

SOC FACE



Un projet de recherche collaboratif entre

Historiens, économistes, démographes

Archivistes

Experts en intelligence artificielle



FranceArchives
PORTAIL NATIONAL DES ARCHIVES

*Archives départementales
et municipales*

TEKLIA

Lionel Kesztenbaum



Journée au vert MIM
24/09/2024



The local face of social change: one century of French social structure seen from the ground, 1836–1936

Collecter, traiter, retranscrire, organiser et analyser l'ensemble des listes nominatives du recensement de 1836 à 1936 (20 recensements).

SocFace produira une base de données complète des individus ayant vécu en France entre 1836 et 1936 et l'utilisera pour analyser le changement social dans la longue durée.

Pourquoi Socface?

- Importance croissante des données micro (individuelles) dans la recherche quantitative en sciences sociales: économie, histoire, sociologie, démographie...
- Progrès considérables dans les technologies de reconnaissance d'écriture manuscrite et de traitement d'images.

**Produire des données
individuelles en masse à
l'échelle nationale**

**Développer les technologies de
traitement de vaste ensemble
de sources historiques**

**Diffuser les informations
obtenues au grand public
comme aux chercheurs**

La démographie historique française (résumé rapide)

- Reconstituer la population française pré-recensement: l'enquête Louis Henry
 - ❖ Reconstitution de population au niveau de la paroisse à partir des actes de baptême/mariage/sépulture.
 - ❖ Base des études de démographie historique sur la France.
- Les limites de l'observatoire local
 - ❖ Les reconstitutions ne fonctionnent qu'au sein d'une paroisse donnée.
 - ❖ Observations partielles pour tous ceux qui migrent.
- Une solution post-révolutionnaire: l'enquête TRA (Jacques Dupâquier)
 - ❖ Tire avantage des tables alphabétiques décennales.
 - ❖ Un échantillon sans limite dans l'espace et le temps.
 - ❖ Problème de représentativité + question des appariements.

Des données individuelles à l'échelle nationale

- Une histoire économique et sociale à la fois individuelle et nationale
 - ❖ Combler un vide entre enquêtes nationales à partir de statistiques agrégées et monographies plus locales.
 - ❖ Une période décisive dans l'histoire de la France: industrialisation, urbanisation, transformations sociales profondes.
- Différentes pistes poursuivies
 - ❖ Sur une année, plusieurs années combinées, avec appariement des individus ou non.
- Et après?
 - ❖ Base d'étude ou d'analyse pour de futures enquêtes historiques: matrice de la recherche en histoire quantitative contemporaine.
 - ❖ Liens avec la période contemporaine.

LES LISTES NOMINATIVES

Défis et objectifs



A small, vertically oriented handwritten list on aged paper with a black border. The text is dense and appears to be a list of names or entries, possibly organized in columns.



A small, vertically oriented handwritten list on aged paper with a black border. The text is dense and appears to be a list of names or entries, possibly organized in columns.



A larger, vertically oriented handwritten list on aged paper with a black border. The text is dense and appears to be a list of names or entries, possibly organized in columns.

DESIGNATION		NUMÉROS, PAR QUARTIER, VILLAGES, HAMEAUX ou VILLES			NOMS	ANNÉE	LIEU	NATIONALITÉ	SITUATION PAR RAPPORT	PROFESSION	REMARQUES	
des quartiers, villages ou hameaux	des rues ou des villages	des numéros	des ménages	des individus	DE FAMILLE	de naissance	de naissance	LITTÉ.	ou chef de ménage			
1	2	3	4	5	6	7	8	9	10	11	12	
				2	Eribout	Lucie	1886	Reims	F	épouse	ans	
				3	Eribout	Senni	1909	Claris	"	enfant	"	
				4	Eribout	Alphons	1912	"	"	"	"	
				5	Eribout	Reni	1916	aubervilliers	"	"	"	
				6	Eribout	Audé	1920	"	"	"	"	
				7	Pierre	Henriette	1881	Reims	"	belle sœur	Magasinier	Boulangerie Valenciennes
				8	Pierre	Liliane	1912	Bagnollet	"	filie	sans	
				9	Pierre	Lucien	1918	Woisylot	"	neveu	sans	
				10	Parmenier	Julave	1877	Mesfiteau	"	chef	maçon	
				11	Parmenier	Lucie	1867	Lambach	"	épouse	sans	
				12	Parmenier	Maximilien	1888	Paris	"	enfant	Confectionnier	Chulman
				13	Parmenier	Emma	1900	W. Denis	"	"	empl.	Minist. Pensions
				14	Parmenier	Ludovic	1902	"	"	"	plum.	Calix
				15	Caliez	Georges	1878	Lille	"	chef	Condamné	Magist.
				16	Caliez	Céline	1884	Taches	"	épouse	sans	
				17	Renauld	Jean	1869	Combrigny	"	chef	chef équipe	Magist.
				18	Renauld	Olivia	1869	Combrigny	"	"	"	

Une liste nominative:

