018 (voir 2.3.1), un « cadre de nécessité et de proportionnalité », et mis en avant l'**approche scientifique** des statisticiens, qui favorise l'état d'esprit de transparence [10,11, 11 bis.]

Le principe de **nécessité** vise à s'assurer qu'il y a un bénéfice pour la société canadienne associé au traitement de données. Pour le principe de **proportionnalité**, les experts de Statistique Canada déterminent comment recueillir uniquement les données nécessaires. Ils tiennent également compte de la nature délicate des données et veillent à ce qu'il n'existe aucune autre approche moins invasive pour atteindre le même objectif. Ces principes sont appliqués de longue date à Statistique Canada, en particulier pour les enquêtes. La communication grand public sur le respect de ces principes pour les données administratives est, elle, en revanche, très récente.

Statistique Canada s'est doté en 2011 d'une directive sur le couplage (appariement) des microdonnées [12] revue en 2017 qui indique que :

« l'article 3 de la *Loi sur la statistique* donne à Statistique Canada le mandat de *recueillir, de compiler, d'analyser, de dépouiller et de publier des renseignements statistiques sur les activités et les conditions commerciales, industrielles, financières, sociales, économiques et générales de la population. Pour remplir ce mandat, Statistique Canada procède à des couplages de microdonnées, une technique importante utilisée pour l'élaboration, la production, l'évaluation et l'analyse de données statistiques afin de fournir des précisions sur d'importants enjeux sociétaux. Même si, en raison de sa nature, le couplage de microdonnées représente une intrusion dans la vie privée, Statistique Canada procède à des couplages de microdonnées dans les cas où des avantages pour le public en ressortent clairement et l'emportent sur les inconvénients liés à l'intrusion dans la vie privée. Cette intrusion tient au fait que des renseignements portant sur une personne donnée sont rassemblés d'une manière généralement inconnue de l'intéressé. Comme c'est le cas pour n'importe quelles microdonnées, la personne pourrait être désavantagée dans certaines situations. La mesure d'atténuation est que, par définition, l'analyse statistique ne vise pas une personne en particulier, seulement des renseignements non-confidentiels agrégés sont produits.* »

Trois types d'appariements sont considérés :

- Type A – Couplage visé par une autorisation générale de couplage de microdonnées
- Type B – Modification d'un fichier couplé existant
- Type C – Nouveau couplage (proposition pour un nouveau fichier couplé ; non inclus dans les types A ou B)

Le champ des couplages de type A est très large. Les couplages de type B doivent être approuvés par le responsable du système d'information. Les couplages C sont présentés sur le site internet de Statistique Canada. La directive liste des précautions à respecter dans les processus pour garantir la bonne sécurité des données.

4. Le projet de répertoires statistiques d'individus et de logements (Résil)

4.1 L'objectif de Résil

Le projet Résil correspond à une volonté stratégique de l'Insee, partagée avec les autres instituts nationaux de statistique, de progresser sur l'utilisation des sources administratives et des appariements de données ainsi que sur la qualité des résultats obtenus. Il s'inscrit dans une tendance à moyen terme, largement partagée à l'international.

Plus conjoncturellement, l'Insee a été confronté en 2020 à l'annonce de la disparition en 2023 de la taxe d'habitation (TH) alors qu'un extrait des données de gestion de cette taxe transmis annuellement à l'Insee constitue le socle de plusieurs opérations statistiques importantes, parmi lesquelles le tirage des échantillons pour les enquêtes réalisées auprès des ménages, la préparation, le contrôle et une partie du calcul des populations du recensement.

Il a été décidé fin 2020 de réagir de ce fait en deux temps : d'abord construire provisoirement à partir des sources fiscales disponibles un fichier aussi comparable que possible à celui de la TH, puis aller plus loin en mobilisant d'autres sources de données administratives, pour construire une source plus solide en cas de disparition de sources administratives (à l'instar de la TH) et plus complète en termes de population couverte, et de l'utiliser également pour faciliter les appariements de sources statistiques diverses.

Une telle base deviendra ainsi la colonne vertébrale des statistiques démographiques et sociales. Pourquoi parler de colonne vertébrale ? Parce qu'on pourra y raccorder toutes les sources d'une part, parce qu'elle contiendra très peu de données d'autre part.

Le projet Résil est constitué de deux **répertoires statistiques, l'un portant sur les individus et l'autre sur les logements**. C'est une liste exhaustive d'entités (des individus, des logements) comportant très peu de variables sur chaque entité. L'objectif est de faciliter la réalisation technique d'appariements en permettant un meilleur taux de réussite de chaque appariement, un meilleur contrôle de sa qualité et une meilleure efficacité (moins de moyens pour les réaliser), tout en permettant de séparer les informations identifiantes des autres données afin d'améliorer la sécurité de l'information. Il ne contiendra que des données d'identification :

- des informations qui permettent d'identifier sans ambiguïté les individus et les logements et d'éviter des doublons ;
- des informations permettant de relier les individus et les logements entre eux sans ambiguïté, pour reconstituer les ménages qui sont l'entité de base pour analyser les niveaux de vie ;
- des informations permettant de relier les individus et les logements sans ambiguïté à des données métier gérées par la statistique publique et stockées dans d'autres bases de données.

Il y a une différence cruciale entre un **répertoire statistique** et un **répertoire administratif** :

- Un répertoire statistique ne sert qu'à l'élaboration de statistiques.
- Il n'est accessible qu'aux statisticiens de la statistique publique.
- Il ne peut pas y avoir pour un individu donné de décision ou d'appréciation individuelle le concernant qui soit prise au vu de ses données stockées dans le répertoire.

Le système de répertoires statistiques d'individus et de logements sera l'élément central d'un système statistique démographique et social plus performant ayant pour principaux objectifs d'améliorer la connaissance de la population résidant en France, de produire plus rapidement des statistiques répondant à un large éventail de besoins et de fournir les informations nécessaires à l'évaluation des politiques publiques, tout en offrant de solides garanties de protection de ses données.

Pour atteindre ces objectifs, Résil aura comme **finalités** :

1. Finalité principale : fiabiliser et rendre moins coûteux les appariements entre sources portant sur les individus, les ménages et les logements³³.

Ces appariements peuvent notamment servir à :

- alléger la collecte d'information par voie d'enquête ou en orienter le questionnement vers des aspects non couverts par les données administratives ;
- produire des statistiques à une échelle fine, impossible à produire sur échantillons ;
- éclairer des aspects méthodologiques particuliers (par exemple la différence entre chômeur au sens du BIT et demandeur d'emploi inscrit à Pôle emploi)
- permettre l'évaluation de politiques publiques (suivi de trajectoire de bénéficiaires).

Ils seront fiabilisés et rendus moins coûteux par la présence d'identifiants communs et la possibilité d'en mesurer la représentativité et la qualité.

2. Evaluer et limiter l'impact de la disparition ou la modification substantielle d'une des sources administratives utilisées pour les statistiques démographiques et sociales.

Cette finalité, importante pour garantir la pérennité des statistiques produites et réduire au minimum les risques de rupture de séries, est apportée par la présence de sources d'origines diverses. Il est en effet peu probable que toutes soient affectées au même moment par des transformations d'ampleur.

- 3. Constituer des bases de sondage pour tirer les échantillons des enquêtes auprès des individus ou logements, en s'appuyant sur davantage de sources qu'actuellement.**
- 4. Préparer et réaliser les enquêtes annuelles de recensement et contribuer à la production des statistiques de population, à l'instar de l'utilisation actuelle des fichiers issus de la taxe d'habitation.**

4.2 Le contenu de Résil

Concrètement, Résil sera constitué de **deux répertoires statistiques distincts** (un **répertoire d'individus** et un **répertoire de logements**) qui seront **mis à jour régulièrement** avec les naissances, les décès et les sources fiscales, sociales, ou d'autres sources apportant une information sur un public spécifique comme, par exemple, les fichiers d'inscription dans l'enseignement supérieur (voir infra tableau 1 et tableau 2).

Point essentiel : on ne retient dans ces sources que des données d'identification (nom, prénoms, dates et lieux de naissance, identifiants propres à chacune des sources) **des individus, des logements, ou d'adresse des logements et des liens entre individus et logements**.

A partir de ces répertoires seront produits annuellement des « photographies » :

- la liste des individus présents sur le territoire national au 1^{er} janvier ;
- la liste des logements situés sur le territoire national au 1^{er} janvier, en caractérisant leur statut (principal, secondaire, vacant) ;
- la liste des ménages et leur composition au 1^{er} janvier, selon deux notions : partage

³³ Les difficultés pour réaliser les appariements sont exposées dans l'encadré 8, partie 3.5.

d'un même logement / partage d'un même budget au sein d'un même logement. Les listes d'individus et de logements serviront de référence pour la statistique démographique et sociale ; les listes de ménages sont indispensables pour la construction de données telles que les niveaux de vie, ou pour la réalisation d'enquêtes.

Les informations contenues dans Résil seront essentiellement des clés d'identification pour assurer le rôle de liste de référence (éviter les oublis, éviter les doublons) et pour réaliser les appariements :

- Des identifiants d'individus : le Code statistique non signifiant (CSNS) créé par la loi pour une République numérique en 2016 pour faciliter les appariements entre sources, un identifiant spécifique à Résil, pérenne pour la gestion de l'historique et qui ne sortira jamais de Résil, les identifiants des sources utilisées dans Résil. **Le numéro d'inscription au répertoire (NIR³⁴) ne sera pas stocké dans Résil.**
- Des identifiants de logement (un identifiant spécifique à Résil pérenne, les identifiants des sources utilisées pour Résil).
- Chaque logement aura un identifiant d'adresse issu du référentiel d'adresses de l'Insee du type « BZ140JD » (et non l'adresse en clair « 8 rue Zéphyrin Brioché à Gleux-lès-Lure, département de la Haute-Saône ») ce qui le rend inexploitable en dehors de l'Insee.
- Les données d'état civil (nom, prénom, date et lieu de naissance).
- Des liens entre individus et leur logement d'habitation, avec, le cas échéant, plusieurs logements possibles pour un même individu selon les sources ; *in fine* une résidence principale sera déterminée pour chaque individu.

Quelques autres variables permettant la gestion du répertoire et la mesure de la qualité :

- date de mise à jour des données
- date des changements de valeur pour certaines variables de Résil
- indicateur de présence sur le territoire français
- présence de la personne ou du logement dans chaque source administrative (oui/non)

Résil ne contiendra aucune autre information. Les données telles que le revenu, l'état matrimonial, la profession, la surface des logements... figureront dans des bases spécifiques, et ne seront mobilisées qu'à la demande (cf. partie sur les appariements) dans le cadre d'un traitement distinct. Résil ne sera donc pas une « méga-base » contenant tout ce que l'on sait sur chaque individu ou chaque logement, même s'il permet d'accéder, au cas par cas, à ces données.

4.3 Durée de conservation des données de Résil

En production :

L'enjeu est de pouvoir réaliser des appariements pour analyser sur une période suffisamment longue des évolutions concernant des individus ou des logements. Il est prévu actuellement une durée de conservation de dix ans en **production (ou base active)**, comme pour d'autres traitements réalisés par l'Insee³⁵ (ex : [panel DSN](#), [Filosofi](#)). Les fichiers de données stockés en base active pour la production de statistiques seront ensuite conservés à l'issue des traitements de production sous forme d'**archives intermédiaires** (voir infra).

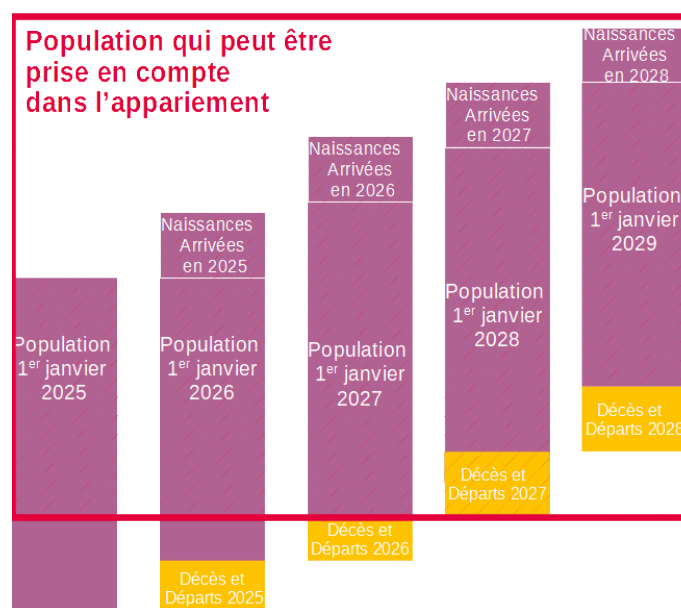
34 Toute personne née en France métropolitaine ou dans les départements d'outre-mer (DOM), ou née à l'étranger mais bénéficiant d'un régime de sécurité sociale, est inscrite au répertoire national d'identification des personnes physiques (RNIPP) et se voit attribuer un numéro d'identification, le NIR (communément appelé « numéro de sécurité sociale » car utilisé comme tel dès la création de la Sécurité sociale). Le NIR est signifiant de par son mode de construction.

