



# Hackathon

« les champs de Sirene »

exemples d'identification difficile



Quelques exemples pour illustrer les problèmes d'identification:

Même si ces exemples réels sont issus de l'enquête Emploi, les problèmes rencontrés avec la base du recensement sont similaires. On collecte la raison sociale et l'adresse de l'établissement employeur. L'appariement est réalisé via une application peu souple qui suppose que l'adresse existe dans la base et que la raison sociale soit exacte. Ces contraintes sont contournées avec le moteur de recherche présent sur <http://www.sirene.fr/sirene/public/accueil>. Dans les exemples qui suivent, selon les cas, en passant par le moteur Sirene on retrouve bien une partie des établissements. Toutefois, il est nécessaire de passer par un moteur de recherche sur Internet.

Dans tous les cas, avec l'application assurant l'appariement automatique, l'**adresse incomplète** empêche a priori l'identification et pour la plupart la **raison sociale collectée** est insuffisante ou ne correspond pas à celle enregistrée dans le répertoire.

cas	Raison sociale collectée	Adresse collectée	Établissement repris	Activité déclarée
1	MAISON DE RETRAITE	59375 MARCHIENNES	MAISON DE RETRAITE EMILE DUBOIS 2 RUE D ORCHIES 59 MARCHIENNES	87 MAISON DE RETRAITE
2	IME	59192 EMERCHICOURT	PAPILLONS BLANCS ARRONDISSEMENT DE DOUAI 7 RUE DE L EGALITE 59 EMERCHICOURT	87 IME
3	DEMECO	62249 COURCELLES-LES-LENS	SARL MERY - ARTDEM TRANSPORTS LOGISTIQUE MERY RUE DE L ABBE POPIELUSZKO 62 COURCELLES LES LENS	49 DEMENAGEMENTS
4	DENISART JEAN PHILIPPE	59387 MARQUETTE-EN-OSTREVANT	GENERATION ESPACE VERT - G.E.V 2 RUE PASTEUR ENTREE 5 59 MARQUETTE EN OSTREVANT	81 ESPACES VERTS
5	TEREOS	RUE D ERES 59122 CAMBRAI	TEREOS FRANCE RUE D'ERRE 59 ESCAUDOEUVRES	10 AGROALIMENTAIRE
6	YMERIS	RTE WAHAGNIES 59462 PHALEMPIN	IMERYS TC - IMERYS TOITURE ROUTE DE WAHAGNIES 59 PHALEMPIN	41 BATIMENT

Les cas 1 et 2 de ce tableau sont retrouvés dans le moteur Sirene. Les cas 3, 4, 5 et 6 ne sont pas retrouvés via le moteur Sirene pour différentes raisons.

Pour le cas 3, il semble que la SARL Mery- Artdem soit une filiale ou travaille pour DEMECO. On retrouve cet établissement en première page d'une recherche Google en tapant Demeco et Courcelles les Lens. Le premier écho de cette recherche donne la société ARTDEM, agence de déménagement à Courcelles les Lens, et l'adresse du site commence par [www.demeco.fr/](http://www.demeco.fr/)[...]

Pour le cas 4, on retrouve cet établissement en première page d'une recherche Google en tapant Jean-Philippe Denisart et Marquette en Ostrevant mais cela demande : d'identifier qu'il y a une erreur d'orthographe sur le nom (en fait, le chef de l'entreprise s'appelle Denizart), de prendre en compte l'activité déclarée (espaces verts). À ce propos, si on tape Génération espaces verts (au lieu de Génération espace vert) dans raison sociale et Marquette en Ostrevant dans commune sur l'API Sirene, on n'a pas d'écho.

Dans le cas 5, on retrouve cet établissement en première page d'une recherche Google en tapant Tereos Cambrai. Plusieurs informations sur cette première page doivent être lues pour arriver à

identifier l'établissement. On s'aperçoit qu'il existe bien une société Tereos près de Cambrai, on remarque que l'on trouve un établissement Tereos, rue erre à Escaudoevres, libellé d'adresse proche du libellé recueilli : rue d'eres. Un contrôle via l'API Sirene permet de vérifier que cet établissement existe bien, et son activité « Fabrication de scure » correspond bien au libellé d'activité recueilli « Agroalimentaire ».

Dans le cas 6, on retrouve cet établissement en première page d'une recherche Google en tapant Ymeris et l'adresse collectée. Cela permet de récupérer la bonne orthographe et de voir que le libellé d'activité recueilli était imprécis : « Bâtiment » mène au la NAF-2 41 alors que la personne travaille dans le secteur 23. C'est une illustration du défaut de qualité d'un code NAF à partir du seul libellé même en NAF2 (voire en NAF-1 dans ce cas, secteur C vs secteur F).