

Développer la culture des données

Quetelet-Progedo-Diffusion

Table ronde de l'assemblée plénière du CNIS

23 janvier 2024

Erik Zolotoukhine, responsable du service de l'ADISP



Les origines de Progedo

Aux origines ... Demandes individuelles des chercheurs auprès de l'INSEE

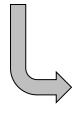


Années 80 : Chercheurs du Lasmas : enquêtes FQP puis enquêtes Emploi

→ Le Lasmas deviendra le Centre Maurice Halbwachs (2004)



Années 90 : Diffusion des données (INSEE) étendue au sein du CNRS



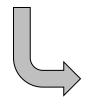
Années 2000: **Diffusion "générale"** via le Réseau Quetelet

- → Champ étendu aux SSM (services stat. ministériels) et autre producteurs "publics"
- → Accès pour l'ensemble de la communauté scientifique (universités, EPST, ...)



2012: Création de Progedo

- → Objectif : développer la culture des données
- → Coordination du Réseau Quetelet

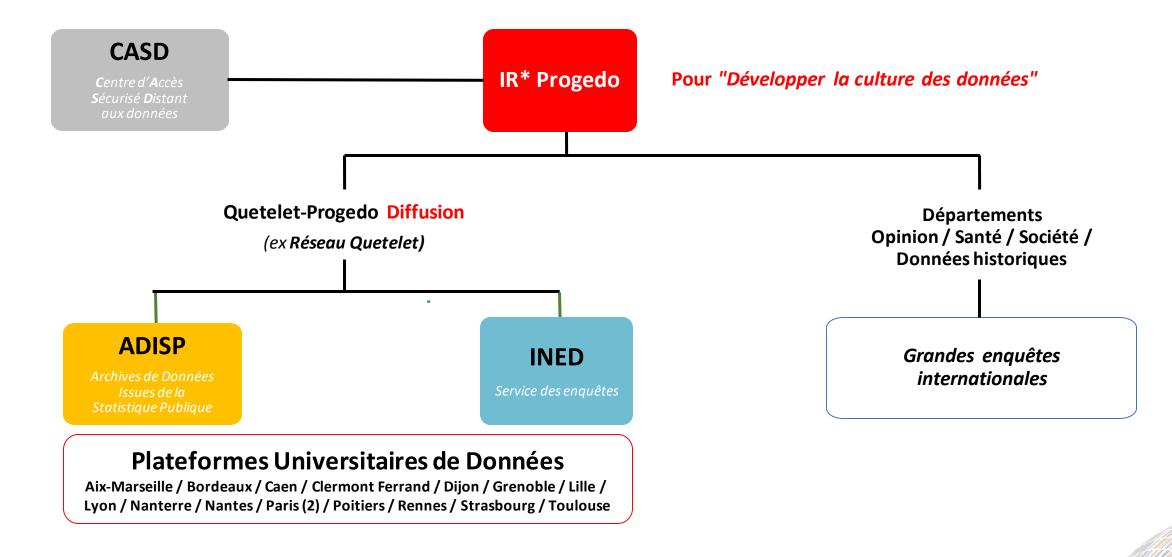


2021 : ouverture aux données de la recherche académique

1999 : Rapport **Silberman**



Progedo aujourd'hui





Les types de fichiers

Données confidentielles



Secure use files



- ✓ NON anonymisées
- ✓ Variables <u>très</u> détaillées (données d'origine, "brutes")



- Accessibles via des accès sécurisés
- > CASD, HDH ...

Données *pseudonymisées*



Terminologie internationale alternative:

Scientific use files



- ✓ FPR (statistique publique)
- ✓ DPR (recherche académique)
- ✓ Variables détaillées



