

Le codage de l'établissement employeur

Hackathon « Les champs de Sirene »

Lucile Vanotti : Division Méthodes et
Traitements des Recensements (DMTR)



Mesurer pour comprendre



Plan de la présentation

- I) Présentation**
- II) La méthode actuelle**
- III) Difficultés rencontrées**
- IV) Opération qualité**

I) Présentation

La France : 67 Millions d'habitants (RP2014 – Cumul des EAR 2012 à 2016) :

- **34** millions en petites communes (PC)
- **33** millions en grandes communes (GC)

Une campagne de recensement (EAR2017) interroge :

- **20%** des petites communes : 6,8 millions
- **8%** des grandes communes : 2,6 millions

→ 9,5 millions de BI

Exploitation complémentaire

- **1/5** des BI dans les petites communes : 1,4 million
- Ensemble des BI des grandes communes : 2,6 millions

→ 4,2 millions de BI (4 compl. + 0,2 EDP)

Codification :

Champ de la codification : actifs occupés (environ 40%):

→ 1,7 million de BI

I) Présentation

Le codage : en quoi ça consiste ?

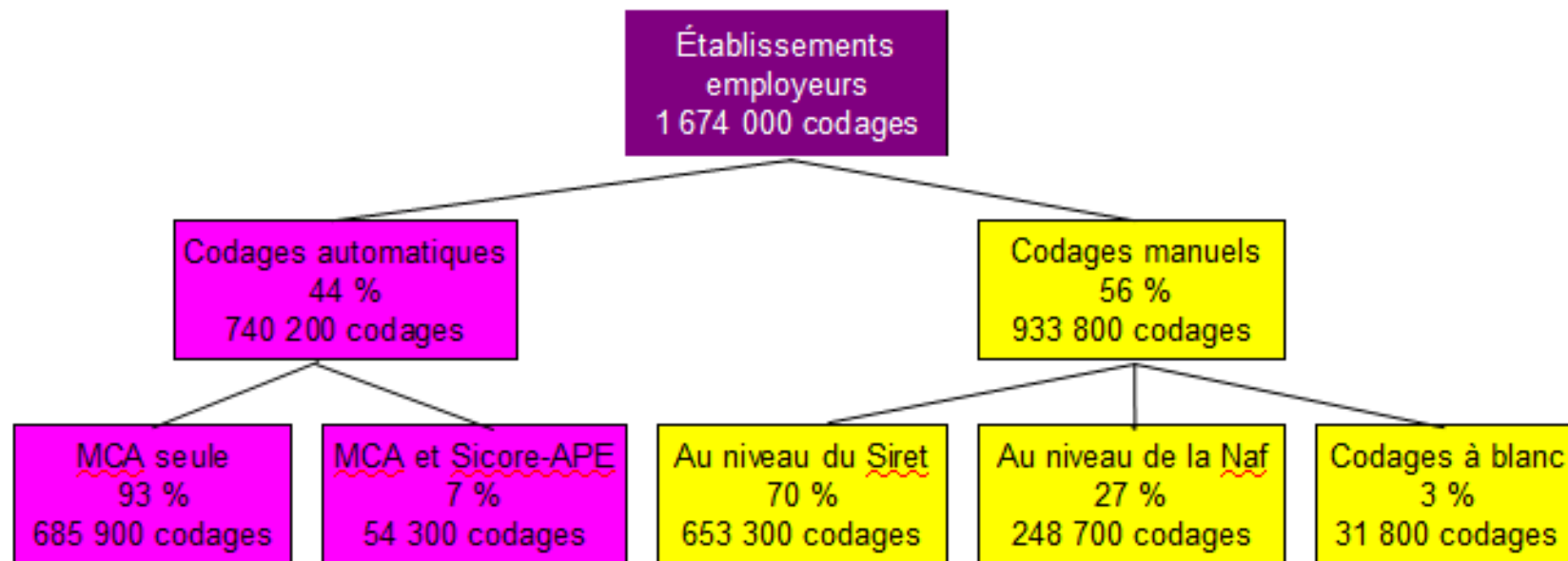
C'est un processus de traitement permettant de traduire un ou plusieurs libellés sous la forme d'un code exploitable appartenant à une nomenclature donnée.

Comment se fait cette codification ?