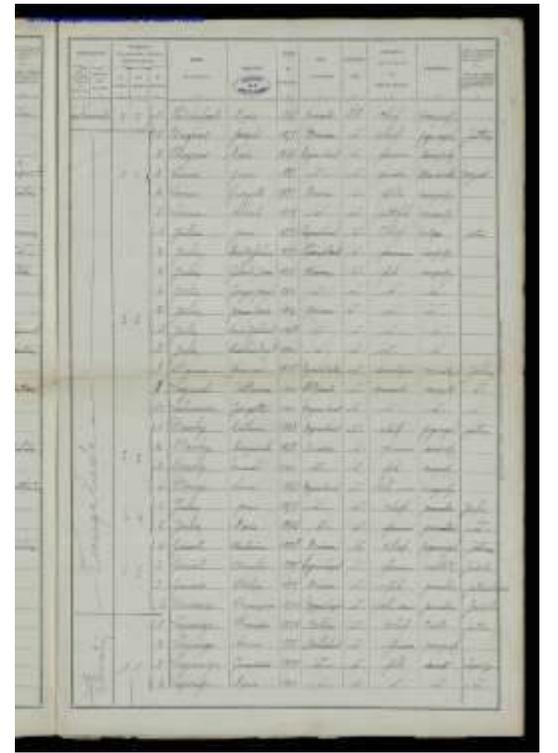
Aubervilliers, 1921

Rue
des
Fillettes

REGISTRATION		NUMEROUS, the number, being shown in it.			NAME	YEAR	SEX	NATIONALITY	SITUATION	PROFESSION	Remarks		
1	2	3	4	5	6	7	8	9	10	11	12		
		II	II	2	Eribout	Lucie	186	Reine	F	épouse	ans		
				3	Eribout	Sami	1909	Clara	"	enfant	"		
				4	Eribout	Alphonse	1912		"	"	"		
				5	Eribout	Reni	1916	Antoinette	"	"	"		
				6	Eribout	Audie	1920	"	"	"	"		
				7	Pierre	Jenneth	1881	Renie	"	bellevue	Magasin	Portefaite	
		8	Pierre	Liliane	1912	Barthé	"	Uice	ans				
		9	Pierre	Lucien	1918	Winglot	"	veuve	ans				

Pourquoi les listes nominatives?

- Une source abondante, simple, stable et standardisée.
 - Une source régulière dans le temps.
 - Déjà photographiée par beaucoup de services d'archives.
 - Source nationale, à l'identique.
 - Permet de construire un échantillon de toute la France (ou presque...).
- Une source idéale pour le passage à l'échelle de la reconnaissance de texte.
 - Une source qui n'a de sens que étudiée à grande échelle.



Défis et obstacles

➤ Collecte des données

- ❖ Les sources et les images sont conservées dans les archives départementales et municipales (100 dépôts d'archives).
- ❖ Très forte hétérogénéité dans la conservation des sources.

➤ Transcription du texte

- ❖ Très grande quantité de scripteurs différents, avec des pratiques spécifiques.
- ❖ Très grande diversité des entrées, en particulier pour les abréviations ('idem').

➤ Appariement des individus à travers le temps et l'espace

- ❖ Important gap in information: whole areas are missing for some periods, no collective dwellings, etc.
- ❖ Faiblesse de l'information individuelle disponible (ainsi souvent il n'y a qu'un seul prénom).

➤ Et en sciences sociales?

- ❖ Information sur la catégorie sociale limitée: uniquement la profession.
- ❖ Important vides: Paris est absente (jusqu'en 1926), de larges zones sont manquantes, etc.

→ Un défi commun: **la taille**, des millions d'images, des centaines de millions d'observations...

Les défis comme objectifs

➤ Collecte des données

- ❖ Un des objectifs du projet est de quantifier précisément l'état des collections en France.
- ❖ Le projet est incitatif pour que les Archives développent et améliorent leurs collections.

➤ Transcription du texte

- ❖ La diversité des cas n'est pas simplement une question d'écriture, mais aussi de pratiques, d'habitudes, etc.
- ❖ C'est un argument fort pour justifier la collaboration entre démographes, historiens et experts en informatique.

➤ Appariement des individus à travers le temps et l'espace

- ❖ Certaines particularités du recensement français peuvent aider à améliorer les appariements (nom de jeune fille).
- ❖ Permet d'évaluer comment la reconnaissance automatique influence l'appariement.

➤ Et en sciences sociales?

- ❖ Permet de déplacer la focale au-delà de zone très étudiées (hors de Paris, banlieue, etc.).
- ❖ Une base de données qui servira de base à d'autres études, autour d'autres sources.

Une source stable et uniforme?

N°	NOM	PROFESION	LIEU DE NAISSANCE	AGE EN ANS					OCCUPATION
				1860	1861	1862	1863	1864	
20	Agostini	Opérateur	...	1	1	1	1	1	73
21	Agostini	Opérateur	...	1	1	1	1	1	71
22	Agostini	Opérateur	...	1	1	1	1	1	60
23	Agostini	Opérateur	...	1	1	1	1	1	70
24	Agostini	Opérateur	...	1	1	1	1	1	71
25	Agostini	Opérateur	...	1	1	1	1	1	71
26	Agostini	Opérateur	...	1	1	1	1	1	71
27	Agostini	Opérateur	...	1	1	1	1	1	71
28	Agostini	Opérateur	...	1	1	1	1	1	71
29	Agostini	Opérateur	...	1	1	1	1	1	71
30	Agostini	Opérateur	...	1	1	1	1	1	71
31	Agostini	Opérateur	...	1	1	1	1	1	71
32	Agostini	Opérateur	...	1	1	1	1	1	71
33	Agostini	Opérateur	...	1	1	1	1	1	71
34	Agostini	Opérateur	...	1	1	1	1	1	71
35	Agostini	Opérateur	...	1	1	1	1	1	71
36	Agostini	Opérateur	...	1	1	1	1	1	71
37	Agostini	Opérateur	...	1	1	1	1	1	71
38	Agostini	Opérateur	...	1	1	1	1	1	71
39	Agostini	Opérateur	...	1	1	1	1	1	71
40	Agostini	Opérateur	...	1	1	1	1	1	71

1836

Cases à cocher

N°	NOM	PROFESION	LIEU DE NAISSANCE	AGE EN ANS					OCCUPATION
				1860	1861	1862	1863	1864	
30	Agostini	Opérateur	...	1	1	1	1	1	73
31	Agostini	Opérateur	...	1	1	1	1	1	71
32	Agostini	Opérateur	...	1	1	1	1	1	60
33	Agostini	Opérateur	...	1	1	1	1	1	70
34	Agostini	Opérateur	...	1	1	1	1	1	71
35	Agostini	Opérateur	...	1	1	1	1	1	71
36	Agostini	Opérateur	...	1	1	1	1	1	71
37	Agostini	Opérateur	...	1	1	1	1	1	71
38	Agostini	Opérateur	...	1	1	1	1	1	71
39	Agostini	Opérateur	...	1	1	1	1	1	71
40	Agostini	Opérateur	...	1	1	1	1	1	71
41	Agostini	Opérateur	...	1	1	1	1	1	71
42	Agostini	Opérateur	...	1	1	1	1	1	71
43	Agostini	Opérateur	...	1	1	1	1	1	71
44	Agostini	Opérateur	...	1	1	1	1	1	71
45	Agostini	Opérateur	...	1	1	1	1	1	71
46	Agostini	Opérateur	...	1	1	1	1	1	71
47	Agostini	Opérateur	...	1	1	1	1	1	71
48	Agostini	Opérateur	...	1	1	1	1	1	71
49	Agostini	Opérateur	...	1	1	1	1	1	71
50	Agostini	Opérateur	...	1	1	1	1	1	71

