

# Normalisation ex-ante/ex-post pour l'IA appliquée aux archives: le cas du projet Socface

Lionel Kesztenbaum, INED

Christopher Kermorvant, TEKLIA

Journée d'étude AFNOR/BnF

# Socface: IA et recensements 1836 - 1936

**Objectif** : étudier la population française à l'échelle de l'individu grâce aux listes manuscrites du recensement entre 1836 et 1936

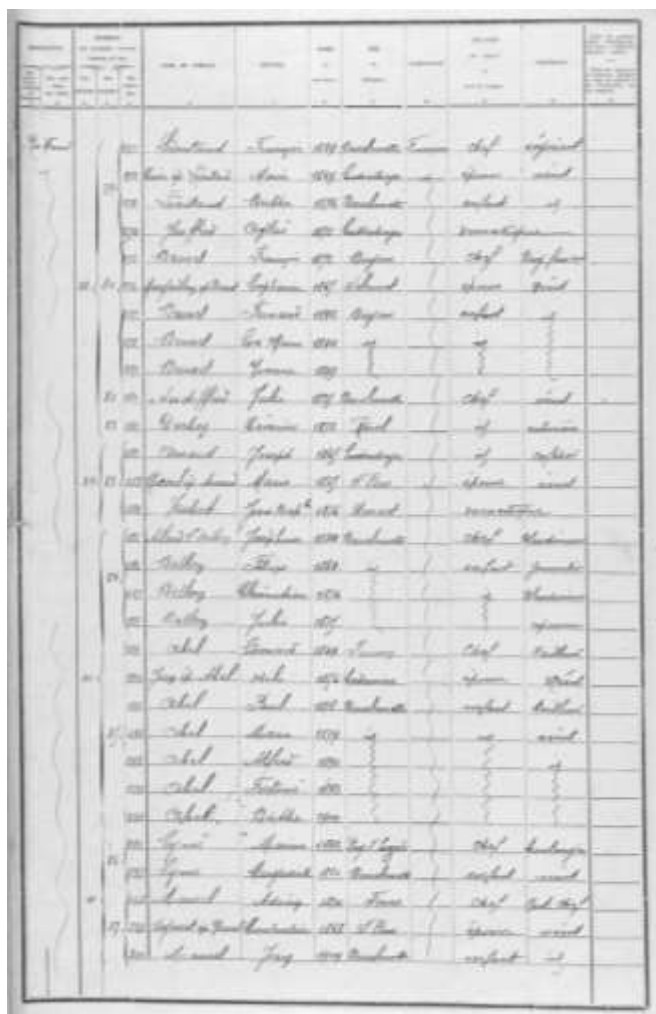
**Moyen** : reconnaissance automatique d'écriture manuscrite (HTR)

**Défis** : 30 millions de pages, des dizaine de milliers de scripteurs, 100 ans d'écritures

**Avantage**: des formulaires normalisés

NUMERO	NOM	AGE	SEXE	PROFESSION	REMARQUES
37	Jouy	27	M	ouvrier	
38	Cressant	12	F	cultivateur	
39	Hellot	47	M	ouvrier	
40	Cressant	12	F		
41	Cressant	7	M		
42	Cressant	26	M	cultivateur	
43	de Cressant	14	M		
44	Cressant	14	M		
45	Cressant	36	M	ouvrier	
46	Cressant	47	M		
47	Cressant	32	M	cultivateur	
48	Jouy	26	M	ouvrier	
49	Cressant	7	M		
50	Cressant	7	M		
51	Cressant	1	M		
52	Cressant	13	M		
53	Chaplat	43	M		
54	Cressant	41	M	cultivateur	
55	Cressant	21	M	cultivateur	
56	Cressant	7	M	ouvrier	
57	Cressant	11	M		
58	Cressant	11	M		
59	Cressant	16	M	ouvrier	
60	Cressant	13	M		
61	Cressant	7	M		
62	Cressant	23	M	cultivateur	
63	Cressant	18	M	ouvrier	
64	Cressant	7	M		
65	Cressant	1	M		

# Modèles génératifs de reconnaissance d'écriture



Structured data output showing extracted information from the handwriting, including names, birth dates, nationalities, and professions, with color-coded fields.

