

Appariements de données individuelles : vers une meilleure qualité et plus de transparence

Après avoir encouragé le recours accru aux données de sources administratives et leurs appariements pour établir des statistiques, le Conseil national de l'information statistique (Cnis) constate désormais une forte croissance de ces usages, non seulement pour produire des statistiques, répondre aux besoins dans le cadre de l'élaboration et de l'évaluation des politiques publiques, mais aussi pour alimenter le débat public.

La rencontre organisée le 28 janvier 2022 avait pour but d'informer et d'échanger sur l'état des pratiques françaises et étrangères en matière d'appariement de données individuelles pour construire des statistiques ainsi que sur les projets de la statistique publique. L'Institut national de la statistique et des études économiques (Insee) a d'abord retracé l'évolution de ces pratiques depuis les années 1950, puis a présenté la mise en place du Code statistique non signifiant (CSNS) ainsi que le projet Résil (Répertoire statistique d'individus et de logements), qui visent à rendre les appariements du service statistique public plus simples à réaliser et de meilleure qualité. Cette rencontre a conclu sur l'importance de renforcer l'information du public sur ces méthodes afin d'améliorer la transparence. A la suite des échanges, un groupe de concertation a été mis en place sous l'égide du Cnis pour poursuivre et approfondir la réflexion sur les garanties que le projet Résil doit offrir.

Fin 2020, l'Insee a présenté dans le cadre du Cnis sa volonté d'intensifier l'utilisation des données administratives au sein du service statistique public, à l'image de ce qui se pratique dans les instituts statistiques étrangers. Il s'agit pour l'Insee de renforcer ses outils et ses méthodes afin d'être en mesure de les mobiliser plus rapidement, tout en améliorant la qualité et la mutualisation au sein du service statistique public, en recourant dans certains cas à des appariements de données. En effet,

l'appariement entre elles de données d'origine administrative utilisées à des fins de statistiques, en les combinant le cas échéant avec des données d'enquêtes, permet d'élaborer des statistiques qui répondent mieux à une demande croissante d'information et de connaissance de la part des pouvoirs publics, de la recherche et de la société civile. [L'Autorité de la statistique publique](#) (ASP) soulignait ainsi en septembre 2021 « la place essentielle qu'est amené à prendre ce troisième type de collecte



de l'information » et exprimait le souhait que « soit simplifié le contexte actuel des opérations d'appariement ».

Afin de dresser un panorama des pratiques en matière d'appariements de données individuelles en France comme à l'étranger, et d'en discuter les avancées méthodologiques récentes, le Cnis a organisé une [rencontre](#) d'une journée le 28 janvier 2022. Intitulée : « Appariements de données individuelles : entre richesse statistique et protection de la vie privée », elle visait à partager l'information sur les pratiques et les projets en la matière, puis à organiser le dialogue entre les producteurs, les utilisateurs, mais aussi les représentants des personnes concernées par ces données. L'Insee y a notamment présenté son projet de répertoire statistique d'individus et de logements (Résil) et le service d'appariement de données qu'il offrira au service statistique public pour des usages de production statistique à compter de 2025. Les débats ont confirmé l'intérêt des utilisateurs pour ce dispositif, mais aussi la nécessité de poursuivre la concertation pour recueillir les interrogations qu'il peut susciter auprès du public ou de parties prenantes, au-delà des seuls utilisateurs de données, et en tirer les conséquences, soit en termes de conception du projet, soit en termes de communication et de concertation « au fil de l'eau ».

Un groupe de concertation rassemblant des expertises très variées (libertés fondamentales, protection des données, juridique, éthique...) a donc été mis en place sous l'égide du Cnis en mai 2022 pour questionner le projet Résil sous ces différents aspects et échanger sur les réponses apportées par l'Insee. Dans le [rapport du groupe](#) paru en novembre 2022, treize recommandations sont adressées à l'Insee et plus largement au service statistique public, mais aussi au Cnis, pour adapter à l'avenir ses missions aux enjeux soulevés par Résil et l'élargissement des possibilités d'appariement, et enfin à l'ASP comme autorité de contrôle de l'indépendance du service statistique public et de la bonne application des principes du code des bonnes pratiques de la statistique européenne.

