

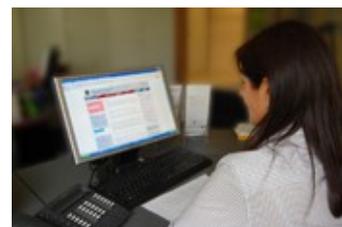
Quels usages possibles des données privées pour les statistiques publiques ?

Enjeux et premières expérimentations

Pauline Givord (DMCSI)



Mesurer pour comprendre



Contexte

- Phénomène « Big Data » ou « déluge de données », intérêt croissant pour l'exploitation des données privées
- Données caractérisées par :
 - De très grands volumes (données issues de l'internet, de capteurs),
 - De formats variés (peu structurées, textes, images...)
 - Des flux importants
- Réflexions des différents INS sur la possibilité d'utiliser ces données dans la production statistique
 - Projets les plus aboutis : données de caisse, données de téléphonie mobile
- Dans un contexte de productions d'indicateurs statistiques par de nouveaux opérateurs (privés)
 - Billion Prices Project
 - ADP
 - Flux Vision (Orange)

Opportunités pour la statistique publique

- Disponibilité immédiate de l'information : réduire les délais de publication
- Nombre d'observations très élevé : publication à des échelles (territoriales, population, produits...) plus fines
- Données « objectives »: compléter les indicateurs existants (réduire coût et charge d'enquête)
 - Exemples : temps de transport, dépenses précises...
- Compléter le dispositif existant (économie numérique, indicateur de développement durable...)

Difficultés identifiées

- Accès aux données, en général privées : doit être encadré par un cadre juridique
- Garantir la confidentialité :
 - Utilisation de données personnelles ; risque de réidentification augmenté par la tentation/nécessité d'appariement des données
- Difficile de qualifier la qualité de ces données en termes de représentativité, mais aussi de **pérennité**
- Nécessité de disposer d'outils (logistiques et statistiques) et de compétences adaptés

Coordination statistique internationale

- Plusieurs groupes de travail auquel le SSP participe
- **Système Statistique Européen**
 - Mémorandum de Scheveningen (2013), avec la mise en place d'une Task Force Big data en 2014
 - Objectif : mutualiser, acquérir de l'expérience, identifier les sources intéressantes
 - « ESSNET » en 2016-2019, sur des projets pilotes (concrets) : offre d'emploi, données mobile, compteurs intelligents...
- **Coopération internationale :**
 - Projet Unece « big data in Official Statistics » : « Sandbox » (2015), plateforme « bac à sable » pour tester les technologies (infrastructures et logiciels) spécifiques

Les expérimentations à l'Insee

- Projets « données de caisse »

Projet lancé en 2015 (après phase expérimentale en 2011), avec objectif d'intégrer dans la production en 2019

Remarque : déjà en production en Norvège, Pays Bas, Suisse, Suède, bientôt Belgique, Danemark, Luxembourg, Pologne

- Projets expérimentaux, en particulier avec la création d'un poste de R&D sur ces questions au département des méthodes statistiques depuis septembre 2014 (renforcé en sept 2016)

Objectif : identifier les compétences nécessaires, les données intéressantes (et à quel niveau d'agrégation minimum)

Utilisation des requêtes Internet pour la prévision économique (« nowcasting »): dossier dans la note de conjoncture de Mars 2015

Données de téléphonie mobiles (partenariat Orange Lab-Eurostat, en cours) : prévision de population présente, modélisation des flux (qualification du territoire)

Données massives et Statistique Publique

Merci de votre attention !

Contact :
Pauline Givord
pauline.givord@insee.fr

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :
www.insee.fr / Contacter l'Insee
+33 9 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00