35 Rubrique Protection des données du site internet Insee <https://www.insee.fr/fr/information/3719162>.

Prenons pour simplifier l'hypothèse d'une durée de conservation des informations relatives aux individus et aux logements dans Résil qui ne serait que de deux ans. Prenons l'exemple d'une évaluation d'une politique publique sur cinq ans (2025 à 2029). Les personnes décédées entre 2025 et 2027 ou ayant quitté le territoire national de 2025 à 2027 ne pourront pas être prises en compte dans un appariement de données sur 2025-2029 (voir figure 1). Cela représente environ 3 % de la population présente en 2025. Un tel pourcentage est trop important pour parvenir à des résultats incontestables.

Avec une évaluation de politique publique sur dix ans et une durée de conservation de 5 ans, la part de la population exclue serait d'environ 6 %. De tels pourcentages ont donc conduit l'Insee à retenir un délai de dix ans pour la durée de conservation des données.

Figure 1 - Population prise en compte dans un appariement basé sur Résil pour une durée de conservation des informations relatives aux individus de deux ans



Pour des raisons de lisibilité du graphique, la part de population non prise en compte n'est pas à l'échelle

En archives intermédiaires

Les traitements statistiques mis en œuvre par l'Insee ont pour finalité la production et la diffusion d'études et de statistiques d'intérêt général. Les fichiers de données stockés en base active qui permettent la production de ces études et statistiques sont conservés à l'issue des traitements sous forme d'**archives intermédiaires**, afin de permettre leur réutilisation dans le cadre de nouveaux traitements statistiques d'intérêt public, au sein de l'Insee ou par d'autres organismes, en application des obligations du livre II du Code du patrimoine et du livre III du Code des relations entre le public et l'administration, et dans le respect des dispositions de la [loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques](#).

La conservation de ces fichiers de données à l'Insee au titre de l'archivage intermédiaire fait l'objet de mesures techniques et organisationnelles strictes qui en garantissent la confidentialité et la protection des intérêts, droits et libertés des personnes qu'elles peuvent concerner. Elle peut atteindre des durées longues (jusqu'à plusieurs dizaines d'années), en fonction des besoins d'accès récurrents et de la profondeur historique que peuvent nécessiter certaines études et analyses statistiques.

Les données mobilisées au sein de l'Insee dans le cadre de la réutilisation de ces archives intermédiaires sont identifiables à partir des sources indiquées dans les publications, tableaux et bases de données diffusées par l'Insee sur son site.

4.4 La mise à jour de Résil

Le répertoire d'individus et de logements sera mis à jour régulièrement, à chaque mise à disposition d'une des sources administratives retenues, dont la liste figure infra.

Concrètement, la mise à jour consistera le plus souvent pour un individu, à indiquer s'il était présent dans telle source, pour tel mois ou à telle date, et, s'il était présent, si c'est à la même adresse ou non qu'à la livraison précédente de la même source. Dans certains cas, (naissances, individus apparaissant pour la première fois), il pourra y avoir création d'un nouvel individu. Les décès enregistrés sur le territoire feront également l'objet d'une mise à jour.

Les sources que l'Insee identifie aujourd'hui comme celles qui doivent alimenter Résil **ont été présentées au groupe avant de recueillir son avis. Parmi celles-ci**, certaines sont déjà traitées ou produites par l'Insee et leur apport pour le répertoire ne fait pas de doute pour la statistique publique; elles sont listées dans le tableau 1 ci-dessous.

Tableau 1 - Les sources retenues pour Résil

Nom des fichiers utilisés	Administration à l'origine de la création des données	Informations utilisées pour Résil
Fichier d'imposition des personnes (Fip) et Fichier permanent des occurrences de traitement des émissions (Pote) qui donne des informations sur l'imposition des personnes Fichier de mise à jour des informations cadastrales (Majic) qui donne des informations de localisation et qui permet de décrire le parc des locaux Fichier « Gérer mes biens immobiliers » (GMBI) nouvelle source pour l'Insee en remplacement de celle de la taxe d'habitation	Direction générale des finances publiques (DGFip) livraison annuelle	Nom, prénoms, date et lieux de naissance, adresses des logements, composition du foyer fiscal, Statut du déclarant (déclarant principal, conjoint, personne à charge, enfant), nature du local (maison, appartement, etc.), nature d'occupation du local (occupé par le propriétaire, occupé par un locataire, vacant, etc.) Pour GMBI on utilisera le statut principal/secondaire du logement et identité de l'occupant du local
Répertoire national d'identification des personnes physiques (RNIPP)	Insee livraison mensuelle ou hebdomadaire	Permet de connaître les noms, prénoms, dates et lieux de naissance, dates de décès
Fichier annuel de référence des allocataires des caisses d'allocations familiales (Caf)	Caisse nationale des allocations familiales (Cnaf) livraison mensuelle	Composition des foyers d'allocataires pour la bonne information sur les ménages (composition, localisation). Date de versement d'une prestation³⁶, noms, prénoms (uniquement pour le déclarant) dates lieux de naissance (pour tous les membres du foyer), adresse

³⁶ Le versement d'une prestation constitue un signe de présence plausible sur le territoire français.

Fichiers annuels d'allocataires	Mutualité sociale agricole (MSA) livraison mensuelle	Donne la composition des foyers d'allocataires, pouvant aider à reconstituer les ménages (composition, localisation). Date de versement d'une prestation, noms, prénoms (uniquement pour le déclarant) dates de naissance (pour tous les membres du foyer), adresse
Recensement Enquête annuelle de recensement (EAR) Répertoire des immeubles localisés(RIL) / Répertoire des communautés ³⁷	Insee	Le répertoire des communautés est utilisé pour compléter le répertoire des locaux. Le RIL et l'EAR seront utilisés pour des vérifications de qualité et possiblement extraire des éléments de repérage des logements
Prélèvement à la source pour les revenus autres (Pasrau) nouvelle source pour l'Insee	Direction générale des finances publiques (DGFip) livraison mensuelle	La fréquence mensuelle de cette source peut permettre une mise à jour plus fiable du répertoire d'individus (durée de présence sur le territoire, complétion des données fiscales). Date de versement d'un revenu, noms, prénoms dates et lieux de naissance, adresse
Déclaration sociale nominative (DSN)	Caisse nationale d'assurance vieillesse (Cnav) et Groupement d'intérêt public-Modernisation des déclarations sociales (GIP-MDS) livraison mensuelle	La fréquence mensuelle de cette source peut permettre une mise à jour plus fiable du répertoire d'individus (durée de présence sur le territoire, prise en compte de jeunes actifs permettant de compléter des données fiscales) et une prise en compte plus précoce des changements d'adresse. Activité salariée sur un mois donné, noms, prénoms, date et lieu de naissance, adresse

D'autres sources demandent encore des investigations techniques, juridiques et d'opportunité ; elles sont listées dans le tableau 2 ci-dessous. La position du groupe quant à ces sources est présentée en partie 5.

³⁷ Une communauté est un ensemble de locaux d'habitation relevant d'une même autorité gestionnaire et dont les habitants partagent à titre habituel un mode de vie commun. Les communautés sont de types très variés : maisons de retraites, cités universitaires, internats, foyers de travailleurs, prisons, communautés religieuses...

Tableau 2 - Les sources encore à l'étude, qui pourraient être utilisées pour alimenter Résil

Nom des fichiers utilisés	Administration à l'origine de la création des données	Informations utilisées pour Résil
Répertoire national commun de la protection sociale (RNCPS) nouvelle source pour l'Insee si l'analyse de l'apport est conclusive au regard de la proportionnalité	Caisse nationale d'assurance vieillesse des travailleurs salariés (Cnav) livraison annuelle ou trimestrielle	Ce répertoire, de portée très large, peut compléter utilement la source fiscale. Date de versement d'une prestation, noms, prénoms dates et lieux de naissance, adresse
Fichier des inscriptions des étudiants (Sise) nouvelle source pour l'Insee si l'analyse de l'apport est conclusive au regard de la proportionnalité	Service statistique public du Ministère de l'enseignement supérieur, de la recherche et de l'Innovation livraison annuelle	Etudiants non présents dans les sources fiscales (par exemple étudiants étrangers ne percevant pas de revenus) ou de préciser leur localisation (logement des parents ou logement étudiant). CSNS, commune de résidence et d'étude si l'analyse de l'apport est conclusive
Fichier des inscriptions des élèves du premier et du second degré Nouvelle source pour l'Insee si l'analyse de l'apport est conclusive au regard de la proportionnalité	Service statistique public compétent en matière d'éducation nationale : Direction de l'évaluation, de la prospective et de la performance (DEPP) livraison annuelle	Pour renforcer la couverture du répertoire sur les jeunes enfants, voire, par l'indication des adultes responsables, de consolider l'information sur la composition des ménages. CSNS, lien avec la personne responsable de l'enfant adresse si l'analyse de l'apport est conclusive
Titres de séjours des étrangers Nouvelle source pour l'Insee (Agdref) si l'analyse de l'apport est conclusive au regard de la proportionnalité	Service statistique public compétent en matière d'immigration livraison annuelle	Pour renforcer la couverture sur les personnes arrivées récemment sur le territoire noms, prénoms dates et pays de naissance, adresse si l'analyse de l'apport est conclusive
Système national de données de santé Données des cartes vitales Nouvelle source pour l'Insee si l'analyse de l'apport est conclusive au regard de la proportionnalité	Service statistique public compétent en matière de santé (Drees) livraison annuelle	Pas de données relatives à la santé : information sur le fait (oui/non) d'avoir utilisé sa carte Vitale sur une période donnée pourrait (sous réserve de tests) être précieuse comme indication de présence sur le territoire. La carte a été utilisée au moins une fois (oui/non), identité, adresse, noms, prénoms dates et lieux de naissance

Une fois Résil mis en place, à la réception de ces sources, les informations seront « triées » et orientées vers des systèmes distincts : les données d'identification de l'individu ou du logement et l'indication de leur présence dans la source viendront mettre à jour le répertoire, les données d'adresse seront orientées vers le répertoire des adresses qui renverra un identifiant non signifiant, tandis que le reste des informations qui constituent les données « métier » (revenu, état

matrimonial, catégorie sociale...) qui sont utilisées pour la production statistique ou les études seront stockées sous forme pseudo-anonymisée dans les systèmes d'information « métier ».

Pour les sources du second tableau, qui ne sont pas déjà utilisées par l'Insee, seules les informations indiquées dans le tableau seront utilisées dans Résil.

L'utilisation de sources différentes permet de renforcer la couverture des répertoires, car aucune source administrative n'est totalement exhaustive sur son champ ; elle peut toutefois mener également à une sur-couverture en apportant des individus qui ne sont plus présents sur le territoire. Ce risque peut être contrôlé par la méthode des signes de présence (voir annexe 7).

L'enjeu est en particulier d'améliorer sa qualité pour les populations difficiles à suivre dans les sources fiscales : par exemple les jeunes étudiants ou jeunes actifs qui sont encore rattachés fiscalement à leurs parents ou les personnes âgées vivant en communauté mais ayant conservé leur domicile. Le suivi de personnes récemment entrées sur le territoire national peut également être un enjeu.

4.5 Le cadre juridique prévu pour Résil

L'Insee souhaite disposer pour Résil d'un encadrement juridique fort. Les dispositions d'ordre législatif qui sont nécessaires figurent déjà dans les lois de 1951 et de 1978 présentées précédemment. L'Insee ne prévoit donc pas de texte de niveau législatif. Un simple arrêté du directeur général de l'Insee, en tant que responsable de traitement, aurait pu apparaître suffisant d'un point de vue juridique, pourvu qu'il soit complété par une étude d'impact et un avis de la Cnil. Il a toutefois semblé à l'Insee que l'importance du sujet imposait un texte de niveau supérieur.

L'Insee s'est donc orienté vers un décret en Conseil d'État, donc du même niveau que le décret qui encadre les traitements utilisant le NIR ou les modalités d'usage du code statistique non signifiant. Un tel décret conférerait une visibilité supplémentaire à l'avis de la Cnil, nécessaire en tout état de cause.

Afin de donner au dispositif toute la transparence nécessaire, tout en permettant au répertoire Résil d'évoluer selon les besoins et les opportunités, l'Insee propose que la liste des sources utilisées pour sa confection et sa mise à jour soit publiée dans un arrêté ou un décret qui serait mis à jour régulièrement après avis de la Cnil.

Le projet de décret instituant Résil, auquel serait joint le projet de décret ou d'arrêté fixant la liste initiale des sources, serait déposé fin 2022 ou début 2023, prenant ainsi en compte les travaux du groupe de concertation.

De tout premiers échanges avec la Cnil ont eu lieu dans le cadre d'une demande informelle de conseil en décembre 2021/janvier 2022.

Résil nécessite par ailleurs, en application des dispositions du RGPD, une analyse d'impact en raison du caractère exhaustif du répertoire et des croisements de données qu'il nécessite.

4.6 La sécurité du système d'information Résil

La politique de sécurité du système d'information de l'Insee s'appuie sur la politique de sécurité des systèmes d'information de l'État (PSSIE) et sur les standards de sécurité édictés par le fonctionnaire de sécurité des systèmes d'information des ministères économiques et financiers. Elle est complétée et mise à jour régulièrement pour tenir compte des évolutions de la réglementation ou de la technique, ainsi que des risques nouveaux pouvant apparaître. Elle est déclinée au niveau de chaque ministère et de chaque administration ; sa mise en œuvre passe notamment par une démarche d'homologation des systèmes d'information, reposant sur une analyse de risques et une évaluation des solutions apportées pour contrôler ces risques ou en

atténuer les conséquences. Cette politique d'homologation s'appuie sur des audits externes (expertises techniques ou tentatives d'intrusions).