Une longue pratique d'appariements dans la statistique publique

L'Insee et les [Services statistiques ministériels](#) (SSM), qui constituent le [service statistique public](#), disposent d'une longue expérience d'appariements de données sur les individus. Apparier des données relatives aux individus consiste à assembler pour une même personne des données qui la concernent et sont issues de différentes sources : enquêtes différentes sur une même période, mêmes enquêtes sur des périodes différentes, enquêtes et données administratives, ou encore plusieurs sources de données administratives. Dès la fin des années 1950, des appariements de données sont à l'origine de l'enquête Revenus fiscaux ([ERF](#)) – renommée plus tard « enquête Revenus fiscaux et sociaux » ([ERFS](#)). Cette « enquête » rapprochait à l'origine des données du recensement de la population et des données fiscales pour un échantillon de personnes recensées, afin de calculer le niveau de vie de la population et d'établir sa distribution. Elle utilisait donc uniquement des données déjà collectées provenant d'une collecte terrain (le recensement) et des données d'origine administrative (fiscales). L'échantillon démographique permanent ([EDP](#)) mis en place en 1968 juxtapose, quant à lui, à l'origine, des données issues des recensements de la population et de l'état civil. L'EDP s'est peu à peu enrichi par appariement avec de nouvelles sources, et a été récemment apparié par le service statistique ministériel chargé de la santé et des solidarités ([Drees](#)) avec des données issues du Système national des données de santé pour construire l'[EDP santé](#). Les appariements permettent ainsi de produire une information enrichie, à un coût raisonnable, qu'il serait impossible de recueillir par une collecte directe auprès des personnes, pour des questions de charge pour l'enquête et de budget.

Au cours des années 1970, différents panels sont mis en place pour suivre les individus dans le temps, reconstituer des trajectoires, en particulier les parcours scolaires dès 1973, puis dans l'emploi avec le panel DADS

« déclarations annuelles de données sociales » à partir de 1976 qui relie pour un large échantillon des données sur les salaires aux périodes d'emploi déclarées par les employeurs, d'abord limité au secteur privé, puis étendu au public ([panel « tous salariés »](#)) et aux non-salariés ([panel « tous actifs »](#)). En 1988, le SSM Santé a constitué l'[échantillon interrégimes de retraités](#) (EIR), qui apparie, pour un échantillon de retraités, des données sur leurs montants de retraite dans les différents régimes afin de reconstituer leurs montants de retraite globale. Plus récemment, la reconstitution des trajectoires individuelles des bénéficiaires de minima sociaux a permis de mieux comprendre l'entrée dans la pauvreté et d'évaluer les dispositifs de protection sociale. Les [panels](#) du SSM de l'éducation ([Depp](#)) ont été mobilisés pour évaluer les réformes éducatives. Avec le SSM en charge du travail ([Dares](#)), les SSM en charge respectivement de l'éducation ([Depp](#)) et de l'enseignement supérieur ([Sies](#)) ont récemment mis au point respectivement les dispositifs [Inserjeunes](#) et [InserSup](#) (ce dernier est encore à l'état de projet) pour évaluer les effets des mesures d'aide à l'emploi sur l'insertion des jeunes passés par différentes filières d'éducation.

