



Université

de Strasbourg

Pauline Boyer

# L'open data dans l'enseignement supérieur et la recherche

## Etat des lieux et perspectives

*Data Coffee – Université de Strasbourg*

*15 mai 2023*



# Qu'est-ce que l'open data ?

Une donnée ouverte est « une donnée, publique ou privée, qui peut être librement utilisée, réutilisée et redistribuée par quiconque - sujette seulement, au plus, à une exigence d'attribution et de partage à l'identique. » (Open Knowledge Foundation)

Pas de consensus sur définition concrète → plusieurs standards :

- Les 8 principes de l'ouverture des données publiques, adoptés lors de la conférence de Sébastopol (2007)
- Les recommandations pour ouvrir des données, du World Wide Web Consortium (W3C ; 2009)
- Le modèle en 5 étoiles, de Tim Berners-Lee (2010)

Ouverture des données = processus et pratiques pour rendre des données utilisables, réutilisables et redistribuables + « mouvement techno-politique » (Dulong de Rosnay, 2017 : 403)

# **1. Le contexte de l'open data**

# Une progressive mise à l'agenda

- Un **mouvement international** avec un rôle de premier plan joué par des acteurs de la société civile (Goeta 2016)
  - Les premières **définitions des standards d'ouverture des données** viennent d'acteurs socio-économiques et d'activistes du savoir ouvert : rédaction de l'Open Definition par l'OKFN (2005), réunion de Sébastopol (2007), publication du modèle en 5 étoiles de Tim Berners-Lee (2010)
  - ... Avec une **progressive mise à l'agenda des gouvernements** : mise en œuvre de l'Open Government Partnership (2011) et adoption d'une charte par les chefs d'Etat du G8 (2013)
  - Et un **rôle clef joué par l'UE** : directive PSI sur la réutilisation des informations du secteur public (2003 ; 2013), directive INSPIRE sur les données environnementales géographiques (2007), directive open data (2019) qui introduit le concept d'ensemble de données de grande valeur avec liste des jdd concernés listés en 2023 + mécanismes incitatifs.
- La **France fait figure de bonne élève** : en 2021, elle est classée 1<sup>ère</sup> dans l'Open Data Maturity Report de la Commission, notamment grâce à la prise d'initiatives législatives contraignantes et précurseures.
  - **Déclaration des droits de l'homme et du citoyen** (1789) : Art. 15 : la Société a le droit de demander compte à tout Agent public de son administration
  - **Loi CADA** (1978) : Liberté d'accès aux documents administratifs pour quiconque en fait la demande
  - Création de la **mission Etalab** auprès du PM + ouverture de la **plateforme nationale** de la donnée ouverte data.gouv.fr (2011)
  - **Loi pour une République numérique** (2016) : Art. 6 fixe le principe d'ouverture par défaut pour toutes les données administratives, dans le respect de la RGPD

## **2. L'open data dans l'ESR en France**

# L'ouverture des données dans l'ESR : outils et acteurs

Plusieurs plateformes

Qui gère ?

Qui peut publier des jeux de données ?

**data.gouv.fr**

Nationale et communautaire

Etalab  
(DINUM)

Tous (entreprise, administration, citoyen) si données d'IG

Dans l'ESR : quelques universités (ULCO, Paris 13, U. Angers, U. Nanterre) + des opérateurs (ANR, ABES)

Des plateformes locales

Open data Hauts-de-Seine, DataGrandEst, ...

Collectivités territoriales

Selon les plateformes : uniquement la coll.terr. ou n'importe qui si données d'IG

**#dataESR**

Plateforme du Ministère de l'ESR  
+ outils de visualisation

SIES  
(MESR)

Uniquement le MESR

# L'ouverture des données dans l'ESR : outils et acteurs

Plusieurs plateformes

Qui gère ?

Qui peut publier des jeux de données ?

**data.gouv.fr**

Nationale et communautaire

Etalab  
(DINUM)

Tous (entreprise, administration, citoyen) si données d'IG

Dans l'ESR : quelques universités (ULCO, Paris 13, U. Angers, U. Nanterre) + des opérateurs (ANR, ABES)

Des plateformes locales

Open data Hauts-de-Seine, DataGrandEst, ...