sumname Léautaud, firstname François, occupation négociant, link\_chef, birth\_date 1839, nationality française, job Barcelonnnette  
sumname Caire ep Léautaud, firstname Marie, occupation néant, link\_épouse, birth\_date 1849, nationality idem, job Enchartrayes  
sumname Léautaud, firstname Berthe, occupation idem, link\_enfant, birth\_date 1876, nationality idem, job Barcelonnnette  
sumname Jauffred, firstname Aglaé, occupation idem, link\_domestique, birth\_date 1875, nationality idem, job Gnetrartrayes  
sumname Brunet, firstname François, occupation brig ferrier, link\_chef, birth\_date 1852, nationality idem, job Bayons  
sumname Margailan ep Brunet, firstname Euphanie, occupation néant, link\_épouse, birth\_date 1865, nationality idem, job Selonnet  
sumname Brunet, firstname Fernand, occupation idem, link\_enfant, birth\_date 1892, nationality idem, job Bayons  
sumname Brunet, firstname Eva Marie, occupation idem, link\_idem, birth\_date 1894, nationality idem, job idem  
sumname Brunet, firstname Yvonne, occupation idem, link\_idem, birth\_date 1899, nationality idem, job idem  
sumname Audiffred, firstname Julie, occupation néant, link\_chef, birth\_date 1825, nationality idem, job Barcelonnnette  
sumname Derbez, firstname Césarine, occupation culurière, link\_idem, birth\_date 1872, nationality idem, job Revel  
sumname Arnaud, firstname Joseph, occupation cafetier, link\_idem, birth\_date 1865, nationality idem, job Enchartrayes  
sumname Risoulép Arnaud, firstname Marie, occupation néant, link\_épouse, birth\_date 1865, nationality idem, job St Pons  
sumname Jaubert, firstname Jean Baptiste, link\_domestique, birth\_date 1876, nationality idem, job Nvnet  
sumname Allard vve Bellon, firstname Joséphine, occupation blanchisseuse, link\_chef, birth\_date 1834, nationality idem, job Barcelonnnette  
sumname Bellon, firstname Félix, occupation journalier, link\_enfant, birth\_date 1864, nationality idem, job idem  
sumname Bellon, firstname Clémentine, occupation blanchisseuse, link\_idem, birth\_date 1876, nationality idem, job idem  
sumname Bellon, firstname Julie, occupation repasseuse, link\_idem, birth\_date 1875, nationality idem, job idem  
sumname Abel, firstname Edouard, occupation tailleur, link\_chef, birth\_date 1849, nationality idem, job Faucon  
sumname Jean ép Abel, firstname Odile, occupation néant, link\_épouse, birth\_date 1856, nationality idem, job Condarnine  
sumname Abel, firstname Paul, occupation tailleur, link\_enfant, birth\_date 1878, nationality idem, job Barcelonnnette  
sumname Abel, firstname Marie, occupation néant, link\_idem, birth\_date 1879, nationality idem, job idem  
sumname Abel, firstname Alfred, occupation idem, link\_idem, birth\_date 1890, nationality idem, job idem  
sumname Abel, firstname Fortuné, occupation idem, link\_idem, birth\_date 1892, nationality idem, job idem  
sumname Abel, firstname Berthe, occupation idem, link\_idem, birth\_date 1900, nationality idem, job idem  
sumname Eymé, firstname Marius, occupation boulanger, link\_chef, birth\_date 1842, nationality idem, job Puy de Euzèbe  
sumname Lymé, firstname Marguerite, occupation néant, link\_enfant, birth\_date 1880, nationality idem, job Barcelonnnette  
sumname Maurel, firstname Adrien, occupation cant, link\_chef, birth\_date 1870, nationality idem, job Fours  
sumname Signout ep Maurel, firstname Constantine, occupation néant, link\_épouse, birth\_date 1868, nationality idem, job S. Pon  
sumname Maurel, firstname Jean, occupation idem, link\_enfant, birth\_date 1904, nationality idem, job Barcelonnmarle

