

Rencontre du Cnis sur les appariements de la statistique publique

Bercy, 28 mai 2025

Synthèse

Ouverte par Jean-Luc Tavernier, directeur général de l'Insee, et trois années après le colloque de janvier 2022 sur le même sujet, cette rencontre a répondu à une recommandation du groupe de concertation qui, sous l'égide du Cnis, avait précisé les outils envisagés par l'Insee pour développer les appariements, gagner en qualité, en harmonisation et en sécurité par rapport à ce qui était pratiqué jusqu'alors. Environ 170 personnes y ont participé.

Les quatre temps de la rencontre

Tout d'abord, Christel Colin et Corinne Prost, respectivement directrice des statistiques démographiques et sociales et directrice de la méthodologie et de la coordination statistique à l'Insee, ont expliqué le rôle joué par la mise en place du code statistique non signifiant (CSNS) dans le développement des appariements au sein du service statistique public (SSP). C'est ce développement qui conduit à une réflexion sur leurs usages, pour notamment éviter de multiplier le nombre de variables dans une même base de données, source de complexité des analyses. Ils permettent d'éclairer certains phénomènes socio-économiques, de compléter des champs d'analyses ou d'améliorer des processus de production, sous réserve de la pertinence et de la qualité des données administratives ou privées appariées. N'ayant pas vocation à répondre à toutes les questions, un appariement doit être ciblé sur l'éclairage souhaité, en respectant le principe de minimisation (collecter uniquement les données nécessaires pour répondre à la finalité).

Pour nourrir la réflexion sur les enjeux liés aux appariements, trois exemples ont été présentés et discutés par un utilisateur externe au service statistique public :

- L'intérêt des appariements et aussi leur complexité ont ainsi été illustrés par Laurent Lequien, sous-directeur des statistiques des transports, au service statistique des ministères de l'Aménagement du territoire et de la transition écologique. L'appariement du système d'immatriculation des véhicules et les résultats des contrôles techniques des véhicules, avec différentes sources de caractérisation de ces véhicules, puis, via le CSNS, avec le fichier démographique sur les logements et les individus, ou encore des données sur les aides à l'acquisition de véhicule peu polluant renseigne finement sur le profil des utilisateurs des véhicules routiers. Ce type de travaux, de longue haleine, permet en outre de mieux cerner les enjeux redistributifs de la transition écologique. Clément Malgouyres, chercheur au Crest et à l'Institut des politiques publiques, a témoigné du saut qualitatif qu'ils ont permis dans la compréhension du parc automobile et des politiques publiques qui affectent ses dynamiques. Il a insisté sur les gains que représente pour les chercheurs la mise à disposition des fichiers enrichis, et plus largement de bases de données administratives non appariées comme celles de l'Agence de services et de paiement.

- Pour illustrer l'apport des appariements méthodologiques, Yves Jauneau, chef de la division Synthèse et conjoncture du marché du travail à l'Insee, a expliqué en quoi ils ont éclairé la divergence récente entre l'enquête Emploi et les estimations d'emploi. Ils ont permis, non seulement de quantifier les taux de sous-déclaration de l'emploi de certaines catégories de personnes en emploi, mais aussi d'étudier l'impact du mode de collecte sur ces taux. Magali Dauvin, économiste à l'Observatoire

français des conjonctures économique, a salué et souligné tout l'intérêt de tels travaux pour l'amélioration des prévisions d'emploi.

- Autre exemple, le dispositif InserSup a été présenté. Pierrette Schuhl, cheffe du service statistique ministériel de l'Enseignement supérieur et de la Recherche a rappelé qu'il répondait à une demande interministérielle de 2021 en application de la loi de programmation de la recherche de 2020. Résultat d'appariements de fichiers Etudiants avec ceux du marché du travail, il fournit des taux d'emploi salarié en débouché de la plupart des diplômés et constitue ainsi une aide à l'orientation des jeunes. Il sera progressivement enrichi (taux d'emploi non salarié d'ici fin 2025, couverture plus large des diplômés...). Il nécessite toutefois d'être complété par des enquêtes, s'agissant des emplois à l'étranger par exemple. Nagui Bechichi, créateur du site Suptracker, site qui valorise ces données, a souligné l'importance de ce dispositif qui relie l'enseignement supérieur et l'insertion professionnelles, tant pour l'orientation que pour l'évaluation des politiques d'offre ou de réforme.

A la suite de ces exemples, Olivier Lefebvre, maître d'ouvrage du projet Résil de l'Insee, a précisé ce que contient ce Répertoire statistique des individus et des logements en insistant sur l'absence du NIR et de toute donnée « métier » (revenu, état matrimonial, PCS...), et le service d'appariement qu'il met à disposition du SSP. L'auditoire a pu mesurer le chemin parcouru depuis la présentation du projet en 2022, dans le respect des dispositions issues du groupe de concertation, et par un encadrement juridique transcrit dans le décret du 7 janvier 2024. Le décret n'autorise pas l'accès direct du service d'enrichissement aux chercheurs ; ils pourront néanmoins en tirer bénéfice via des partenariats avec des entités du SSP, s'ils débouchent sur la production de statistiques publiques, comme cela existe pour co-construire des enquêtes ; ils bénéficieront par ailleurs de l'extension de l'offre de fichiers statistiques.