**En automatique si possible,
Sinon, de manière manuelle dans les
établissements régionaux**

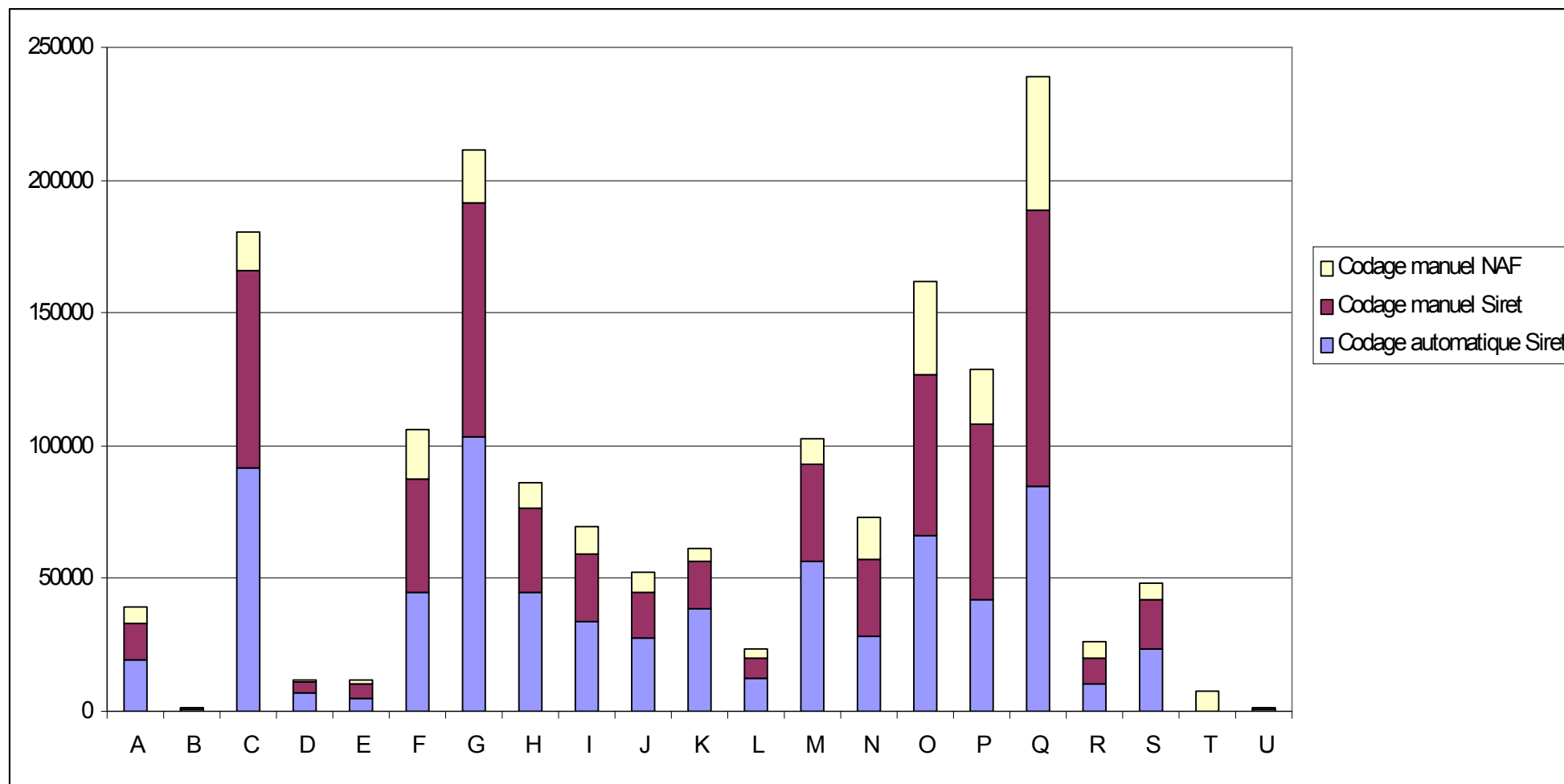
I) Présentation

Taux de codage de l'établissement employeur (campagne 2017)



Le codage manuel occupe 70 personnes pendant 5 mois.
Objectif : réduire le nombre de codages manuels.

