



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

depp Direction de l'évaluation,
de la prospective
et de la performance

ENJEUX STATISTIQUES DU SYSTÈME D'INFORMATION INSERJEUNES SUR L'INSERTION PROFESSIONNELLE

Nathalie Caron & Loic Midy - DEPP

JMS – Mars 2022

Sommaire

- **Les finalités initiales d'InserJeunes**
- **La genèse d'InserJeunes et la situation avant InserJeunes**
- **Ses finalités vont au-delà de celles prévues initialement**
- **Les principaux enjeux du dispositif InserJeunes**
 - Deux enjeux statistiques : **les appariements et le calcul d'une valeur ajoutée pour cerner l'effet établissement**
 - Les autres enjeux
- **Conclusion**

Les finalités initiales d'InserJeunes

▪ Sa finalité initiale

- Permettre, par fusion de bases administratives « scolarité » sur les inscriptions des élèves et apprentis et la base « emploi » MMO (mouvement de main d'œuvre) qui repose sur la DSN (Déclaration Sociale Nominative), de calculer chaque année au niveau établissement :
 - Le taux d'insertion professionnelle **des sortants de l'établissement des niveaux CAP à BTS**
 - La valeur ajoutée de l'établissement (permettant de mesurer « l'effet propre »)
- **Le SI permet également de calculer deux autres indicateurs :**
 - Le taux de poursuite d'études
 - Le taux d'interruption en cours de formation

La genèse d'InserJeunes

Article 24 de la loi pour la liberté de choisir son avenir professionnel N°2018-771 du 5 septembre 2018 :

Art. L. 6111-8 (code du travail) - Chaque année, pour chaque centre de formation d'apprentis et pour chaque lycée professionnel, sont rendus publics quand les effectifs concernés sont suffisants :

- Le taux d'obtention des diplômes ou titres professionnels
- Le taux de poursuite d'études
- Le taux d'interruption en cours de formation
- Le taux d'insertion professionnelle des sortants de l'établissement concerné, à la suite des formations dispensées
- La valeur ajoutée de l'établissement

Pour chaque centre de formation d'apprentis, est également rendu public chaque année le taux de rupture des contrats d'apprentissage conclus

La situation avant InserJeunes

Des enquêtes d'insertion (enquêtes IVA/IPA) menées par la DEPP auprès de tous les jeunes sortants :

- **Exhaustives et très lourdes** à gérer pour les chefs d'établissements et les différents acteurs intervenant dans la collecte (équipes statistiques DEPP et rectorat, ...)
- **coûteuses** (envoi questionnaires, relances,)
- **un seul point** : insertion à 7 mois après la sortie
- **taux de réponse de l'ordre de 60%** : les enquêtes ne permettent pas de fournir des taux d'insertion par établissement ni même au niveau national pour toutes les spécialités

InserJeunes : ses finalités vont au-delà

▪ avec :

- La construction d'indicateurs à un niveau très fin (croisement établissement et niveau de formation et spécialité voire diplôme) en se rapprochant de l'exhaustivité
- Des indicateurs à différents moments après la sortie du système éducatif (6 mois, 12, 18 et 24 mois)
- Un délai de mise à disposition de ces taux plus rapide qu'avant
- La réalisation d'études sur les conditions d'emploi des jeunes sortant de formation, l'adéquation formation/emploi, avec la base de données très riche adossée à InserJeunes et disponible pour les chargés d'études DEPP/ DARES ainsi que pour les chercheurs

Les principaux enjeux du dispositif InserJeunes

Des enjeux statistiques

- Les **appariements** : environ une vingtaine de bases appariées.
- Le calcul d'une **valeur ajoutée** pour les taux d'emploi permettant de mesurer l'effet propre de l'établissement, les taux d'emploi dépendant en partie de facteurs extérieurs (profil des jeunes, ...)