Certaines données sont de meilleure qualité lorsqu'elles proviennent de données administratives que lorsqu'elles sont collectées par voie d'enquête déclarative : c'est notamment le cas du revenu des ménages qui est parfois lourd à restituer pour un enquêteur, par exemple en cas de pluriactivité, ou lorsqu'on cherche à reconstituer les revenus pour des années passées. Depuis les années 2000 et le traitement par la statistique publique de l'ensemble des déclarations socio-fiscales, les enquêtes auprès des ménages sont systématiquement enrichies par des données individuelles couvrant les différents types de revenu, ce qui réduit d'autant la charge d'enquête pour la personne enquêtée. Inversement, certains concepts ne peuvent pas être mesurés directement par des données administratives, c'est le cas notamment du chômage au sens du Bureau international du travail (BIT) que seule une enquête peut mesurer. Plus généralement,

répondre au mieux aux besoins des utilisateurs, notamment à leur demande croissante de données plus pointues pour cerner la complexité des situations, nécessite souvent de coupler les différentes sources d'information : des enquêtes auprès des personnes, des données administratives variées, en s'appuyant sur leurs qualités et apports respectifs.

Les bases exhaustives permettent également d'obtenir des informations à des niveaux géographiques fins, bien en deçà de l'échelle régionale à laquelle seuls les plus gros échantillons d'enquête peuvent descendre. Ainsi, par exemple, le fichier localisé social et fiscal (*Filosofi*) fournit des informations sur les revenus et la pauvreté au niveau local – communes, quartiers, mailles d'un kilomètre carré – par le rapprochement de fichiers fiscaux et de prestations sociales exhaustifs, et via l'imputation de certains revenus.

Certains appariements permettent d'appréhender des phénomènes complexes et de couvrir des périmètres qu'une seule source ne peut aborder : les bases « [tous salariés](#) » ou « [tous actifs](#) » offrent ainsi une vision complète des situations d'emploi des personnes à une date donnée, ou sur une année. En appariant des sources de données thématiques différentes, il est possible d'étudier des phénomènes situés à la croisée de plusieurs domaines : par exemple la mortalité ou la fécondité en fonction du niveau de diplômes, de la catégorie sociale, ou du niveau de vie, et étudier ainsi les inégalités sociales de santé. Enfin, les appariements peuvent permettre de reconstituer des trajectoires (de formation, d'emploi...) beaucoup plus facilement que les enquêtes, pour des raisons de mémoire et de charge de réponse pour la personne enquêtée.

Sur le plan méthodologique, les appariements permettent également de mieux comprendre les forces et faiblesse d'une source, en analysant les écarts entre des sources concernant les mêmes concepts ou des concepts voisins. Ainsi, l'appariement de

l'enquête Emploi¹ de l'Insee, qui mesure le chômage au sens du BIT, et du fichier historique des demandeurs d'emploi de Pôle emploi a permis de mieux comprendre le concept du chômage et les écarts qui existent entre ses [différentes définitions](#).

Le cadre juridique encadrant la collecte et les appariements de données individuelles s'est progressivement enrichi au fil des années. La loi de 1951 sur l'obligation, la coordination et le secret en matière de statistique comporte ainsi des articles ajoutés entre 1986 et 2016 encadrant la mobilisation et la protection de données administratives ou provenant d'opérateurs privés. La loi relative à l'informatique, aux fichiers et aux libertés de 1978 et ses décrets d'application ont encadré les modalités de traitement informatiques de ces données, dont les appariements reposant sur le NIR (numéro d'inscription au répertoire, plus communément appelé numéro de sécurité sociale), et a été profondément modifiée pour tenir compte du répertoire général sur la protection des données (RGPD). La loi encadrait des possibilités, mais fixait également des garanties, des obligations, des « garde-fous ».

Parallèlement aux évolutions de ce cadre juridique, les statisticiens, en France comme à l'étranger, ont progressivement formalisé et enrichi leur réflexion sur les aspects déontologiques de ces traitements et sur la nécessité de disposer d'un « mandat social » en plus du « mandat juridique ».