I) Présentation



Répartition des différents types de codage en fonction des secteurs d'activité
Source : EAR 2017

II) La méthode actuelle : les données brutes disponibles

19 Quel est le nom de l'établissement qui vous emploie ou que vous dirigez ?

Si vous êtes **intérimaire**, précisez le nom de l'établissement où vous faites votre mission. Si vous êtes à **votre compte**, inscrivez le nom de l'entreprise ou votre nom.

RS_X

20 Quelle est l'activité de cet établissement ?

Soyez très précis (par exemple : « RÉPARATION AUTOMOBILE »).
S'il s'agit d'une **exploitation agricole**, précisez également l'orientation des productions (vigne, élevage de volailles, etc.).

ACTET_X

21 Quelle est l'adresse de votre lieu de travail ?

Indiquez l'endroit où vous commencez habituellement votre travail (exemple : 18, boulevard Pasteur).

Si cet endroit n'est pas fixe, notez « variable ».

Si vous travaillez à votre domicile, notez « à domicile ».

Si vous travaillez chez un particulier, notez « particulier ».

NUMVOI_X BISTER_X TYPEVOI_X NOMVOI_X

CPLADR_X

VARDOMPART_X

Est-ce dans la commune où vous résidez ?

(ou dans l'arrondissement pour Paris, Lyon, Marseille)

Oui ☐ 1 Non ☐ 2

ILT_X

Si non, indiquez la commune où vous travaillez :

CLT_X

commune (et arrondissement pour Paris, Lyon, Marseille)

DLT_X

département n° DOM

PLT_X

pays pour l'étranger

VARDOMPART_X :

1 = variable

2 = domicile

3 = particulier

II) La méthode actuelle : les données Sirius

Sirus : système d'identification au répertoire des unités statistiques

Le champ de SIRUS comporte :

- l'ensemble des unités légales (identifiées par un SIREN) productives marchandes de Sirene ainsi que l'ensemble de leurs établissements 'filles' (identifiés par un SIRET)
- l'ensemble des unités légales employeuses avec au moins un salarié des unités légales non marchandes ainsi que l'ensemble de leurs établissements 'filles'.

La division Sirius nous fournit une base contenant l'ensemble des établissements actifs au moins une journée entre le 1^{er} janvier et le 31 mars de l'année de collecte.

II) La méthode actuelle : les outils de codage

SICORE = Système informatisé de codage des réponses aux enquêtes

➡ Codage de l'activité

MCA = Mise en Concordance Automatique

➡ Recherche de l'établissement employeur

II) La méthode actuelle

Retrouver l'établissement employeur :

- MCA seule :

A partir de la raison sociale et de l'adresse

On recherche dans le répertoire, l'établissement le plus probable par rapport à ces deux informations.

en cas de doute :

- MCA et SICORE-APE :

On utilise en plus le libellé activité, qui permet de confirmer les échos « douteux » retenus par la MCA après codage par SICORE-APE de l'activité déclarée sur le BI

II) La méthode actuelle : Sicore-APE

La codification se fait en comparant le libellé d'activité issu du questionnaire aux libellés de référence associés à plusieurs codes, puisés dans une base de connaissance.

Exemples :

Libellé	Code retour	NAF1	%	NAF2	%	NAF3	%	NAF4	%	NAF5	%	nb_tot	nb5
ELECTRICIEN PLAQUISTE	C_C												
AGRICULTURE	CCS	0111Z	23	0141Z	14	0142Z	11	0150Z	11	*****	***	10 967	7649
ENTRETIEN WAGON	C_R	4910Z	41	3317Z	37	3020Z	18	*****	***	*****	***	49	49
AGROALIMENTAIRE	CCS	1011Z	9	1013A	7	7820Z	6	1085Z	6	1012Z	5	8266	2715

Lecture : Le libellé « AGRICULTURE » collecté dans les EAR 2006-2007 a donné 4 codes NAF issus de l'appariement Sirene représentant plus de 10 % des cas.

L'ensemble de ces libellés représente 10 967 occurrences dont 7 649 pour les 5 codes NAF les plus fréquents

Sicore-APE renvoie, au plus, les 5 codes NAF les plus fréquents. Il ne renvoie pas les codes avec une fréquence inférieure à 10 %, sauf si le premier est déjà inférieur à 10 %.

