

Cadre de référence pour les appariements

Projet présenté au bureau du CNIS (12 mars 2025)

Préambule

Les sources de données pour la production de statistiques et d'études sont plurielles. Au-delà des enquêtes statistiques, le statisticien public dispose également de fichiers administratifs qui, dans certaines conditions, peuvent devenir des intrants de production de statistiques. Ces fichiers sont de plus en plus nombreux, plus « normalisés », documentés, donc plus facilement mobilisables à cette fin.

Les appariements permettent de créer des statistiques mixtes en combinant par exemple des données d'enquêtes et des données administratives, ou en combinant des données administratives issues de sources différentes. Ces opérations permettent d'accroître l'information disponible en mettant en relation des informations variées qui se trouvent dans des univers différents, et, par ailleurs, l'appariement permet de compléter utilement la collecte par voie d'enquêtes sans accroître la charge auprès des répondants. Des exemples d'usages sont donnés dans le billet de blog Insee [« Les appariements de données de la statistique publique : des analyses enrichies, un cadre juridique protecteur »](#).

Le développement d'outils ou de services (Code statistique non signifiant, service d'appariement développé dans le cadre du programme Résil¹ de l'Insee), et les capacités techniques croissantes rendent les appariements plus simples à réaliser d'un point de vue technique, y compris sur des volumes importants de données. Or, l'appariement n'est pas neutre en ce sens qu'il crée de nouvelles données, cette facilité confère des responsabilités étendues aux destinataires de ces données, notamment sur le plan déontologique.

La demande de données et d'analyses permettant de décrire finement des populations, d'étudier des trajectoires, d'évaluer les politiques publiques va croissant. Et les appariements jouent un rôle central pour la réponse à ces attentes. En témoignent notamment les avis du moyen terme 2024-2028 du Cnis, par exemple **[DQS. 3]** : Observer les inégalités sociales de santé, **[EQRT 1]** et **[SERPU 3]** : suivre les parcours des jeunes dans le système éducatif jusqu'à leur transition vers le monde professionnel, **[SERPU 4]** : Suivre les trajectoires des jeunes en prise avec la justice.

Le cadre juridique a également évolué pour prendre en compte ces nouvelles opportunités et ces nouveaux besoins, tout en respectant les grands principes de protection des données. Il confère des responsabilités explicites aux acteurs des appariements.

Ce cadre juridique est indispensable, mais pas suffisant : il convient de le compléter par un cadre déontologique permettant à la communauté de la statistique publique de rappeler ses différentes valeurs et la manière dont elles s'appliquent, dans ce contexte marqué à la fois par une demande accrue d'informations, des opportunités croissantes en termes de données disponibles et de manière de les traiter, de vigilance sur la protection des données. Que fait-on et pourquoi, jusqu'où peut-on aller, à quelles conditions, en matière d'appariements ? Au-delà des engagements pris, c'est un enjeu fort de transparence, qui répond à une attente croissante de la société et de ses représentants et qui permet de renforcer la confiance.

1 Répertoire statistique des individus et des logements.

C'est l'objectif du présent cadre, qui répond à **l'avis du moyen terme du Cnis n°10**², ainsi qu'aux recommandations formulées par la Cnil lors de l'examen du programme Résil. Ce cadre s'applique aux appariements réalisés à des fins de statistique publique. Le cas des appariements à des fins exclusives de recherche doit faire l'objet de travaux spécifiques.

Un bilan de sa mise en œuvre sera réalisé au bout de deux ans, et ses conclusions permettront de procéder aux ajustements qui seront jugés nécessaires. Le cadre s'applique à tous les appariements, mais les modalités d'appui du Cnis ne concernent que les nouveaux appariements.

1. Les engagements du SSP pour la pratique des appariements

Les appariements sont réalisés dans le respect des principes suivants :

- Le principe de **nécessité** vise à s'assurer qu'il y a un bénéfice pour la société (l'aide à la décision publique, l'alimentation du débat public et l'information des différents acteurs) associé au traitement envisagé des données.

- Les principes de **proportionnalité et de minimisation** visent à recueillir, lors des traitements, uniquement les données nécessaires à l'analyse en tenant compte de la nature de certaines données et en veillant à ce qu'il n'existe aucune autre approche moins invasive pour atteindre le même objectif.

Enfin, le RGPD³ impose à tout responsable de traitement des obligations qui s'articulent autour de 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. Ces règles relatives à la transparence sur le traitement des données permettent aux citoyens d'avoir des garanties sur le bon usage des données et de rendre compte du respect des principes de nécessité et de proportionnalité.

La finalité première des appariements réalisés par le SSP est la production de statistiques publiques ou d'analyses fondées sur ces statistiques. Comme toutes les données traitées par les services statistiques publics, les données résultant de ces appariements font l'objet de mesures de protection informatique adaptées et sont soumises aux règles de confidentialité en matière de diffusion (secret statistique). Les fichiers produits par appariement à des fins de statistique publique peuvent être mis à disposition des chercheurs, dans les conditions en vigueur pour l'accès aux fichiers détail produits par la statistique publique.

Les appariements sont proportionnés aux finalités recherchées ; cette proportionnalité doit notamment s'apprécier en matière de millésimes mobilisés, de prise en compte totale ou partielle du contenu des sources, aussi bien en termes d'informations (sélection des seules données essentielles au traitement) qu'en termes de population utilisée (recours ou non à des échantillons selon la granularité de la diffusion souhaitée).