1886

âge

N°	NOM	PROFESION	LIEU DE NAISSANCE	AGE EN ANS					OCCUPATION
				1860	1861	1862	1863	1864	
60	Agostini	Opérateur	...	1	1	1	1	1	73
61	Agostini	Opérateur	...	1	1	1	1	1	71
62	Agostini	Opérateur	...	1	1	1	1	1	60
63	Agostini	Opérateur	...	1	1	1	1	1	70
64	Agostini	Opérateur	...	1	1	1	1	1	71
65	Agostini	Opérateur	...	1	1	1	1	1	71
66	Agostini	Opérateur	...	1	1	1	1	1	71
67	Agostini	Opérateur	...	1	1	1	1	1	71
68	Agostini	Opérateur	...	1	1	1	1	1	71
69	Agostini	Opérateur	...	1	1	1	1	1	71
70	Agostini	Opérateur	...	1	1	1	1	1	71
71	Agostini	Opérateur	...	1	1	1	1	1	71
72	Agostini	Opérateur	...	1	1	1	1	1	71
73	Agostini	Opérateur	...	1	1	1	1	1	71
74	Agostini	Opérateur	...	1	1	1	1	1	71
75	Agostini	Opérateur	...	1	1	1	1	1	71
76	Agostini	Opérateur	...	1	1	1	1	1	71
77	Agostini	Opérateur	...	1	1	1	1	1	71
78	Agostini	Opérateur	...	1	1	1	1	1	71
79	Agostini	Opérateur	...	1	1	1	1	1	71
80	Agostini	Opérateur	...	1	1	1	1	1	71

1936

Année de naissance

Des « normes » locales

DÉSIGNATION	NUMÉROS			NOMS	PRÉNOMS	ÂGE	NATIONALITÉ	PROFESSION	POSITION	OBSERVATIONS
	DES QUARTIERS, VILLES, BOULEVARDS, etc.	DES RUES	DES FAMILLES							
				371	Jeanjean	Philippine	48	français, sans profession	chef de ménage	
				372	Crissant	Victor	50	id	cultivateur	chef de ménage
				373	Sellet	Louise	47	id	sans profession	sa femme
89	111			374	Crissant	Leopold	13	id	id	son fils
				375	Crissant	Léon	7	id	id	id
				376	Crissant	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, BOULEVARDS, 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
					Chamoin	Jean Louis	18		ouvrier	chef
					Benoit	Léon	41	- id -	- id -	époux
					Chamoin	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*

Lecture en 2 dimensions

312 — 4 —

DÉSIGNATION		NUMÉROS, PAR QUARTIER, VILLAGES, HAMEAUX ou TERRES			NOMS		ANNEE	LIEU	NATIONAL.	SITUATION	PROFESSION		
des villages ou hameaux	des quartiers ou villages	des maisons	des maisons	des maisons	DE FAMILLE	PRÉNOMS	de NAISSANCE	de NAISSANCE	LIBRE	en chef de ménage			
1	2	3	4	5	6	7	8	9	10	11	12	13	
Rue des Filleuls	115	11	2	2	Coribout	Lucie	1886	Reims	F	épouse	ans		
			3	3	Coribout	Léoni	1909	Paris	"	enfant	"		
			4	4	Coribout	Alphonse	1912	"	"	"	"		
			5	5	Coribout	Reni	1916	subisillies	"	"	"		
			6	6	Coribout	Audré	1920	"	"	"	"		
			7	7	Pierre	Léonide	1881	Reims	"	bell. loeur	Magasin	Parabonnes Ménages	
			8	8	Pierre	Liliane	1918	Baguette	"	Uicia	ans		
			9	9	Pierre	Lucien	1918	Winghote	"	revenu	ans		
			10	10	Parmenier	Juliane	1877	Weschateau	"	chef	maçon		
			11	11	Parmenier	Lucie	1867	Lambach	"	époux	ans		
18	18	18	5	5	Parmenier	Maximilien	1885	Paris	"	enfant	Confectionner	Chulman	
			6	6	Parmenier	Emma	1900	W. Denis	"	"	empl.	Merisier	Paris
			7	7	Parmenier	Ludovic	1902	"	"	"	chef	Calix	
			8	8	Caliez	Jorges	1878	Lille	"	chef	Confectionner	Mag.	
			9	9	Caliez	Celine	1884	Stade	"	époux	ans		
10	10	Renault	Jean	1869	Combrigny	"	chef	chef équipe	Mag.				