CCS : réussite

C_C : échec

C_R : éventuel codage pouvant être considéré comme douteux

II) La méthode actuelle

Note_mca_nom_c : note calculée par la MCA sur la raison sociale

De 1000 à 800 : Bien

De 799 à 200 : Assez bien

De 199 à 0 : Mauvais

Note_mca_adr_c : note calculée par la MCA sur l'adresse

10 ou 11 : Bien

20, 21 ou 22 : Assez bien

31, 32, 33 ou 35 : Mauvais

A partir de ces deux mentions on détermine la **classe** :

Adresse \ Nom	Nom		
	B	AB	M
B	1	4	7
AB	2	5	8
M	3	6	9

II) La méthode actuelle

La MCA trie les échos trouvés selon la note_nom, puis la note_adresse, puis par tranche d'effectif (décroissant), et retiens le 1^{er}.

Si cet écho fait partie :

- des classes 1 ou 2 : l'écho est validé
- des classes 6 à 9 : l'écho est rejeté
- des classes 3 à 5 : selon les cas, l'écho peut être validé, rejeté ou considéré comme douteux

Les cas douteux : appel à Sicore-APE

Si un des 5 premiers codages du libellé d'activité = APE de l'écho => écho validé

Si les codages du libellé d'activité \neq APE de l'écho => écho rejeté

On conservera uniquement les informations concernant le 1^{er} écho.

II) La méthode actuelle : les variables codées

- I_mca_c : indicateur de codage de l'établissement par MCA
blanc : non codé (la MCA n'a trouvé aucun écho potentiel)
C : codé
D : code douteux (vérification Sicore APE)
E : non codé avec au moins un écho
- I_reprise_act :
V : l'établissement n'a pas été codé automatiquement
Blanc : l'établissement a été codé automatiquement
- I_siret_m : lorsque l'établissement n'a pas été codé par MCA
1 : l'établissement a été codé manuellement
0 : l'établissement n'a pas été codé manuellement

II) La méthode actuelle : les variables codées

Siretc : siret de l'établissement choisi par MCA (qu'il soit validé ou non)

Siretm : siret de l'établissement codé manuellement

Actet_c_c : APET de l'établissement

Actet_c : code NAF finalement attribué à l'enquêté

Si un établissement employeur a été trouvé (automatiquement ou manuellement), il s'agit de son apet. Sinon, il s'agit du code saisi manuellement.

Clc_c_c : commune de l'établissement choisi par MCA (qu'il soit validé ou non) s'il y en a un. Sinon commune codée à partir des données du BI (CLT_X et ILT_X).

III) Difficultés rencontrées par les gestionnaires

- Incohérence d'adresse (rue ou ville)
- L'activité déclarée sur le BI qui ne correspond à celle de la société
- Les raisons sociales, sigles ou dénominations enregistrés légalement peuvent différer du nom usuel (pour les supermarchés par exemple)
- Les réponses d'enseignants pas assez détaillées : ils mettent éducation nationale comme employeur, n'écrivent pas le type d'établissement (maternelle, primaire, collège, lycée) et/ou l'adresse

Pages jaunes :

Les gestionnaires trouvent souvent l'établissement, même si :

- c'est le nom commercial ou l'enseigne qui a été renseigné au lieu du nom réel (supermarchés par ex.)
- c'est la ville voisine qui a été renseignée (cas courant surtout avec les BI d'Île de France)

III) Difficultés rencontrées : exemple n°1

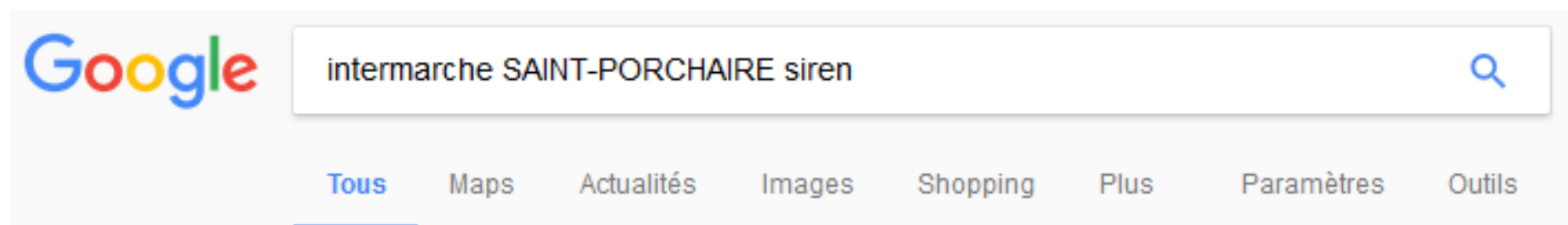
Infos renseignées :

