

DATA SHS 8 décembre 2020 - UPN

*L'accès aux données administratives de l'enseignement secondaire
et supérieur*

Leïla Frouillou
MCF Paris Nanterre, CRESPPA-GTM
Chercheuse associée ESCOL, CIRCEFT
Leila.frouillou@gmail.com

Introduction : un témoignage situé sur les données scolaires

Un témoignage situé : quelques années de relations avec le SIES sur les universités, un postdoc en 2016-2017 sur les collèges du 93

1. Quelques bases de données utiles pour décrire les parcours scolaires
2. Éléments sur la collaboration entre services statistiques et chercheurs
3. Enjeux éthiques : anonymisation *versus* open data
4. Conclusion : CPESR, un projet collaboratif pour outiller les enquêtes sur le supérieur

I. Quelques bases de données sur les parcours scolaires

Bases secondaire	Description	Accès locaux / nationaux
Base Centrale Scolarité	Inscriptions administratives des élèves	Quêtelet-PROGEDO / DEPP
Base Centrale Etablissement	Etablissements (statuts)	DEPP
SIECLE	Etapes d'orientation	Etablissement / DEPP
Affelnet	Affectation en lycée	Rectorat / DEPP
APAE	Pilotage - Gestion	Etablissement / DEPP
OCEAN > CYCLADES	Concours, examens	
FAERE	Appariement inscriptions et résultats DNB Bac + cohortes	DEPP
Panels 1995-2005 ; 2007-2013	Suivi de cohorte (35 000 élèves en 6 ^e)	Quêtelet-PROGEDO / DEPP

- En ligne : archives et données documentaires + R&RS annuels

<https://archives-statistiques-depp.education.gouv.fr/Default/>

I. Quelques bases de données sur les parcours scolaires

Bases sur le sup.	Description	Accès locaux / nationaux
APB2016 et 2017	Vœux, candidats, formations	SIES (opendata)
Parcoursup 2018 et 2019	Vœux, candidats, formations	SIES (opendata)
SISE (inscription / résultats)	Inscriptions des étudiant.es	Quêtelet-PROGEDO / SIES
AGLAE	Bourses CROUS	SIES
Apogée	Inscriptions des étudiant.es et notes	Etablissement
OVE	Echantillon (national), tous les quatre ans, sur conditions de vie	Quêtelet-PROGEDO / OVE

- Une base de gestion des établissements du supérieur ?
- Des panels nationaux ?

2. Éléments sur la collaboration entre services et chercheur.es

- Constitution d'un comité scientifique pour la mise à disposition des données APB

Extrait de notes prises lors de la réunion APB Stats du 25 avril 2017

« Intro Kabla-Langlois [...] »

Intérêt croissant pour l'enseignement sup.

Challenge SIES : assurer la meilleure diffusion des données de l'ESR. Diffuser plus, plus vite et mieux. > chercheurs mais aussi donner des stats plus tôt.

SISE : notes flash plus nombreuses et plus rapidement. Depuis 2013 infos tous les ans sur les VA réussite en Licence.

Meilleur accès aux chercheurs : SISE résultats et inscription sur Quételet.

Opendata de l'ESR (72 jeux de données)

Première publi stats APB sur SIES : 2016 donc récent. Note flash sur les vœux en juin 2016 puis sur les affectations en novembre.

