

# *Les exemples européens d'utilisation de données privées et les perspectives*

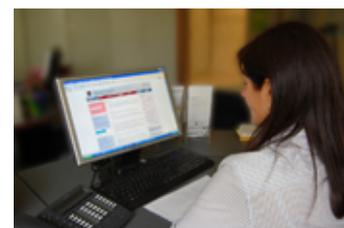
---

Commission Cnis  
"Entreprises et stratégies de marché"

F. Dupont



Mesurer pour comprendre



# Différents types de données

---

## Phénomène de « Big Data » en général

- **De nouvelles sources de données**
  - Grands volumes de données (issues d'internet, de capteurs)
  - Formats variés (peu structurés, textes, images)
- **Des méthodes de traitement nouveaux** (statistique, informatique)

## En pratique pour la statistique publique

- **Sources nouvelles** (réseaux sociaux, moteur de recherche, téléphonie mobile) ou **sources plus faciles à traiter**, d'accès plus facile, avec un plus grand détail (SNIIRAM, données satellites, compteurs intelligents d'électricité)
- **Publiques gratuites** : données satellites, réseaux sociaux
- **Privées** : données de caisse, téléphonie mobile, cartes bancaires

# Des réflexions au niveau européen

---

## Groupe de travail européen (Essnet) Big Data

- **20 pays** dont la France
- Février 2016- Mai 2018
- **Explorer le potentiel pour la statistique publique sur des données réelles**
- Sources choisies en fonction des priorités et de la disponibilité
  - Données téléphonie mobile
  - Portails d'offre d'emploi
  - Données compteurs intelligents d'électricité
  - Données de position des navires
  - Données des sites des entreprises

# Les cartes bancaires

---

**Les données sont collectées pour la ligne voyage de la balance de paiements en France**

- **Données agrégées**
- **Non nominatives**

**Comité européen des statistiques monétaires financières et de la balance des paiements CMFB (BCE, Banques centrales, Instituts de statistiques Eurostat..)**

- **Recommande un état des lieux pour second semestre 2016.**

**Source non retenue dans le groupe de travail Big Data Eurostat en raison des difficultés d'accès à court terme**

# Données de la téléphonie mobile

---

**Des expérimentations sur les données menées en partenariat sont nécessaires pour statuer sur le potentiel et préciser les demandes aux opérateurs**

- **Belgique** Proximus, **Espagne** Telefonica, **France** Orange

**Mais les accès sont difficiles**

- **Limites Régulation Télécom**
- **Limites Autorités de Protection des Données, données sensibles**
- **Limites en raison du marché développé par les opérateurs**

**Discussions sur la bonne façon d'avancer**

- **Eurostat, DG Connect, opérateurs, instituts de statistique 22, 23 sept**
- **Recommandations ONU pour l'accès aux sources privées**
- **Eurostat : revue des aspects législatifs et éthiques (2015-2017)**

# Données des sites internet d'offre d'emploi

---

## Objectif : améliorer les statistiques d'offre d'emploi

- **Source** : offre d'emploi proposées sur des portails spécialisés
- **Atout** : plus fréquent, information sur les compétences, détail géographique
- **Différentes sources**
  - Site d'offre d'emploi
  - Portail de moteur de recherche d'emploi agrégateur
  - Utilisation des données d'un acteur public ?
  - Offres d'emploi sur les sites des entreprises
- **Différents modes d'accès**
  - Webscrapping
  - Accord avec les portails

# Données des sites internet d'offre d'emploi

---

## Problèmes à traiter

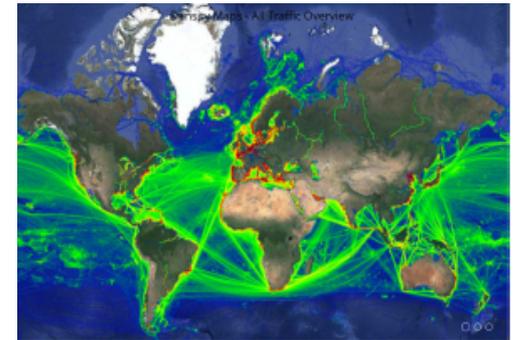
- **Ecarts de définition** : une annonce pour plusieurs lieux, plusieurs emplois, pas forcément vacant,
- **Doublons** : une annonce parue sur plusieurs sites
- **Champ couvert variable** (régional/national, tous secteurs/spécialisé)
- **Codification de l'information** : les informations utilisent des classifications de profession, des compétences non normalisées
- Suivre les évolutions des plateformes (nouvelles, contenu, ...)