RS : Intermarché

Adresse : rue du gros chêne, SAINT-PORCHAIRE

Actet_x : grande distribution

La MCA n'a trouvé aucun écho



Environ 2 010 résultats (0,69 secondes)

PATECY (SAINT PORCHAIRE) Chiffre d'affaires, résultat, bilans sur ...

<https://www.societe.com/societe/patecy-448677989.html>

PATECY à SAINT PORCHAIRE (17250) RCS, SIREN, SIRET, bilans, statuts, chiffre d'affaires, dirigeants, cartographie, alertes, annonces légales, enquêtes, ...

III) Difficultés rencontrées : exemple n°1

Entreprise du secteur privé		SIREN : 448 677 989		Entreprise : PATECY	
		Identification		Informations activités	
		Effectifs		Liens entre ULs	
		Champ d'observation		Données financières	
		Démographie			
Nombre d'établissements		Dénomin. : PATECY			
Actifs Sys. Prod. : 1		Nom comm. :			
Expl. actifs : 1		Sigle :			
Non expl. X actifs : 0		APE : 4711D Supermarchés			
Employeurs : 1		CJ : 5710 SAS, société par actions simplifiée			
Prés cessés : 0		ESS :			
		Adresse			
		Voie : RUE DU GROS CHENE			
		Code postal : 17250 SAINT PORCHAIRE			
		Loc. géo. : 17387			
		Etat : Active			
		Début validité : 21/05/2003 Source : ***			
		Etat admin. : Active			
		Employeur : Oui Date de début : 26/11/2010			
		Création UL : 21/05/2003 Entrée rep. : 27/05/2003 Dernier trait. : 18/09/2015 Dernier évé. : 30/06/2015			
		Exploitant : Exploitant			
		Diff. com. : Fait partie de la diffusion commerciale			
		Champ ESA/EAP : Non			

☐ Restruct. en cours

Imprimer

Historique Ent.

Sélection Etab.

Historique Etab.

Infos Sirus : Dénomination : Patecy
Adresse : rue du gros chêne, SAINT-PORCHAIRE
APE : 4711D Supermarchés

III) Difficultés rencontrées : exemple n°2

L'activité déclarée ne correspond pas à celle de l'établissement

Infos renseignées

RS : CPDS

Adresse : GRENOBLE

Actet_x : Service traiteur (5621Z)

MCA avait trouvé le bon écho mais notes insuffisantes : rejet (sans tester l'activité)

L'établissement est une association pour adultes et jeunes handicapés donc codé en 8810C Aide par le travail.

Identification	Autres adresses	Informations activités	Effectifs	Liens entre étabs	Champ d'observation	Démographie
NIC : 00178 Catégorie : Etablissement secondaire non RCS						
Enseigne : CAT CPDS APET : 8810C Aide par le travail						
Nom comm. :						
Adresse						
Voie : 21 RUE DES TREMBLES						
Code postal : 38100 GRENOBLE Loc. géo. : 38185 6655						

III) Difficultés rencontrées : exemple n°3

Le nom de commune renseignée est inexact

Infos renseignées

RS : RESTAURANT CHINOIS WOWO

Adresse : 366 rue du 18 août 1944, ANNEMASSE

Actet_x : Restauration

Le meilleur écho trouvé par la MCA est un autre restaurant à Annemasse (rejeté car notes insuffisantes)

Recherche Sirene

Aucun écho

Siren : NIC : CJ : APET :

Dénom./Nom :

Prénoms : Date nais. :

Nom d'usage :

Voie : du au Code :

Commune : ANNEMASSE

III) Difficultés rencontrées : exemple n°3

L'établissement est situé sur la commune d'Étrembières, commune voisine. Recherche google :