1. Rue
2. Maison
3. Ménage
4. Individu

COLLECTE

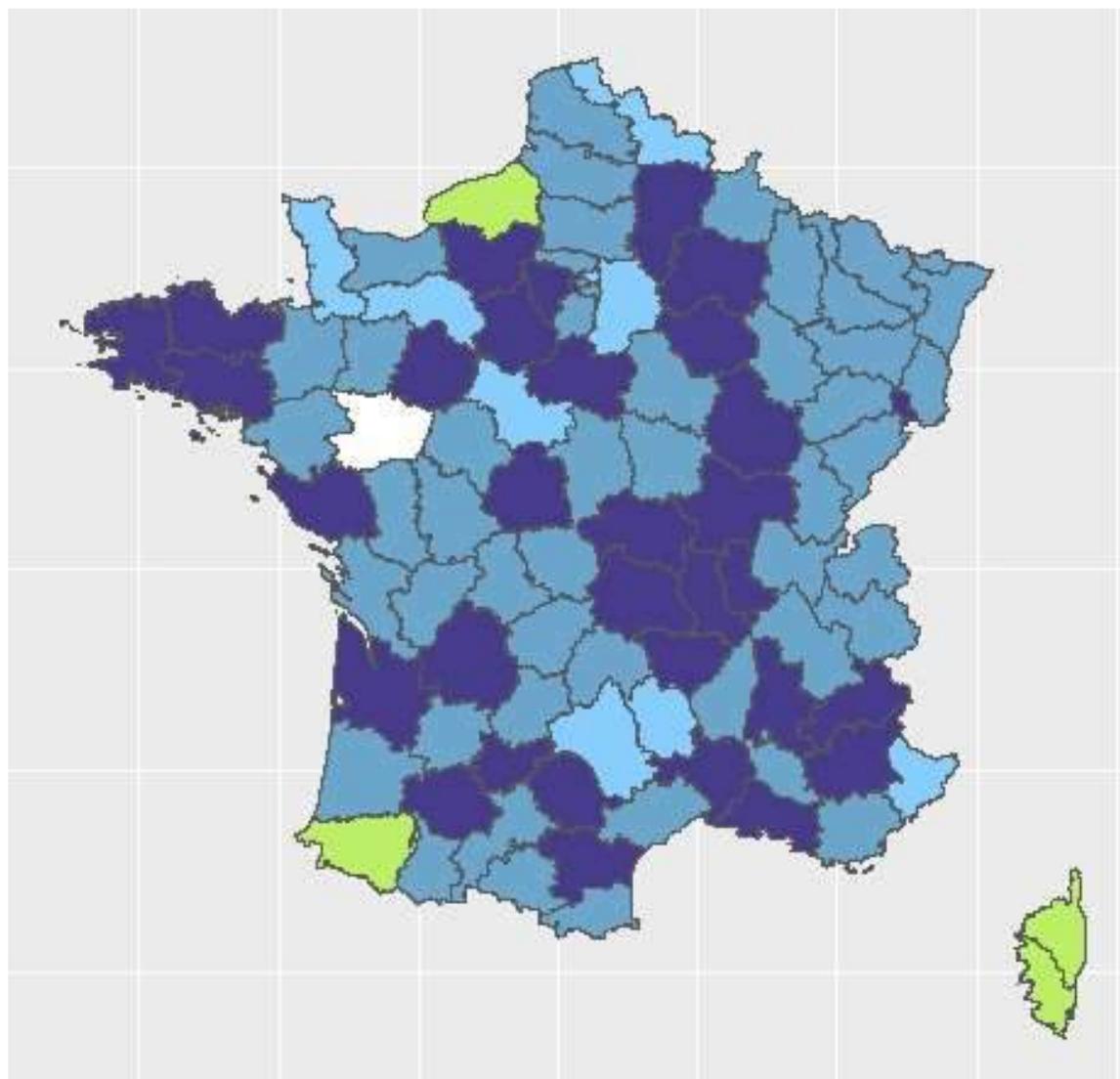
Obtenir et organiser beaucoup d'images





Etat de la collecte au 23/09/2024

Etat de la collecte	AD	(AM)
 En cours de collecte	9	(4)
 Images collectées	49	(4)
 Images dans Arkindex	33	(2)
 Ne participe pas	4	(1)
 Pas d'images	1	



Population couverte

➤ En théorie: 20 recensements x 36 000 communes sur 100 ans

- ❖ Entre 10 et 20 millions d'images (doubles).
- ❖ Entre 400 et 700 millions d'entrées (lignes).

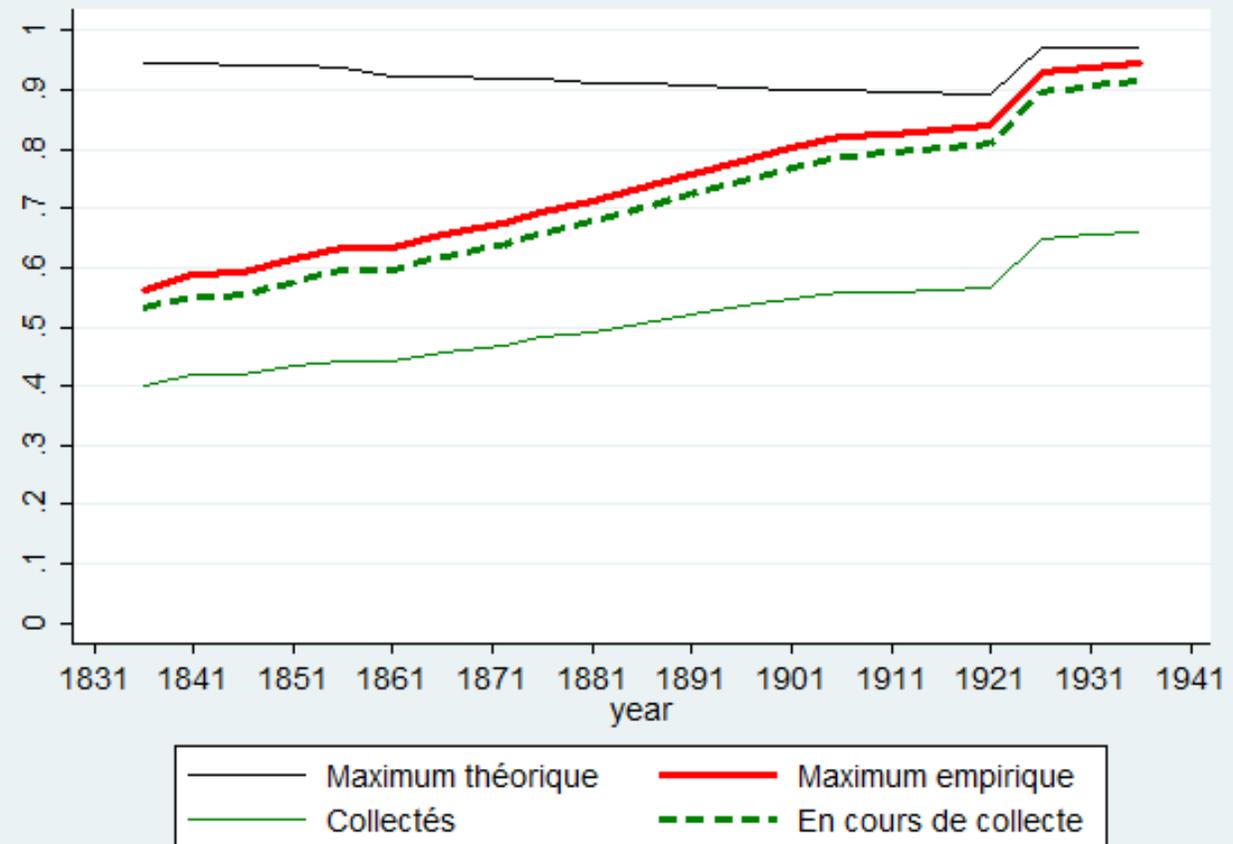
➤ Réalité (à ce jour):

- ❖ Environ 11 millions d'images.
- ❖ Environ 550 millions d'entrées (de lignes).