Et d'autres types d'enjeux

- La **sécurité et la traçabilité des traitements** avec la mobilisation de multiples tables de données avec DCP

1^{er} enjeu statistique : Les appariements au cœur d'InserJeunes

- Entre les sources « **scolarité** » pour isoler les **jeunes sortants** du système éducatif par comparaison sur 2 années
 - Appariement sur l'identifiant national élève propre à l'Education Nationale (l'INE) (sauf privé hors contrat)
- Pour enrichir la **base des sortants avec les sources sur les résultats des examens**
- Entre **le fichier des sortants et la MMO/DSN** :
 - **pas d'identifiant commun entre sources scolarité (INE) et sources emploi (NIR) !**
 - Nom, prénom, date et commune de naissance, sexe

→ sur les 20 appariements, 10 sont sur identifiants indirects

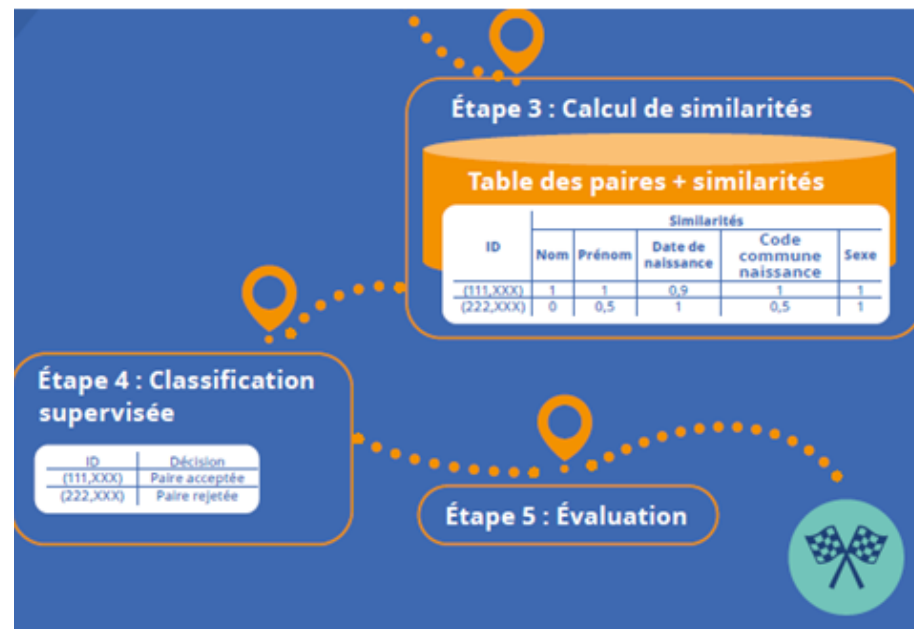
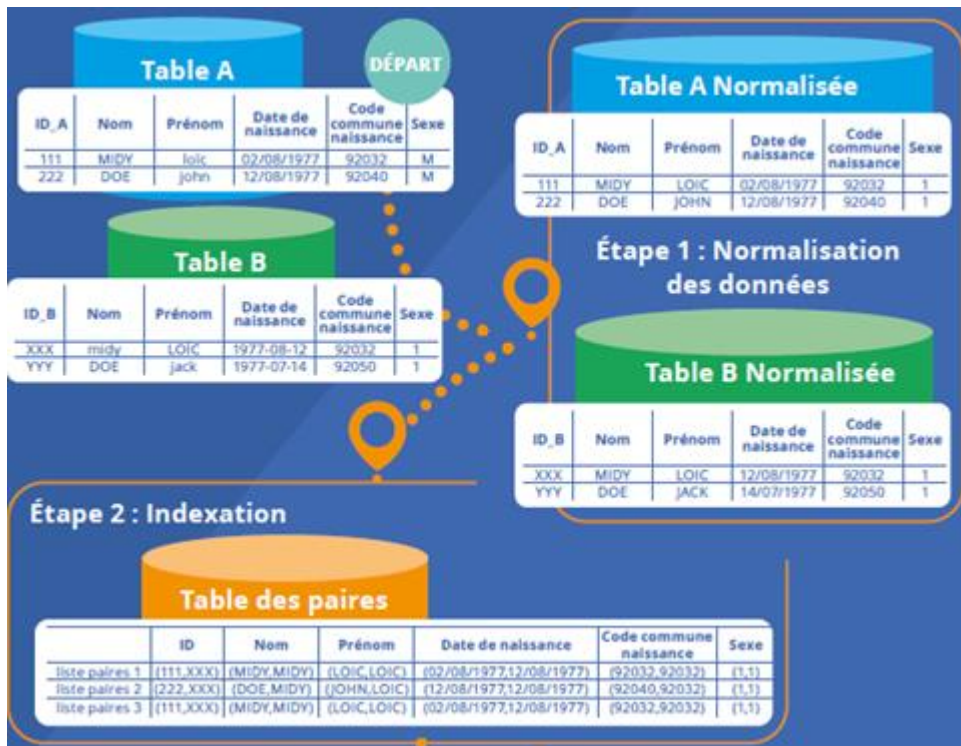
Les appariements : l'outil utilisé

