
Erfi 2 / FamEmp

Appariements à des données administratives

Géraldine CHARRANCE, Paul COCHET, Thomas MERLY-ALPA

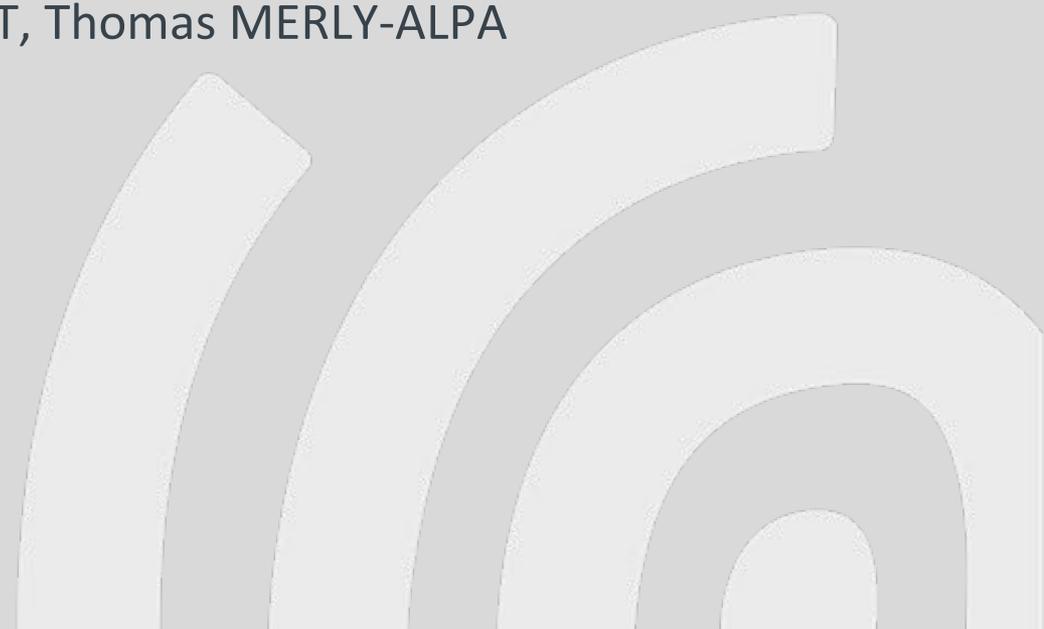
A decorative graphic in the bottom right corner consisting of several overlapping, curved, light gray bands that create a sense of depth and movement.

Table des matières

1. Contexte
2. Flux de fichiers
3. Calendrier

1. Contexte



Contexte

- On souhaite apparier les données des enquêtes Erfi 2 et FamEmp à différentes sources :
 - Sources fiscales et sociales : **POTE** (fichier d'imposition sur les revenus) et **FILOSOFI** (Dispositif sur les revenus localisés sociaux et fiscaux), propriétés de la **DGFIP**. L'appariement sera réalisé par le pôle Revenus Fiscaux et Sociaux (**pôle RFS**)
 - Sources relatives à l'Emploi : Base Tous Actifs (**BTA**) qui résulte de la fusion de la Base Tous Salariés (BTS) et la Base Non salariés (BNS), toutes deux propriétés du Département de l'Emploi et des Revenus d'Activité de l'Insee (**DERA**).

Contexte

Selon les sources à appairer, l'identifiant à utiliser comme clef d'appariement varie :

- Le **SPI** pour POTE/FILOSOFI,
- Le **NIR** pour la BTA.

Le SPI (aussi appelé NIF, numéro d'identification fiscale) est disponible dans la base de sondage Fidéli mobilisée pour ces deux enquêtes.