Les serveurs sur lesquels sont stockées et traitées les données sont situés sur le territoire national ; ils sont administrés par des agents de l'Insee ; aucun accès extérieur n'est autorisé.

Outre les règles générales relatives à la sécurité des locaux et aux conditions d'utilisation des ressources informatiques, des règles spécifiques sont applicables en fonction de la sensibilité des informations traitées. Le répertoire statistique sur les individus et les logements relèvera du niveau de protection le plus élevé. Les accès au répertoire seront ainsi restreints à la dizaine de personnes sélectionnées pour assurer la gestion de ce dernier. Ils seront tous retracés dans un journal.

4.7 Les services que rendra le répertoire Résil

Le paragraphe précédent a décrit le traitement permettant de constituer le répertoire Résil et de le mettre à jour. Les suivants couvrent les usages du répertoire ainsi créé.

4.7.1 Appariements de données réalisés grâce à Résil

Résil sera utilisé pour réaliser des appariements à la demande d'un responsable de traitement de la statistique publique qui sera le responsable de traitement au regard du RGPD et le seul propriétaire final de la base résultante. Ce dernier est désigné par le terme « **utilisateur** » dans les paragraphes qui suivent. Ce traitement devra obligatoirement être à finalité statistique (analyse, étude méthodologique ou processus de production statistique).

Quand un utilisateur voudra réaliser un appariement de données, il sollicitera les équipes en charge de Résil en fournissant un fichier A à enrichir en l'appariant avec un autre fichier B détenu par un producteur de l'Insee. Ce fichier A pourra contenir le CSNS ou alternativement des éléments permettant d'identifier les personnes contenues dans le fichier comme nom, prénom, date de naissance, lieu de naissance, adresse de résidence.

Pour un appariement donné, seules les données correspondant aux unités statistiques sélectionnées et aux variables demandées seront fournies à l'utilisateur. Elles seront mobilisées à la demande et après accord de leur propriétaire.

L'utilisateur aura par ailleurs décrit son traitement et déclaré ses **finalités** avant le traitement conformément aux obligations du RGPD ; il devra s'assurer du respect des principes de **minimisation** et de **proportionnalité**.

Pour assurer la **transparence**, les traitements et leurs finalités seront rendus publics par l'Insee sur son site internet dans la rubrique [protection des données personnelles](#).

Les utilisations des services d'appariement de Résil - dont chacun représentera un traitement différent et différent du traitement qui constitue le répertoire Résil - seront placées sous la responsabilité de chaque service demandeur. Elles seront décrites au cas par cas sur le site du Conseil national de l'information statistique dans les programmes de travail annuels des unités du service statistique public.

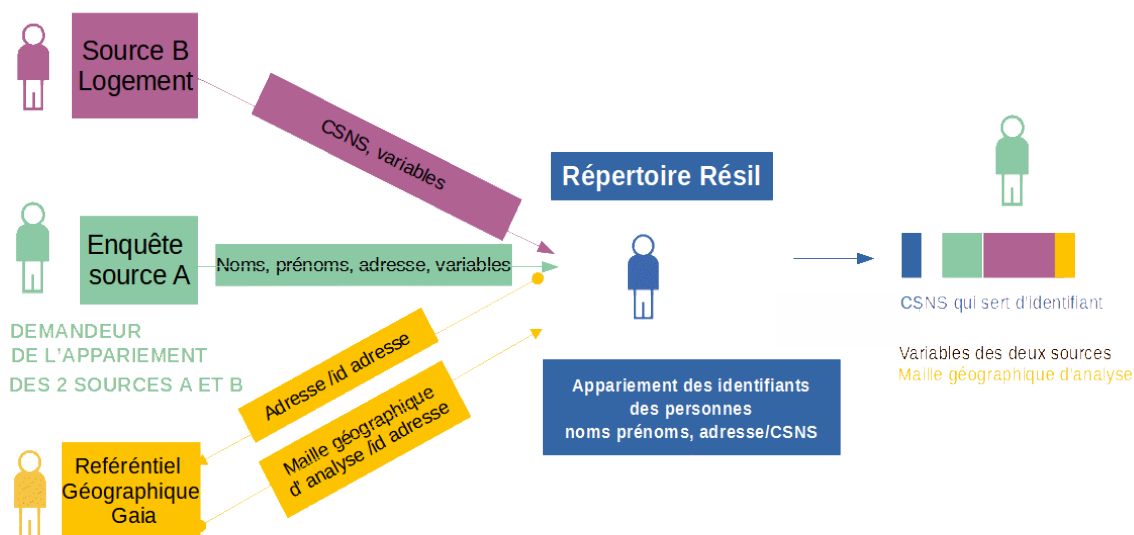
La production de ce fichier, réalisée sous la responsabilité de gestionnaires de Résil, découlera des étapes suivantes (cf. schéma ci-dessous) :

- **validation** de la demande d'appariement sur la base d'une convention. Inscription dans la liste des appariements réalisés par Résil tenue par la maîtrise d'ouvrage de Résil.
- **identification** : recherche dans Résil des individus concernés par l'appariement du fichier A et du fichier B.

- **recueil** auprès du propriétaire de la base B des variables demandées et relatives aux individus concernés par l'appariement. Ce recueil est conditionné à l'accord du propriétaire du fichier B, matérialisé par l'ouverture ponctuelle d'un accès au fichier B, ou à la transmission de l'extrait demandé.
- **constitution** du fichier résultant de l'appariement, qui comprend donc, pour chaque individu, les variables demandées ainsi que des descripteurs géographiques demandés issu du référentiel géographique Gaia (en cours de construction). Il n'y aura pas de données d'identification supplémentaires par rapport à ce dont le responsable de traitement dispose déjà.
- **transmission** du fichier apparié à l'utilisateur. Les gestionnaires de Résil fourniront des éléments permettant d'apprécier la qualité de l'appariement réalisé (taux d'appariement réalisé, représentativité du résultat).
- **destruction** du fichier résultat, du fichier A de demande, et du fichier B par les équipes en charge de Résil, une fois le fichier résultat reçu par l'utilisateur.

Figure 2 - Appariement des fichiers A et B à la demande du propriétaire du fichier A

Les couleurs matérialisent les différents propriétaires des données impliqués dans l'appariement. L'appariement est du point de vue du RGPD sous la responsabilité du propriétaire de l'enquête (source A).



Encadré 10 : Le référentiel (répertoire) géographique de l'Insee en cours de construction

L'Insee construit comme d'autres pays un référentiel géographique pour l'horizon 2024 qui s'appellera Gaia. Il contiendra toutes les adresses existantes, leurs coordonnées géographiques³⁸ associées ainsi que les périmètres géographiques (carreau, Iris³⁹, quartiers prioritaires de la politique de la ville, commune, regroupement de communes, département, régions...) utilisés pour la production de statistiques et leur diffusion ainsi que pour les études statistiques. Les adresses auront un identifiant unique non significatif qui sera la seule information géographique stockée par Résil.

³⁸ Latitude et longitude

³⁹ L'IRIS constitue la brique de base en matière de diffusion de données infra-communales dans les communes de plus de 5000 habitants. Elle correspond à 2000 habitants environ.

4.7.2 Base de sondage pour tirer les échantillons des enquêtes réalisées auprès des personnes

Pour la réalisation des enquêtes auprès des personnes, Résil pourra produire dans un premier temps un fichier de référence (base de sondage), permettant de sélectionner aléatoirement l'échantillon des personnes à enquêter. Cette base de sondage sera produite à partir des listes d'individus ou de logements de Résil, enrichies de quelques variables permettant d'assurer la représentativité et d'optimiser la taille de l'échantillon (principalement lieu d'habitation, PCS, revenu, sexe, âge). D'une enquête à l'autre, les variables utilisées pour optimiser l'échantillon pourront être différentes car l'enjeu est d'optimiser l'échantillon en fonction des thématiques spécifiques de l'enquête.

Une fois le tirage de l'échantillon de personnes à enquêter réalisé par l'unité responsable des sondages, les « données de contact » permettant la réalisation de l'enquête sur le terrain (liste des individus et ménages concernés, adresse, mail, téléphone) seront ajoutées via un second appariement, pour les seules personnes retenues dans l'échantillon.

4.7.3 Préparation et contrôle des enquêtes de recensement, calcul de la population des communes.

Le premier besoin est que Résil puisse fournir les mêmes informations que celles fournies par le fichier de la TH : évolution du nombre de logements, dont résidences principales, par commune, évolution du nombre de personnes vivant dans ces résidences principales.

Ces données servent d'une part à préparer les enquêtes de recensement, d'autre part à réaliser des contrôles de qualité de la collecte, enfin à calculer les populations légales des communes de moins de 10 000 habitants.

Si ces informations sont fournies plus rapidement que celles de la TH, ou sont de meilleure qualité (en raison de l'utilisation de plusieurs sources), on pourrait envisager de les utiliser pour des estimations plus précoces que celles réalisées actuellement.

5. Le point de vue du groupe de concertation

5.1 Considérations générales

Les auditions et les exemples d'utilisation présentés pour Résil et pour les appariements ont permis de mettre en évidence un certain nombre d'opportunités.

Les discussions ont aussi fait apparaître des interrogations, voire des craintes, dans la lignée de celles qui avaient été soulevées lors de la Rencontre du 28 janvier 2022. La discussion a porté d'une part sur Résil proprement dit et d'autre part sur les appariements qui seront réalisés grâce à Résil. Elle a évoqué les craintes d'atteinte au respect de chacun à sa vie privée.

Encadré 11 : Risques d'atteinte au respect de chacun à sa vie privée

Au-delà du fait que ce respect est une valeur déontologique forte pour le statisticien public, on sait que le respect de la vie privée est expressément protégé par l'article 8 de la Convention européenne des droits de l'homme et des libertés fondamentales, l'article 9 de la Charte des droits fondamentaux de l'Union européenne et, dans le droit interne, par l'article 9 du code civil. Le Conseil constitutionnel l'a érigé en droit que la Constitution garantit, en le rattachant à l'article 2 de la Déclaration des droits de l'Homme de 1789 (en particulier, Conseil constitutionnel, 23 juillet 1999, n° 99-416 DC).

Quelle est la portée de ce droit ? « Dans son acception originelle, il fournit d'abord une

protection contre une intrusion dans l'intimité des personnes, qu'elle soit le fait de l'État ou des tiers. Ainsi conçue, la vie privée est étroitement liée au domicile, à la correspondance, aux relations intimes, dont il faut préserver le secret » écrit Vincent Mazeaud (*Les nouveaux Cahiers du Conseil constitutionnel*, juin 2015).

On y adjoint – sans qu'il puisse y avoir entière confusion – le droit à la protection des données personnelles, renforcée en 2016 par le règlement européen de protection des données (RGPD) et, en France, par la loi du 20 juin 2018 ayant en particulier modifié celle du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

Il est possible de porter atteinte au droit au respect de la vie privée. Mais seulement dans l'hypothèse où une telle mesure est nécessaire à l'intérêt général, par exemple, indique la Convention européenne des droits de l'homme, au « bien-être économique du pays », dont fait partie l'établissement de statistiques publiques. Il faut qu'une telle atteinte soit proportionnée au but légitime recherché.

Le groupe a, de manière très générale, à plusieurs reprises, questionné une tendance de notre temps qui est celle de la « **gouvernance par le nombre** ».

Il existe depuis plusieurs années une remise en cause profonde des modalités de gouvernement des années d'après-guerre, reposant fréquemment sur des indicateurs économiques et sociaux chiffrés et des modèles macro-économiques, jusqu'aux résultats de « performance » du moindre établissement médico-social. Ces modalités trouvent en effet leurs limites dans l'ignorance des faits sociaux qui ne peuvent être quantifiés : le bien-être, la haine, la négociation...[34].

Renforcer la précision des analyses quantifiées n'est évidemment pas sans liens avec la propension du responsable politique à s'en satisfaire et, par conséquent, à « gouverner par le seul nombre ». A ce titre, Résil peut en accroître la possibilité. Le groupe de concertation y a été sensible.

Le groupe est cependant convaincu que cela ne doit pas interdire au statisticien de chercher à améliorer ses outils d'analyse.

Toutefois, le groupe de concertation reste sensible à la nécessité pour l'Insee, accrue dans le cas de Résil, de bien clarifier la portée et les limites des résultats obtenus par appariement.

Les participants se sont également beaucoup interrogés, en faisant le lien avec des précédents historiques datant de la Seconde Guerre mondiale, sur la possibilité d'un **détournement de finalités du répertoire, dans un contexte politique particulier** (régime plus autoritaire ou besoins de sécurité accrus...) se traduisant par une demande d'accès aux données à des fins non statistiques, en contradiction avec les obligations légales et éthiques de la statistique publique.

Face à ce risque, une attention particulière au principe de minimisation doit être apportée, ainsi qu'une vigilance continue de la part de tiers extérieurs à la statistique publique, comme l'Autorité de la statistique publique. Celle-ci peut être saisie par les représentants de la Nation et du gouvernement, par le président du Conseil national de l'information statistique, par le service statistique public, le directeur général de l'Insee agissant alors au titre de ses attributions de coordination, par un pouvoir d'autosaisine, par un lanceur d'alerte⁴⁰.

Les participants se sont également fortement interrogés sur les risques de mésusages de données à la suite d'une cyber-attaque.