Trois exemples

• L'échantillon national interrégimes d'allocataires de compléments de revenus d'activité et de minima sociaux (Eniacrams)

Comme d'autres panels, l'*Eniacrams* est construit à partir d'appariements. Ce panel vise depuis 2001 à suivre les bénéficiaires des minima sociaux actifs jusqu'à l'âge de la retraite. Il n'est pas exhaustif, le principe de minimisation des données ayant conduit à opter pour un échantillon. Il comprend fin

2020 environ 170 000 allocataires ou conjoints d'allocataires du RSA (revenu de solidarité active), 90 000 bénéficiaires de l'AAH (allocation aux adultes handicapés), 25 000 allocataires de l'ASS (allocation de solidarité spécifique) et 410 000 allocataires ou conjoints d'allocataires de la prime d'activité, ce qui permet de mener des analyses très fines et un suivi des différentes trajectoires d'allocataires, qui peuvent sortir, puis revenir dans la sphère des minima sociaux.

Construit à partir d'appariements de données relatives aux prestations sociales provenant des caisses verseuses (Caisse nationale des allocations familiales – Cnaf, Mutualité sociale agricole – MSA, et Pôle emploi) et de données sur les individus et leurs emplois provenant des panels « tous salariés » et « tous actifs » de l'Insee, l'*Eniacrams* permet d'observer l'ensemble du parcours de l'emploi, pendant et après la sortie des minima sociaux, ainsi que de nombreuses caractéristiques sociodémographiques. De plus, l'échantillon basé sur le jour de naissance partage une caractéristique commune avec de nombreux panels, notamment l'échantillon démographique permanent (EDP), desquels il peut être rapproché.

• Appariement entre l'enquête Emploi et le fichier historique de Pôle emploi pour comprendre les différences entre nombres de chômeurs et de demandeurs d'emploi

L'idée d'apparier l'enquête Emploi et le fichier historique de Pôle emploi est née du constat que le nombre de chômeurs au sens du BIT, mesuré chaque année dans l'enquête Emploi, s'écartait du nombre de demandeurs d'emploi inscrits à Pôle emploi et classés en catégorie A (sans emploi), deux mesures dont les définitions sont *a priori* proches. Situés à un niveau équivalent vers 2008 ces deux indicateurs ont progressivement divergé jusqu'à aboutir à un écart de plus d'un million de personnes. L'attention a été portée sur la période 2013-2017, celle

¹ Enquête sur l'emploi, le chômage et l'inactivité.



qui a vu cet écart se creuser, avec près de 200 000 personnes supplémentaires inscrites dans la catégorie A de Pôle emploi et environ 200 000 chômeurs en moins au sens du BIT. Pour réaliser cet appariement entre l'enquête Emploi et le fichier historique de Pôle emploi, des équipes de l'Insee, de la Dares et de Pôle emploi ont œuvré ensemble.

Un chômeur au sens du BIT est une personne sans emploi qui recherche activement un travail et qui est disponible dans les deux prochaines semaines. Cette définition du chômage permet de réaliser des comparaisons internationales. Ainsi, des questions très précises sont posées dans l'enquête Emploi pour distinguer les chômeurs en fonction de cette définition. Inversement, le demandeur d'emploi en fin de mois inscrit à Pôle emploi en catégorie A est sans emploi et est tenu d'en rechercher un. Néanmoins, aucun critère ou question ne permet de mesurer le caractère actif de leur recherche d'emploi. Ainsi, une partie des chômeurs au sens du BIT ne sont pas inscrits à Pôle emploi. Dans l'autre sens, une partie des chômeurs inscrits à Pôle emploi ne sont pas considérés comme des chômeurs dans la définition du BIT.

Bien que l'appariement n'ait pas été simple à réaliser (pas de patronyme dans l'enquête Emploi qui échantillonne des logements, adresse de contact de Pôle emploi parfois différente du lieu d'habitation...), les résultats ont bien retracé les évolutions globales des années 2013-2017 et permis d'observer des évolutions et de mieux comprendre la divergence touchant les deux indicateurs de chômage. Les données rapprochées permettent *in fine* de relativiser la proximité entre deux concepts qui semblaient similaires au départ. Il sera utile de renouveler ce rapprochement à intervalles réguliers si l'on veut notamment observer les divergences postérieures à la crise du covid-19.