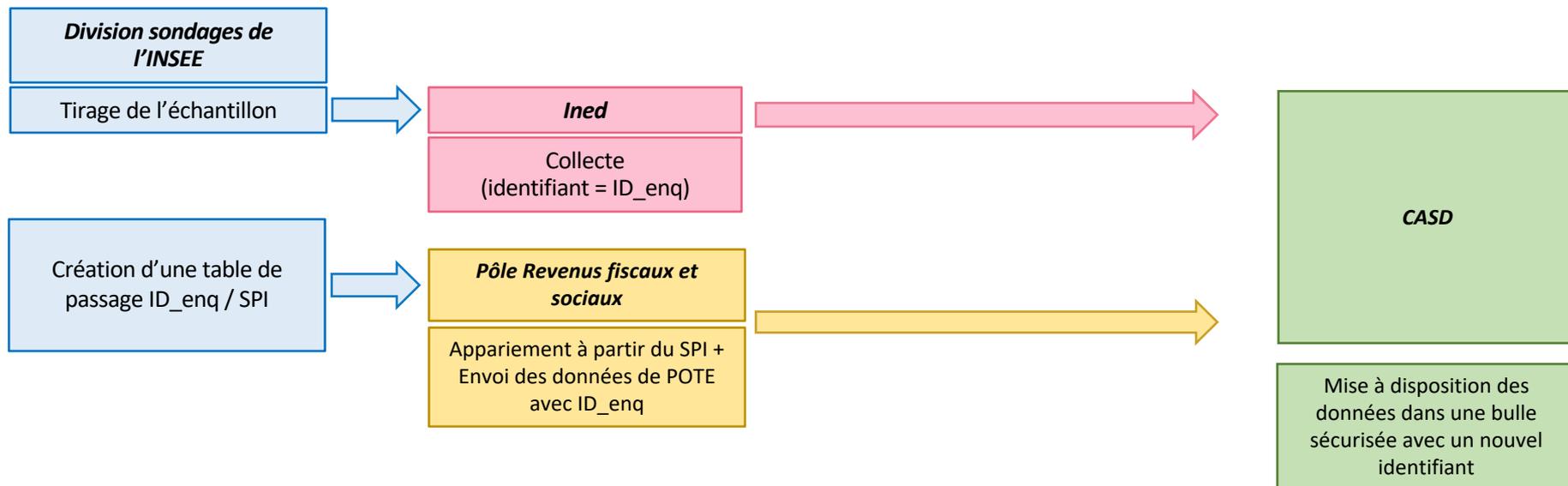
Le NIR (Numéro d'Inscription au Répertoire ou numéro de sécurité sociale) doit être reconstitué à partir des informations de Fidéli (nom, prénom, date de naissance, lieu de naissance...).

2. Flux de fichiers



Appariement POTE/FILOSOFI

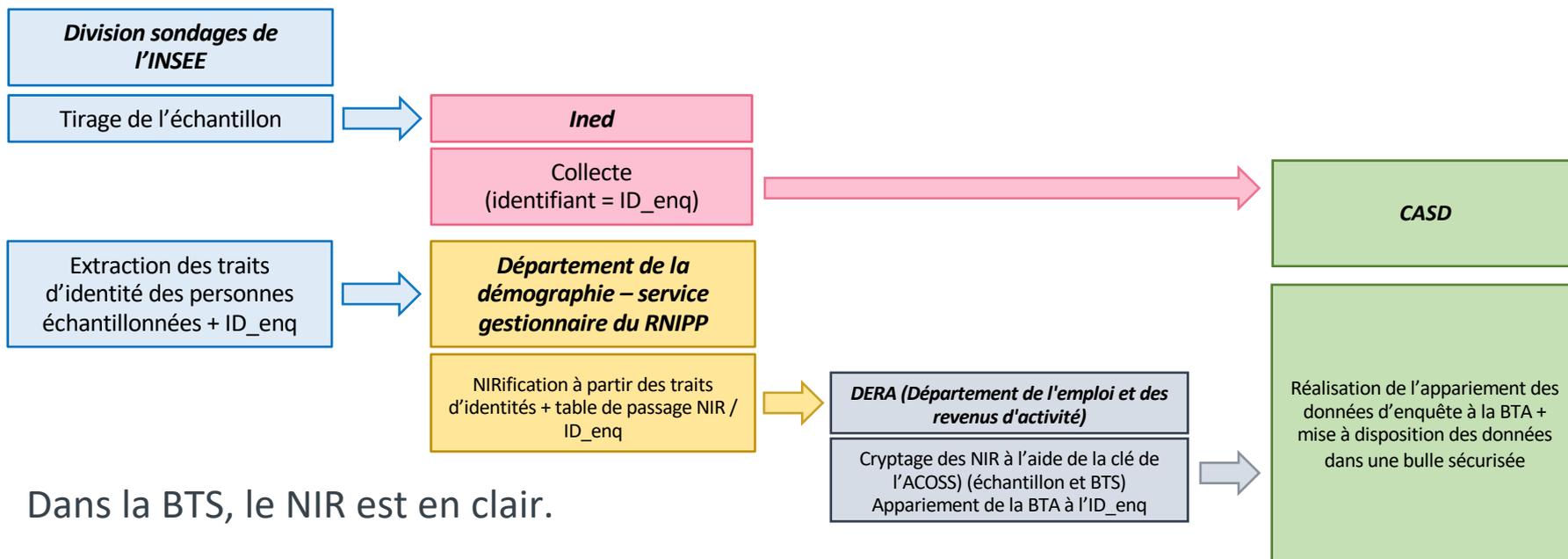
- Pour que l'appariement se fasse, nécessité de coordonner cinq acteurs : division Sondages, Ined, Pôle RFS, CASD, DGFIP (pour la convention uniquement)
- Les données d'enquête appariées aux données administratives ne sont disponibles que via une bulle sécurisée avec un nouvel identifiant.