# Des formulaires normalisés ?

N°	NOM	PRENOM	LIEU DE NAISSANCE	AGE					OCCUPATION
				AN	MOIS	JOUR	SEMAINE	ANNÉE	
20	Agostini	Joseph	Genève	1	1	1	1	1	73
21	Agostini	Joseph	Genève	1	1	1	1	1	71
22	Agostini	Joseph	Genève	1	1	1	1	1	60
23	Agostini	Joseph	Genève	1	1	1	1	1	70
24	Agostini	Joseph	Genève	1	1	1	1	1	70
25	Agostini	Joseph	Genève	1	1	1	1	1	70
26	Agostini	Joseph	Genève	1	1	1	1	1	70
27	Agostini	Joseph	Genève	1	1	1	1	1	70
28	Agostini	Joseph	Genève	1	1	1	1	1	70
29	Agostini	Joseph	Genève	1	1	1	1	1	70
30	Agostini	Joseph	Genève	1	1	1	1	1	70
31	Agostini	Joseph	Genève	1	1	1	1	1	70
32	Agostini	Joseph	Genève	1	1	1	1	1	70
33	Agostini	Joseph	Genève	1	1	1	1	1	70
34	Agostini	Joseph	Genève	1	1	1	1	1	70
35	Agostini	Joseph	Genève	1	1	1	1	1	70
36	Agostini	Joseph	Genève	1	1	1	1	1	70
37	Agostini	Joseph	Genève	1	1	1	1	1	70
38	Agostini	Joseph	Genève	1	1	1	1	1	70
39	Agostini	Joseph	Genève	1	1	1	1	1	70
40	Agostini	Joseph	Genève	1	1	1	1	1	70

1836

Cases à cocher

N°	NOM	PRENOM	LIEU DE NAISSANCE	AGE	OCCUPATION	ANNÉE DE NAISSANCE				
						AN	MOIS	JOUR	SEMAINE	ANNÉE
31	Agostini	Joseph	Genève	73	1	1	1	1	1	1
32	Agostini	Joseph	Genève	71	1	1	1	1	1	1
33	Agostini	Joseph	Genève	60	1	1	1	1	1	1
34	Agostini	Joseph	Genève	70	1	1	1	1	1	1
35	Agostini	Joseph	Genève	70	1	1	1	1	1	1
36	Agostini	Joseph	Genève	70	1	1	1	1	1	1
37	Agostini	Joseph	Genève	70	1	1	1	1	1	1
38	Agostini	Joseph	Genève	70	1	1	1	1	1	1
39	Agostini	Joseph	Genève	70	1	1	1	1	1	1
40	Agostini	Joseph	Genève	70	1	1	1	1	1	1

1886

âge

N°	NOM	PRENOM	LIEU DE NAISSANCE	AGE	OCCUPATION	ANNÉE DE NAISSANCE				
						AN	MOIS	JOUR	SEMAINE	ANNÉE
61	Agostini	Joseph	Genève	73	1	1	1	1	1	1
62	Agostini	Joseph	Genève	71	1	1	1	1	1	1
63	Agostini	Joseph	Genève	60	1	1	1	1	1	1
64	Agostini	Joseph	Genève	70	1	1	1	1	1	1
65	Agostini	Joseph	Genève	70	1	1	1	1	1	1
66	Agostini	Joseph	Genève	70	1	1	1	1	1	1
67	Agostini	Joseph	Genève	70	1	1	1	1	1	1
68	Agostini	Joseph	Genève	70	1	1	1	1	1	1
69	Agostini	Joseph	Genève	70	1	1	1	1	1	1
70	Agostini	Joseph	Genève	70	1	1	1	1	1	1

1936

Année de naissance

# Des « normes » locales

DÉSIGNATION	NUMÉROS			NOMS	PRÉNOMS	ÂGE	NATIONALITÉ	PROFESSION	POSITION	OBSERVATIONS
	DES QUARTIERS, VILLES, BOULEVARD, etc.	DES RUES	DES FAMILLES							
				371	Jeanjean	Philippine	48	français, sans profession	chef de ménage	
				372	Creissant	Victor	50	id	cultivateur chef de ménage	
				373	Sellet	Loise	47	id	sans profession sa femme	
89	111			374	Creissant	Leopold	13	id	id	son fils
				375	Creissant	Léon	7	id	id	id
				376	Creissant	Marc Antoine	86	id	cultivateur son père	
				377	de Lore	Gustave	14	id	id	berger

DÉSIGNATION	NUMÉROS			NOMS	PRÉNOMS	ÂGE	NATIONALITÉ	PROFESSION	POSITION	OBSERVATIONS
	DES QUARTIERS, VILLES, BOULEVARD, etc.	DES RUES	DES FAMILLES							
				39	Lambert	Anton	21	f	ouvrier	enfant
					Lambert	Antoine	20	- id -	- id -	
				41	Blanc	Alphonse	19		ouvrier	chef
					Ranc	Charles	41	- id -	- id -	époux
					Chamois	Jean Louis	18		ouvrier	chef
					Benoit	Léon	41	- id -	- id -	époux
					Chamois	Paul	29	- id -	- id -	enfant