• Mieux connaître l'insertion des jeunes : le système d'information InserJeunes

InserJeunes est un tout nouveau système d'information situé à la croisée des sphères

de l'éducation et de l'emploi. Sous co-maîtrise d'ouvrage de la Depp et de la Dares, ce système d'information repose sur une vingtaine d'appariement de bases de données individuelles. InserJeunes permet en premier lieu de mesurer le taux d'emploi des jeunes qui sortent de formations professionnelles de niveau CAP à BTS, par voie scolaire ou par voie d'apprentissage. Ce dispositif permet de qualifier les caractéristiques de cette insertion à travers différents indicateurs tels que le type de contrat, les secteurs dans lesquels les jeunes s'insèrent, l'adéquation entre la formation et l'emploi, ou encore le salaire.

La mesure du taux d'insertion se base sur des indicateurs déclinés par formations au niveau national et régional, jusqu'au niveau des établissements. Utile aux rectorats pour produire des éléments de cadrage national et régional, il l'est aussi pour mettre à disposition des familles et des jeunes ces données d'insertion, pour aider les élèves à s'orienter en fin de troisième ou en terminale. En outre, cette fusion permet aussi de calculer deux autres indicateurs : le taux de poursuite d'études et le taux d'interruption en cours de formation.

Mise en œuvre des appariements, et projets

La mise en œuvre concrète des appariements ne s'avère pas toujours facile. Il peut être difficile de vérifier que les individus en question sont bien exactement les mêmes dans les différentes sources rapprochées. Le cas le plus simple est celui où les sources à appairer disposent l'une et l'autre d'un code en commun qui identifie sans ambiguïté les personnes. Il peut notamment s'agir d'un identifiant certifié, comme le numéro d'inscription au répertoire (NIR) géré par l'Insee, ou encore l'identifiant national des étudiants et des élèves (INE) géré par le service statistique en charge de l'éducation, la Depp. L'identifiant (ou les traits d'identité) utilisé pour l'appariement peut être supprimé du fichier de résultats une fois l'opération réalisée.

Pratiqué depuis plusieurs années, et encouragé par la CNIL, le recours à des identifiants

non significatifs dédiés à de tels usages se développe pour appairer des données. C'est le cas du « code statistique non significatif » (CSNS), obtenu par une transformation non réversible du NIR.

La loi de 2016 « pour une République numérique » encadre les appariements qui mobilisent le NIR à des fins de statistiques ou de recherche. La procédure prévue pour la recherche partage des similitudes avec celle qui est attribuée à la statistique publique, en s'appuyant sur des « tiers de confiance ». Par ailleurs, pour leur sécurité, les données doivent rester confinées au sein d'une « bulle sécurisée ». Le Centre d'accès sécurisé aux données (CASD) est un tiers de confiance mettant à disposition un équipement conçu pour permettre aux utilisateurs de travailler à distance sur des données fournies par des producteurs dans des conditions de sécurité élevées. Les chercheurs peuvent y travailler sur des données en toute indépendance vis-à-vis des organismes détenteurs des données. Ils peuvent y mener des opérations conjointes avec le service statistique public. Enfin, les chercheurs peuvent travailler pour le compte d'organismes tels que les SSM sur la base de crédits d'études et de recherche.