Face à ces deux risques, une attention particulière sur le principe de minimisation des données est également nécessaire, assorti d'un niveau de sécurité satisfaisant.

40 Décret n° 2022-1284 du 3 octobre 2022 : <https://www.legifrance.gouv.fr/eli/decret/2022/10/3/2022-1284/jo/texte>

5.2 Les principes au cœur de Résil

Le groupe s'est ensuite, de manière plus précise, interrogé sur les principes au cœur de la constitution de Résil.

5.2.1 Sur la nécessité du répertoire et ses modalités d'alimentation

Comme il a été indiqué ci-dessus, le projet Résil comporte en réalité deux répertoires (d'individus et de logements) qui sont, comme le terme même le signifie, à vocation exhaustive, c'est-à-dire regroupant la totalité de la population résidente et la totalité du parc immobilier français.

Interrogé sur le point de savoir si l'exhaustivité du champ sur la population et les logements était nécessaire, en d'autres termes si un échantillon de la population (actualisé) ne pouvait être substitué au répertoire, l'Insee a expliqué de manière convaincante au groupe de concertation qu'un tel échantillon ne pouvait rendre des services identiques, tant en outils disponibles qu'en qualité des résultats.

Le groupe de concertation a également relevé que d'autres pays avaient recours à des instruments de cette nature pour des besoins statistiques, comme la Nouvelle-Zélande et le Canada. Il a été sensible aux arguments tirés de l'expérience de ce dernier pays à travers l'audition du représentant de Statistique Canada.

Il a, en outre, noté que les pays qui n'avaient pas recours aux répertoires étaient, pour beaucoup d'entre eux, des Etats pourvus de registres de population. Une telle orientation paraît au groupe devoir être résolument écartée dans le cas de la France. L'exemple, détaillé ci-avant, du registre finlandais, lequel mêle à la fois des données nominatives et des données comportementales de personnes (par exemple des mentions de faits de délinquance ou de jugements de tribunaux) semble, d'un point de vue français, particulièrement inconcevable parce que susceptible de porter une atteinte manifestement excessive à la vie privée. Il a bien été noté, *a contrario*, que les informations inscrites dans les répertoires de Résil, si elles sont, comme dans les registres, exhaustives, sont en nombre très limité.

Il a aussi été sensible au fait que les inscriptions au répertoire d'individus ne faisaient jamais utilisation du NIR. Les garanties de sécurité, légales ou de fait, qui assortissent la définition et l'usage de Résil (par exemple le petit nombre de personnes qui en autorisent l'accès, la tenue d'un journal de ces accès...) lui sont, enfin, apparues solides.

Enfin, le groupe a pris acte qu'il n'y avait pas d'alternative satisfaisante à la solution proposée sur la base des arguments présentés par l'Insee :

- un répertoire sur échantillon ne permettrait pas de fournir de base de sondage suffisamment précise, pas plus que les données nécessaires au recensement de la population ou à la création de fichier permettant des analyses sur des domaines assez fins ;
- un système reposant sur un nombre plus réduit de sources (par exemple uniquement des sources fiscales) pour l'alimentation serait moins exhaustif (certaines populations comme des jeunes étudiants, des jeunes actifs ou des personnes âgées en Ehpad sont en effet moins bien couvertes ou moins bien localisées, ce qui créerait des biais dans l'analyse). Il serait également plus fragile en cas de transformation de la source dominante ou de problèmes de qualité sur celle-ci ;
- un système ne reposant pas sur des répertoires rendrait les appariements plus difficiles à réaliser, et rendrait leur contrôle de qualité incomplet.

Par ailleurs, le groupe a noté que l'Insee indiquait ne retenir que les seules variables nécessaires à la bonne mise en œuvre des fonctionnalités de Résil et qu'un effort pour diminuer au maximum l'information identifiante était donc entrepris. A cet égard, le fait que ne figure dans Résil aucune donnée permettant de caractériser les personnes ou les logements est apparu comme une mise en œuvre du principe de minimisation.

Il n'a, en définitive, pas émis d'objection à l'utilisation de principe de répertoires, qui lui est apparu conforme au principe de *proportionnalité* exigé des textes en vigueur.

5.2.2 Sur l'utilisation de Résil pour réaliser des appariements pour d'autres utilisateurs

Le groupe s'est ensuite interrogé sur le fondement juridique de la notion d'appariement, qui est au cœur de Résil.

Il a notamment été conduit à se demander s'il n'est pas opportun de dissiper toute ambiguïté et de faire apparaître, soit dans la loi de 1978, soit dans la loi de 1951 (où il ne figure explicitement qu'incidemment et pour les seules données de santé, comme il a été dit) le concept d'appariement et sa définition et d'en tirer expressément une disposition selon laquelle, dès lors qu'ils sont réalisés à « des fins exclusives d'établissement de statistiques », ces opérations sont autorisées.

Cette solution aurait l'avantage de clarifier les perspectives de travail de l'Insee et de sécuriser le projet Résil, mais est-elle pour autant indispensable ?

Si, en effet, le législateur (européen ou national) n'a pas mentionné de manière explicite la possibilité d'appariement, il ne l'a pas écartée expressément. Il a défini un régime particulier pour les travaux statistiques (ainsi qu'à fin d'archives et de recherches scientifiques), pourvu que soient respectées certaines garanties tenant à la protection des données personnelles (cf. le titre I de l'article 89 du RGPD). Cette intention incontestable n'a pas, et ne pouvait pas définir toutes les modalités d'établissement de statistiques dont l'appariement est un aspect parmi d'autres.

En outre, et surtout, la lecture qui est faite par la statistique publique de la notion d'interconnexion, laquelle est expressément autorisée par la loi en vigueur, conduit à penser qu'appariement et interconnexion constituent une seule et même notion juridique de rapprochement de fichiers.

Dès lors que cette manière de voir est juridiquement confirmée, on peut considérer que, même si les méthodes d'appariement ne sont pas explicitement prévues par la loi en vigueur, elles sont bien autorisées.

Il reste qu'il serait plus judicieux d'employer dans les textes législatifs, rédigés à des époques et pour des finalités différentes, un seul et même terme pour qualifier une opération identique. Le législateur devrait y remédier dans les mois à venir.

En attendant, il paraît cependant utile d'introduire, mieux qu'elle ne l'est aujourd'hui, la notion d'appariement dans la réglementation en vigueur. C'est pourquoi le groupe de concertation suggère que, dans le décret fondant Résil, soit définie la notion d'appariement et qu'il soit précisé dans le décret créant le Cnis que le Cnis « délibère et donne son avis » sur les projets d'appariements tels que définis dans le décret fondant Résil.

Le groupe s'est ensuite interrogé sur l'usage des appariements.

Il a constaté, d'une part, que les rapprochements de fichiers selon cette modalité étaient relativement anciens ; d'autre part, que les résultats des appariements constituaient bien une *nécessité* pour la définition de politiques publiques plus précises à la fois dans le temps (suivi de cohortes) et dans l'espace (découpage territorial plus fin) ; enfin que de nombreux pays étaient déjà engagés dans cette voie recommandée en outre par des instances internationales compétentes.

Le groupe de concertation a bien noté l'intérêt que présentent l'usage des sources administratives et les appariements pour la production statistique en complément des enquêtes et l'enrichissement de la connaissance, ainsi que l'apport que représente Résil pour faciliter et rendre ces opérations plus efficaces et mieux en piloter la qualité.

La constitution de Résil et les appariements qu'il permettra répondent à un besoin d'intérêt général de meilleure connaissance des phénomènes économiques et sociaux.

Le groupe relève également que l'usage des appariements est une tendance largement partagée au sein des instituts nationaux de statistique, même si elle s'inscrit dans des contextes administratifs, culturels, techniques propres à chaque Etat.

Face à ces besoins, l'Insee a la conviction que disposer d'une infrastructure permettant de renforcer la sécurité de ces traitements et d'éviter la duplication des données représenterait un point fort à encourager. Ce point est important pour le groupe.

L'argument de l'Insee faisant valoir la possibilité grâce à Résil d'une meilleure visibilité des appariements est également apparu comme une opportunité positive, au regard notamment du besoin de transparence vis-à-vis de nos concitoyens.

Il a relevé aussi, et surtout, que le législateur, dans la période récente, avait défini des outils destinés à faciliter et à sécuriser la réalisation d'appariements. C'est la loi n° 2016-2031 du 7 octobre 2016 pour une République numérique qui a, notamment dans son article 34, défini le code statistique non signifiant (CSNS), utilisé à cette fin, ainsi qu'il a été mentionné ci-dessus.

Pour ces raisons, le groupe de concertation n'a pas émis d'objection au principe des appariements, conscient que la mise en œuvre du projet soumis à son examen conduira à les faciliter, par conséquent à en élargir l'usage à des fins statistiques.

5.3 L'architecture du projet

Le groupe, en deuxième lieu, s'est demandé si l'architecture du projet pouvait faire naître des interrogations.

5.3.1 les dérogations accordées à la statistique publique

Il a estimé d'abord que, conformément à la lecture juridique qu'il en a faite, le projet rentrait bien dans les garanties applicables aux opérations de statistique publique. Autrement dit, les éléments dérogatoires prévus, tant dans le RGPD que dans la loi dite « Informatique et libertés » de 1978, s'appliquent aux opérations que permettra Résil. Compte tenu des précautions prises dans le recueil et le traitement des données et, surtout, de l'absence dans ceux-ci de toute opération pouvant conduire à quelque utilisation que ce soit à fin de décision administrative ou d'opinion portée sur une personne, le contrôle des usagers sur l'usage des données qui les concernent (sous forme de droit d'accès par exemple) n'a pas lieu de s'appliquer, comme il en va pour les autres travaux de statistique publique. Résil permettra de rassembler les données qui disparaîtront avec la taxe d'habitation et d'accroître sans aucun doute les données disponibles, mais la finalité qui leur est donnée sera exactement la même que celle des opérations statistiques existantes. Autrement dit, il n'y a pas lieu de prévoir une *dérogation au régime dérogatoire* institué par la loi et le RGPD au bénéfice de la statistique dans la mesure où certains droits rendent impossible ou entravent sérieusement la réalisation des finalités de statistique⁴¹.

5.3.2 les craintes concernant la vie privée

Le groupe s'est fait l'écho de craintes concernant la tendance à constituer des bases très riches qui dévoilent une partie trop importante de la vie privée (modes de vie / trajectoires de vie) et/ou portent sur une part trop importante de la population ont été exprimées. En effet, plus l'information construite est riche, plus l'impact des éventuels usages malveillants exposés pour Résil plus haut serait préjudiciable.

Un usage sans précaution de Résil remettrait gravement en cause les deux aspects du respect du droit à la vie privée (encadré). En effet, les appariements auxquels il est procédé font

⁴¹Si les personnes peuvent effacer le fait qu'elles ont perçu une prestation sociale, comment calculer la proportion de bénéficiaires ? Comment connaître l'évolution des bénéficiaires de prestations ?

apparaître des comportements individuels dont les choix et les manifestations peuvent révéler la vie privée des personnes, dans ce qu'elle a d'intime. La mise à disposition sous une forme non agrégée de ces données à des personnes non habilitées pourrait amener à révéler des situations propres à un individu ou à un ménage donné.

Le groupe de concertation a été très attentif au risque accru que faisait porter sur la vie privée la multiplication d'appariements. Toutefois, les conditions de traitement des données ainsi appariées, qui sont celles appliquées à toutes les enquêtes statistiques publiques, ont paru suffisamment solides pour écarter le risque d'atteinte à la vie privée dans un usage conforme aux règles qui doivent encadrer Résil. Il revient à l'Insee de prévenir tout mésusage de l'instrument, en écartant toute possibilité que soient détournées ou dérobées des données. Le groupe note que l'institution présente sur ce point de solides garanties, pour maintenir le risque à un niveau suffisamment réduit pour qu'il lui paraisse acceptable.

5.3.3 les craintes relatives au risque de stigmatisation d'un groupe social

De même, une demande a été formulée à plusieurs reprises d'être particulièrement vigilant, compte tenu de la finesse des informations qui découlent des appariements de données, à ne pas produire de statistiques qui conduiraient à décrire trop finement, voire profiler ou stigmatiser un groupe social. Cette préoccupation était déjà présente dans le code de déontologie professionnelle du statisticien à destination du secteur public et privé adopté en 2010⁴² [4] par l'Association internationale des statisticiens. Elle est prise en compte dans les critères de diffusion de certaines données statistiques qui imposent des tailles minimales pour les unités diffusées, en nombre d'habitants ou en superficie. L'Insee doit être vigilant sur les données qu'il diffuse sur la base d'appariements réalisés à partir de Résil, comme sur la base de tout autre collecte (peut-être plus encore s'agissant d'une information plus riche) pour éviter qu'elles ne soient utilisées ni à des fins de stigmatisation de tel ou tel groupe social, ni de « profilage » des individus.

5.3.4 Risque de trop grande rigidité du système

A l'inverse, une crainte a été formulée qu'une trop grande rigidité du système n'amène à limiter la richesse des données accessibles pour des finalités de recherche ou de production statistique. Il y a donc un équilibre à trouver en s'appuyant autant que possible sur des regards complémentaires à celui du statisticien, sur la consultation et sur l'écoute du point de vue sur le sujet des appariements des citoyens ou de leurs représentants ainsi que et des experts notamment juridiques, des libertés publiques. Cet équilibre doit être assorti d'une communication transparente sur les bénéfices, risques et garanties pour le citoyen.