Chef de ménage Français id

Chef f - id- ou " ou |

Plus que la diversité de l'écriture (forme), c'est la diversité de l'usage (norme) qui pose problème

➤ S'adapter à un usage local mais stable, qui devient une *norme locale*

# Comment naissent les « normes » locales?

Jusqu'en 1876: état civil en case à cocher

DÉSIGNATION	NOM					TITRES, QUALIFICATIONS, etc.	ÉTAT CIVIL DES HABITANTS						ÂGE	NATIONALITÉ	OBSERVATIONS
	DE FAMILLE						MARIAGE			NÉCESSITÉ					
	Prénoms	Noms	Prénoms	Noms	Prénoms		Noms	Prénoms	Noms	Prénoms	Noms	Prénoms			
Bord (Puy-de-Dôme)	151	Martin	Liane	prop. cult.	ch. de m.	1						44	français	français	
	152	Martin	Marguerite	sa femme					1			44	français	française	
	153	Martin	Elisabeth	sa sœur					1			44	français	française	
	154	Coste	Zulma	domestique		1						19	français	française	
	155	Margu	Henriette	domestique								18	français	française	
	156	Mellissier	Antoine	ch. de m.			1					54	D		
	157	Mellissier	Marie	sa fille					1			16	D		
	158	Mellissier	Marie	prop. cult.	ch. de m.				1			64	D		
	159	Mellissier	Françoise	sa femme						1		56	D		
	160	Guy	Françoise	sa femme						1		41	D		
161	Carandier	Marie	sa femme						1		44	français	française		
162	Chapuis	Marie	sa femme						1		38	français	française		

1881: le tableau évolue, mais les habitudes persistent

DÉSIGNATION	NOM					TITRES, QUALIFICATIONS, etc.	ÂGE	PROFESSION	SITUATION	OBSERVATIONS
	DE FAMILLE									
	Prénoms	Noms	Prénoms	Noms	Prénoms					
Bord (Puy-de-Dôme)	26	27	Pugi	Antoine	48	cultivateur	ch. de m.	sa femme	sa femme	
	26	28	Chapuis	Marguerite	66	sa femme				
	26	29	Guy	Françoise	46	D		domestique	cultivateur	
	26	30	Explicmann	Marie	51	cultivateur	ch. de m.	sa femme	sa femme	
	26	31	Chapuis	Marie	38	sa femme				
	26	32	Explicmann	Marie	18	D		sa fille	cultivateur	
	27	33	Explicmann	Mariette	11	D		sa fille		
	27	34	Explicmann	Françoise	6	D		sa fille		
	27	35	Explicmann	Éugénie	2	D		sa fille		
	27	36	Martin	Priscille	40	cultivateur	ch. de m.	sa femme	sa femme	
28	29	37	Martin	Marguerite	40	sa femme				
		38	Martin	Elisabeth	60	D		sa sœur		

## Origines des normes locales:

- Diversité d'interprétation des règles (inhérent au flou des règles).
- Maintien d'habitudes anciennes (comme ici).
- Pratiques scripturales différenciées (id ou " ou |).

# Normaliser par des référentiels *internes*

Référentiels internes au document :