• Le code statistique non significatif (CSNS) : une clé de chiffrement pérenne sur 10 ans

Le code statistique non significatif (CSNS) permet à l'ensemble du service statistique public de faciliter le processus juridique des appariements de données individuelles non sensibles mobilisées par le service statistique public dans le cadre de ses missions, réalisés pour des finalités statistiques exclusivement et au sein du seul service statistique public. Il permet de valoriser ainsi davantage les sources de données disponibles, dans un cadre juridique très strict, et renforce la sécurité des données en évitant de recourir au NIR. Le CSNS se fonde sur une composante technique, avec la production d'une clé d'appariement, ainsi que sur un aspect organisationnel. La production d'une clé d'appariement permet de lier un code à un individu. Le calcul de cette clé doit produire le même code, quelle que soit la source où

l'on retrouve cet individu. Cette clé de chiffrement peut donc être utilisée pour apparier les informations de fichiers différents. Cette clé est non signifiante : elle ne doit comporter aucune information sur la personne en question. Par ailleurs, cette clé demeure pérenne sur dix ans. Au bout de ces dix années, une procédure de renouvellement est menée. En cas de faille de sécurité, cette période de dix ans peut également être raccourcie. La composante organisationnelle du CSNS consiste à proposer un protocole d'usage qui comprend une phase de conventionnement entre les différentes parties. En outre, une application informatique dédiée permet aux utilisateurs de déposer des fichiers pour lesquels ils souhaitent obtenir un CSNS. Ils trouveront en retour sur cette plateforme les résultats du calcul du CSNS produits par l'Insee.

Le calcul du CSNS s'effectue à partir du NIR, potentiellement obtenu après une étape d'identification statistique sur la base de variables d'état civil. Une opération de hachage et de chiffrement du NIR est ensuite effectuée. Le hachage permet l'irréversibilité de l'opération, il n'est donc pas possible de retrouver un NIR une fois qu'il a été haché. Le service de l'Insee délivrant le CSNS ne conserve aucune donnée, dès que le traitement est validé par le demandeur.

• Le répertoire statistique d'individus et de logements (Résil) : un projet pour 2025

L'Insee a engagé un projet de répertoire statistique d'individus et de logements (Résil) dont l'horizon est 2025 pour sécuriser, fiabiliser et faciliter les appariements, dans le respect des principes juridiques et déontologiques qui régissent la statistique publique. Le répertoire sera construit grâce à l'appariement de plusieurs sources administratives, pour assurer la meilleure prise en compte possible de la population, sans dépendre exclusivement d'une source en particulier pour limiter les chocs liés à un changement

dans les sources², tout en restant conforme aux principes de nécessité et de proportionnalité dans le choix des sources. Progresser dans l'utilisation des sources administratives et des appariements implique la mobilisation d'outillages techniques et méthodologiques.

Un répertoire constitue une liste exhaustive d'objets – en l'occurrence des individus et des logements –, à la fois très large, car il contient potentiellement toutes les observations d'un champ, mais aussi très étroit, car il ne renferme que peu de variables. Les variables insérées dans un répertoire doivent permettre d'identifier sans ambiguïté les unités qu'il contient, notamment pour éviter des doublons et permettre de les relier avec d'autres éléments du système d'information. Le répertoire joue ainsi le rôle d'une colonne vertébrale du système d'information.

Résil répond à plusieurs finalités exclusivement statistiques :

- Il permettra d'offrir, au service statistique public uniquement, un service d'appariement de données³ 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.

Lors de sa mise en œuvre, prévue en 2025, Résil sera constitué de deux répertoires statistiques exhaustifs. Le premier répertoriera les individus résidant sur le territoire national et l'autre les logements. Tous deux seront mis à jour en continu avec les naissances, les décès et différentes sources administratives, en ne retenant que des données d'identification. En effet, les autres données seront transmises aux systèmes d'information métier qui les intégreront dans leurs processus de production. Ces deux répertoires constitueront donc des entités vivantes avec des mises à jour régulières pour vérifier que les personnes répertoriées sont bien toujours résidentes sur le territoire. Ils permettront la production de « photos » annuelles donnant la situation au 1^{er} janvier.