+ portail de l'U. Evry

Collectivités territoriales

Selon les plateformes : uniquement la coll.terr. ou n'importe qui si données d'IG

**#dataESR**

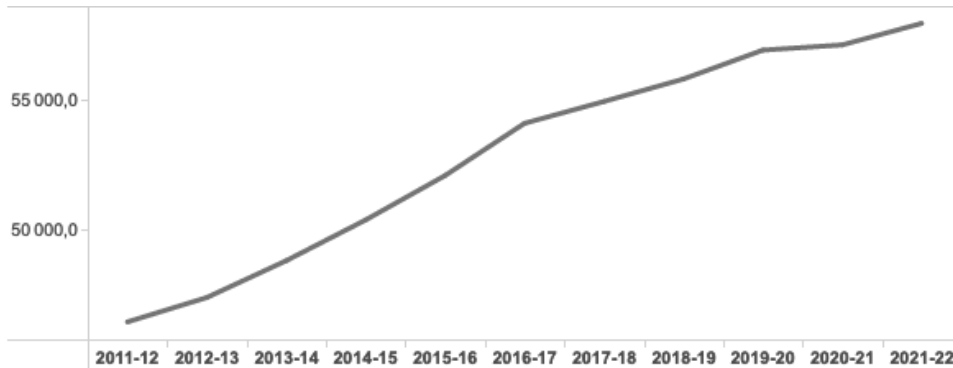
Plateforme du Ministère de l'ESR  
+ outils de visualisation

SIES  
(MESR)

Uniquement le MESR

# Zoom sur ... la plateforme ministérielle #dataESR

**164** jeux de données open data sur l'enseignement supérieur et la recherche



### Sélectionner un indicateur

- 1. Nombre total d'inscriptions
- 2. Nombre d'étudiants inscrits avec doubles inscriptions CPGE...
- 3. Nombre d'étudiants inscrits hors doubles inscriptions CPGE ...
- 4. Nombre de nouveaux bacheliers inscrits
- 5. % de nouveaux bacheliers parmi les étudiants inscrits
- 6. % de doubles inscriptions CPGE parmi les nouveaux bachel...
- 7. Effectifs d'étudiants inscrits dans les diplômes nationaux
- 8. % d'étudiants inscrits dans les diplômes nationaux
- 9. Effectifs d'étudiants inscrits dans les diplômes d'établisseme...
- 10. % d'étudiants inscrits dans les diplômes d'établissement
- 11. Effectifs d'étudiants inscrits en 1er cycle
- 12. % d'étudiants inscrits en 1er cycle

### Valeur moyenne et médiane

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22
<b>Moyenne / Médiane</b>	11 726,4 8 002,0	11 748,0 7 583,5	11 986,6 7 116,0	12 326,9 7 039,5	13 189,8 7 777,0	13 221,6 7 328,5	13 370,1 5 795,0	13 586,1 5 768,0	15 860,1 9 499,0	16 367,1 10 436,0	

### Graphique de dispersion :



### Rentrée universitaire

- 2011-12
- 2012-13
- 2013-14
- 2014-15
- 2015-16
- 2016-17
- 2017-18
- 2018-19
- 2019-20
- 2020-21
- 2021-22

<https://data.enseignementsup-recherche.gouv.fr/pages/home/>

Plateforme développée par la start-up OpenDataSoft et gérée par la Sous-direction des systèmes d'information et des études statistiques du Ministère de l'ESR

164 jeux de données (csv, json, excel) interopérables

Des métadonnées

Des datavisualisations (tableaux de bord interactifs, cartes, ...)

Des API

Un périmètre incomplet (étabs. publics sous tutelle du MESR) et en constante recomposition (fusions, EPE, ...)

Une couverture temporelle réduite



# Zoom sur ... la plateforme de l'Université d'Evry



ACCUEIL **DONNÉES** CARTES GRAPHIQUES VISUALISATIONS

## 22 jeux de données

Tri des jeux Alphabétique  
Modifié Populaires **A-Z**

## Filtres

Trouver un jeu de données...

## Vue

Analyse 22  
Carte 5  
Vue personnalisée 2

## Modifié

2021 12



### Aides financières CVEC

Aides financières attribuées aux étudiants

Producteur Université d'Evry, DVEC

Territoire France

Vie étudiante aides soutien CVEC

Tableau  
Analyse  
Export



### Associations étudiantes

Associations étudiantes de l'Université

Producteur Université d'Evry, DVEC

Territoire France

Vie étudiante associations

Tableau  
Analyse  
Export

<https://univevry.opendatasoft.com/pages/accueil/>

22 jeux de données (csv, json, excel)

Des métadonnées

Des datavisualisations (graphiques, cartes, synthèses, ...)

Des thématiques variées : aides financières CVEC, assos étudiantes, consommation en électricité et eau par bâtiments, étudiants, localisation des points de réparation de vélos, de restauration, des distributeurs de protections périodiques, ...