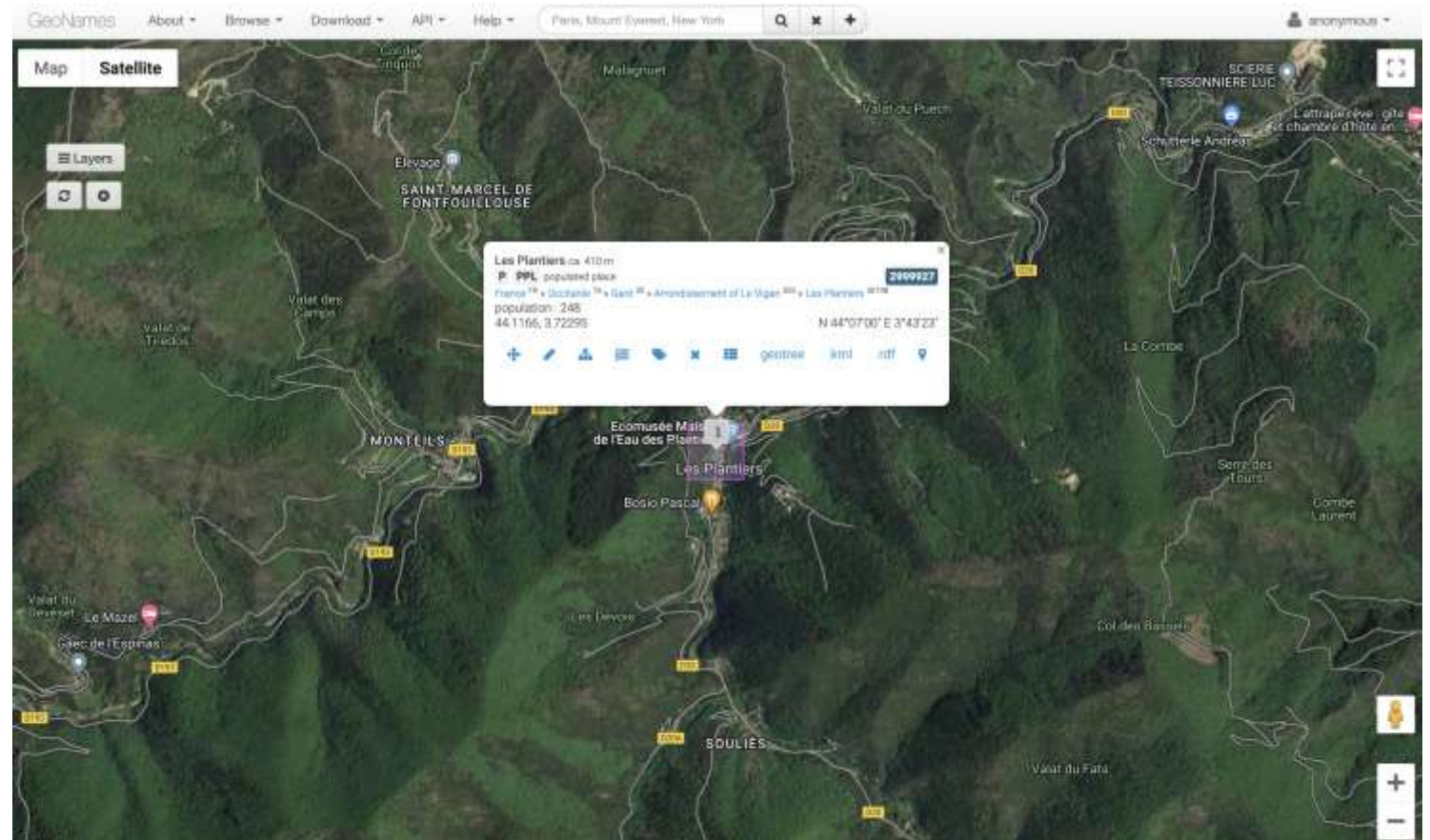
regroupement des ménages , résolution des idems

83	104	367	Berty	Mélanie	64	id	ménagère	chef de ménage
	108	368	Berty	Antoine	66	w	<u>cultivateur</u>	<u>chef de ménage</u>
		369	Luithard	Martial	50	id	id	w
84	106	370	Ducros	Delfphine	50	id	sans profession	sa femme
		371	Luithard	Eugène	22	id	cultivateur	son filz
		372	Luithard	Charley	19	w	w	w
		373	Lilere	Jules	16	id	id	locyer

# Normaliser par des référentiels *externes*

## Référentiels géographiques

📍 Cassini ID	33149
📍 Census years	1836, 1841, 1846, 1851, 1856, 1861, 1866, 1872, 1876, 1881, 1886, 1891, 1896, 1901, 1906, 1911, 1921, 1926, 1931, 1936
📍 Department	30
📍 INSEE ID	30198
📍 Name 1794	Saint Marcel de Fonfouillouse
📍 Name 1801	Saint-Marcel-de-Fouillouse
📍 Name 99	Plantiers (Les)
📍 Pseudo INSEE ID	30198