5.3.5 La gestion des identifiants et de l'identifiant pérenne

Le groupe a bien noté que le NIR ne figurerait pas dans le répertoire d'individus de Résil ; il considère que c'est un point fondamental au regard de la problématique de protection des données.

L'utilisation dans Résil d'un identifiant technique pérenne pour les individus, strictement interne à Résil et destiné à la gestion informatique de la base, apparaît justifiée, dès lors que soit définie une procédure de changement de cet identifiant en cas de compromission, à l'instar de ce qui est prévu pour le CSNS.

42 « Nous respectons les groupes auprès desquels nous collectons des données et prenons garde aux torts qui pourraient leur être faits par un usage inadéquat ou impropre des résultats. » et « Les statisticiens se doivent de protéger autant que faire se peut les sujets de leurs travaux, comme individus et comme groupes, contre des effets potentiellement nuisibles de leur participation aux investigations menées. Cette responsabilité n'est en rien diminuée par l'accord explicite de participation donné par les sujets ou si cette participation résulte d'une obligation légale.

Le principe de la gestion des identifiants de logements, qui sera fondée sur le même principe d'un identifiant technique non signifiant, ne permettant pas de faire de lien immédiat avec tel logement (88, avenue Verdier 2^e étage gauche) ou communauté (Couvent des oiseaux à Gleux-les-Lure) sans informations complémentaires contenues dans le référentiel géographique de l'Insee Gaia a été également approuvé par le groupe.

5.3.6 Le groupe de concertation s'est aussi interrogé sur le délai de conservation des données et sur l'exercice du droit à l'effacement

Il lui a été indiqué que le délai de dix ans décrit ci-dessus se définissait comme le délai pendant lequel les deux répertoires étaient maintenus dans une base active, c'est-à-dire mis à jour. Autrement dit, au-delà de ce délai, les personnes pour lesquelles il n'y a pas de signes qu'elles appartiennent à la population résidente (et les personnes décédées bien entendu) disparaîtront du répertoire. Il a noté qu'un délai inférieur apparaissait trop court, en portant atteinte à la qualité des résultats, comme indiqué dans la partie descriptive de ce rapport.

Le droit d'effacement ne s'applique pas aux données collectées dans le cadre de traitements à des fins de statistiques publiques (cf partie 3, les dérogations accordées à la statistique publique) ; cette dérogation n'entraîne pas de préjudice pour les intéressés, puisque les données ne sont jamais diffusées sous forme individuelle et que des précautions sont prises lors de la diffusion de données agrégées pour protéger la confidentialité des données initiales ; par ailleurs, la mise en œuvre du droit d'effacement rendrait impossible le suivi de trajectoires ou l'évaluation de politiques publiques, puisqu'on ne pourrait plus reconstituer la population statistique d'intérêt.

5.3.7 Concernant les données contenues dans les répertoires et les informations utilisées pour les alimenter

Le groupe formule les réflexions suivantes :

Outre l'absence de toute référence au NIR, le groupe de concertation a également relevé l'absence de toute donnée sensible au sens du RGPD, au stade actuel du projet. Il lui paraît nécessaire de maintenir cette caractéristique ultérieurement.

Précisément sur ce point, il a bien été pris note de ce que les fichiers de données administratives susceptibles d'être utilisés pour alimenter Résil étaient rangés en deux listes distinctes : ceux dont il était certain qu'ils seraient utilisés, et ceux qui, sous réserve d'investigations ultérieures sur l'intérêt de leur utilisation, étaient susceptibles de l'être. Se pose néanmoins la question de l'utilisation de fichiers contenant des données sensibles, même si les informations extraites de ces fichiers pour alimenter Résil sont très limitées (identification, localisation, existence d'une interaction avec l'administration).

Les participants se sont également interrogés sur l'éventualité d'ajout d'autres fichiers, contenant des informations également jugées sensibles ou révélatrices de la vie privée (comportements de consommation, données de santé, données relatives à la sécurité), si d'autres sources étaient mobilisées à l'avenir pour alimenter le répertoire.

Sur ce point, l'Insee a précisé sa position : les fichiers comportant des données de sécurité, de santé, de consommation, ne devraient pas, sauf exception, être utilisés pour alimenter le répertoire d'individus ou de logements. Toute exception devrait être subordonnée à une instruction sur le plan technique de la qualité des données, de leur apport à la couverture des répertoires, à une analyse juridique et déontologique et à l'avis de la Cnil (consultée sur tout projet de texte modifiant la liste des sources utilisées pour les répertoires). Relèveraient ainsi du régime des exceptions :

- Les fichiers comportant l'information binaire oui/non « a utilisé sa carte Vitale sur le territoire national au cours des 3 derniers mois (ou des 12 derniers mois) » ;

- Parmi les fichiers de consommation, des informations provenant d'opérateurs en charge d'une mission de service public (eau, électricité, gaz) : liste et localisation des logements et, pour les fluides, une donnée qualitative permettant d'apprécier si le logement est occupé à titre permanent, intermittent (résidence secondaire ou occasionnelle) ou s'il est vacant ;
- Parmi les fichiers de sécurité, ne serait utilisée que la liste des personnes vivant en établissement pénitentiaire, dans l'hypothèse où aucune autre source ne permettrait de les prendre en compte et de les localiser.

Le groupe prend note de cette position. Il est également conscient du fait qu'en vertu du principe d'indépendance professionnelle de la statistique publique, lequel est garanti par l'Autorité de la statistique publique, aucun ajout de source ne saurait être imposé « de l'extérieur ».

Une option pourrait consister à interdire a priori certains types de fichiers (santé, sécurité, consommation) pour l'alimentation de Résil. Elle a été écartée parce qu'une telle pratique laisserait penser que tout ce qui n'est pas interdit serait possible.

Toutefois, en l'état des informations données, si la première liste de fichiers appelés à faire l'objet d'appariements qui figure au point 4.4. du présent rapport, n'appelle pas de commentaires particuliers de la part du groupe, la seconde, qui présente des sources sur lesquelles l'Insee s'interroge, a suscité des réserves de sa part, relatives à la présence du fichier des titres de séjour des étrangers (AGDREF, pour « Application de gestion des dossiers des ressortissants étrangers en France »), d'une part, à la carte VITALE, d'autre part, au RNCPS (« Répertoire national commun de la protection sociale ») enfin.

AGDREF contient des informations relatives au séjour des étrangers en France (titres de séjour) et des données fournies à l'occasion des demandes de titre, notamment des empreintes digitales. L'Insee envisage de recourir aux données de cette application pour avoir des éléments sur l'entrée d'un étranger en France. Compte tenu de la sensibilité qui s'attache, depuis trop longtemps, à la question de la présence étrangère en France et à celle, connexe, de savoir qui possède ou non parmi les étrangers un titre de séjour, il est apparu au groupe que l'utilisation de cette information n'est pas opportune.

A l'égard des signes d'une appartenance à la population résidente, la carte VITALE est à coup sûr beaucoup plus performante. Toutefois, elle contient des données sensibles, relatives à la santé (consultations, usage de médicaments...). Certes, l'Insee entend n'utiliser que ce qui est relatif au simple usage de la carte, sans aller au-delà. Toutefois, le risque de confusion, soit dans les faits (par suite d'erreurs) soit dans l'opinion, percevant mal la séparation des deux natures de données, conduit le groupe à la même conclusion que pour le fichier des étrangers. Il semble au groupe que l'utilisation de cette source pose un problème de proportionnalité qui peut porter atteinte au projet Résil lui-même.

Il en va de même pour le RNCPS dès lors qu'outre l'identification des bénéficiaires des organismes de Sécurité sociale (une centaine), il comprend les données relatives aux titres qui leur confèrent cette qualité et celles concernant les prestations servies (allocations sociales diverses, prestations de soins...). De ce point de vue, il comprend aussi des données sensibles. Mais il y a plus : le RNCPS est lui-même un regroupement de fichiers. Si les appariements sont acceptables dès lors qu'ils consistent à rapprocher un fichier constitué par une administration pour ses besoins d'un autre fichier d'une seconde administration, ils posent une question d'une autre nature dès lors qu'un des deux fichiers (*a fortiori* les deux) sont des regroupements de fichiers antérieurement existants. Or le RNCPS, institué voici une douzaine d'années par la loi à fin de lutte contre la fraude sociale, est bien un regroupement de fichiers détenus par de multiples organismes différents. Pour ces deux raisons, le groupe de concertation est défavorable à l'introduction future du RNCPS au nombre des fichiers pouvant être mobilisés pour alimenter Résil.

5.3.8 La sécurité des données

En ce qui concerne la sécurité des données, le groupe formule les observations suivantes :

La fiabilité du dispositif repose sur une bonne utilisation des techniques permettant de rendre non identifiables les personnes dont les données figurent dans les travaux conduits grâce au projet Résil. Sur ce point, le groupe a exprimé des inquiétudes renouvelées, portant en particulier sur les groupes restreints (soit à une petite échelle géographique soit à raison de caractères précis – composition du ménage par exemple).

Sur ce point, l'Insee a apporté les informations suivantes :

- Tout d'abord, il n'y aura pas de variables sensibles au sens du RGPD dans Résil.
- Les identifiants des logements, des individus, des ménages et des lieux répertoriés ne seront pas signifiants⁴³
- Les personnes vivant en communauté figureront bien dans Résil, mais il n'y aura pas d'information concernant la catégorie de communauté (hôpital, établissement pénitentiaire, EHPAD, internat, communauté religieuse....). Le code associé à l'adresse de la communauté sera non signifiant, il ne permettra donc pas d'identifier la commune dans laquelle elle se trouve. Il ne sera donc pas possible à partir du seul répertoire Résil de retrouver une personne en communauté, ni de connaître la catégorie d'une communauté.
- Les personnes sans abri seront prises en compte dans Résil dans la mesure où elles apparaîtront dans une source administrative (prestation sociale par exemple). Pour ces personnes on n'aura pas de logement associé, mais on cherchera à attribuer une commune en s'appuyant là encore sur un code non signifiant.
- Dans le fichier donnant la liste des ménages et leur composition, on n'aura que des identifiants non signifiants d'individus. L'identité des individus figurera dans un fichier distinct. La liste des ménages comme le fichier contenant les identités bénéficieront du niveau de protection le plus élevé techniquement possible.
- Les agents de l'Insee ayant accès aux tables d'identifiants de Résil seront en nombre très restreint.
- Les fichiers construits par appariements de données autour des répertoires de Résil, mobilisant des variables issues d'autres fichiers, feront l'objet d'un haut niveau de protection et de restrictions quant à la finesse des mailles de diffusion, dans l'esprit de ce qui est pratiqué sur la diffusion de données au sein du service statistique public.

Le groupe de concertation a été finalement convaincu que les techniques mises en œuvre, certaines issues des prescriptions de la loi et du règlement (CSNS), d'autres non, permettaient de dresser un obstacle réel à l'association des résultats d'un appariement avec des individus déterminés.

Le groupe s'est interrogé aussi sur la sécurité du dispositif prévu, à la fois en interne (service statistique public) et en externe (accès par des tiers). Il a conscience que le dispositif prévu repose en interne – on l'a déjà mentionné – sur un petit nombre de personnes dont le travail doit être strictement délimité (en particulier l'usage des codes nécessaires doit être impérativement contrôlé pour n'être pas trop facilement partagé, comme il arrive souvent en matière de traitements) ; qu'il fait l'objet d'évaluations et de « mises à jour » périodiques, ainsi qu'il a été décrit. Par ailleurs, les obligations légales et déontologiques pesant sur les statisticiens publics sont fortes et les risques de fuite intempestives, comme le montre l'expérience, bien faibles. A l'externe, aucun système ne peut aujourd'hui prétendre à être dans une situation inexpugnable. Si les pénétrations du dispositif pour extorsion de fonds (« rançongiciel ») apparaissent peu probables (pour des motifs tenant aux finalités du fichier et

43 Il n'est pas possible de remonter aux individus, aux logements, aux adresses avec ces seuls identifiants sans autres données complémentaires.

aux moyens limités des rançonneurs au regard des dispositifs de protection de l'Insee qui rendent quasiment impossible qu'une attaque de ce type atteigne les bases de répertoires), elles restent possibles de la part de collectivités étrangères pour des motifs de renseignement ou, comme l'a indiqué le représentant de l'Anssi, de déstabilisation. Toutefois, d'une part, il apparaît que peu de systèmes nationaux statistiques (et pas l'Insee) ont fait l'objet d'attaques de cette nature et, en tout état de cause, sur ce point, le projet Résil ne se sépare pas des autres données détenues par l'Institut, dont certaines sont d'ailleurs beaucoup plus sensibles que le dispositif projeté. Il a été noté qu'en cas de fuite de données, l'Insee est tenu par le RGPD de la signaler à la Cnil dans un délai de 72 heures. L'Insee est également tenu de signaler une fuite de données à l'Anssi par la Politique de sécurité des systèmes d'information de l'État.

En outre, la discussion au sein du groupe a évoqué différentes techniques permettant de réduire le risque en cas d'attaque (forme du stockage, anonymisation des données (« *differential privacy* »), mise à jour régulière des mesures de sécurité pour tenir compte de l'évolution des risques et type d'attaques).