# Données de position des navires

---

**Objectif : améliorer les statistiques de trafic existantes, en proposer d'autres en particulier pour estimer les émissions de CO2**

- Pays Bas, Grèce, Norvège, Danemark
- **AIS (automatic identification system)**
  - identifiant, position, trajet, vitesse...
  - par satellite ou stations terrestres
- **Trouver le bon accès**
  - **International** : Marine traffic données payantes depuis 2009
  - **Européen** : European Maritime Safety Authority (EMSA)  
Dirkzwager : entreprise privée
  - **National** : administrations côtières, accord possible mais seulement pour les données nationales



# Données sur les compteurs intelligents

Des différences importantes en terme de déploiement pour 2020 et d'ici 2020 et des possibilités d'accès aux données très différentes :

NSI	Plans to explore smart meters?	Legal obstacles	Data hub
Sweden	Yes	Yes	No
<b>Norway</b>	<b>Yes</b>	<b>No</b>	<b>No</b>
Hungary	Yes	No	No
Lithuania	No	No	No
Cyprus	No	No	No
Bosnia and Herzegovina	No	No	No
Poland	Yes	Yes	No
Belgium	No	No	No
Germany	Yes	Yes	No

NSI	Plans to explore smart meters?	Legal obstacles	Data hub
Portugal	No	No	No
Luxembourg	No	NA	No
The former Yugoslav Republic of Macedonia	No	NA	No
<b>Denmark</b>	<b>Data received</b>	<b>No</b>	<b>Yes</b>
<b>Estonia</b>	<b>Data received</b>	<b>No</b>	<b>Yes</b>
<b>Austria</b>	<b>No</b>	<b>Yes</b>	<b>No</b>
Greece	Yes	Yes	No
Spain	No	No	No

# Données des compteurs intelligents

---

**Objectif : améliorer les statistiques de consommation d'électricité des ménages, des entreprises, et celles sur les logements vacants**

- **Estonie, Danemark, Autriche , Suède**
- **A ce stade, essentiellement une difficulté d'accès aux données**
  - Estonie données de 2013 -2014 (1ere demande en 2013)
  - Danemark données 2013 faible taux de couverture
  - Suède, Autriche pas de données

Country	Deployment strategy	Implementation speed	Penetration rate by 2020	Data refresh rate
<b>Austria</b>	Mandatory roll-out	<b>2012-2019</b>	<b>95%</b>	15 minutes
<b>Denmark</b>	Mandatory roll-out	<b>2014-2020</b>	<b>100%</b>	15 minutes
<b>Estonia</b>	Mandatory roll-out	<b>2013-2017</b>	<b>100%</b>	1 hour
<b>Sweden</b>	Voluntary	<b>2003-2009</b>	<b>100%</b>	1 hour

# Données des sites d'entreprises

---

**Objectif : utiliser les informations disponible sur les sites d'entreprises pour répondre à certaines questions de l'enquête européenne TIC**

- **Italie, Pays Bas, Pologne, Bulgarie, Suède, RU**
- **Source** : sites des entreprises + webscrapping
- **Variables**
  - Pratique le e commerce
  - Offre des emplois
  - Présence dans les médias sociaux

# Conclusion

---

**S'appuyer sur les pays plus avancés pour évaluer le potentiel d'une source**

**Mener des expérimentations nationales en raison des spécificités nationales qui jouent sur l'existence et la qualité d'une source**

**Echanger au niveau international sur les aspects de protection des données, les aspects éthiques**