2 *Avis de moyen terme n° 10 du Cnis : Appariements de données individuelles*

Encouragé de longue date par le Cnis, le recours aux informations issues de sources administratives et aux appariements de données individuelles constitue aujourd'hui un moyen de développer et d'enrichir le système d'information statistique avec de multiples bénéfices pour la connaissance (croisement de thématiques, suivi longitudinal, précision augmentée...).

Le Conseil invite à harmoniser et standardiser les méthodes d'appariements dans un objectif de qualité. Dans un contexte de facilité technique croissante des appariements de sources exhaustives, qui peuvent présenter des risques pour la sécurisation des données personnelles, il recommande, de définir et de mettre en place un cadre de référence et une déontologie adaptés à ces traitements de données, pour en assurer un haut niveau de sécurité et une bonne information du public, dans la suite des recommandations du groupe de concertation Résil.

3 Règlement général sur la protection des données.

Les responsables des appariements en informent le public, a fortiori les personnes concernées directement par ces traitements, sous une forme explicite, pédagogique et accessible, sur leur site internet ou via les « lettres avis » des enquêtes statistiques, le cas échéant.

Un appariement doit répondre de sa conformité aux obligations du Règlement général sur la protection des données et de la loi Informatique et libertés ainsi qu'aux règles déontologiques de la statistique publique. Ces règles s'inscrivent dans différents corpus de valeurs, par exemple celui formalisé récemment par l'Unece⁴, qui retient 6 valeurs : pertinence, impartialité (qui recouvre une notion de neutralité, mais aussi de couverture de l'ensemble du champ), indépendance professionnelle, respect de la vie privée, transparence, collaboration (pour aboutir, notamment à des statistiques comparables).

La demande d'appariement est faite par un responsable de traitement issu de la statistique publique qui est le responsable de traitement au regard du RGPD et le seul propriétaire final de la base résultante. Il lui appartient de décrire son traitement et déclarer ses finalités, conformément aux obligations du RGPD. Il établit si nécessaire une Analyse d'impact pour la protection des données et indique comment s'assurer du respect des principes de **minimisation** et de **proportionnalité**.

2. Comment le Cnis peut-il venir en appui de la mise en œuvre de ce cadre ?

En matière de transparence, le Cnis donnera une visibilité supplémentaire aux appariements réalisés dans le cadre du SSP. Une rubrique spécifique du site internet du Cnis sera consacrée aux appariements, à la fois pour en décrire les grands principes et les usages et pour diffuser la liste des appariements réalisés chaque année au sein du SSP, avec une description succincte de leurs principales caractéristiques : responsable de traitement, finalités, principaux objectifs, sources utilisées, et justification des choix faits en ce qui concerne la population étudiée et les données retenues.

Pour ce qui concerne les principes de nécessité et de proportionnalité, le Cnis pourra apporter un avis, qui peut intervenir en amont d'autres consultations (Cnil, Cesrees⁵...) et ainsi en éclairer certains aspects, selon les modalités suivantes :

- dans le cas des enquêtes dont le protocole de collecte ou les exploitations futures prévoient un appariement avec des données administratives, l'avis d'opportunité rendu par la commission compétente porte sur l'ensemble du dispositif, donc intègre l'appariement ; la fiche de présentation des enquêtes sera aménagée pour rendre plus visibles les appariements et les décrire de manière plus normalisée⁶ ;
- dans le cas des appariements de sources administratives mobilisées via les demandes d'accès au titre de l'article 7bis de la loi de 1951, les demandeurs doivent mentionner dans leur demande d'accès les appariements envisagés et leurs finalités⁷ ; l'avis de cession est rendu en intégrant cette dimension, au regard des principes de nécessité et de proportionnalité ; un examen supplémentaire en commission du Cnis n'est pas nécessaire ;
- pour les autres opérations, par exemple la construction de fichiers mobilisant des sources diverses dont l'appariement n'était pas prévu initialement, un avis du Cnis sera demandé, qui s'apparentera aux avis d'opportunité rendus pour les enquêtes. L'avis s'appuiera sur un document de présentation succinct indiquant les objectifs précis de l'appariement et les données mobilisées (sources et thématiques, recours ou non à l'échantillonnage) ; le Cnis pourra procéder par consultation électronique si nécessaire, par exemple pour traiter les demandes les plus simples ou prendre en compte des demandes urgentes. Dans le cas d'opérations récurrentes, l'avis sera rendu pour plusieurs années.

4 Commission économique pour l'Europe des Nations unies.

5 Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé.

6 Présentation des données retenues et explication de leur utilité.

7 Idem.

Par dérogation à cette disposition, les appariements menés à titre exploratoire (à des fins méthodologiques ou d'étude ponctuelle) ne nécessitent pas d'avis du Cnis, mais une simple information.

- sur des projets spécifiques de par leur ampleur (population concernée, nombre de sources mobilisées, multiplicité des thématiques) ou les variables concernées (présence ou non de variables sensibles au sens du RGPD par exemple), et si cela n'a pas été fait par ailleurs⁸, une instruction spécifique visant notamment à croiser différents regards peut être menée, sur proposition du Bureau du Cnis, d'un président de commission ou du responsable du traitement. Elle peut par exemple prendre la forme d'un débat approfondi en commission ou d'un groupe de travail ad hoc. Elle débouche sur un avis du Cnis.

Les suivis statistiques de cohortes internes à un ministère, fondées sur un identifiant unique et ne nécessitant pas de mobiliser d'autres données externes au ministère, ne nécessitent pas d'avis spécifique du Cnis.

8 Par exemple par le Cesrees pour les appariements mobilisant des données de santé, mais un tel avis peut aussi être rendu postérieurement à un avis du Cnis, qu'il intègrera dans sa réflexion.