Pour poursuivre les échanges de la journée, Bertrand du Marais, président du Cnis, a animé la dernière séance dédiée aux apports d'un cadre de référence pour la réalisation d'appariements. Ce cadre, présenté par Corinne Prost, est voulu par le SSP afin que les appariements respectent les principes de nécessité (à l'image des avis d'opportunité pour les enquêtes statistiques), de proportionnalité et de minimisation des données utilisées ainsi que le principe de transparence de l'ensemble des appariements réalisés. Corinne Prost a insisté sur les enjeux de coordination des pratiques au sein du SSP, et de la mise en visibilité des pratiques.- Anthony Guérout, représentant l'association des maires de France au Cnis, a évoqué les besoins des collectivités locales, productrices mais aussi grosses consommatrices de données fiables, précises et à jour pour élaborer, mettre en œuvre et suivre localement l'impact des politiques publiques. Outre les enjeux de sécurité informatique, il souligne l'indispensable protection des libertés individuelles ; il note à ce propos que lors du recensement de la population des citoyens se montrent surpris, plutôt qu'inquiets, de la collecte de données individuelles les concernant, pensant que la mairie en dispose déjà. Au nom des services statistiques ministériels, Christelle Minodier a souligné la grande utilité d'avoir un cadre commun pour une meilleure harmonisation des pratiques et pour organiser la diffusion aux chercheurs au Centre d'accès sécurisé aux données. Elle a mentionné l'usage précieux de Résil pour disposer d'une population de référence lors de calculs de taux (de prévalence, de bénéficiaires, de non-recours, etc.). François Clanché, directeur de l'Ined, a souhaité que le cadre de référence pour la statistique publique permettant le développement des appariements ouvre des portes pour la recherche, encore faut-il donner suffisamment de visibilité aux nouvelles productions. Il a porté la demande des chercheurs de bien clarifier, sans trop tarder, ni les limiter, quelles seraient leurs possibilités d'accès aux appariements. Dominique Meurs, professeure à l'Université Paris Nanterre, a renforcé la demande en témoignant de l'apport pour les travaux de recherche sur données administratives que constituent aux Pays-Bas les facilités d'appariements, avec des dispositifs d'accès aux données souples, dans des délais courts particulièrement adaptés pour les jeunes docteurs.

Des échanges avec les participants tout au long de la journée, il est ressorti un vif intérêt pour le potentiel qu'offrent le service Résil et plus largement les appariements réalisés par la statistique publique, que ce soit à des fins de connaissance, d'allègement de la charge d'enquêtes ou pour des travaux méthodologiques.

TSVP

S'ils permettent d'alléger la charge d'enquêtes, en limitant le nombre de questions – cas d'usage préconisé par le code de bonnes pratiques de la statistique européenne – ils soulèvent des questions éthiques que justifie le besoin de veiller au respect des principes de transparence, de proportionnalité et de minimisation. L'apport du cadre de référence pour les appariements de la statistique publique est donc reconnu. S'agissant de la transparence, le Cnis y contribuera en affichant sur son site la liste des appariements, qui seront détaillés sur les sites du SSP, à l'instar de l'Insee qui publiera les appariements réalisés avec Résil. S'agissant de la minimisation, il s'agit de faire preuve de discernement dans le choix des variables retenues lors de l'enrichissement d'un fichier par appariement avec d'autres sources, car « toute donnée n'a pas de valeur » et le coût d'analyse est élevé. Ce principe fait cependant craindre aux chercheurs d'être limités dans leurs travaux, ne pouvant pas forcément identifier ex ante les variables les plus pertinentes pour leurs analyses. Les chercheurs sont par ailleurs intéressés à ce que la richesse des fichiers statistiques issus d'appariements ne conduise pas à réduire la mise à disposition au centre Progedo-Quetelet (fichiers de production pour la recherche – FPR) au profit uniquement du CASD dont l'accès peut être onéreux. Pour préserver le secret statistique des FPR, une solution existe qui est de mettre à disposition des échantillons extraits des fichiers exhaustifs, voie déjà empruntée. La conférence a rappelé à cette occasion que garantir la sécurité informatique des données représentait un coût, contrepartie du maintien de la confiance du public dans la statistique publique.

Le débat a également porté sur la façon dont le principe de proportionnalité, que les producteurs sont engagés à respecter comme les y oblige le RGPD, peut être apprécié lors d'un examen au Cnis, d'une part pour un projet donné et d'autre part au vu d'un ensemble d'appariements.

La journée a également rendu compte des limites des appariements, c'est pourquoi ils allègent plus souvent le questionnaire d'une enquête qu'ils ne s'y substituent. L'opportunité qu'ils offrent pour renforcer la comparabilité des statistiques européennes, dans la mesure où ils permettent de mieux cerner les populations couvertes par une source de référence lorsqu'elle est appariée avec une base plus complète, reste à étudier.

En conclusion de la rencontre, Bertrand du Marais, président du Cnis, a proposé que le Cnis fasse un bilan de la mise en œuvre du cadre de référence pour les appariements de la statistique publique dans deux ans, et qu'il réfléchisse à une proposition de cadre de référence pour la réalisation d'appariement par les chercheurs eux-mêmes.