

## Projet de Rencontre du Cnis sur les appariements - 28 janvier 2022-

### Les appariements de données administratives sur les individus pour améliorer les connaissances statistiques dans le domaine économique et social

#### 1. Contexte et objectifs

Le programme RÉSIL 2020 – 2025 (**R**épertoires **S**tatistiques d'**I**ndividus et de **L**ogements) vise à construire un socle (exhaustif en couverture mais réduit en nombre de variables) de données relatives aux individus, aux ménages et aux logements, point d'appui pour l'échantillonnage des enquêtes, les redressements, les contrôles, la mesure de la couverture des enquêtes et des sources administratives.

Pérenne mais évolutif, ce socle de répertoires serait mis à jour à partir de diverses sources administratives. Construit par l'Insee, il serait au service de l'ensemble du Service statistique public. Le programme RÉSIL permettra également de faciliter les appariements de sources pour les individus, les ménages ou les logements pour produire des données multisources. Le nombre de variables stockées sera limité à celles strictement nécessaires à l'identification, aux appariements avec les bases "métier" et au suivi de la qualité des répertoires.

Dans le cadre de la mise en place du programme RÉSIL, des concertations sont envisagées avant mi-février 2022, date de dépôt d'un dossier de décret en conseil d'État. Compte tenu du caractère potentiellement sensible du dossier, l'Insee souhaite associer assez largement des acteurs du monde économique, social et académique pour que les points de vue s'expriment et soient intégrés à la préparation du programme et du texte.

Une rencontre du Cnis sur le sujet plus global des appariements, qui englobe la plupart des questions que peut soulever RÉSIL (identifiants, traitement et rapprochements de données à grande échelle, potentiellement sur l'ensemble de la population...), semble un cadre adapté pour mener une large concertation. La rencontre aurait pour objectifs de faire un point sur les pratiques d'appariements, d'informer sur le projet RÉSIL, d'évoquer les questions d'identifiants pour ces appariements, ainsi que les questions éthiques et la légitimité de procéder aux appariements pour recueillir des informations. Elle doit également permettre de donner des perspectives aux utilisateurs sur les bénéfices qu'ils peuvent en retirer et sur les garanties à apporter (sécurité des données, information, etc.). Cette rencontre se tiendrait au début de l'année 2022. Elle pourrait ainsi s'inscrire dans la préparation du nouveau moyen terme du Cnis (2024-2028), qui devrait inclure à nouveau un point sur les appariements de sources de données, dans le prolongement de l'orientation 7 du moyen terme actuel (2019-2023) :

**« Avis n° 7 Développer les appariements entre sources de données**

*Le Conseil demande à l'ensemble des producteurs de la statistique publique de développer les appariements entre sources de données afin d'enrichir l'analyse des liens entre différents thèmes, en veillant au strict respect de la confidentialité lorsque les appariements reposent sur des informations identifiantes. »*

Les appariements de données administratives ont déjà été évoqués plusieurs fois au Cnis en 2020, et tout particulièrement au bureau de décembre 2020, où ont été présentées les conclusions du rapport de l'IG sur les appariements. L'avancement du projet CSNS y a été évoqué, ainsi que le lancement du programme RÉSIL.

Complémentairement, en Assemblée plénière, le 27 janvier 2021, l'**Orientation n° 2 : Donner un statut aux statistiques construites sur des données administratives**<sup>1</sup> a été commentée dans le cadre de la discussion sur les programmes de travail.

*Encouragés notamment par le Conseil dans son programme à moyen terme, ces derniers ont vocation à devenir, pour les statistiques publiques en France, comme à l'étranger, « un mode de collecte comme un autre », dans un cadre technique et juridique adapté.*

Il a été jugé utile d'associer des pairs, c'est-à-dire quelques instituts étrangers, d'autant que certains ont pris de l'avance et que le sujet se pose partout. Cela permettra d'inscrire le projet explicitement dans la tendance générale des instituts de statistiques, lui donnant ainsi une légitimité et des points de repère supplémentaires, même si chaque pays a son contexte spécifique. Il est aussi prévu d'associer l'ASP et d'inviter la Cnil.

### **Exemples de questions abordées durant la rencontre :**

- Quelle est la situation au niveau international en matière d'appariements, d'identifiants et de répertoires d'individus ?
- Comment fonctionnent les registres à l'étranger ?
- Quels ont été les appariements réalisés jusqu'à présent ? Quels appariements sont difficiles à réaliser aujourd'hui dans l'état de l'art ?
- Comment est gérée l'information des personnes ? Y a-t-il une demande de consentement ? Dans quelles conditions en faudrait-il une ?
- Positionner Fideli dans l'histoire des appariements
- Pourquoi le projet RÉSIL en plus du CSNS ? Articulation entre RÉSIL et CSNS
- Quel est le lien entre le projet RÉSIL et le recensement de la population (en creux ou explicitement) ?
- Gouvernance et information des utilisateurs sur les projets d'appariement pour leur permettre de réagir sur les priorités de ces nouvelles sources comme sur les projets de nouvelles enquêtes
- Transparence sur les méthodes et libre accès aux outils d'appariements utilisés par l'Insee
- Quelle ouverture du service d'appariement RÉSIL ? Quel service d'appariement pour les chercheurs ?
- Quel accès possible aux fichiers résultant des appariements pour les chercheurs ?
- Comment recenser les évaluations des politiques publiques reposant sur des données appariées pour un utilisateur extérieur ?

**Date de la rencontre :** 28 janvier 2022 à Paris Bercy.

**Format :** une journée.

**Organisation :** un comité de suivi, composé de 4 représentants du SG Cnis et de 5 représentants de l'Insee, a été mis en place pour la conception et l'organisation de la rencontre.

---

<sup>1</sup> Référence : Avis généraux 4, 7 et 8 du moyen terme du Cnis 2019-2023 sur les programmes de travail 2021

## 2. Le programme (provisoire)

### **9H 1. Introduction (15 minutes)**

**Patrice Duran**, président du Cnis

### **9H15-10H15 2. Rappel historique et état des lieux (1h)**

Insee : Les pratiques d'appariements de la statistique publique dans le temps (30 mn)

CASD : Etat des lieux des appariements réalisés par les chercheurs (15 mn)

**Questions / interventions de la salle (15 mn)**

**Pause 10H15-10H30**

### **10H30-11H30 3. Des exemples d'appariements (1h)**

**Echanges avec la salle**

### **11H30-12H30 4. Les projets d'avenir : le CSNS et RésIL**

Avancement actuel du projet CSNS (dont fichiers pivots)

Enjeux du programme RésIL

Communications faites au Cnis sur les appariements

**Echanges avec la salle**

**12h30-14H Cocktail**

### **14H-15h15 5. Table ronde : Quels appariements pour quels usages ?**

5 intervenants maximum

idées : Experts INS étrangers, Eurostat, chercheurs en évaluation des politiques publiques, France stratégie, chercheur géographe, représentant de politiques publiques.

**Echanges avec la salle**

### **15h15-16H30 6. Table ronde : quelle transparence, quelle information du public ?**

5 intervenants maximum

Idées : Statistique canada, Cnil, un responsable éthique, un représentant européen ou du comité d'éthique anglais, un représentant association, syndicat, un journaliste spécialisé.

**Echanges avec la salle**

### **16h45 7. Conclusion de la rencontre : Jean-Luc Tavernier, DG Insee**