



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

*Liberté
Égalité
Fraternité*



Les enjeux de la donnée publique dans l'ESR

Politique des données, des algorithmes et des codes sources

Isabelle Blanc

Administratrice ministérielle des données, des algorithmes et des codes sources

<https://www.linkedin.com/in/isabelleblanccatala/>

isabelle.blanc@recherche.gouv.fr



Au programme

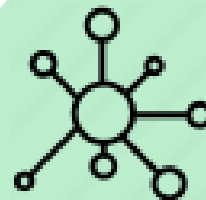
Un peu d'histoire et
de cadre
réglementaire

Les données pour le
pilotage :
De quoi parle-t-on ?
Existent-elles ?

Comment faire pour
que la donnée
circule ?

Ouverture et circulation des données publiques une obligation légale

- 1789 ● Article 15 de la Déclaration des droits de l'homme de 1789 prévoyait que « **la société a le droit de demander compte à tout agent public de son administration** »
- 1978 ● Loi du 17 juillet 1978 intitulé « **de la liberté d'accès aux documents administratifs** »
- 2014 ● Création de la fonction **d'administrateur général des données** auprès premier ministre
- 2016 ● **Ouverture des données publiques** - Loi pour une république numérique 8 octobre 2016
Passage d'une logique de la demande d'un accès à une logique de l'offre de données publiques
- 2020 ● Création de la fonction **d'administratrice ministérielle des données** au MESR, pilotage et mise en œuvre de la politique et de la gouvernance des données, des algorithmes et des codes sources dans ESR
- 2021 ● **Une politique publique de la donnée** déclinée dans chaque ministère
- 2023 ● Partage d'informations entre administration devient obligatoire : « **dites-le nous une fois** »
- 2024 ● **Administration proactive** grâce à la circulation des données
Le partage entre administrations devient obligatoire, traitement créé par la loi
https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000045213315
<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000047541339>
<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000047541361>





Les données, les algorithmes et les codes sources au service de 3 enjeux

1

Innovation

Ouvrir les données

Pour **favoriser leur réutilisation** par
les chercheurs, les enseignants, les
citoyens et par les entreprises

Pour qu'elles deviennent **créatrices
de valeur
scientifique et économique**

2

Confiance

Amplifier l'ouverture des données,
des algorithmes et des codes sources
pour plus de **transparence de l'action
publique**

3

Simplification

Faire de la circulation des données
un outil pour **simplifier et rendre
plus efficaces les actions et les
processus** administratifs

« Dites-le nous une fois ! »

Une politique des données, des algorithmes et des codes sources de l'ESR



Données et
codes sources
de la recherche

Données et
codes sources
de
l'enseignement
supérieur

Données
administratives

Données pour
le pilotage

<http://www.esr.gouv.fr/politique-donnee/>

Bilan de la feuille de route – janvier 2024 : 50/53 actions réalisées ou en cours

Focus sur quelques actions phares



Simplifier la vie des étudiants

- API statuts étudiants et statuts étudiants boursiers pour favoriser le « Dites-le nous une fois »

Transparence, bien commun, pilotage

- Ouverture données des politiques publiques – Data ESR et ScanR
- Ouverture des algorithmes : et données Parcoursup
- Inersup, insertion professionnelle
- Pilotage de l'offre de formation
- Recherche Data Gouv – ouverture et partage des données de la recherche
- Rapport au parlement sur la production et la valorisation du logiciel de recherche

Simplifier la vie des chercheurs et des laboratoires

- Stratégie des Identifiants uniques et pérennes pour le partage et la circulation de la donnée => éviter la resaisie et faciliter le pilotage
- Portail des appels à projet recherche : visibilité de des appels de 27 financeurs de la recherche
- Un référentiel national univoque des structures

Acculturation - formation

- Qualification des métiers de la donnée
- Formation et parcours sur les métiers de la donnée
- Dispositifs d'accompagnement sur le territoire des équipes de recherche dans la gestion et la diffusion des données et codes sources de recherche

Données pour le pilotage : De quelles données parle-t-on ?

Données

Finances
RH
Scolarité
Formations
Diplômes
Insertion professionnelle
Contrats
Partenariats
Immobiliers
Unités de recherche
Publications
Brevets
Données de la recherche
...

