

---

# ACCÉDER AUX DONNÉES ADMINISTRATIVES : NOUVELLES DONNÉES, NOUVEAUX OUTILS

---

Laurent Toulemon, Ined

Statistique appliquée en sciences sociales :  
Révolution des données, évolution des pratiques  
Jeudi 31 mai 2018, Ined



# Accéder aux données administratives : nouvelles données, nouveaux outils

---



- I - Les données deviennent plus accessibles
  - Un nouveau cadre légal
  - Une « révolution » des données
- II - Les modalités pratiques de l'accès
  - Les acteurs : producteurs, diffuseurs, utilisateurs
  - Les produits : fichiers, documentation, autorisations
- III - Associer producteurs et utilisateurs
  - Une nécessité nouvelle
  - Un exemple: le projet Big Stat

# I - Les données deviennent plus accessibles



- Un nouveau cadre légal
  - La loi sur la république numérique d'octobre 2016 rend les fichiers administratifs disponibles
    - Obligation de mise à disposition
  - Appariements de données fondés sur le NIR
    - NIR chiffré, deux tiers de confiance, l'un qui chiffre et l'autre qui apparie
- Une « révolution » des données
  - enrichissement des fichiers, notamment de l'EDP
    - Données fiscales, données de santé
  - Appariement des données administratives

## II - Les modalités pratiques de l'accès



- **Les producteurs**

- L'Insee enrichit ses fichiers statistiques à partir de données administratives

- Les administrations doivent transformer des fichiers administratifs en données statistiques

- Un changement de finalité

- Gestion individuelle, exploitation statistique
    - Unité de traitement (individu, ayant droit)

- Construire de nouveaux fichiers à partir des besoins pour la recherche

- Panélisation, échelle de temps, suivi de quelles entités ?

## II - Les modalités pratiques de l'accès



- **Les producteurs**
- Produire des fichiers utiles et documentés
  - Valider la qualité des données
    - Exactitude des informations
    - Définition du champ, exhaustivité
    - Pertinence des informations pour la recherche
    - Pondérations, imputations ?
  - Quelle documentation produire ?
    - Précision, complétude
    - Validations externes ?
  - Mises à jour des fichiers et de leur documentation

## II - Les modalités pratiques de l'accès

---



- **Les diffuseurs**
- Interface entre producteurs et utilisateurs
  - Expliciter (et harmoniser?) les règles d'accès
- En amont des demandes
  - Respecter le cadre légal pour les procédures
- En aval
  - Respect des règles du secret
  - Organiser les accès
- Accès spécifiques
  - Appariements (tiers de confiance)

## II - Les modalités pratiques de l'accès



- **Les diffuseurs**
- **Comment contrôler les accès ?**
  - Accès dans les locaux des producteurs
  - Diffusion des données à l'utilisateur
  - Accès sécurisé à distance
- **Le CASD**
  - Un équipement d'excellence financé jusqu'en 2019
  - Forme juridique, modèle économique ?

## II - Les modalités pratiques de l'accès



- **Les utilisateurs**
- Des besoins à examiner avec les producteurs
  - Compétence statistique, coût technique d'entrée
  - Adéquation des données aux questions posées
- Des compétences spécifiques
  - Recherche internationale, ouverture
  - Questions inédites, temps à leur disposition
- **Les garants du respect des procédures**
  - Comité du secret statistique, autres comités
  - Commission nationale informatique et libertés