# Pourquoi s'engager dans l'open data au niveau établissement ?

*Pourquoi ouvrir des données qui sont peu ou pas comparables au niveau national (thématiques, variables, niveaux d'agrégation et référentiels utilisés varient selon les établissements) ?*

## 3 motivations principales

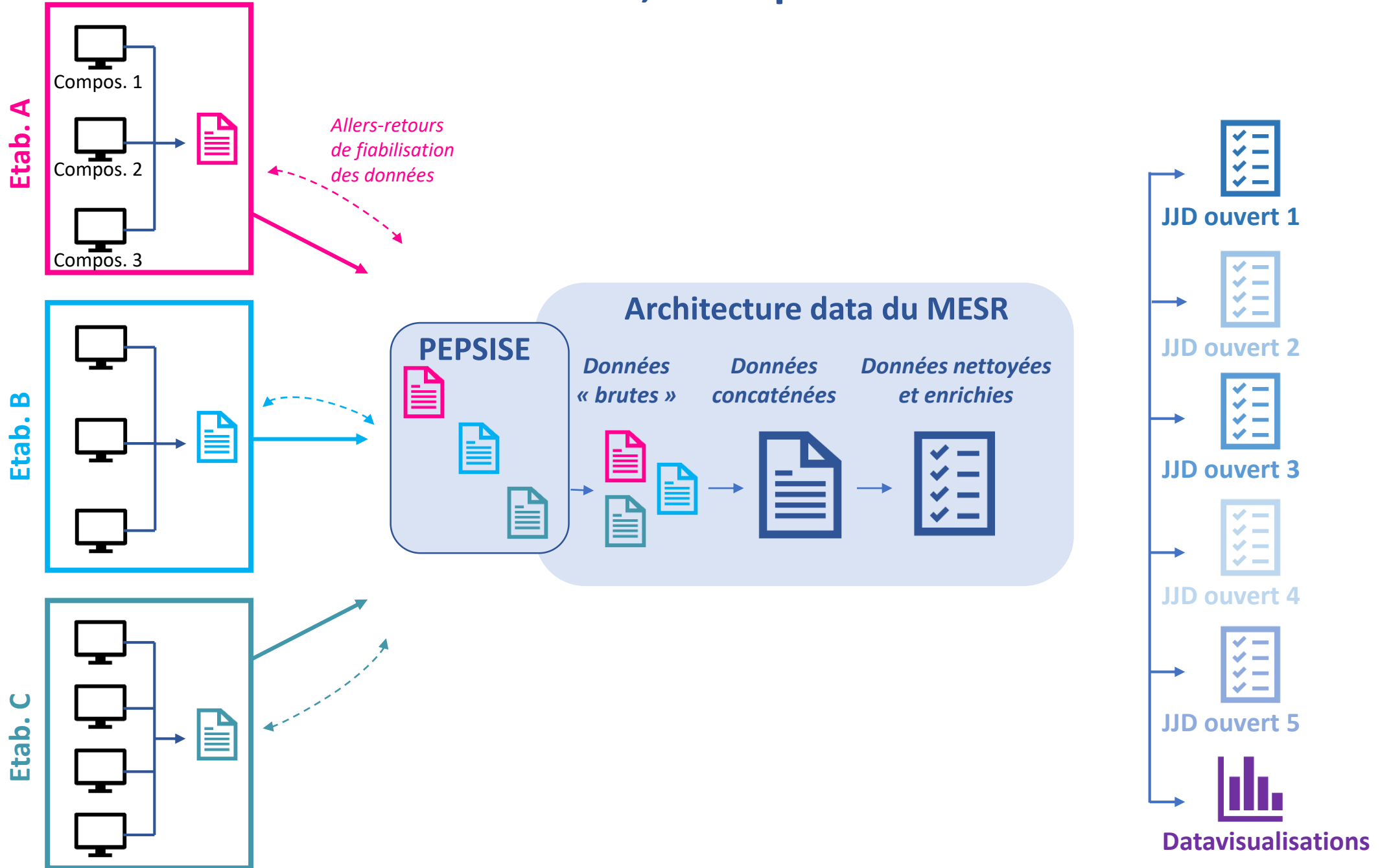
- **Réaffirmer l'université comme un acteur du territoire**
- **Mieux connaître et fiabiliser des données utiles au pilotage de l'établissement**
- **Ouvrir, c'est choisir**

## Quelques recommandations

- **Assumer un portage politique du projet**
- **Intégrer l'initiative open data aux missions d'un·e ou de plusieurs agent·e·s**
- **Travailler de concert avec les acteur·rice·s du territoire**
- **Ne pas sous-estimer les moyens**