➤ Réalité (à venir, estimations grossières):

- ❖ Autour de 13 millions d'images (doubles).
- ❖ Sans doute 20 millions de simple page à traiter.
- ❖ Soit 600 millions d'entrées.

Couverture de la population par SocFace
(France métropolitaine)



Une page: Orléans, 1891

Rue de Gervaise

NO	NOM	PRENOM	AGE	PROFESSION	REMARQUES
11	11	11	11	11	11
12	12	12	12	12	12
13	13	13	13	13	13
14	14	14	14	14	14
15	15	15	15	15	15
16	16	16	16	16	16
17	17	17	17	17	17
18	18	18	18	18	18
19	19	19	19	19	19
20	20	20	20	20	20
21	21	21	21	21	21
22	22	22	22	22	22
23	23	23	23	23	23
24	24	24	24	24	24
25	25	25	25	25	25
26	26	26	26	26	26
27	27	27	27	27	27
28	28	28	28	28	28
29	29	29	29	29	29
30	30	30	30	30	30
31	31	31	31	31	31
32	32	32	32	32	32
33	33	33	33	33	33
34	34	34	34	34	34
35	35	35	35	35	35
36	36	36	36	36	36
37	37	37	37	37	37
38	38	38	38	38	38
39	39	39	39	39	39
40	40	40	40	40	40
41	41	41	41	41	41
42	42	42	42	42	42
43	43	43	43	43	43
44	44	44	44	44	44
45	45	45	45	45	45
46	46	46	46	46	46
47	47	47	47	47	47
48	48	48	48	48	48
49	49	49	49	49	49
50	50	50	50	50	50

Filter by type...

Hide

Double_page 19

A+

page 19-1

page 19-2



METADATA

- ark_lienark__ark_name <https://archives.orleans-metropole.fr/ark:24625/s00527b7f738738d/1>
- cote:coteTexte 1F35
- Spider ID 02016754-ad5c-4bb4-9f7e-360064a0d417

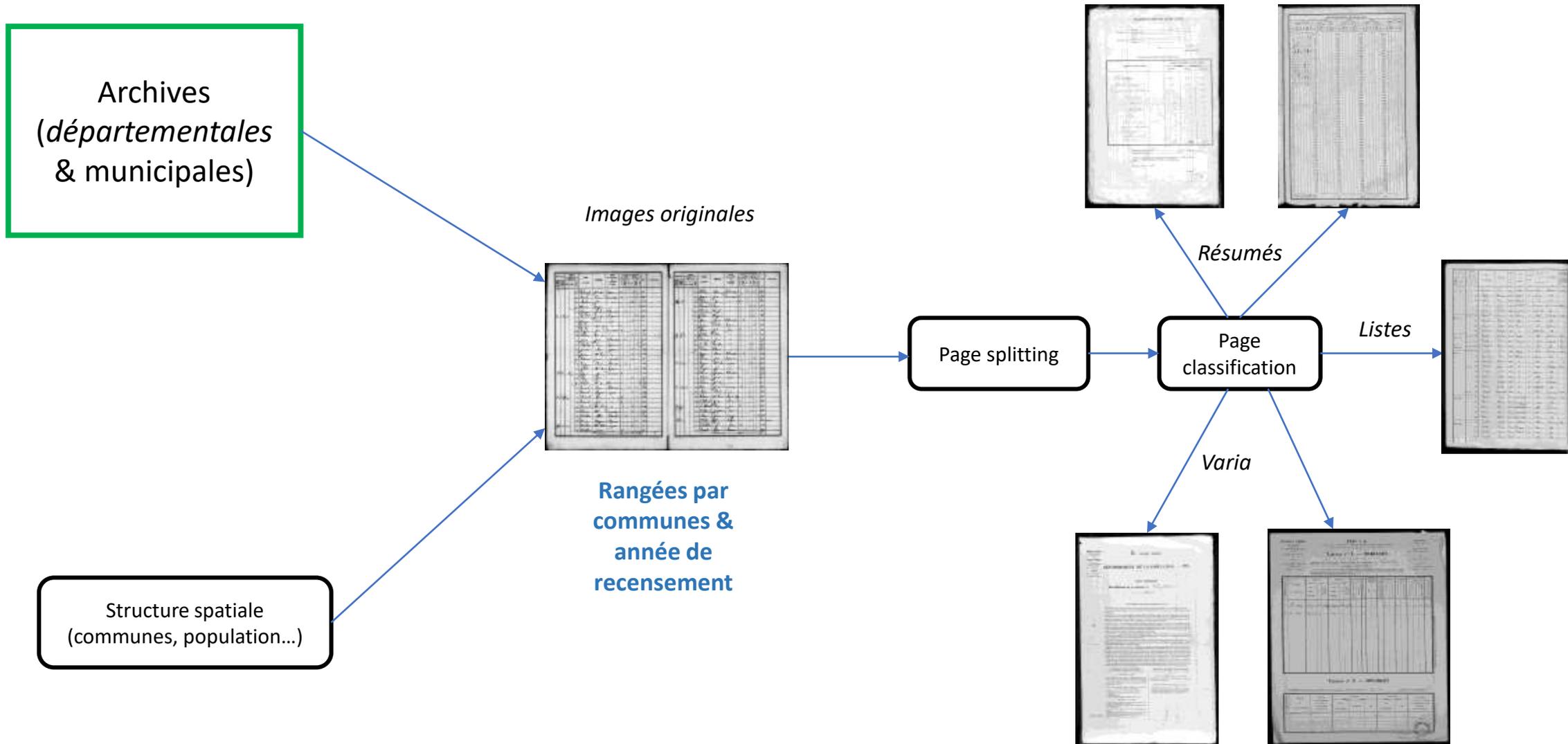
A small, square thumbnail image of a handwritten ledger page. The page is filled with dense, cursive handwriting in a grid-like format, typical of a financial or administrative record.