Les données seront pseudonymisées et les identifiants seront non significatifs ; la table qui contiendra les identités sera isolée du reste du répertoire et fera l'objet de mesures de protection adaptées. La « *differential privacy* », qui consiste à brouiller les données individuelles ou à agréger des postes de nomenclatures sans perturber les agrégats ne peut pas convenir pour des identifiants car leur brouillage les rendrait inopérants. La technique n'est donc pas adaptée à Résil en lui-même. L'usage de la « *differential privacy* » pourrait toutefois être utile pour élever le niveau de protection des fichiers enrichis par appariements en aval de Résil dans une perspective de diffusion. Elle fait partie des différentes techniques étudiées par l'Insee et ses homologues étrangers pour ces usages en aval.

Les participants ont soulevé la question des risques d'un usage malveillant des données résultant d'une fuite de données du répertoire Résil, par le fait de sous-traitants peu scrupuleux. Il a été possible d'écarter ces risques. En effet, l'Insee n'aura recours à la sous-traitance ni pour le développement des applications de gestion de Résil (l'Insee garde la possibilité de s'appuyer sur des experts externes, mais gardera le contrôle de son développement et de la maintenance des applicatifs), ni sur leur exploitation. L'Insee garde la maîtrise de l'administration de ses serveurs.

5.3.9 Concernant l'accès aux données par les chercheurs

En revanche, s'agissant de l'ouverture à l'extérieur, le projet doit évidemment bénéficier aux chercheurs, dans les conditions rappelées dans la description de Résil.

L'Insee a indiqué que les chercheurs pourraient bénéficier des accès aux fichiers de données pseudonymisés⁴⁴ résultant des appariements demandés par des responsables de traitement pour des finalités statistiques, sous réserve de l'accord du comité du secret. Des procédures d'appariements réservées aux chercheurs sont possibles [18]. Il n'est pas prévu d'ouvrir dans le projet le service d'appariement de Résil directement aux chercheurs. Les prestations d'appariements utilisant Résil seront réalisées par l'Insee et seront réservées au service statistique public. Les chercheurs bénéficieront toutefois de l'enrichissement de la production de données par le service statistique public ; on peut en effet considérer que davantage de fichiers pourront être versés au CASD et leur seront accessibles, selon les conditions en vigueur.

5.3.10 Impact de Résil sur le recensement de la population

Enfin, le groupe s'est interrogé sur **l'impact de Résil sur le recensement de la population**. L'Insee a confirmé qu'il y aurait des interactions entre Résil et le recensement de la population, en dehors de la fourniture directe de données évoquée plus haut.

- Le recensement sera utilisé dans les analyses qualité de Résil.

⁴⁴ Tous les fichiers mis à disposition des chercheurs sont pseudonymisés. Ils ne contiennent jamais le CSNS dont l'usage est réservé au service statistique public.

- Résil pourrait également servir à produire des estimations avancées de population, de plus en plus demandées par l'Europe.

Contrairement à d'autres pays où l'objectif premier de la mise en place de tels répertoires est de modifier en profondeur le recensement, il est trop tôt pour imaginer des liens plus ambitieux que ceux décrits à ce stade. Il faut d'abord mettre en place ce nouveau système de répertoires et éprouver sa qualité dans le temps, y compris à la maille très fine que sont les communes (35 000 communes en France dont la moitié ont moins de 400 habitants).

Sous réserve des remarques qui précèdent le groupe de concertation n'a pas émis d'objection quant à l'architecture du système.

5.4 Les procédures par lesquelles Résil doit être approuvé et suivi

En troisième lieu, le groupe de concertation s'est efforcé de déterminer les procédures qui pourraient permettre d'approuver, d'une part le projet dans son état actuel, d'autre part ses évolutions ultérieures. Il lui paraît en effet que les appariements permis par le projet Résil relèvent des dispositions de l'article 35 du RGPD (traitement faisant courir des risques particuliers aux droits et libertés des personnes physiques) et doit, en conséquence, être traité comme tel.

Le système d'information Résil et les services d'appariement qu'il pourra rendre permettront aux statisticiens publics de disposer, pour un usage strictement statistique, d'une masse importante d'informations sur les personnes, les ménages, les logements.

Ces informations devront être strictement protégées, à la fois sur le plan technique (protection contre des attaques informatiques, par exemple) et juridiques (fondement du traitement, inscription dans le droit national et européen (RGPD, loi de 1951), encadrement de la liste des utilisateurs directs de l'application et des bénéficiaires de ses services, finalités exclusivement de statistique publique).

5.4.1 Le dispositif juridique

Le dispositif juridique encadrant Résil, à sa création et lors de ses évolutions, doit être d'un niveau suffisant pour lui conférer la visibilité nécessaire et en garantir l'examen par des acteurs extérieurs à la statistique publique (Conseil d'État, Cnil).

La solution envisagée par l'Insee : un décret en Conseil d'État fondant le traitement, complété par un décret ou un arrêté pris après avis de la Cnil et fixant la liste des informations mobilisées pour alimenter les répertoires, a paru adaptée au groupe, qui a par ailleurs préconisé que le Cnis puisse être consulté en amont de ces textes.

5.4.2 Un examen d'opportunité par le Cnis pour les appariements réalisés via Résil ?

Selon le groupe de concertation, la sensibilité de la méthode le conduit à proposer que le Cnis soit informé en amont de tous les projets d'appariement via Résil, qu'il puisse formuler un avis préalable à la réalisation de ces traitements, et que ses avis soient relativement contraignants, par analogie avec les avis d'opportunité des enquêtes⁴⁵.

Il y aurait de surcroît avantage à informer régulièrement les formations spécialisées du Cnis de l'utilisation effective de Résil.

45 Décret N°2009-318, article 2, « programmes... établis en tenant compte des avis formulés par le Conseil national de l'information statistique »

La question principale toutefois est de savoir si la composition actuelle du Cnis garantit l'efficacité de l'examen préalable et du suivi.

Le groupe de concertation constate que la composition du Cnis (représentants du monde économique et social et personnes qualifiées) vise à permettre la confrontation la plus stimulante entre l'Insee (et les services statistiques ministériels) et les utilisateurs de statistiques pour deux finalités distinctes : mieux connaître les besoins de ces derniers, y compris en faisant des propositions de production de données nécessaires pour des études et d'enquêtes, et, ce faisant, les convaincre (en particulier les entreprises) de livrer les informations nécessaires. Pour parfaitement légitime qu'elle soit, au regard de la mission du Cnis définie dans la loi de 1951⁴⁶, cette composition ne reflète pas la préoccupation aussi légitime de la protection des droits et libertés des personnes physiques, même si des échanges sur ce point n'ont pas manqué à l'occasion de l'examen de certains projets.

C'est pourquoi, afin de donner à la consultation et à l'information du Cnis tout son intérêt, le groupe de concertation suggère que sa composition soit élargie (ou modifiée à nombre égal) pour que puissent y siéger (quatre ou cinq membres) des associations, des universitaires ou des juristes spécialistes des droits et libertés des personnes et de la protection des données. Ces évolutions pourraient nécessiter la modification du décret décrivant les missions et les travaux du Cnis.

5.5 Concertation et communication autour du projet Résil

En dernier lieu, le groupe s'est interrogé sur les actions à entreprendre pour poursuivre l'accompagnement du projet Résil afin qu'il soit bien compris par les citoyens dans sa définition comme dans son usage, en notant qu'il s'agissait bien d'écouter les réactions des citoyens et de leur donner des garanties dans la construction du dispositif.

L'usage des sources administratives à des fins statistiques, tout comme la possibilité de les combiner entre elles ou avec des données d'enquêtes, sont mal connus du public. L'Insee et le service statistique public doivent accompagner le développement de ces formes de collecte d'information d'un effort d'information et de transparence sur ces techniques et ces usages. Les premiers travaux réalisés par l'Insee et les services statistiques ministériels vont tout à fait dans ce sens et doivent se poursuivre, en s'inspirant de pratiques d'autres instituts de statistique. Là encore, tout doit être fait pour expliquer et montrer que l'Insee et le service statistique public fonctionnent dans la transparence.

Reste que le grand public, peu sensible aux données techniques qui sont importantes dans la définition du projet, est en revanche attentif à des questions pratiques comme l'usage qui peut être fait de sa carte Vitale, ou de savoir si son orientation de genre est connue de l'administration. Comme l'a rappelé D. Agacinski dans son audition devant le groupe, les questions des experts et celles des citoyens ne coïncident guère et tout, par conséquent, doit être entrepris pour ne pas limiter la communication aux premiers et la faciliter pour les seconds.

Dans ce domaine, un certain nombre d'expériences ont été conduites en France dans des domaines bien différents de la statistique : on pense ici aux conférences de consensus conduites dans le domaine médical ou encore de la justice, aux réunions d'échantillons de population, prévues par la loi, sur les questions de bio-éthique, ou encore à la convention citoyenne sur le climat.

Le projet Résil devrait permettre à l'Insee, dont de nombreux travaux impliquent la collecte de données de tous ordres issues des citoyens ordinaires, de s'interroger sur le point de savoir si le temps n'est pas venu d'associer non seulement les utilisateurs de statistique, comme le fait depuis soixante-dix ans le Cnis, mais aussi ceux qui en sont la source même.

Des formules très diverses peuvent être envisagées, que le mandat du groupe de concertation n'a pas permis d'explorer longuement. Ce groupe a toutefois été sensible à la concertation mise en œuvre par Statistique Canada en 2018, et aussi par cette expérience d'une fondation britannique qui, sur le thème de la protection des données (et aussi de l'intelligence artificielle),

46 Article 1 bis : « I. - Le Conseil national de l'information statistique est chargé, auprès de l'Institut national de la statistique et des études économiques, d'organiser la concertation entre les producteurs et les utilisateurs de la statistique publique. »

réunit des « jurys » d'une cinquantaine de personnes, composés sur la base d'un échantillon représentatif de l'ensemble de la population, afin qu'ils se prononcent sur les modalités des politiques d'usage et de protection des données.

De tels processus sont sans aucun doute coûteux, en temps et en crédits budgétaires et peut-être faut-il rechercher d'autres partenaires pour les mener à bien. Toutefois, une mise en œuvre à propos du projet Résil, et élargie ensuite à d'autres travaux statistiques posant des questions du même ordre, paraît au groupe de concertation particulièrement utile. Il est à peine besoin d'ajouter qu'une telle consultation n'a nullement pour vocation de se substituer aux nécessaires discussions avec les partenaires habituels de l'Insee, y compris les partenaires sociaux.

6. Les préconisations du groupe de concertation

Le groupe de concertation, au vu de ce qui précède, a formulé les préconisations suivantes :

Poursuivre l'information et la concertation sur le projet Résil tout au long du projet et jusqu'à son aboutissement (2025)

La démarche d'information et de concertation engagée sur Résil, avec notamment la Rencontre du 28 janvier 2022 et la création de ce groupe de concertation, doivent se prolonger tout au long de la vie de ce projet et être étendue à une information « grand public », gage de transparence sur ces traitements, dans l'esprit du RGPD.

De manière générale, tout doit être fait pour assurer :

- que l'Insee fonctionne dans la transparence ;
- que Résil permette d'appréhender mieux des situations qui préoccupent les Français et n'a d'autres finalités que les finalités statistiques ;
- que Résil n'est pas un fichier général, ni un échange de données administratives entre administrations (pas de retour d'informations individuelles de l'Insee vers les administrations) ;
- que l'Insee et la statistique publique mettent en œuvre les dispositions assurant que Résil ne menace pas les libertés.

1	Insee	Développer une rubrique « programme Résil » sur le site internet de l'Insee, orientée « grand public » et y faire sur le sujet ; le document de communication fera l'objet d'une présentation à la presse ; cette rubrique sera prolongée après 2025 pour rendre compte des évolutions du dispositif (notamment l'évolution de la liste des sources utilisées ou plus généralement du cadre juridique, s'il y a lieu) et de ses principaux usages
2	Insee	Informier régulièrement le bureau du Cnis et les commissions compétentes des avancées du programme Résil
3	Cnis, Insee	Organiser une rencontre sur les appariements et la mise en œuvre de Résil au 1 ^{er} trimestre 2025, à l'image de la Rencontre qui s'est tenue le 28 janvier 2022

Assurer à Résil un fondement juridique de haut niveau, conforté par des regards extérieurs (Conseil d'État, Cnil, Cnis)

4	Insee	Assurer le fondement juridique de Résil par un décret en Conseil d'Etat, pris en application des lois de 1951 et 1978, et du RGPD ; rappeler dans ce texte que les finalités de Résil sont purement statistiques et que les destinataires des informations tirées de Résil ne peuvent être
---	-------	--

		que des agents du service statistique public ; préciser dans ce texte ce que recouvre la notion d'appariement
5	Insee	Faire figurer la liste des sources alimentant le répertoire dans un texte (décret simple ou arrêté, le groupe privilégie un décret simple) pris après avis consultatifs du Cnis et de la Cnil ; le texte et les avis devront figurer sur le site de l'Insee ; l'avis de la Cnil et du Cnis sera demandé pour toute évolution de la liste des sources
6	Insee	Le RNCPS, le fichier d'utilisation de la carte Vitale et le fichier AGDREF ne devront pas faire partie de la liste des sources utilisées pour alimenter Résil

Garantir la sécurité informatique de Résil

7	Insee	Dans le cadre des démarches d'homologation, et plus souvent si nécessaire, faire auditer la sécurité du système d'information Résil à sa création puis régulièrement par un intervenant externe
---	-------	---

Apporter dans la durée, à Résil et aux appariements qu'il permet, la transparence attendue par les citoyens et les utilisateurs

8	Insee SSP, Cnis	Développer une communication active à destination d'un public non statisticien sur l'utilisation de données administrative et les appariements, en s'inspirant des bonnes pratiques des instituts nationaux de statistique étrangers
9	Insee, SSP Cnis	Publier la liste des appariements réalisés avec mention de leur responsable de traitement, des sources mobilisées, des finalités des traitements et des populations concernées, de manière à rendre compte du respect des principes de nécessité et de proportionnalité

S'appuyer pour cela sur une consultation « directe » des citoyens

10	Insee, SSP, Cnis	Constituer un <i>focus group</i> pour préparer la communication sur Résil
11	Insee, SSP, Cnis	Constituer un panel citoyen pour émettre un avis sur les évolutions de certains dispositifs statistiques, en intégrant le point de vue du « citoyen observé et du citoyen concerné par les résultats »

Elargir les missions du Cnis pour les adapter aux enjeux soulevés par Résil et l'élargissement des possibilités d'appariement

La concertation engagée à l'occasion de la construction du programme Résil doit se poursuivre et s'étendre à la question des appariements qui seront entrepris en régime de croisière, à l'image de ce qui se pratique sur les enquêtes. Cette concertation devra notamment prendre en compte les questions éthiques et celles liées aux droits et libertés des personnes. Cette concertation devrait se faire sous l'égide du Cnis, dont la composition et les missions devraient être élargies à cette fin.