Appariement BTA : NIRification

- L'appariement à la BTA nécessitant le **NIR**, une opération de NIRification (reconstitution du NIR à partir des informations issues de Fidéli) est nécessaire.
- Cette étape sera réalisée par le **Département de la Démographie** de l'INSEE.
- Cette opération coûte 50 000€ si le volume d'individus à NIRifier dépasse les 100K individus. Pour éviter ces coûts, il est prévu de NIRifier et appairer uniquement une sous-partie des non-répondants.
- En plus de la NIRification, le département de la Démographie réalisera chaque année une actualisation du statut vital des individus échantillonnés, ce qui nous permettra d'identifier les personnes décédées entre 2 vagues

Appariement BTA



Dans la BTS, le NIR est en clair.

Dans la BNS, le **NIR est crypté par une clé propriété de l'ACOSS.**

Cela nécessite une étape supplémentaire : le DERA devra crypter les NIR à l'aide de la clé de chiffrement de l'ACOSS pour obtenir une table de passage ID_enq/NIRh.

Cette opération concernera les NIR des échantillonnés, et également les NIR de la BTS.

L'appariement sera donc réalisée sur le NIRh, qui sera ensuite supprimé pour la mise à disposition des données au CASD.

Opposition aux appariements

Lors de l'appariement des données d'enquête avec les données des sources d'emploi et sociales dans la bulle sécurisée du CASD, il faudra exclure les personnes ayant exprimé leur opposition aux appariements.