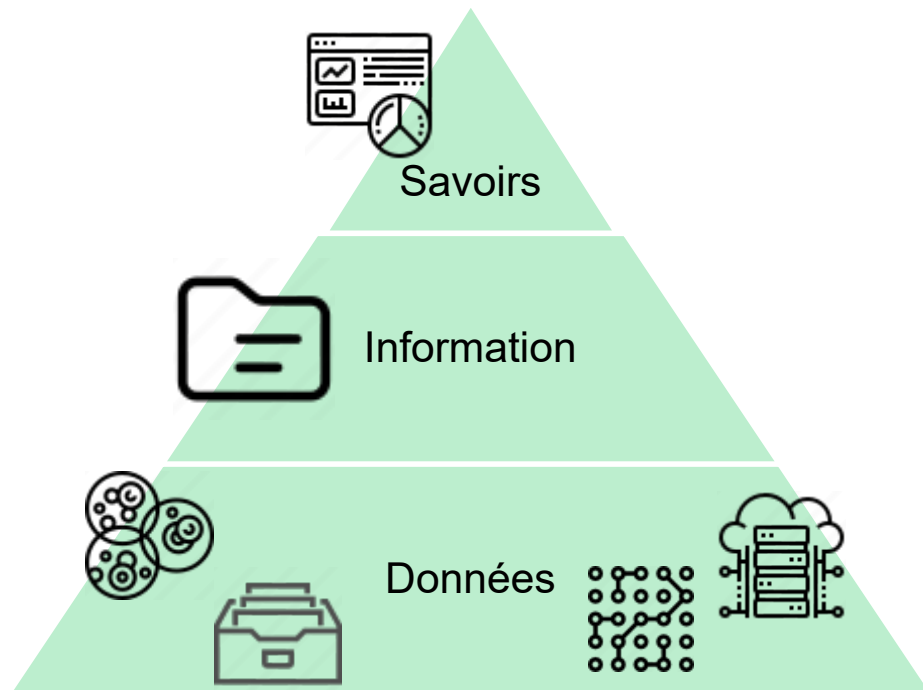


Partenaires

- Universités – Ecoles
- Organismes recherche
- Industries

Données

Les données sont définies comme des enregistrements factuels (chiffres, caractères, textes, images et sons), qui constituent les fondations sur lesquelles l'information et le savoir sont créés



De la donnée ... à l'information

Données

Les données sont définies comme des enregistrements factuels (chiffres, caractères, textes, images et sons), qui constituent les fondations sur lesquelles l'information et le savoir sont créés

Métadonnées

Une métadonnée est une information structurée qui décrit, explique, localise une ressource informationnelle ou la rend facile à trouver, utiliser ou gérer

3 principaux types de métadonnées :

Descriptives : métiers

Structurelles : format, vocabulaires contrôlés, référentiels

Documentaires ou administratives : condition d'utilisation, de préservation, d'accès

Une donnée n'a de valeur que par son usage

Une donnée de gestion prend de la valeur lorsqu'elle est utile au pilotage

Comment piloter l'offre de formation en France ?

A-t-on des données disponibles ?

Ces données sont-elles les mêmes dans chaque établissement ?

A-t-on les moyens de consolider rapidement les données d'une même formation dispensée dans différents établissements ?

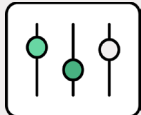
Connaît-on le nombre de diplômés d'une même formation suivie dans différents établissements ?

Connait-on le taux d'insertion professionnelle d'une formation ?

...

Quelle solution ?

Faire de la donnée un bien commun à tous les métiers et toutes les fonctions



Gouvernance des données regroupe une politique, une organisation, des processus, normes, standards

(vocabulaires communs, données de références)

Organisation globale des données produites
ou détenues pour

- Les connaître
- Les décrire
- Les qualifier du point de vue de leur sensibilité

Visibilité sur les données d'intérêt détenues
par d'autres « administrations »



Capital commun de données fiables et compréhensibles

Assurer qualité, fiabilité,
interopérabilité, disponibilité
facilitant l'utilisation et la
meilleure circulation possible
Des données univoques



Accès aux données Tableaux de bord Visualisation à façon Exploitation

Selon les métiers, les structures,
les fonctions et les besoins de
chacun

Pour résumer



Des données pensées à l'origine pour être (ré)utilisables

Des données nativement réutilisables, c'est-à-dire des données décrites selon des vocabulaires partagés

Les données doivent sortir de leur silo métier

Des données univoques

Une gouvernance de l'ensemble des données des différentes entités d'un établissement

Un cadre de référence

Un capital de données commun

Il ne s'agit pas d'harmoniser les outils mais de relier les données

Cadre de confiance partagé entre producteurs et réutilisateurs de données

Des données réutilisées deviennent fiables et valorisent la fonction de production

Des métiers complémentaires à faire travailler ensemble

Les chantiers entamés pour que la circulation des données devienne réalité

La circulation des données sur la recherche entre les tutelles des unités de recherche et pour les unités de recherche



Fluidité des échanges de données

Faciliter la circulation d'informations entre les systèmes d'information des établissements / opérateurs sans imposer un système unique



Réduction de la redondance administrative

Éviter aux chercheurs et unités de fournir plusieurs fois une information
Consolider des informations multi-sources pour une unité



Création d'un capital commun de données

Consolider des données multi-sources
Restituer automatiquement et en temps réel des informations sous forme de tableaux de bord dynamiques aux unités



Adoption d'identifiants uniques et pérennes (PIDs)

Utilisation de PIDs internationaux pour garantir l'unicité et la persistance des données associées, facilitant ainsi leur échange et leur traçabilité

Circulation et partage de la donnée pour produire des indicateurs sur l'insertion professionnelle

CONTEXTE ET ENJEUX

Les données administratives d'insertion professionnelle des sortants du supérieur ne sont pas pleinement exploitées au bénéfice des étudiants.

La production d'indicateurs d'insertion professionnelle des sortants du supérieur permettra de mieux éclairer le choix de formation des étudiants et de contribuer au pilotage de l'offre de formation des établissements.

DEMARCHE

Production d'indicateurs basée sur le rapprochement des fichiers étudiants et de la déclaration sociale nominative

Collaboration entre le ministère de l'enseignement supérieur et de la recherche et la DARES du Ministère du Travail, du Plein emploi et de l'Insertion

INDICATEURS TAUX D'INSERTION

- Décembre 2023 : Taux d'insertion relatifs aux licences pro et master mis à disposition dans Parcoursup, Monmaster et en opendata

Prochaines étapes :

- Mi 2024 : licences générales
- Fin 2024 : Ecoles d'ingénieurs et de commerce
- DUT >> BUT (Bachelor) dès 2025 après sortie de 1^{ère} promotion en 2024
- Doctorat : 2026

Insup prolonge Inserjeunes où figure déjà les taux d'insertion pour les CEP, BEP et les bacs pro