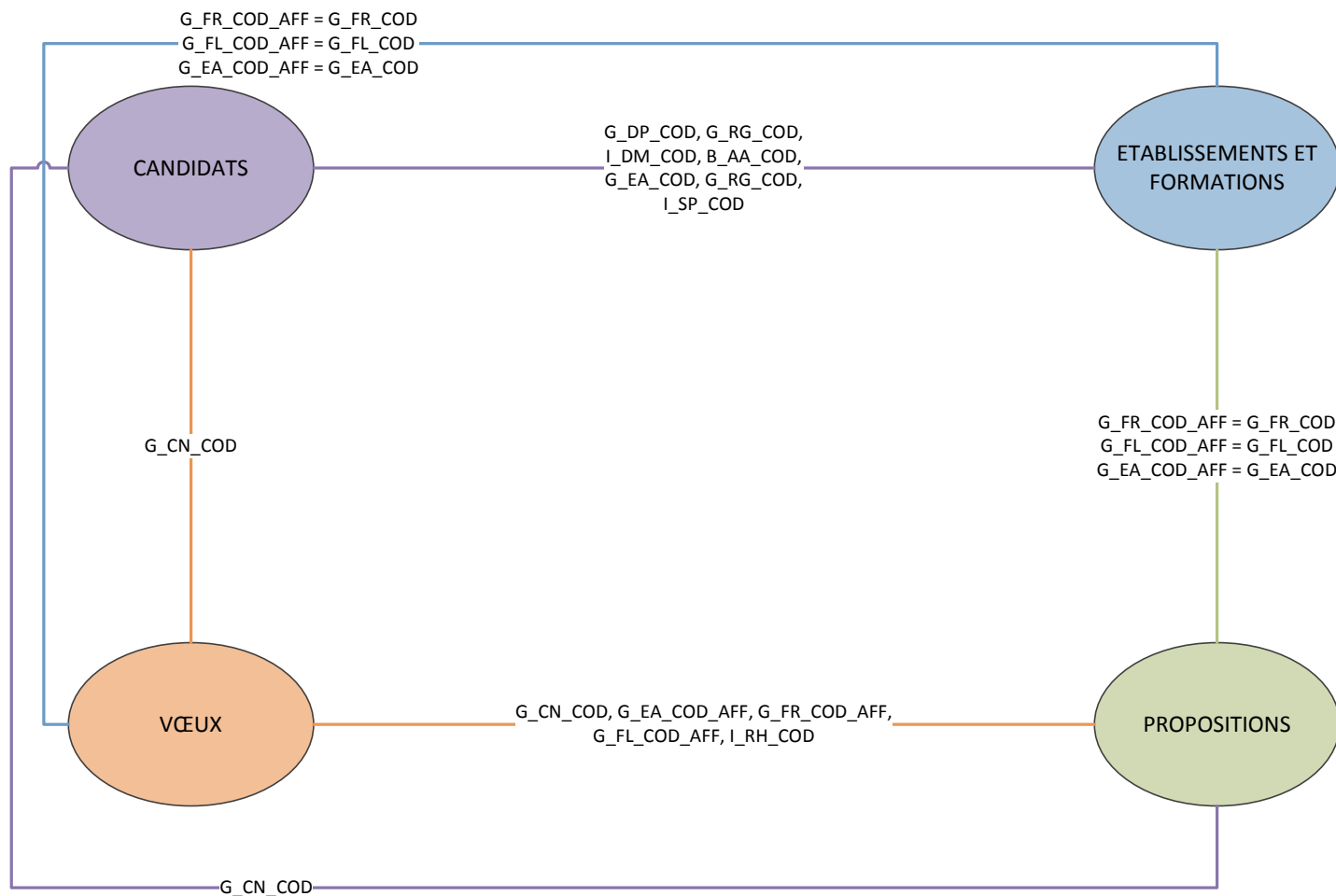
APB d'abord et avant tout un outil d'expression des vœux pour organiser leur transmission et l'affectation. [...]

> Préciser les données et leurs modalités de livraison / appui pour finaliser ces données pas de valorisation collective des données prévue auj > automne prochain. »

2. Éléments sur la collaboration entre services et chercheurs

- Un travail colossal
 - Deux enjeux liés aux données : recherche + opendata
 - 4 Fichiers pour la recherche : établissements, candidats, vœux, propositions

2. Éléments sur la collaboration entre services et chercheurs



2. Éléments sur la collaboration entre services et chercheur.es

- Un travail colossal
 - Deux enjeux liés aux données : recherche + opendata
 - 4 Fichiers pour la recherche : établissements, candidats, vœux, propositions
 - Penser l'archivage des données
 - Place des chercheur.e.s : choix des variables / test des premières bases de données
 - Question du (sous)dimensionnement du SIES au regard de la DEPP
- Le travail au ministère (conventions, places)
- Pas de ParcoursupStats...

3. Enjeux éthiques entre anonymisation et opendata

- Quelle mise à disposition des données pour la recherche ?
 - Des engagements individuels et un travail sur place : la contrainte de l'anonymisation
 - Le pas de temps de données disponibles, l'archivage et les limites de la CNIL

Durée de conservation 18 mois nominatives, 10 ans pour anonymes.

Appariements avec d'autres fichiers (ex impôts) ? Création de variables sur les quartiers de résidence avant effacement de l'adresse ? Conservation de données agrégées par établissements ? Supprimer les petits effectifs plutôt que le détails des variables ?

>>> Un accès limité pour les chercheur.e.s

3. Enjeux éthiques entre anonymisation et opendata

- Les projets en cours : suivis de trajectoire et identifiant national

Extrait de notes prises lors de la réunion APB Stats du 25 avril 2017

« Projet INES identifiant national dans l'enseignement supérieur, prend appui sur RNIE (répertoire national des identifiant élèves et apprenti). Ce dernier a pour vocation de mettre en place un identifiant unique national dans le scolaire (avant académique). Ce projet INES vise à mettre en place une plateforme pour les établissements pour contrôle d'INE et création d'INE (ex étudiant étranger). Incitation à la remontée individuelle pour ESR. >> vrai suivi de parcours depuis la seconde vers sup. »

> Un identifiant élève national mis en place en 2017 et qui va se propager ensuite dans le supérieur

3. Enjeux éthiques entre anonymisation et opendata

- Opendata
 - Loi sur le numérique et la transparence de l'action publique

NEXT  IMPACT

ACTUS HIGH-TECH

#LEBRIEF

COMMUNAUTÉ

BONS PLANS

Alors que la loi Numérique impose désormais aux administrations de faire la transparence sur les algorithmes qu'elles utilisent pour prendre des décisions individuelles, aucune ne semble pour l'heure s'être conformée à ses nouvelles obligations légales.

