

# Le projet de construction d'un répertoire d'entreprises en Palestine

*Ph. Brion (inspecteur général honoraire Insee)*

*Séminaire répertoires d'entreprises et classifications  
économiques*

*Libourne, 10 juillet 2017*

# Plan

- 1. Contexte
  - 2. Fabrication d'une version « zéro » du répertoire
  - 3. Mise en place d'un système « permanent »
  - 4. Les utilisations en tant que répertoire statistique
- *NB: cette présentation n'a pas fait l'objet d'une relecture par le PCBS*

# 1. Contexte

# 1.1. Contexte

- Le PCBS (*Palestinian Central Bureau of Statistics*) construit son répertoire statistique à partir du recensement économique réalisé tous les cinq ans, à partir d'une opération de terrain exhaustive, et le met à jour à partir de ses enquêtes statistiques
- Le problème posé par l'utilisation de ce fichier en tant que base de sondage pour les enquêtes économiques est celui de son actualisation :
  - créations d'entreprises
  - cessations
  - mise à jour des informations relatives aux entreprises

## 1.2. Contexte

- Par ailleurs, le PCBS s'est vu confier la mission de présider un comité national chargé de développer un répertoire national d'entreprises partagé avec l'ensemble des administrations
- La France fournit une contribution à ce projet, en termes d'assistance technique (Expertise France, Insee)

## 1.3. Contexte (suite)

- Les principaux partenaires concernés sont, en plus du PCBS :
  - Le Ministère de l'Économie
  - Le Ministère des Finances
  - Les Municipalités
  - Les Chambres de commerce
  - *Il faut noter qu'il n'y a pas d'organisme de sécurité sociale*

## 1.4. Contexte (suite)

- Chaque administration possède son répertoire, pour ses propres besoins
- Cependant, du fait que les finalités ne sont pas les mêmes, différentes questions sont posées pour la mise en commun d'informations :
  - les différents fichiers ne sont pas forcément complets
  - pas forcément à jour
  - les unités utilisées (en particulier les unités économiques enregistrées) sont différentes d'un fichier à l'autre, et au sein d'un même fichier ...

## 1.5. Contexte (suite)

- Questions posées par les différents fichiers (suite):
  - les unités peuvent y être dupliquées
  - Les variables d'identification utilisées ne sont pas les mêmes d'un répertoire à l'autre
  - La qualité des informations renseignées est variable (par exemple sur les noms, les adresses, les numéros de téléphone, les identifiants « numéro de carte d'identité », « numéro de compagnie »)

## 1.6. Contexte (suite)

- Aboutir à la mise en place d'un répertoire partagé alimentant les besoins de chacun des partenaires présente des avantages :
  - exhaustivité
  - « qualité » des variables caractérisant les unités

## 1.7. Contexte (suite)

- Remarques liées aux « points forts » correspondant aux finalités des fichiers des organismes partenaires :
  - Ministère de l'Économie : fournit un identifiant spécifique nécessaire pour exporter → quasi exhaustif sur les grandes entreprises
  - Ministère des Finances : enregistre des contribuables ; existence de liens entre contribuables, entreprises et établissements

## 1.8. Contexte (suite)

- Remarques liées aux finalités des fichiers des organismes partenaires (suite) :
  - Chambres de commerce : enregistrent les entreprises de leur champ
  - Municipalités : très bon suivi de terrain, en particulier en vue de collecter des taxes locales ; éventuellement, enregistrements dupliqués (quand taux de taxation différents, quand mauvais repérage de la continuité entre deux suivis annuels)

# 1.9. Contexte (suite)

- Deux préalables à la mise en place d'un répertoire partagé:
  - Accord sur les unités (et sur la capacité à lier les niveaux entreprise, établissement, propriétaire), et sur les variables qui servent à les identifier (afin que chaque partenaire les collecte lors de ses opérations propres de gestion de son fichier)
  - Accord sur le contenu du répertoire
- Ces deux sujets ont fait l'objet d'une concertation au sein de groupes de travail animés par le PCBS

# 1.10. Contexte (suite)

- Concernant les variables d'identification, elles sont de différents ordres :
  - Deux numéros sont collectés de manière plus ou moins poussée par les différents partenaires
    - Le numéro de la carte d'identité du propriétaire dans le cas d'une entreprise individuelle
    - Le numéro donné par le Ministère de l'économie aux entreprises qu'il enregistre (environ 15%, essentiellement les plus grandes)
  - Le numéro de téléphone (en fait deux possibles : fixe, portable)
  - Le nom de l'entreprise (en fait deux : le nom commercial, le nom du propriétaire)

# 1.11. Contexte (suite)

- Concernant les variables contenues dans le répertoire, on envisage, en plus des variables d'identification :
  - Éléments d'adresse (dont site web)
  - Activité principale
  - Variable de taille
  - Liens entre unités