Geoname



# Normaliser par des référentiels *externes*

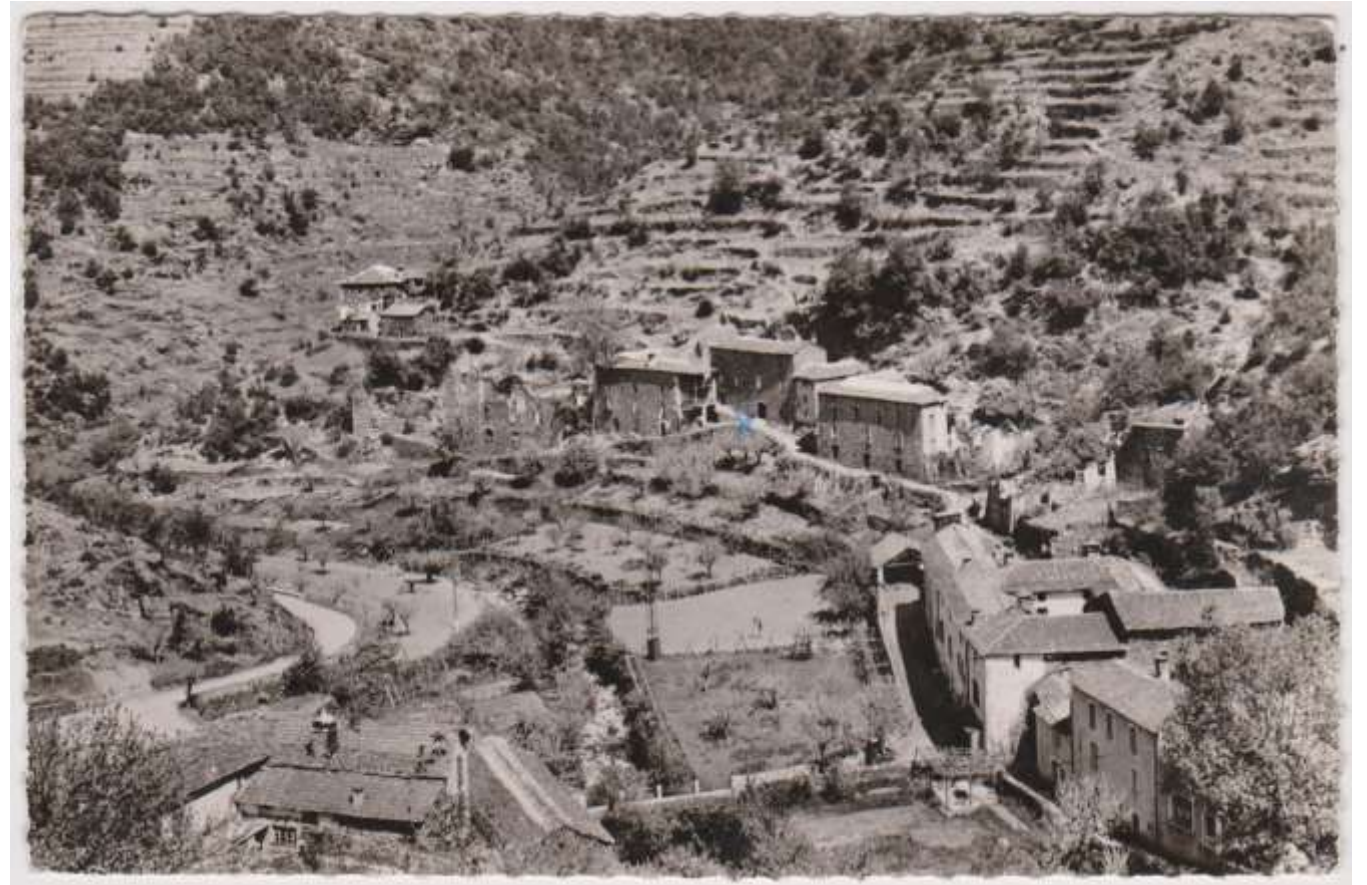
Référentiels géographiques : limites

Les Plantiers, **hameau de la Hierle**

➤ dans aucun référentiel

Saint Martin

➤ 255 communes



# Normaliser par des référentiels *externes* : *professions* (1)