- 12** Cnis Conférer au Cnis la mission explicite de donner un avis d'opportunité sur les appariements réalisés par la statistique publique, avis prenant en compte les dimensions éthiques et celles liées aux droits et libertés des personnes.
Ceci passera par des évolutions organisationnelles et, si nécessaire, réglementaires lui permettant de :
- se prononcer en opportunité sur les modalités d'alimentation de Résil et appariements qu'il permettra ;
 - renforcer dans l'examen des projets la prise en compte de critères éthiques (nécessité, proportionnalité, transparence, équité, loyauté vis-à-vis des individus et des groupes d'individus) et liés aux droits et libertés des personnes.

S'appuyer sur l'Autorité de la statistique publique (ASP) pour confirmer que les usages de Résil s'inscrivent bien dans les principes du Code de bonnes pratiques de la statistique européenne

- 13** Insee, Cnis, ASP Rendre compte annuellement à l'ASP de la mise en œuvre, des évolutions et des usages du programme Résil.
- Expliciter et faire connaître le rôle que peut jouer l'ASP en cas de dysfonctionnement ou de mésusage du dispositif.

Références bibliographiques

Gouvernance de la statistique publique

- [1] Isabelle ANXIONNAZ et Françoise MAUREL, 2021, Le Conseil national de l'information statistique : la qualité des statistiques publiques passe aussi par la concertation, *Courrier des statistiques* n° 6 - Juillet
- [2] Dominique BUREAU, 2020, L'autorité de la statistique publique : dix ans d'activité, pour une statistique indépendante et de qualité, *Courrier des statistiques* n° 5 - Décembre
- [3] Marc CHRISTINE, Nicole ROTH, 2020, Le Comité du label : un acteur de la gouvernance au service de la qualité des statistiques publiques, *Courrier des statistiques* n° 5 – Décembre

Aspects éthiques

- [4] Déclaration d'éthique professionnelle adoptée par le conseil de l'Institut International de Statistique (ISI) 22-23 juillet 2010
- [5] [Principes fondamentaux de la Statistique Officielle](#), 1994, ONU
- [6] [Déclaration d'Éthique Professionnelle de l'Institut international de statistique](#), 2010
- [7] [Code de bonnes pratiques de la statistique européenne](#), 2017
- [8] Grandes bases de données et protection des personnes, Journal de la société française de statistique, tome 146, no 3 (2005), p. 89-93, René Padieu
- [9] Comité consultatif pour la statistique publique et la recherche anglaise : détail de la structure placée sous l'égide de l'autorité de la statistique, principes éthiques, guide d'autoévaluation. <https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/> accédé le 5 août 2022
- [10] Principe de nécessité et de proportionnalité de Statistique Canada, 2019, <https://www.statcan.gc.ca/fr/confiance/reponse>
- [11] Eric RANCOURT, 2019. 'The Scientific Approach as a Transparency Enabler Throughout the Data Life-cycle'. *Statistical Journal of the IAOS*, vol. 35, n°4, pp. 549-558, 2019
- [11bis] Eric RANCOURT, 2020, An Introduction to Statistics Canada's Necessity and Proportionality Framework, October 14
- [12] Statistique Canada, Directive sur le couplage de microdonnées de 2017 <https://www.statcan.gc.ca/fr/enregistrement/politique4-1>
- [13] Youssr YOUSSEF, René Carmille, 2021, un hacker sous l'occupation, documentaire, octobre
- [14] Claude POULAIN, 2022, SAFARI a existé de 1941 à 1944, *open edition journals*, 132-133,
- [14bis] C. Poulain, 2022, "SAFARI 1970-1974", revue [Terminal](#), 134 (à paraître en novembre 2022)
- [15] Stéphane JUGNOT, 2021, les opportunités créées par la facilitation et le développement des utilisations de données administratives à des fins de recherche et de production statistique, *Cereq working paper*, numéro 9
- [15bis] Jean-Pierre AZÉMA, Raymond LÉVY-BRUHL, Béatrice TOUCHELAY. Mission d'analyse historique sur le système statistique français. [Rapport de recherche] Insee. 1998. <hal-02426370>

Appariements de données et utilisation de données administratives

[15] Stéphane JUGNOT, 2021, les opportunités créées par la facilitation et le développement des utilisations de données administratives à des fins de recherche et de production statistique, *Cereq working paper*, n°9

[16] Autorité de la Statistique Publique, 2021, *délibéré* sur les appariements de données individuelles au sein du service statistique public (SSP), septembre.

[17] Antoine BOZIO & Pierre-Yves GEOFFARD, 2017, L'accès des chercheurs aux données administratives, État des lieux et propositions d'actions, *Rapport du groupe de travail du Cnis*, mars

[18] Kamel GADOUCHE, 2019, Le centre d'accès sécurisé aux données (CASD), un service pour la data science et la recherche scientifique, *Courrier des statistiques* n° 3 - Décembre

[19] Pierre LAMARCHE et Stéfan LOLLIVIER, 2021, Fidéli, l'intégration des sources fiscales dans les données sociales, *Courrier des statistiques* n° 6 - Juillet

[20] Claire-Lise DUBOST, Aude LEDUC, 2020, L'EDP-Santé : un appariement des données socio-économiques de l'échantillon démographique permanent au Système national des données de santé (DREES), *Les dossiers de la Drees*, n° 66 • septembre

[21] Alexis HAMEAU, Sylvain LARRIEU, Anis MARRAKCHI, Alexis MONTAUT, Christophe DIXTE, Yohan CODER et Sophie HAMMAN, 2019, Appariement entre l'enquête Emploi et le fichier Historique de Pôle emploi sur la période 2012-2017, Méthode et premiers résultats, *document de travail Insee* n° 1904, juillet

[22] Loïc MIDY, 2021, Un outil d'appariement sur identifiants indirects : l'exemple du système d'information sur l'insertion des jeunes, *Courrier des statistiques* n° 6 - Juillet

[22bis] Hugues LERMECHIN, 2016, Quel est l'apport de la taxe d'habitation à l'extrapolation du nombre de résidences principales au recensement de la population ? *Économie et statistique* n° 483-484-485

[22ter] Alain GODINOT, 2005, Pour comprendre le recensement de la population, *Insee Méthodes*, hors série, mai.

Expériences étrangères sur les registres de population

[23] Michel POULAIN, Anne HERM, 2013, Le registre de population centralisé, source de statistiques démographiques en Europe. *Population*, 2013/2 Vol. 68 | pages 215 à 247. INED.

[24] BAKKER B., van ROOIJEN J., van TOOR L., 2014, The System of social statistical datasets of Statistics Netherlands : An integral approach to the production of register-based social statistics, *Statistical Journal of the IAOS* 30, 411–424

[25] WALLGREN A, WALLGREN B. 2007, Register-based Statistics : Administrative Data for Statistical Purposes. New York : Wiley.

[26] Mathew PAGE, Clara EATHERLEY, Christine BYCROFT: Stats NZ, Shane TA'ALA, 2022, A Statistical Person Register in New Zealand: Progress and challenges, *Statistical Journal of the IAOS*, à paraître

[27] Julie TRÉPANIÉ, 2022, The Canadian experience of building a privacy-responsible integrated statistical register infrastructure, *Statistical Journal of the IAOS*, à paraître

[28] Arno BENS, Stefan SCHUKRAFT, 2019, Modernisation des registres administratifs en Allemagne Développements actuels et enjeux pour la statistique publique, *Courrier des statistiques* n° 2.

[28bis] Diana BELTADZE, 2020, Developing methodology for the register-based census in Estonia, *Statistical Journal of the IAOS* 36 (2020) 159–164 159.

[29] UNECE, 2022, Handbook on Registers-Based Population and Housing
<https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3e-Handbook-E.pdf>

Communication et concertation sur les accès aux données, les appariements et le projet Résil

[30] Françoise DUPONT, Stéphane GREGOIR, 2016, La réutilisation par le système statistique public des informations des entreprises, *Rapport du Cnis*, mars

[31] Compte rendu de la rencontre du 28 janvier 2022 sur les appariements de données, présentations et vidéos <https://www.Cnis.fr/evenements/appariements-de-donnees-individuelles-entre-richeesse-de-linformation-statistique-et-respect-de-la-vie-privee/?category=1067>

[32] Compte rendu de la réunion du bureau du Cnis du 17 mars 2022 et pièces préparatoires sur le bilan de la rencontre du 28 janvier 2022 et proposition d'un groupe de concertation

[33] Rubrique sur site web de Stat Can sur l'utilisation de données administratives ou privées <https://www.statcan.gc.ca/fr/nos-donnees>, accédé août 2022

[34] Daniel AGACINSKI, 2018, Expertise et démocratie, Faire avec la défiance, rapport de France Stratégie, décembre

Annexes

Annexe 1 : Mandat du groupe de concertation

Annexe 2 : Liste des membres du groupe de concertation

Annexe 3 : Liste des personnes auditionnées

Annexe 4 : Liste des signes et termes utilisés

Annexe 5 : Situation dans les pays étrangers en matière de registres administratifs de population

Annexe 6 : La méthode des signes de présence pour statuer sur la présence d'une personne sur le territoire national

Annexe 7 : le cadre de nécessité et de proportionnalité de Statistique Canada

Annexe 1 : Mandat du groupe de concertation



Mandat du groupe

« Concertation préalable à la mise en place du répertoire statistique des individus et des logements »

Contexte

L'Insee est engagé dans la construction d'un système de répertoires statistiques d'individus et de logements (Résil) pour le bénéfice de toute la statistique publique avec une mise en service prévue pour 2025.

Bâtir un référentiel statistique sur les individus et les logements, qui permettra d'enrichir l'information statistique disponible au sein du Service Statistique Public en combinant plusieurs sources d'information issues de différents services, est un projet structurant pour les statistiques démographiques et sociales. Il répond à trois finalités :

- Résil offrira un service d'appariement de données permettant de gagner en efficacité, en qualité et en harmonisation des méthodes.
- Résil permettra, en constituant un point de référence pour la comparaison, de mesurer la qualité des sources administratives qui constituent une des ressources principales de la statistique publique.
- Résil permettra, dans le prolongement des dispositifs existants, de construire la base de sondage dans laquelle tirer des échantillons pour les enquêtes réalisées par le SSP auprès des ménages, avec une couverture encore mieux assurée.

Une rencontre¹ organisée par le Conseil national de l'information statistique, le 28 janvier dernier, a permis de présenter ce projet aux utilisateurs potentiels ainsi qu'aux représentants de la société dans le cadre d'une journée consacrée plus largement aux appariements de données de la statistique publique sur les données socio-démographiques.

Le projet de répertoires statistiques Résil a été présenté à la CNIL en décembre 2021 et janvier 2022 dans le cadre d'une demande de conseil pour préparer l'analyse d'impact et consolider l'assise juridique du projet. Un décret en Conseil d'État est en préparation pour disposer d'un cadre légal adapté début 2023, de manière à engager les développements en s'appuyant sur une base assurée.

Objectifs

La journée du 28 janvier a fait apparaître des interrogations sur les conditions dans lesquelles les appariements sont réalisés ainsi qu'une demande de poursuite des discussions, en croisant des points de vue variés sur le sujet pour mieux l'expliquer, en approfondir les conséquences et accompagner sa mise en œuvre.

Dans ce contexte, afin de favoriser la communication et la transparence sur le projet RESIL et d'asseoir une bonne acception de la démarche de la statistique publique par la société civile, le Conseil National de l'Information Statistique lance un groupe de concertation qui poursuit la démarche entreprise avec la rencontre du 28 janvier dernier.

¹ <https://www.cnis.fr/evenements/appariements-de-donnees-individuelles-entre-richesse-de-linformation-statistique-et-respect-de-la-vie-privee/?category=1067>

Le groupe aura pour objectifs :

- de lister les interrogations que suscite ce projet
- de présenter à la discussion les réponses apportées par l'Insee
- de proposer des modalités concrètes pour poursuivre les réflexions tout au long de la construction du projet et le cas échéant au-delà.