- Diffusion à la communauté scientifique
- > Quetelet-Progedo-Diffusion

Données **Standards**



Public use files



- ✓ Anonymisées
- ✓ Variables agrégées *

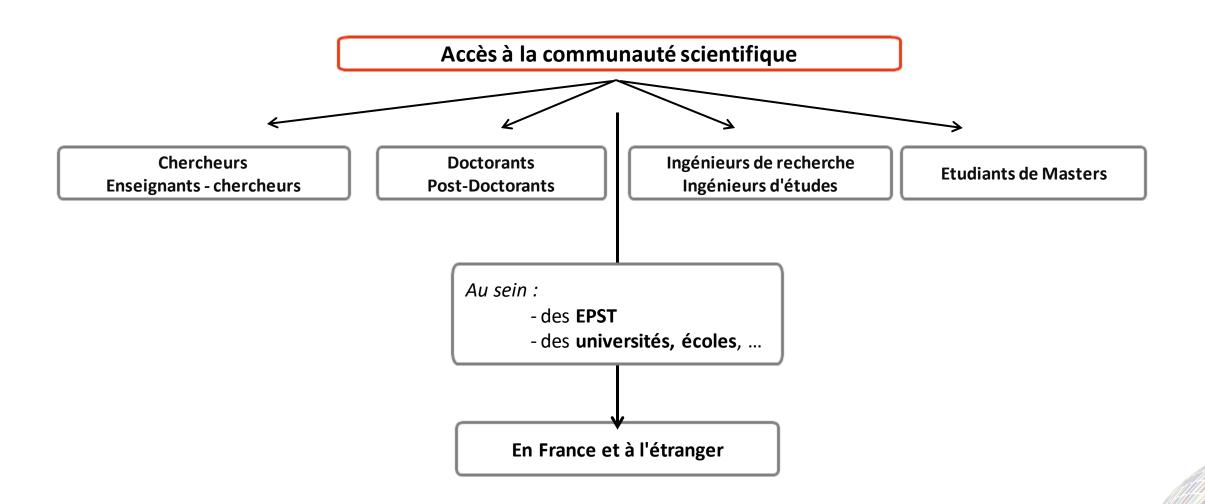


- > Diffusion en Open data
- > Insee.fr, data.gouv.fr, ...

(*) Agrégations le plus souvent opérées sur les variables géographiques, de professions et autres nomenclatures ; ainsi que sur des variables numériques synthétisées en tranches



Les publics de Quetelet-Progedo-Diffusion





Le catalogue de l'ADISP

les producteurs









MINISTÈRE DE L'ÉDUCATION NATIONALE

DEPS

Ministère de la Culture











ISSP







CENTRE POUR LA RECHERCHE ÉCONOMIQUE ET SES APPLICATIONS





les données

plus de 1500 références :

(Insee: 75% / SSM: 15% / Autres: 10%)

- essentiellement des données de la statistique publique
- quelques données produites par la recherche académique

Le catalogue de l'INED

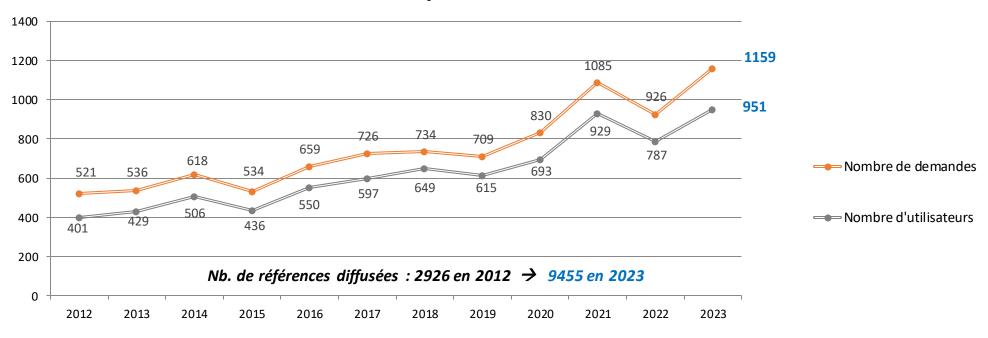
70 références

(enquêtes (co-)produites par l'Ined)



Quelques statistiques de diffusion

Nombre de demandes / d'utilisateurs



Les utilisateurs (en 2023) :

Statut

chercheurs: 30%
doctorants: 20%
ingénieurs: 5%
étudiants: 45%

Discipline

économie: 42%
sociologie: 30%
géo. / démog.: 7%
sciences pol.: 5%

- ...

<u>Origine</u>

Ile de France : 52%
 région : 37%
 étranger: 11%