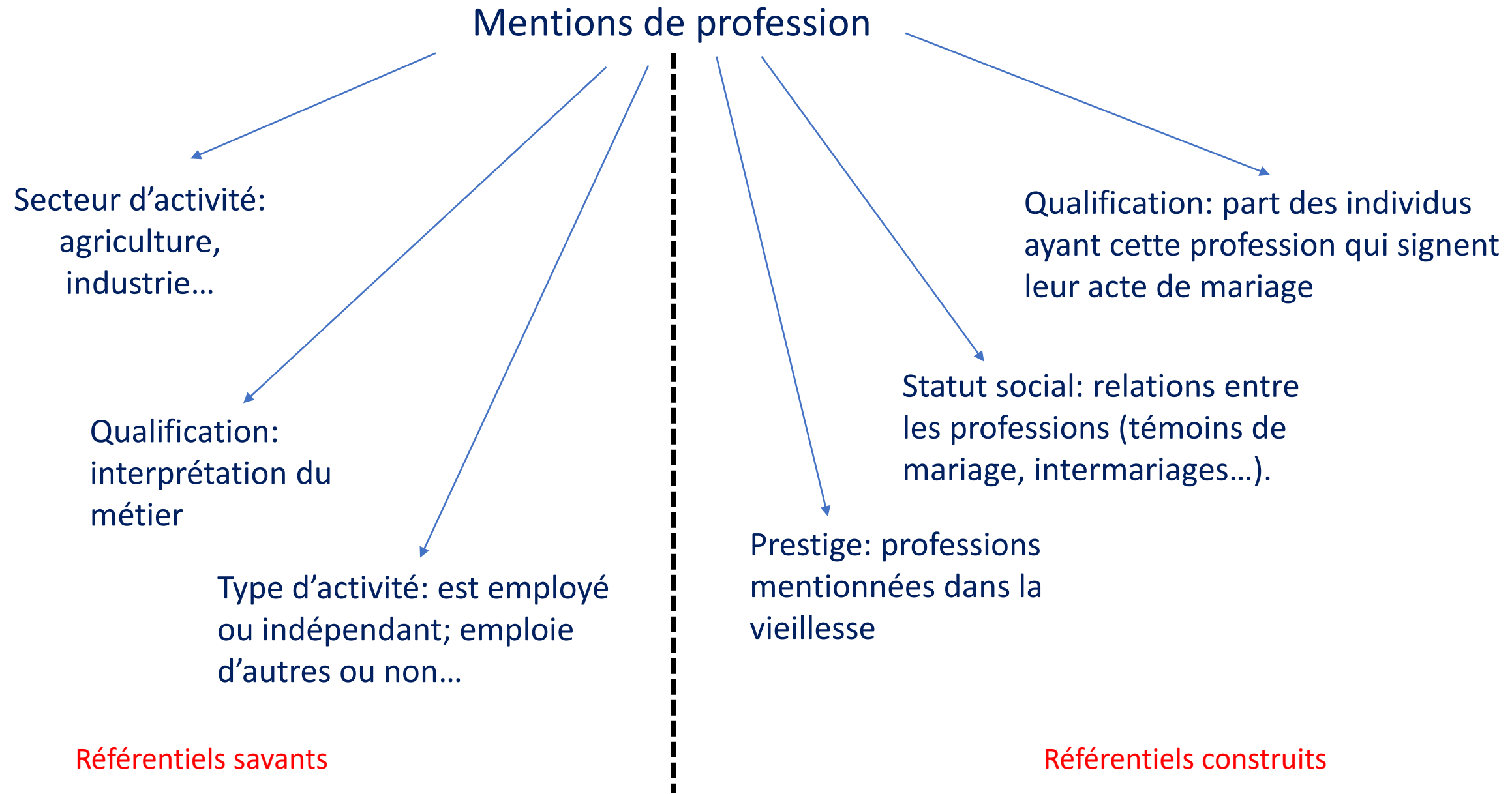
Dans l'enquête TRA (registres de mariage et tables de successions):  
237 termes renvoient à « cultivateur »

- Erreurs orthographiques.
- Abréviations.
- Multi-activité.
- Ancienneté (ancien, ex...).
- Statut (propriétaire cultivateur).
- ...