TRAITEMENT

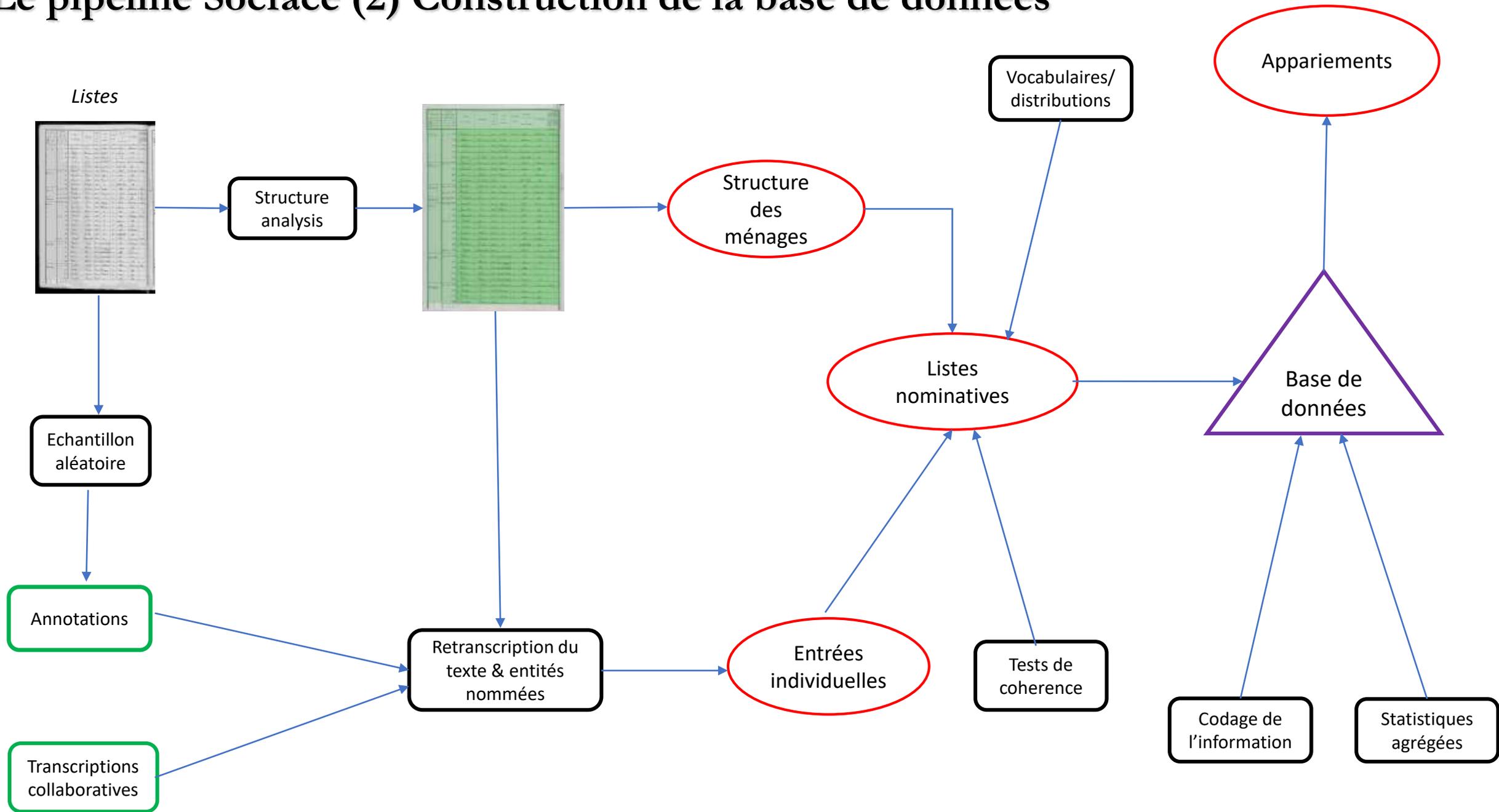
Extraire l'information des listes

A vertical thumbnail image of a handwritten ledger page, similar to the one in the top left. It shows a grid of handwritten entries in a cursive script.A larger, vertical thumbnail image of a handwritten ledger page. The handwriting is very dense and fills most of the page, organized into a clear grid structure.

Le pipeline Socface (1) collecte et analyse des images



Le pipeline Socface (2) Construction de la base de données



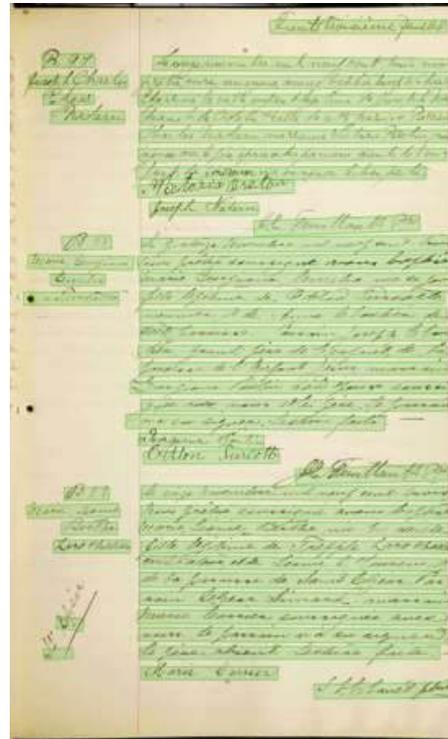
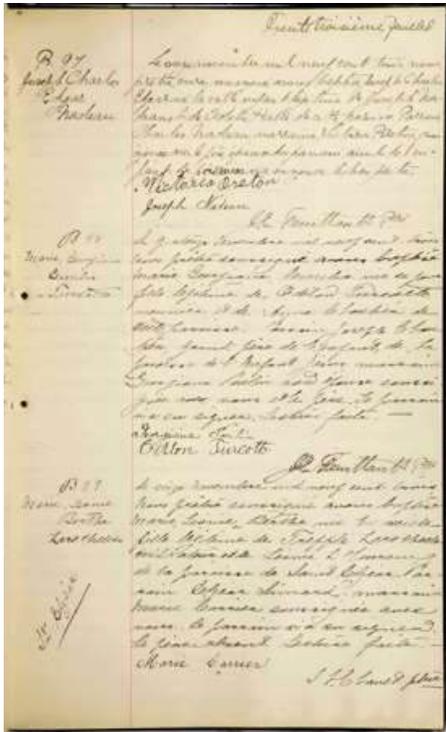
Chaîne de traitement automatique