Montant des impôts, attribution d'allocations familiales ou de bourses scolaires,... **Depuis le 1er septembre**, toutes ces décisions individuelles prises « *sur le fondement d'un traitement algorithmique* » doivent être accompagnées d'une « *mention explicite* » informant l'utilisateur qu'un programme informatique est donc venu s'immiscer dans le calcul de ses APL, de sa taxe d'habitation, etc.

Avec cette petite ligne, rendue obligatoire par le nouvel article **L311-3-1** du Code des relations entre le public et l'administration, chaque acteur public est tenu de rappeler au citoyen qu'il a le droit d'obtenir la communication des « *règles* » et « *principales caractéristiques* » de mise en œuvre du traitement algorithmique utilisé.

L'objectif, à terme, est de permettre au public de mieux comprendre comment fonctionnent ces traitements automatisés. Les codes sources étaient en effet d'ores et déjà considérés comme des documents administratifs « *communicables* » au citoyen, sur demande. Il s'agit donc ici de venir mettre des mots sur des lignes de code...

Par [Xavier Berne](#)

le mercredi 08 novembre 2017 à 16:15

3. Enjeux éthiques entre anonymisation et opendata

- Opendata
 - Loi sur le numérique et la transparence de l'action publique

De nouvelles obligations en vigueur depuis le 1er septembre

Dorénavant, l'administration est ainsi tenue de fournir (sur demande), « *sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi* », les informations suivantes :

- Le degré et le mode de contribution du traitement algorithmique à la prise de décision
- Les données traitées et leurs sources
- Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé
- Les opérations effectuées par le traitement

Dans l'**étude d'impact** annexée au projet de loi Numérique, le gouvernement de Manuel Valls expliquait que cette réforme permettrait par exemple d'y voir plus clair sur le fonctionnement d'**Admission Post-Bac** (finalement appelé à disparaître). « *Le recours à ce logiciel, reposant sur des traitements algorithmiques, peut susciter des interrogations sur les mécaniques et les règles de fonctionnement qui conduisent à un résultat décisif pour l'avenir des étudiants : comment ce système est-il paramétré ? Quelle est la part de tirage au sort dans la procédure d'affectation pour les filières les plus demandées ? Comment s'assurer qu'il n'est pas possible de « tricher » avec le système ?* »

Par [Xavier Berne](#)

le mercredi 08 novembre 2017 à 16:15

3. Enjeux éthiques entre anonymisation et opendata

- Opendata
 - Loi sur le numérique et la transparence de l'action publique
 - Publication du code de Parcoursup, quelques jours avant les résultats du 22 mai 2018
 - A qui sert l'opendata scolaire ? Agences de coaching privé, classement des établissements par les journaux, parents stratèges et pilotage d'établissements mis en concurrence...

Conclusion : CPESR

- CPESR : « La Conférence des praticiennes et praticiens de l'enseignement supérieur et de la recherche (CPESR) est une organisation non militante œuvrant à la production et à la diffusion de connaissances sur l'ESR. »

<https://cpesr.fr/>

- Coordonnée par J. Gossa, collaborations diverses :

- Collaboratif :

- **Les commentaires sur les articles** : pour les signalements ponctuels, comme des références manquantes par exemple.
- **Le serveur Discord** : pour une participation dans le temps, par exemple pour demander des références spécifiques ou pour travailler à de nouvelles représentations de données.
- **Le groupe Zotero** : pour utiliser ou participer aux références et dossier.
- **Le groupe GitHub** : pour contribuer aux travaux DataESR ou signaler des problèmes en rapport.
- **Un formulaire de contribution** : pour proposer son aide pour des actions plus ciblées, comme la collecte de données dans les établissements.

Conclusion : CPESR

- Sources des données : R&RS + bilans sociaux +

- **Portail DΣPPaDoc** : Centre de documentation de la Direction de l'évaluation, de la prospective et de la performance (DEPP)
- **Portail SIES** : Notes et rapports du service d'analyse statistique du MESRI
- **Portail #dataESR** : Les datas de l'ESRI
- **Portail opendata** : Jeux de données open data sur l'enseignement supérieur, la recherche et l'innovation
- **Portail data.gouv.fr** : Partagez, améliorez et réutilisez les données publiques
- **Tableau de bord financier** : Suivi financier des établissements, et anticipation des éventuels risques liés à leur équilibre financier.
- **Fiches démographiques du CNU** : L'objectif de ces fiches est de présenter certaines caractéristiques démographiques de la section du Conseil national des universités concernée et de permettre une comparaison avec son groupe ou sa grande discipline d'appartenance.
- **Notes RH du MESRI** : Concours, emploi et carrières dans l'enseignement supérieur.
- **Site de l'AMUE** : Agence de mutualisation des universités et établissements

Conclusion : CPESR

- Sources des données : R&RS + bilans sociaux + opendata

Extrait du dernier tableau de bord pour Université Savoie-Mont-Blanc