Une concertation élargie sous l'égide du Cnis

Les débats qui ont eu lieu le 28 janvier 2022 dans la table ronde sur la transparence et l'information du public, dont les intervenants découvraient la richesse des appariements ainsi que le projet Résil, ont souligné que cette réflexion sur Résil et les appariements de la statistique publique devait se poursuivre sur une plus longue période en associant un panel d'expertise adapté aux enjeux de protection des libertés fondamentales que soulèvent ces opérations de traitement de données personnelles.

L'Insee a donc choisi de poursuivre la concertation avec un groupe placé sous l'égide du Cnis, réunissant des expertises variées (protection des libertés fondamentales, protection des données personnelles sur les plans juridique et informatique, transformation numérique, éthique, recherche en sciences sociales, etc.) situées pour la plupart en dehors des interlocuteurs habituels de la statistique publique. Celui-ci s'est réuni de mai à septembre 2022 sous la présidence de Jean-Marie Delarue pour dresser la liste des interrogations suscitées par le projet et

² Comme celui lié à la suppression de la taxe d'habitation en 2023.

³ Ces appariements peuvent porter sur des données administratives entre elles, ou sur des données administratives et des enquêtes statistiques.

échanger sur les réponses apportées par la maîtrise d'ouvrage (Insee). Ce groupe a mené des auditions auprès d'un juriste, de membres d'associations ou de *think tanks*, de l'Agence nationale de la sécurité des systèmes d'information, d'un chercheur utilisateur des données d'Eurostat, d'un responsable des réflexions éthiques dans un institut statistique étranger, pour multiplier les angles de vue, réduire les zones d'ombre, poser toutes les questions relatives aux finalités, au cadre juridique, déontologique et technique, aux risques aussi qu'un tel projet fait courir. Le groupe a examiné la liste des sources qui pourraient alimenter Résil (RNIPP⁴, sources fiscales et sociales...) en interrogeant systématiquement les principes de nécessité et de proportionnalité. Il a enfin recommandé des modalités de concertation et de communication aux différentes étapes du projet jusqu'en 2025 et au-delà, lorsque Résil sera pleinement opérationnel.

Si le groupe de concertation a été « très attentif au risque accru que faisait porter sur la vie privée la multiplication d'appariements », il

considère dans son [rapport](#)⁵ que le projet Résil est légitime et proportionné à ses finalités statistiques. Il accorde sa confiance à l'Insee sur les plans technique et déontologique pour respecter les bonnes pratiques en matière de répertoire et d'appariement et « prévenir tout mésusage de l'instrument en écartant toute possibilité que soient détournées ou dérobées des données ».

Le cadrage juridique envisagé par l'Insee, à savoir un décret en Conseil d'Etat fondant le traitement, complété par un texte pris après avis de la Cnil et fixant la liste des informations mobilisées pour alimenter les répertoires, a paru adapté au groupe, qui recommande que la finalité purement statistique de Résil soit rappelée et la notion d'appariement précisée dans le décret.

Le groupe préconise notamment de développer une communication efficace à destination d'un vaste public sur l'utilisation de données administratives et les appariements, et notamment d'en publier la liste

avec leurs principales caractéristiques de façon à rendre compte du respect des principes de nécessité et de proportionnalité.

Il recommande enfin de confier au Cnis la mission explicite de donner 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 libertés des personnes.

Une fois l'avis de la Cnil rendu, le dossier sera déposé au Conseil d'Etat et les échanges se poursuivront pour préparer la mise en œuvre des préconisations du groupe et faire que ce projet se construise dans la confiance. Le Cnis poursuivra l'information et la concertation sur ce projet, afin de veiller au respect de la vie privée dans l'utilisation statistique de données personnelles ainsi qu'à la sécurité du dispositif. ■ ■ ■

Françoise Dupont,
François Guillaumat-Tailliet et
Cristina D'Alessandro

⁴ Répertoire national d'identification des personnes physiques.

⁵ Le projet de répertoire statistique d'individus et de logements « Résil » - Rapport du groupe de concertation du Cnis, [Rapport du Cnis](#) n° 161, novembre 2022.