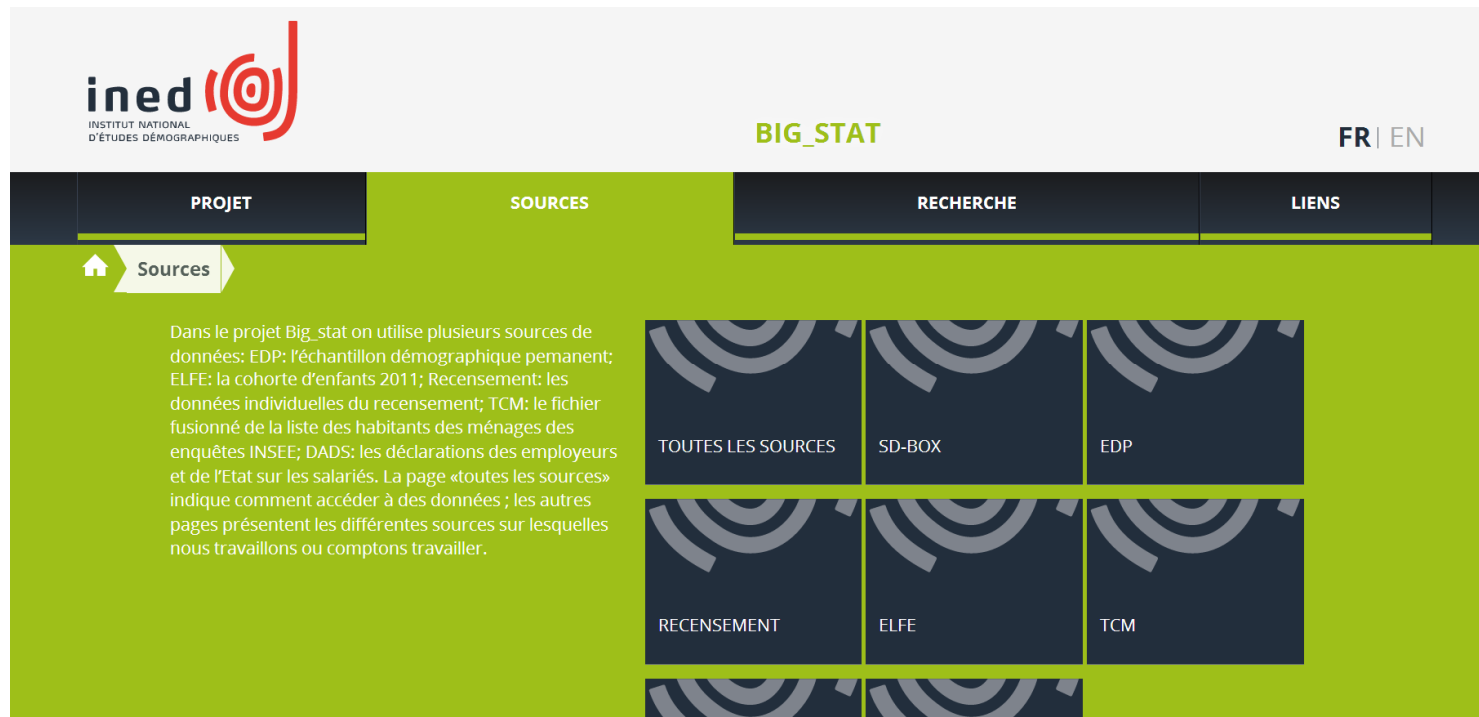
### III - Associer producteurs et utilisateurs

---



- Les producteurs ne peuvent pas tout faire
  - Enrichissement et complexité des données
  - Les producteurs ne peuvent ni valider proprement ni documenter complètement leurs fichiers
  - Un fardeau très (trop ?) important
- Les utilisateurs peuvent avoir un rôle actif
  - Valider les données, compléter la documentation
  - Homogénéiser les modalités d'accès
  - Organiser les retours d'information vers les producteurs

### III - Un exemple concret : le projet Big Stat



- Un projet financé par l'ANR, soumis avec l'Insee
  - Relation entre utilisateurs (Ined, universitaires) et producteurs (Insee)
  - [https://big\\_stat.site.ined.fr](https://big_stat.site.ined.fr)

# Le contenu du projet BigStat

- Recherche  
→ population et situations familiales, jeunes adultes, enfants de parents séparés
- Formation des utilisateurs
- Diffusion et documentation des fichiers
- Retours vers les producteurs, collaborations  
→ Mini-sites dédiés aux sources



# Le contenu du projet BigStat



- Données Insee
  - A l'Insee: Sébastien Durier, Céline Leroy, Stéfan Lollivier, Gaël De Peretti, Isabelle Robert-Bobée
  - Mise à disposition des données par l'Insee
  - Travaux en commun, retour sur les fichiers
  - EDP, Recensement et EAR, Fideli, TCM
- D'autres données
  - Cohorte Elfe : antérieur à Big Stat (Lidia Panico)
  - Cnaf : en cours de mise en place
- Certains sites sont actifs, d'autres pas encore

# Conclusion

---

- Des situations très diverses
  - L'EDP s'enrichit et se complexifie
    - Données fiscales, données de santé
  - Les fichiers Elfe ont un coût d'entrée très important
  - La Cnaf s'est dotée d'un comité spécifique
- Les utilisateurs peuvent et doivent être utiles
  - Complémentarité entre producteurs et utilisateurs
  - Partage du fardeau, prise en charge commune
  - Regrouper les utilisateurs

---

# Merci de votre attention

Laurent Toulemon, [toulemon@ined.fr](mailto:toulemon@ined.fr)

---