Wo Wo
3,7 ★★★★★ 77 avis Google
Restaurant

Site Web

Itinéraire

Adresse : 366 Rue du 18 Août 1944, 74100 Etrembières

Horaires : Ouvert aujourd'hui · 12:00–14:30, 19:00–22:30

Téléphone : 04 50 31 20 20

Recherche Sirene

1 écho

Siren : NIC : CJ : APET :

Dénom./Nom :

Prénoms : Date nais. :

Nom d'usage :

Voie : du au Code :

Commune :

M3 - Personne morale: recherche communale

N°	Dénomination	Voie
1	SARL WOWO	366 RUE DU 18 AOUT 1944

III) Difficultés rencontrées : exemple n°4

La raison sociale renseignée est inexacte

Infos renseignées

RS : CARRIERE DE JAUMONT

Adresse : Écart Saint-Hubert, MALANCOURT LA MONTAGNE

Actet_x : Conducteur d'engins

Les deux échos trouvés ne correspondent pas :

- Jaumont finances : 6450Z Activité de sociétés de Holding
- Jaumont, Pierre de culture du pays Lorrain : 9499Z Autre organisation fonctionnant par adhésion volontaire

III) Difficultés rencontrées : exemple n°4

La carrière Jaumont est exploitée par l'entreprise VAGLIO SAS et...
Recherche Google :



MENTIONS LÉGALES

IDENTIFICATION

Le présent site **www.jaumont.fr**, (ci-après "le Site") est édité par la société VAGLIO SAS.

Établissement principal :

VAGLIO SAS

Les Ecartis Saint-Hubert

BP 50010

F-57361 Malancourt-la-Montagne

Tél. : +33 (0)3 87 53 51 00

Fax : +33 (0)3 87 53 47 84

Conception / réalisation du site : **Infinirouge & Ealys**

Hébergeur du site : **Ealys**

III) Difficultés rencontrées : exemple n°4

... le nom de commune est inexact.

Recherche sur société.com :

VAGLIO Société : 302638424

Fiche entreprise : chiffres d'affaires, bilan et résultat

[Ajoutez le logo de votre société](#)

VAGLIO
ECART MALANCOURT LA MONTAGNE
SAINT HUBERT
MALANCOURT LA MONTAGNE
57360 AMNEVILLE
FRANCE

Documents Officiels

Identité

Carto

III) Difficultés rencontrées : exemple n°5

Difficultés à trouver le Siren des établissements municipaux

Infos renseignées

RS : Crèche municipale Saint Jean de Vedas

Adresse : Place du puits de Gau, Saint Jean de Vedas

Actet_x : Accueil des enfants de 3 mois à 4 ans

L'écho trouvé par la MCA correspond à la mairie.

III) Difficultés rencontrées : exemple n°5

Après recherche sur la mairie, il est possible de retrouver les établissements dont la crèche.

Nombre d'établissements

Actifs Sys. Prod. : 15

Expl. actifs : 15

Non expl. X actifs : 0

Employeurs : 1

Prés cessés : 0

Restruct. en cours

Imprimer

Historique Ent.

Sélection Etab.

Historique Etab.

Etablissements

Liste ☒ Détail ☐

00018 - Siège

N° - Nic - DepCom

03 - 00042 - 34270

04 - 00059 - 34270

05 - 00067 - 34270

06 - 00075 - 34270

07 - 00083 - 34270

08 - 00091 - 34270

09 - 00109 - 34270

10 - 00117 - 34270

Identification

Informations activités

Effectifs

Liens entre ULs

Champ d'observation

Données financières

Démographie

Dénomin. : COMMUNE DE SAINT-JEAN-DE-VEDAS

Sigle :

Appart. sect. admin. publ. : Collectivités territoriales

APE : 8411Z Administration publique générale

CJ : 7210 Commune et commune nouvelle

ESS :

Adresse

Comp. adresse : HOTEL DE VILLE

Voie : 4 RUE DE LA MAIRIE

Code postal : 34430 SAINT JEAN DE VEDAS

Loc. géo. : 34270

Etat : Active

Début validité : 01/01/1982

Source : ***

Etat admin. : Active

Employeur : Oui

Date de début : 25/12/1996

Création UL : 01/01/1982

Entrée rep. :