Image

Détection des lignes

Reconnaissance

Extraction d'entités



Le onze novembre mil neuf cent trois , nous prêtre , curé soussigné , avons baptisé Joseph Charles Edgar né la veille , enfant légitime de Joseph na deau et de Odibs Hallé , de cette paroisse . Parrain Charles Nadeau , marraine Victoria Bleton , vu n ' a signé avec le père époux du parrain , Etaient de l ' en - Vant . Le parrain n ' a su signer . Lechoe faite . J . Victoria Arston Joseph Nadeau J . E . Feuilteault Ptre Le quatorze Novembre mil neuf cent trois nous prêtre soussigné avons baptisé Marie Georgiana , Emilia née ce jour fille légitime de Odilon Turcotte meunier , et de Anna Cloutier de cette paroisse . Parrain Joseph Clau tier , grand père de l ' enfant de la paroisse de l ' Enfant Jésus ; marraine Georgianna Pinaulin son épouse , soussi gné avec nous . et le père . Le parrain n ' a su signer . Lecture faite



Le [onze novembre mil neuf cent trois](#) , nous prêtre , curé soussigné , avons baptisé [Joseph Charles Edgar](#) né [la veille](#) , enfant légitime de [Joseph na deau](#) et de [Odibs Hallé](#) , de cette paroisse . Parrain [Charles Nadeau](#) , marraine [Victoria Bleton](#) , vu n ' a signé avec le père époux du parrain , Etaient de l ' en - Vant . Le parrain n ' a su signer . Lechoe faite . J . Victoria Arston Joseph Nadeau J . E . Feuilteault Ptre Le [quatorze Novembre mil neuf cent trois](#) nous prêtre soussigné avons baptisé [Marie Georgiana](#) , Emilia née [ce jour](#) fille légitime de [Odilon Turcotte meunier](#) , et de [Anna Cloutier](#) de cette paroisse . Parrain [Joseph Clau tier](#) , grand père de l ' enfant de la paroisse de l ' Enfant Jésus ; marraine [Georgianna Pinaulin](#) son épouse , soussi gné avec nous . et le père . Le parrain n ' a su signer . Lecture faite

Tous les modèles sont entraînés à partir d'exemples

Integration de HTR et NER



Approche séquentielle

DAN or PyLaia en HTR
SpaCy for NER

Approche intégrée

DAN for HTR+NER
Des tokens sont utilisés pour les entités

Un exemple avec retranscription et reconnaissance des entités nommées

NOM		PRÉNOM		AGE	SEX	PROFESSION	ÉTAT
4	Robert	Juliette	8	F	enfant	1. fille	
5	Robert	Albertine	3	F	-	-	
6	Johan	Jules	19	M	domestique	chauger	
7	Belouet	Jeanne	22	F	-	-	
8	Beignet	Charles	32	M	J les	chef	
9	Bezault	Hermance	27	F	-	épouse	
10	Bugnet	René	6	M	-	1. fils	
11	Biignet	Juliette	3	F	-	1. fille	
1	Portin	Jean	50	M	J leer	chef	
2	Gronni	Marie	46	F	-	épouse	
3	Fortin	Juliette	19	F	-	sa fille	
4	Portin	Georgette	12	F	-	-	
5	Portin	Maurice	10	M	-	-	
6	Vappereau	Andrée	1	F	-	sa fille	
7	Grangeau	Philomèn	47	F	meunier	chef	

surname Robert, firstname Juliette, occupation néant, link l. fille, age 8, nationality française
 surname Robert, firstname Albertine, occupation idem, link idem, age 3, nationality idem
 surname Johan, firstname Jules, occupation domestique, link chauger, age 19, nationality idem
 surname Belouet, firstname Jeanne, occupation idem, link idem, age 22, nationality idem
 surname Beignet, firstname Charles, occupation J les, link chef, age 32, nationality idem
 surname Bezault, firstname Hermance, occupation idem, link épouse, age 27, nationality idem
 surname Bugnet, firstname René, occupation idem, link l. fils, age 6, nationality idem
 surname Biignet, firstname Juliette, occupation idem, link l. fille, age 3, nationality idem
 surname Portin, firstname Jean, occupation J leer, link chef, age 50, nationality idem
 surname Gronni, firstname Marie, occupation idem, link épouse, age 46, nationality idem
 surname Fortin, firstname Juliette, occupation néant, link sa fille, age 19, nationality idem
 surname Portin, firstname Georgette, occupation idem, link idem, age 12, nationality idem
 surname Portin, firstname Maurice, occupation idem, link idem, age 10, nationality idem
 surname Vappereau, firstname Andrée, occupation idem, link idem, age 1, nationality idem
 surname Grangeau, firstname Philomèn, occupation meunier, link chef, age 47, nationality idem

LES LISTES NOMINATIVES

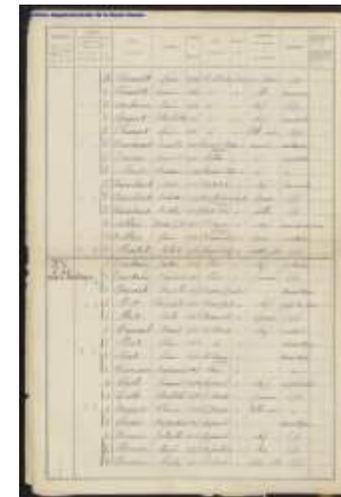
Pour faire quoi?



A small, vertically oriented table with multiple columns and rows of text, likely a nominative list. The text is dense and appears to be a list of names or entries.