- Etude de plusieurs outils et méthodes existantes : outil DEPP, « matchID », programmes SAS, méthode « tours de clé », RELAIS (INS italien),...
- Décision de faire un outil en utilisant **Python** et des librairies existantes
 - À un **coût raisonnable** (environ 60 jours/hommes)
 - **Facile à installer**
 - **Adaptable** à d'autres appariements sur identifiants indirects
 - **Rapide**

Dans notre cas, existence d'un module **qualité** sur l'appariement : taux d'appariement de 97% entre la source « recensement des apprentis » et MMO/DSN (sources avec des variables bien renseignées)

- Cet outil sera mis en **open source** avec un guide d'utilisation

Cet outil est basé sur 5 étapes



Graphique extrait de l'article de : Midy, L., 2021, « un outil d'appariement sur identifiants indirects : l'exemple du système d'information sur l'insertion des jeunes », *Courrier des statistiques* N6 - 2021, Insee

2^{ème} enjeu statistique : le calcul d'une valeur ajoutée

Pour cerner l'action propre d'un établissement en termes d'accompagnement à l'insertion, les taux observés ne suffisent pas :

- Le taux d'emploi global de l'établissement s'accompagne d'une valeur ajoutée pour contrôler les facteurs sur lesquels le lycée ou le CFA a peu ou pas de prise et qui ne mesurent pas son action propre
- On calcule ainsi pour chaque établissement un taux d'emploi attendu qui prend en compte le profil d'élèves accueillis, l'offre de l'établissement ainsi que le taux de chômage local.
- Dans une logique similaire à celle développée par la DEPP avec les IVAL, on définit $\text{valeur ajoutée} = \text{taux observé} - \text{taux attendu}$

Comment calculer le taux attendu ?

Utilisation d'un modèle **Logit à 3 niveaux** pour tenir compte des effets de contexte.

Pour un élève (i), établissement (j) et zone d'emploi (k) de la **commune de résidence de l'élève**, la probabilité d'avoir un emploi P_{ijk} est modélisée par :

$$\text{Logit}(P_{ijk}) = \beta_0 + \beta_1 X_{ijk} + \beta_2 W_j + \beta_3 Z_k + \alpha_j + \gamma_k$$

Avec

- β_0 = constante
- X_{ijk} = variables au niveau individu (sexe, âge, diplômé, spécialité formation...).
- W_j = variable au niveau établissement j
- Z_k = taux de chômage de la zone d'emploi k
- α_j = résidu pour l'établissement j et γ_k = résidu pour la zone d'emploi k.

Et **Taux attendu** = moyenne des probabilités individuelles sans le résidu α_j

Comment calculer le taux attendu ?

Les grands déterminants de l'insertion :

- Niveau de diplôme (CAP ou équivalent, Bac, Brevet professionnel et BTS)
- Filière de formation (apprentissage et voie scolaire)

→ en tout 9 modèles (dont 1 spécifique pour les
apprentis en formation dans les lycées)

Chaque modèle est élaboré par ajout successif de variables : niveau individus, niveau établissement (% d'élèves en situation de handicap), niveau ZE (taux de chômage).