Dernier trait. : 31/05/2017

Dernier évé. : 25/12/2016

Exploitant : Exploitant

Diff. com. : Fait partie de la diffusion commerciale

Champ ESA/EAP : Non

N°	NIC	Dep	Com	N°voie	IR	Type	Lib. voie	APET	Eff. 31/12
	00018	34	270	4		RUE	DE LA MAIRIE	8411Z	148
1	00026	34	270			RUE	FEDERICO GARCIA LORCA	8510Z	
2	00034	34	270	2		ALL	VAL DES GARRIGUES	8510Z	
3	00042	34	270	1		RUE	DES ESCHOLIER	8520Z	
4	00059	34	270			RUE	DES ECOLES	8891A	
5	00067	34	270	2		ALL	DU VAL DES GARRIGUES	8520Z	
6	00075	34	270	9		RUE	ENGABANAC	8510Z	
7	00083	34	270	9		RUE	ENGABANAC	8520Z	
8	00091	34	270					3700Z	
9	00109	34	270	21		RUE	AUGUSTE RENOIR	9101Z	
10	00117	34	270			RTE	DE BEZIERS	8424Z	
11	00125	34	270			ALL	DU TERRAL	9004Z	
12	00133	34	270			PL	DU PUIT DE GAUD	8891A	
13	00141	34	270	1		CHE	DE LA COMBE	9329Z	
14	00158	34	270			AV	DE LIBRILLA	8110Z	

III) Difficultés rencontrées : exemple n°6

Les enseignants

Infos renseignées

RS : Ministère de l'éducation nationale

Adresse : rue Louise Leriget, Angoulême

Actet : Enseignement

L'écho trouvé par la MCA ne correspond pas à un établissement scolaire

Dénomin. :	DIRECTION DES SERVICES DEPARTEMENTAUX DE L'EDUCATION NATIONALE DE LA CHARENTE		
Dénom. cond. :	DIR SCES DEP EDUCATION NAT CHARENTE		
Sigle :	DSDEN	Appart. sect. admin. publ. :	Organisme de sécurité sociale
APE :	8412Z	Administration publique (tutelle) de	CJ : 7172 Service déconcentré de l'État à co
Comp. adresse : CI ADM CHAMP DE MARS BAT B			
Voie :	RUE RAYMOND POINCARE		
Code postal :	16000	ANGOULEME	Loc. géo. : 16015 3533
Cedex :	16023	ANGOULEME CEDEX	

III) Difficultés rencontrées : exemple n°6

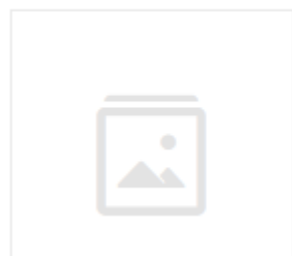
Les enseignants

Recherche pages jaunes :



☐ A proximité





1 **Lycée Marguerite De Valois** [+ d'infos](#)

12 r Louise Lériget, 16000 ANGOULÊME
lycées d'enseignement général et technologique publi...

[Ecrire un avis](#)
[+ d'adresses](#)

SIRET	19160003000016
Code NAF	8531Z
Effectif de l'établissement	100 à 199 salariés
Typologie de l'établissement	Siège

IV) Opération qualité

1. On sélectionne un sous-échantillon de BI parmi ceux où l'établissement employeur a été codé (automatiquement ou manuellement).
2. On effectue un second codage, manuel, de l'établissement.
3. Pour les cas où le second codage ne correspond pas au codage initial, on effectue un arbitrage.
Le codage jugé « bon » peut alors être le codage initial, le second codage, ou aucun des 2.

Cette opération a été réalisée sur les données 2014.

IV) Opération qualité : Résultats

Résultats selon le mode de codage initial :

	Convergence	Arbitrage en faveur du codage initial	Codage initial jugé faux ou incertain
Codage MCA seul	86 %	5 %	9 %
Codage MCA + Sicore-APE	86 %	9 %	5 %
Codage manuel	67 %	21 %	12 %

Conclusion

Les solutions trouvées seront examinées plus précisément après le hackathon.

On privilégiera les pistes en fonction du rapport « apport / coût de développement ».

La mise en œuvre de la proposition retenue, se fera en fonction des moyens informatiques disponibles.

Le codage de l'établissement employeur

Merci de votre attention

Avez-vous des questions ?