### **3. Circulation et transformation de la donnée de scolarité, de la production à l'ouverture**

# La donnée de scolarité, de la production à l'ouverture



Transformation de l'information	Nombre de variables collectées	Exemples
<p><b>Transparence totale.</b> Pas de perte d'information entre la collecte et l'ouverture, à l'exception de celle liée à l'agrégation des données. La variable est conservée à l'identique ou scindée en plusieurs variables en basculant d'une variable de type ouvert (plusieurs valeurs possibles) à fermé (la valeur est le nombre d'étudiants correspondant à la caractéristique indiquée par la variable).</p>	2	<p>La variable 'Sexe' se scinde en deux variables : 'Sexe : M' et 'Sexe : F'  La variable 'Inscription première' se scinde en deux variables : 'Nombre d'étudiants inscrits (inscriptions principales)' et 'Nombre total d'étudiants inscrits (inscriptions principales et secondes)'</p>
<p><b>Transparence partielle.</b> Perte d'information entre la collecte et l'ouverture. La variable collectée est utilisée pour calculer un indicateur</p>	9	<p>L'information contenue dans la variable 'Nationalité' est transformée en 4 variables : 'Nombre d'étudiants inscrits issus des pays membres de l'UE27', 'Nombre d'étudiants inscrits issus des pays membres de l'OCDE', 'Nombre d'étudiants inscrits issus des pays engagés dans le processus de Bologne' et 'Nombre d'étudiants inscrits issus des pays composant le BRICS'.  La variable 'Département d'obtention du baccalauréat' devient la variable 'Attractivité départementale (département d'obtention du baccalauréat)' avec comme valeurs possibles : 'Même département', 'Départements limitrophes', 'Autres départements français', 'Etranger', 'Non bachelier', 'Non renseigné'.</p>
<p><b>Non-transparence.</b> La variable collectée n'est pas ouverte et l'information qu'elle contient a complètement disparu et n'est pas utilisée, même marginalement, dans le calcul d'un indicateur.</p>	48	<p>Notamment toutes les variables permettant une réidentification des étudiants : numéro d'inscription, INE, jour de naissance, mois de naissance, pays, commune et code postal de résidence de l'étudiant, ...  Les variables sur les catégories socio-professionnelles des parents</p>

# Pour aller plus loin ...


La conférence des praticiens de l'ESR (CPESR), une organisation qui produit et diffuse des connaissances sur l'ESR et dont les travaux se basent notamment sur l'exploitation de données ouvertes

<http://cpesr.fr/>

**CPESR** CONFÉRENCE DES PRATICIENS DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

Dossiers ▾ Références ▾ Veille ▾ DataESR ▾ CPESR ▾

🔍 📧 ⚙️ 🐦

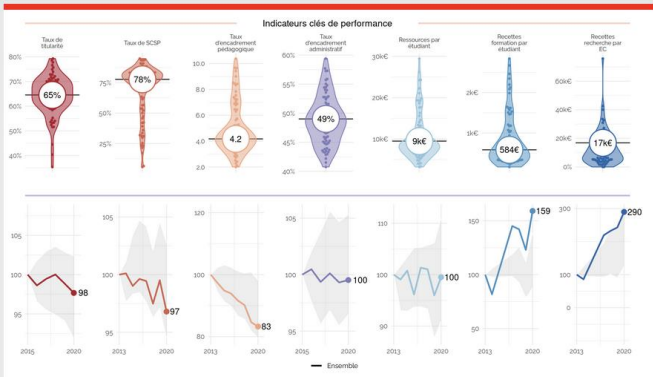


Enquêtes

## BAROMÈTRE DE L'ESR 2023

Ce baromètre vise à mesurer un état de l'ESR au travers de la perception de ses acteurs.

Indicateurs clés de performance

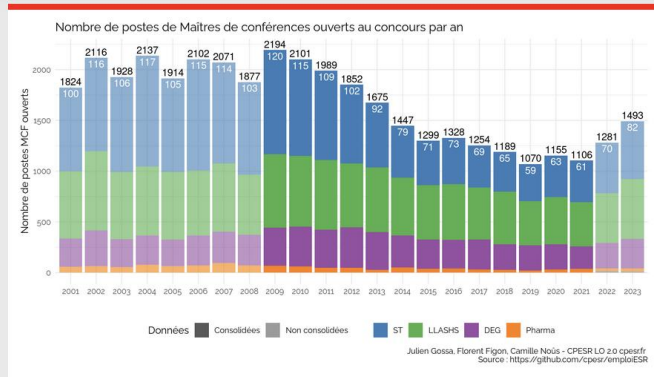


DataESR

## TABLEAU DE BORD DE L'ESR – EDITION 2023

Une sélection d'indicateurs et de représentations graphiques permettant de saisir l'évolution des établissements français.

Nombre de postes de Maîtres de conférences ouverts au concours par an



DataESR

## L'ENTRÉE DANS LA CARRIÈRE DES ENSEIGNANTS-CHERCHEURS – EDITION 2023

Ce document présente les évolutions de différents indicateurs relatifs à l'entrée dans la carrière des enseignants-chercheur.