Il contribuera ainsi à l'analyse d'impact sur la protection des données prévues par le règlement général sur la protection des données (RGPD).

Composition du groupe

Le groupe est présidé par : Jean-Marie Delarue, membre honoraire du Conseil d'État.

La rapporteuse est : Françoise Dupont, programme Résil, Insee

Le groupe sera composé de personnalités qualifiées aux profils divers (juristes, sociologues, informaticiens, élus, journalistes...) permettant d'éclairer l'acceptabilité sociale du projet et de personnalités ayant en charge la protection des données et des libertés ou la représentation de groupes de citoyens (ligue des droits de l'homme, défenseur des droits, associations...).

Fonctionnement

Le groupe de travail tiendra environ cinq séances d'une demi-journée. Pour mener à bien ses travaux, le groupe de travail bénéficiera d'exposés introductifs lors de la première séance sur le projet.

Il pourra auditionner en tant que de besoin des personnalités qualifiées susceptibles d'apporter un éclairage complémentaire sur les interrogations que peut susciter le projet et sur les mesures permettant de limiter son impact. Une bibliographie sera fournie.

Calendrier

Le groupe de travail présentera l'avancée de ses travaux en juin et rendra ses conclusions sous la forme d'un rapport fin septembre 2022.

Annexe 2 : Liste des membres du groupe de concertation

Président : Jean-Marie Delarue (Conseiller d'Etat honoraire)

Rapporteuse : Françoise Dupont (Insee)

Accompagnement du groupe :

François Guillaumat-Tailliet (secrétaire général adjoint du Cnis)

Cédric Afsa (inspecteur général honoraire de l'Insee)

Membres du groupe⁴⁷

Maryse Artiguelong*, vice-présidente de la Ligue des droits de l'Homme

Jennyfer Chrétien*, Renaissance Numérique, déléguée générale

Valérie Gayte, CFE-CGC, représentante du CESE au Cnis

Mark Hunyadi, Université de Louvain, philosophe. Pour des raisons de disponibilité, Mark Hunyadi n'a pas pu participer aux travaux du groupe au-delà de la première séance

Alexandre Léchenet*, la Gazette des communes, data journaliste (journaliste spécialiste des données)

Olivier Lefebvre, Insee, maître d'ouvrage de Résil

Georges Louis, délégué national CFE-CGC, CESE

Michelle Meunier, sénatrice de Loire-Atlantique, vice-présidente du comité de déontologie du Sénat

Benjamin Nguyen, INSA Centre - Val de Loire, chercheur, directeur du Laboratoire d'informatique fondamentale d'Orléans

Emmanuelle Roux, Chaudron.io, dirigeante

Marion Selz, Société française de statistique, bureau Statistique et Enjeux publics

Roxane Silberman, directrice de recherche émérite CNRS, Centre d'accès sécurisé aux données (CASD), conseillère scientifique

Bernard Sujobert, CGT, représentant au Cnis

Cécile de Terwangne, Université de Namur, Centre de recherche Information, droit et société (CRIDS), directrice de recherche

Youssr Youssef*, le Figarodata, journaliste spécialiste des données

⁴⁷ Les fonctions mentionnées concernent la période des travaux du groupe de début mai à fin septembre 2022. Quatre membres du groupe signalés par un * ont depuis lors changé de fonction.

Annexe 3 : Liste des personnes auditionnées

Par l'ensemble du groupe	
Patrick Redor , responsable de l'Unité Affaires Juridiques et contentieuses, Insee	Le cadre juridique de la statistique publique
Antoine Bozio , chercheur, directeur de l'Institut des politiques publiques (IPP), également président de la commission <i>Services publics et services aux publics</i> du Cnis	L'utilisation des appariements et leur valeur ajoutée
Eric Rancourt , directeur des méthodes statistiques et de la science des données, Statistique Canada	L'éthique des appariements à Statistique Canada
Ani Todorova , responsable de l'Unité F.2 — Population et migration, Eurostat	Le projet de règlement européen ESOP, état des lieux des pratiques de répertoires de population dans les États membres
Françoise Maurel , secrétaire générale du Cnis	Le rôle du Cnis dans l'accès aux données administratives, les appariements de données
Audition de Serge Slama , professeur de droit public, membre du Centre de recherches juridiques (CRJ), Université Grenoble-Alpes,	Le projet Résil et la protection des libertés fondamentales et plus largement regard sur l'encadrement légal du projet Résil
Audition de Grégoire Lundi , coordinateur sectoriel finances, Agence nationale de la sécurité des systèmes d'information.	Le rôle de l'ANSSI pour les politiques de sécurité des administrations
Audition de Marie-Claire Martel , vice-présidente du Conseil économique, social et environnemental (CESE) en charge de la participation citoyenne et membre du groupe des Associations du Cese	L'expérience et le rôle du Cese en matière de concertation. L'audition prévue n'a pas pu avoir lieu.
Par une partie du groupe	
Claude Poulain , retraité de l'Insee	Le projet Safari et ses enseignements
Daniel Agacinski , délégué général à la médiation auprès de la Défenseure des droits, sollicité au titre de son expertise personnelle	Quelles modalités de dialogue et de communication avec le grand public pour un projet comme Résil ?
Carly Kind , directrice de l'Institut Adalovelace (UK)	Ethical aspects of statistical population register and matching personal data

Annexe 4 : Liste des sigles et termes utilisés

CASD Centre d'accès sécurisé aux données

CNAF Caisse nationale des allocations familiales

CNAV Caisse nationale de l'assurance vieillesse des travailleurs salariés

Cnis Conseil national de l'information statistique

Cnil Commission nationale de l'informatique et des libertés

CSNS Code statistique non signifiant

DEPP Direction des études, de la prospective et de la performance (ministère de l'Education nationale)

DGFIP Direction générale des finances publiques

Drees Direction de la recherche, des études, de l'évaluation et des statistiques (service statistique public compétent en matière de santé)

DSN Déclaration sociale nominative

EDP Echantillon démographique permanent

ESOP European Statistics on Population

GMBI Gestion de Mes Biens Immobiliers

INS Instituts nationaux de statistique

PASRAU Prélèvements à la Source pour Revenus Autres

MSA Mutualité sociale agricole

NIR Numéro d'inscription au répertoire national d'identification

Résil Répertoire statistique d'individus et de logements

RGPD Règlement général sur la protection des données

RNIPP Répertoire national d'identification des personnes physiques

SIES Sous-direction des systèmes d'information et des études statistiques (ministère de l'Education nationale)

SSM services statistiques ministériels qui, avec l'Insee, composent le service statistique public

UNECE United Nations Economic Commission for Europe

Annexe 5 : Situation dans les pays étrangers en matière de registres administratifs de population

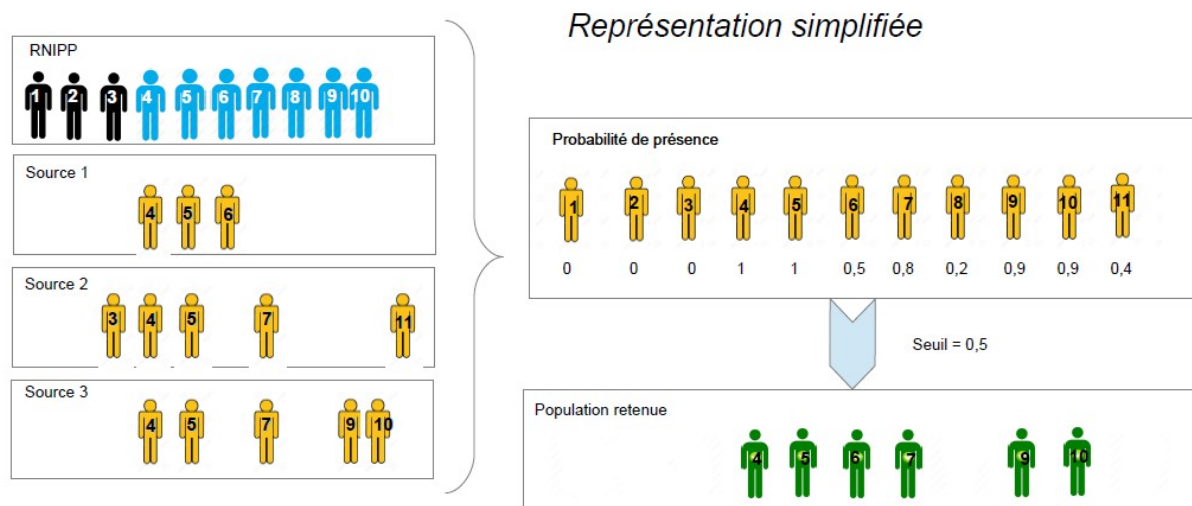
Tableau 1. Nom officiel du registre centralisé de population, administration responsable et année de création, dans 30 pays européens

Pays		Nom du registre	Type de registre *	Autorité responsable	Année de création
Allemagne	DE	Melderegister	RP locaux. RP centralisé dans quelques Länder. Registre fédéral des étrangers	Communes et Ministère de l'intérieur de chacun des Länder	–
Autriche	AT	Zentrale Melderegister (ZMR)	RP centralisé	Ministère fédéral de l'intérieur	2002
Belgique	BE	Registre national des personnes physiques	RP centralisé	Ministère de l'intérieur	1985
Bulgarie	BG	Naselenie Esgraon	RP centralisé	Ministère du développement régional et des travaux publics	1977
Chypre	CY	Archeio Plithismou	RC locaux	Ministère de l'intérieur	–
Danemark	DK	Det Centrale Personregister	PIS	Ministère de l'intérieur	1968
Espagne	ES	Fichero de Coordinación de los Padrones municipales (Padrón Continuo)	RP centralisé	Institut national de statistique (INE)	1996
Estonie	EE	Rahvastikuregister	Central PIS	Ministère des affaires intérieures	2002
Finlande	FI	Väestötietojärjestelmä	Central PIS	Ministère des affaires intérieures	1970
France	FR	–	–	–	–
Grèce	GR	Dimotologio	RC locaux	Ministère de l'intérieur	–
Hongrie	HU	Népszámlálórendszer	RP centralisé	Ministère de l'intérieur	1975
Irlande	IE	–	–	–	–
Islande	IS	Thjóðskrá	RP centralisé	Statistics Iceland	1953
Italie	IT	Anagrafe della popolazione residente	RP locaux et registre centralisé en préparation	Ministère de l'intérieur et municipalités	–
Lettonie	LV	Latvijas Republikas iedzīvotāju reģistrs	RP centralisé	Ministère de l'intérieur	1992
Lituanie	LT	Gyventojų registro tarnyba prie Vidaus reikalų ministerijos	RP centralisé	Ministère de l'intérieur	1992
Luxembourg	LU	Répertoire général des personnes physiques	RP centralisé	Centre informatique de l'État	1979
Malte	MT	Public Register	Registre civique séparé à Malta et Gozo	Ministry for Justice and Home Affairs	–
Norvège	NO	Det sentrale folkeregister	Central PIS	Directorat des taxes au ministère des finances	1964
Pays-Bas	NL	Gemeentelijke Basisadministratie persoonsgegevens	RP locaux reliés électroniquement	Communes et ministère des affaires intérieures	1994
Pologne	PL	CBD PESEL (Powszechny Elektroniczny System Ewidencji Ludności)	RP centralisé	Voivodes et ministère de l'intérieur et de l'administration	1979
Portugal	PT	–	–	–	–
République tchèque	CZ	Centrální registr obyvatelstva	RP centralisé	Ministère de l'intérieur	1980
Roumanie	RO	Registrul permanent de evidență populației	RP centralisé	Ministère de l'intérieur	1996

Annexe 6 : La méthode des signes de présence pour statuer sur la présence d'une personne sur le territoire national

L'indicateur de résidence sur le territoire français représente la probabilité (entre 0 et 1) d'être considéré comme résidant au 1^{er} janvier d'une année donnée sur le territoire français. Elle dépend de sa présence dans la source administrative qui témoigne des droits qui ont été demandés. Ensuite, une synthèse des probabilités de présence dans les différentes sources administratives est réalisée à l'aide de seuils. La méthode a été développée par l'Estonie [28bis] et elle est peu à peu adoptée par d'autres pays. A noter que l'entrée est plus facile à capter que la sortie du territoire dans les fichiers administratifs

Prenons un exemple :



Le RNIPP (Répertoire national d'identification des personnes physiques) donne l'information sur les décès et les naissances. Dans notre exemple, on constate au 1^{er} janvier que les trois premières personnes en noir sont décédées dans l'année qui précède.

Les personnes 4 et 5 sont dans toutes les sources. Les signes de présence sur le territoire sont suffisamment fiables pour acter leur présence.

La personne 6 est dans une seule source assez fiable pour acter la présence sur le territoire.

La personne 7 est dans deux sources. L'information est suffisamment fiable pour acter sa présence sur le territoire.

La personne 8 n'est dans aucune source, mais elle était sur le territoire l'année précédente et il n'y a pas de signe d'un départ, elle a donc une très faible probabilité d'être encore sur le territoire. Ce n'est cependant pas suffisant pour acter sa présence sur le territoire.

Les personnes 9 et 10 sont dans une seule source qui est très fiable.

La personne 11 n'est présente que dans une seule source qui n'est pas assez fiable. Elle n'est pas considérée comme présente sur le territoire.

Annexe 7 : le cadre de nécessité et de proportionnalité de Statistique Canada