## 1.12. Contexte (suite)

- Dans un premier temps, il a été décidé de limiter l'exhaustivité du répertoire partagé aux établissements de plus de 20 salariés, sauf sur certaines zones (par exemple Ramallah) où l'on va d'emblée essayer d'avoir un répertoire exhaustif (le fichier de la Municipalité étant jugé de bonne qualité, du point de vue de l'exhaustivité de son ratissage de terrain annuel)

## 2. Fabrication d'une version « zéro » du répertoire

## 2.1. Fabrication de la version « zéro »

- Travail préalable de « nettoyage » sur les différents fichiers, en particulier :
  - Sur les noms
  - Sur les n° de téléphone

## 2.2. Fabrication de la version 0 (suite)

- Appariement : utilisation du programme DUKE
- Pas d'identifiant commun facilement utilisable
- Donc travail à partir de clés « primaires », dans un premier temps ... :
  - N° carte identité
  - Identifiant donné par le Ministère de l'économie
- ... « Secondaires », dans un deuxième temps
  - Nom du propriétaire et nom commercial
  - N° de téléphone

## 2.3. Fabrication de la version 0 (suite)

- Le programme DUKE fournit des propositions d'appariement à partir de fonctions de distance (ici distance de Levenshtein)
- ... qui sont ensuite validées de façon manuelle
- Il est nécessaire de procéder à des réglages « par itérations » du programme pour réduire les appariements « oubliés », ainsi que d'avoir trop d'appariements proposés à tort
- Le travail se fait de façon séquentielle : on commence par les cas les plus simples à valider, et on travaille ensuite sur les cas complexes

## 2.4. Fabrication de la version 0 (suite)

- Le travail d'appariement réalisé a permis d'avoir un retour sur chacun des fichiers utilisés en entrée
  - Nécessité de remplir des « fiches qualité » pour chacun d'entre eux (nombre d'unités, contenu en termes de variables, % de données manquantes, informations sur le travail de nettoyage nécessaire, etc.)

## 2.5. Fabrication de la version 0 (suite)

- Résultats obtenus sur Ramallah, sur les « compagnies »
  - 601 compagnies ont pu donner lieu à un appariement avec au moins deux fichiers ...
  - ... dont 328 appariées pour les 4 fichiers (Municipalité, Ministère de l'Économie, Ministère des Finances, Chambre de Commerce) et 119 dans au moins 3 fichiers

### 3. Mise en place d'un système « permanent »

# 3.1. Mise au point d'un répertoire « vivant »

- A partir de la version « zéro », nécessité de mettre à jour, avec un circuit d'échanges d'informations :
  - contenu et fréquence des échanges
  - Stocks ou flux ?
  - règles concernant les arbitrages nécessaires en cas de « conflits »
- Réflexions en cours sur l'identifiant à mettre en place

## 3.2. Mise au point d'un répertoire « vivant »

- Également, mise au point d'un ensemble de métadonnées :
  - Source de l'information
  - Date d'arrivée de l'information
  - Date de l'évènement signalé
  - ...
- Difficultés à sensibiliser les partenaires concernant les dates des évènements

## 4. Les utilisations en tant que répertoire statistique

## 4.1. Les utilisations en tant que répertoire statistique

- A l'heure actuelle, c'est le Recensement des entreprises réalisé tous les cinq ans qui sert de base de sondage, mais il ne peut prendre en compte les créations d'entreprises ...
- Même s'il y a utilisation d'estimateurs intégrant des données de l'enquête emploi pour redresser les résultats des enquêtes entreprises, celle-ci s'est révélée peu efficace...
- *Référence : Using household surveys to maintain the statistical business register, Saleh Alkafri, PCBS, papier à présenter au congrès de l'IIS 2017, Marrakech*

## 4.2. Les utilisations en tant que répertoire statistique (suite)

- Le répertoire ainsi construit permet d'adosser un répertoire statistique
  - attention, certains unités disponibles uniquement dans le fichier du recensement du PCBS ne sont pas intégrées dans le répertoire inter-administratif
- Au lieu d'un point de référence tous les cinq ans, ce répertoire bénéficiera des actualisations du répertoire partagé. Il pourra également être enrichi par les enquêtes statistiques
- Les variables « activité » et « taille » seront très utiles pour stratifier les plans de sondage

## 4.3. Les utilisations en tant que répertoire statistique (suite)

- On peut penser que la couverture du répertoire constitué à partir du répertoire inter-administratif sera de bien meilleure qualité : pour les grands établissements, immédiatement, pour les autres au fur et à mesure de l'introduction des zones couvertes exhaustivement ...
- Des travaux d'évaluation de cette couverture (en particulier pour les zones couvertes exhaustivement) devraient cependant être menés (problème du poids du secteur informel)

# Questions ?

Philippe.brion55@gmail.com