3. Éléments de calendrier

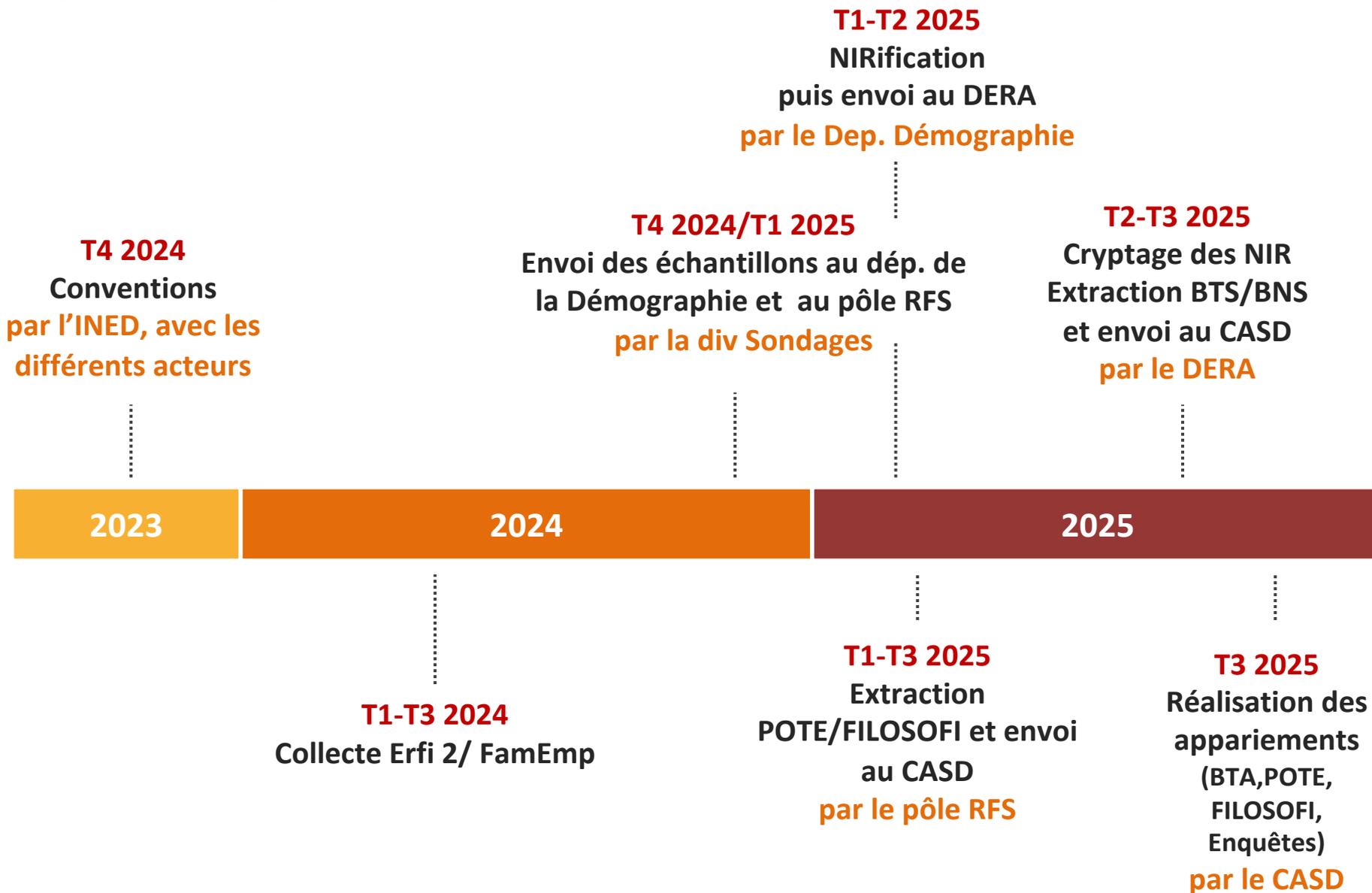


Disponibilité des sources

- Le millésime N de POTE est disponible fin N+1 (millésime 2023 disponible fin 2024)
- Le millésime N de FILOSOFI est disponible mi N+2 (millésime 2023 disponible mi 2025)
- Les bases pour un millésime N sont disponibles à l'été N+2 (ou un peu avant). Pour chaque millésime, on dispose également des données de l'année N-1. Dans la BTS 2024, on aura aussi les données portant sur l'année 2023.

Une fois les sources disponibles, le temps des opérations est relativement court.

Calendrier



Feuille de route

Prochaines étapes :

- Récupérer les engagements écrits des différents acteurs impliqués
- Rencontrer le responsable de l'appariement final et de l'hébergement des données (ici, le CASD) pour leur présenter les schémas et flux de fichiers et leur demander un premier devis
- Conventionner en conséquence (et le faire bien en amont des opérations d'appariements). Dans notre cas (Erfi 2 / FamEmp) , a priori deux conventions à prévoir :
 - Une convention tripartite CASD/Insee/Ined pour les appariements POTE
 - Une convention tripartite (ou quadripartite) CASD/Ined/Insee(/ACOSS)
- Prévoir du temps humain pour superviser et coordonner les opérations des différents acteurs.

→ Premier retour d'expérience de l'Ined à la prochaine AG

Merci de votre attention