ADJOINT AU MAIRE ET CULTIVATEUR	CULTIVATEUR ET COQUETIER
ANCIEN CULT	CULTIVATEUR ET EX SOLDAT
ANCIEN CULTIVATEUR	CULTIVATEUR ET EXPERT SUPPLEANT DU JUGE
ANCIEN MILITAIRE ET CULTIVATEUR	CULTIVATEUR ET FERMIER
ANCIEN PROPRIETAIRE CULTIVATEUR	CULTIVATEUR ET GARCON MEUNIER
CUKLTIV ET MD DE CHARBON	CULTIVATEUR ET GARDE CHAMPETRE
CULITVATRICE	CULTIVATEUR ET GARDE CHASSE
CULIVATEUR	CULTIVATEUR ET HUILIER
CULT	CVR
CULT PROP	CVR ET MARCHAND DE CHEVAUX
CULTIV	EX CULT
CULTIV. / AUBERGISTE	MARCHAND CULTIVATEUR
CULTIV. /DOMESTIQUE	PP CULT
CULTIVATEUR - MOBILISE AU MANS	PPT CULT
CULTIVATEUR AUBERGISTE	PPT CULTIV
CULTIVATEUR CHARRON	PRO CVR
CULTIVATEUR CONSEILLER MUNICIPAL	PROPRIETAIRE CULTIVATEUR
CULTIVATEUR DEBUTANT	S'OCCUPANT à LA CULTURE
CULTIVATEUR ENTREPRENEUR	S'OCCUPANT DE LA CULTURE
CULTIVATEUR ET ARDOISIER	SOLDAT CULTIVATEUR
CULTIVATEUR ET AUBERGISTE	SOLDAT ET CULTIVATEUR
CULTIVATEUR ET CAFETIER	SP ANCIEN CULTIVATEUR
CULTIVATEUR ET CHARPENTIER	SULTIVATEUR
CULTIVATEUR ET CHARRON	VR (1)
CULTIVATEUR ET CHAUFournIER	VULTIVATEUR
CULTIVATEUR ET CONSEILLER MUNICIPAL	

Source: enquête TRA, extrait

# Normaliser par des référentiels *externes* : *professions* (2)



# Normaliser par des référentiels *externes* : professions (3)

Cultivateur

108	368	Besty	Antoine	66	id	cultivateur	chef ménage
	369	Quilhand	Martial	50	id	id	id

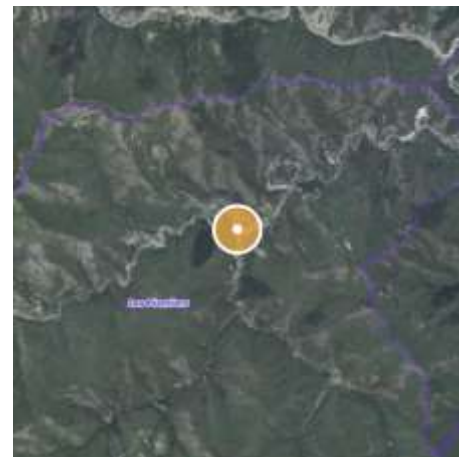
84	106	Ducros	Delfina	50	id	son profession	sa femme
	371	Quilhand	Eugène	22	id	cultivateur	son filz
	372	Quilhand	Charly	19	id	id	id
	373	Lilore	Jules	16	id	id	berger

Cultivateur

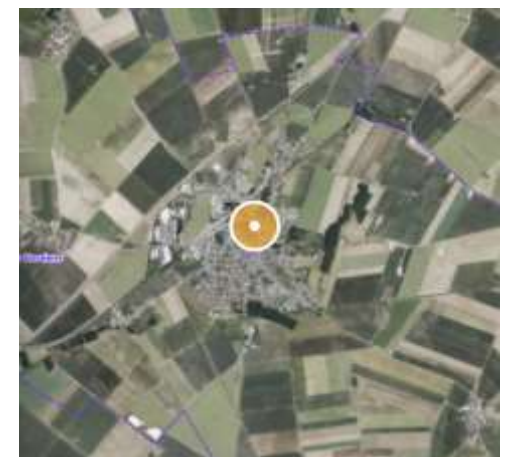
1311	Maugars	Louis Denis	74	id	chef ménage	cultivateur	patron
		arrivé					

1312	Maugars	Armine Paul	39	id	son fils	aidé à culture	son père
1313	Maugars	Caroline Josephine	36	id	sa fille		
3	1314	Maugars	Charlotte Pauline	4	id	sa petite fille	
	1315	Boucher	Edmond Alst	29	id	deux curiers	charrettes M. Maugars
	1316	Faud	Guillaume	20	id	id	domestique
	1317	Veclin	Anastasi Pierre	14	id	id	domestique
							de femme

Gard  
Les plantiers  
Forêt



Eure et Loire  
Voves  
Beauce



# Normalisation et IA

L'IA produit des données brutes, pour produire de la connaissance, il faut les normaliser

Normaliser = lier à un référentiel

Mais normaliser signifie aussi perte d'information

Référentiel interne : lié à la procédure de production humaine des données (historiques)

Référentiel externe:

- Géo-codage: évidence du référentiel mais complexité de la référence
- Profession : évidence de la référence mais complexité du référentiel

**T E K L I A**

[kermorvant@tekliia.com](mailto:kermorvant@tekliia.com)