Les variables retenues niveau individus : **spécialités de formation, activité de l'établissement (pour les apprentis)**, âge, sexe, PCS, obtention diplôme et résultat, *handicap*, ...

Comment calculer le taux attendu ?

Taux d'emploi attendu = moyenne des probabilités d'être en emploi après avoir soustrait le résidu établissement α_j

Il s'interprète comme le taux d'emploi moyen des élèves accueillis dans des établissements comparables en termes de profil des élèves, de formations dispensées et domiciliés dans une zone d'emploi au taux de chômage comparable.

Niveau de diplôme	niveau CAP APP	niveau CAP SCO
Taux d'insertion 6 mois après la sortie	54%	26%
Nombre d'élèves	41 248	26 213
Nombre d'établissements	619	1 637
Nombre d'élèves (établissements avec effectif >=20)	40 986	24 978
Nombre d'établissements (avec effectif >=20)	581	1 487

Critères ajustement	deviance (-2 log vraisemblance)	53 285	26 819
	AIC	53 349	26 869
Estimation variance	variance établissement	0,05	0,11
	variance zone emploi	0,02	0,04
	total	0,07	0,14

filtre 20 sortants minimum par établissement			
Valeur Ajoutée	écart type VA	13,3	16
	Distribution écart absolu VA		
	[0,5[39%	29%
	[5,10[30%	25%
	[10,15[13%	18%
	[15,25[12%	18%
	[25,max[6%	10%

Les autres enjeux

Nombreuses bases de données avec DCP identifiantes (nom, prénom,...) et indirectement identifiantes (INE)

- Enjeux juridiques (RGPD, AIPD,...)
- Enjeux importants de sécurité
 - échange sécurisé des fichiers entre la Depp et la Dares
 - suppression au fur et à mesure du processus des DCP lorsqu'elle ne sont plus nécessaires, espace restreint à quelques personnes à l'espace des données avec DCP
 - méthode de « double hachage » pour l'INE
- Traçabilité des traitements (journal de bord)

Conclusion : les perspectives

- Etendre à l'emploi salarié public à l'horizon 2023
- Intégrer les formations en lycées agricoles
 - Expérimentation début 2022, en production en 2023 ?
- Atteindre la production en rythme de croisière (4 points par cohortes de sortants et deux cohortes en simultané)
- Analyser la richesse des données disponibles

Bibliographie

Sur la méthodologie

- Caron N., Midy, L., 2022, « Enjeux statistiques du système d'information InserJeunes sur l'insertion professionnelle », JMS 2022
- Christen P., 2012, data matching concepts and technique for record linkage, entity resolution and duplicate detection, Springer
- Midy, L., 2021, « un outil d'appariement sur identifiants indirects : l'exemple du système d'information sur l'insertion des jeunes », Courrier des statistiques N6 - 2021, Insee
- Midy, L., 2021, « InserJeunes – taux d'interruption en cours de formation », document de travail série « Méthodes » n°2021-M02, DEPP-MENJS
- Midy, L., Deschamps G., 2021, « InserJeunes – calcul de valeur ajoutée », document de travail série « Méthodes » n°2021-M01, DEPP-MENJS

Sur les résultats

- Antoine R., Collin C., Marchal N. (DEPP), Fauchon A. (Dares), 2021 "Insertion professionnelle des apprentis du niveau CAP à BTS : 6 mois après leur sortie du système éducatif en 2020, 61 % sont en emploi salarié en janvier 2021", Note d'information, n° 21.43, décembre, DEPP
- Antoine R., Collin C., Marchal N. (DEPP), Fauchon A. (Dares), 2021 "Insertion professionnelle des lycéens du niveau CAP à BTS : 6 mois après leur sortie du système éducatif en 2020, 36 % sont en emploi salarié en janvier 2021", Note d'information, n° 21.42, décembre, DEPP
- Le site InserJeunes : <https://www.inserjeunes.education.gouv.fr/diffusion/accueil>

Merci pour votre attention