A small, vertically oriented table with multiple columns and rows of text, likely a nominative list. The text is dense and appears to be a list of names or entries.



A larger, vertically oriented table with multiple columns and rows of text, likely a nominative list. The text is dense and appears to be a list of names or entries.

Différents angles d'analyse

➤ Changement de structure dans la durée

- ❖ Evolution du marché du travail: situation selon l'état marital, le genre, etc.
- ❖ Hétérogénéité spatiale des phénomènes.

➤ Chocs et modifications ponctuelles

- ❖ Conséquences à court, moyen et long terme.
- ❖ Par exemple: crise du phylloxera; Première guerre mondiale.

➤ Structuration de l'espace et des activités économiques

- ❖ Croisement des informations individuelles avec des données spatiales: structuration de l'espace.
- ❖ Effet de la transition entre agriculture et industrie à l'échelle locale.



Le défi des appariements: étudier la mobilité

1886

31	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																																		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

1896

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

1906

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

Le défi des normalisations: mesurer la position socio-économique

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

Un exemple: les professions

- Extraire les professions pour 2 départements: 29 & 71.
- 17.8M mentions (lines) correspondent à 535 000 professions différentes.
 - Dont 6,58M lines sans profession (NA).
 - La profession la plus fréquente est “idem” avec 3M lignes (17% du total).

Première homogénéisation: quelques exemples

occupation	n()	occupation_clear	statut	Colonne1
emp ch de fer	403	employé au chemin de fer	1	0.002251817
emp ch fer	354	employé au chemin de fer	1	0.001978023
emp ch. de fer	103	employé au chemin de fer	1	0.000575526
emp. ch fer	210	employé au chemin de fer	1	0.001173403
emp. ch. de fer	231	employé au chemin de fer	1	0.001290744
emp. ch. fer	315	employé au chemin de fer	1	0.001760105
empl. ch fer	131	employé au chemin de fer	1	0.00073198
empl. ch. de fer	455	employé au chemin de fer	1	0.002542374
empl. ch. fer	684	employé au chemin de fer	1	0.003821942
empl. de ch. de fer	113	employé au chemin de fer	1	0.000631403
employé au ch. de fer	219	employé au chemin de fer	1	0.001223692
employé au chemin de fer	1446	employé au chemin de fer	1	0.00807972
employé ch fer	158	employé au chemin de fer	1	0.000882846
employé ch. de fer	135	employé au chemin de fer	1	0.000754331
employé ch. fer	108	employé au chemin de fer	1	0.000603465
employé de ch. de fer	125	employé au chemin de fer	1	0.000698454
employé de chemin de fer	399	employé au chemin de fer	1	0.002229466

L'infinite des entrées: contremaître

occupation	n()	occupation_clear	statut	Colonne1	Colonne2
contre maître	448	contremaitre	1	0.002503261	83.35591017
contre maitre	563	contremaitre	1	0.003145839	83.35905601
contremaitre	192	contremaitre	1	0.001072826	83.36012884
coutre maitre	229	contremaitre	1	0.001279568	83.60030216
coutre maître	214	contremaitre	1	0.001195754	83.60149791

occupation	n()	occupation_clear	statut	Colonne1
contremacte	1	contremacte	0	5.59E-06
contremactré	1	contremactré	0	5.59E-06
contremadre	1	contremadre	0	5.59E-06
contremaide berval	1	contremaide berval	0	5.59E-06
contremaide du cateur	1	contremaide du cateur	0	5.59E-06
contremailleur	1	contremailleur	0	5.59E-06
contremaître brate	1	contremaître brate	0	5.59E-06
contremaite	5	contremaite	0	2.79E-05
contremaite d'usine	1	contremaite d'usine	0	5.59E-06
contremaitee	1	contremaitee	0	5.59E-06
contremaiteuse	1	contremaiteuse	0	5.59E-06
contremaitr	6	contremaitr	0	3.35E-05
contre maître	448	contremaitre	1	0.002503261
contre maitre	563	contremaitre	1	0.003145839
contremaitre	192	contremaitre	1	0.001072826
coutre maitre	229	contremaitre	1	0.001279568
coutre maître	214	contremaitre	1	0.001195754
contremaître	7	contremaître	0	3.91E-05
contremaitre adricole	1	contremaitre adricole	0	5.59E-06
contremaitre agricole	1	contremaitre agricole	0	5.59E-06
contremaitre an port	1	contremaitre an port	0	5.59E-06
contremaitre au port	1	contremaitre au port	0	5.59E-06
contremaitre aulutannerie	1	contremaitre aulutannerie	0	5.59E-06
contremaitre bis	1	contremaitre bis	0	5.59E-06
contremaitre cardinier	1	contremaitre cardinier	0	5.59E-06
contremaître carrier	1	contremaître carrier	0	5.59E-06
contremaitre coriguetier	1	contremaitre coriguetier	0	5.59E-06
contremaitre d'atelier	1	contremaitre d'atelier	0	5.59E-06

occupation	n()	occupation_clear	statut	Colonne1
contremaitre risseur	1	contremaitre risseur	0	5.59E-06
contremaitre secage	1	contremaitre secage	0	5.59E-06
contremaitre sure	1	contremaitre sure	0	5.59E-06
contremaitre tatelier	1	contremaitre tatelier	0	5.59E-06
contremaitre tindine	1	contremaitre tindine	0	5.59E-06
contremaitre tissage	1	contremaitre tissage	0	5.59E-06
contremaitresse	1	contremaitresse	0	5.59E-06
contremaits	1	contremaits	0	5.59E-06
contremant	1	contremant	0	5.59E-06
contremante d'usine	1	contremante d'usine	0	5.59E-06
contremantier	2	contremantier	0	1.12E-05
contremantre	15	contremantre	0	8.38E-05
contremantre mécanicien	1	contremantre mécanicien	0	5.59E-06
contremantresse	1	contremantresse	0	5.59E-06
contremantrine	1	contremantrine	0	5.59E-06
contremartier	1	contremartier	0	5.59E-06
contremartre	3	contremartre	0	1.68E-05
contrematre	2	contrematre	0	1.12E-05
contrematre centremerie	1	contrematre centremerie	0	5.59E-06
contremattre	9	contremattre	0	5.03E-05
contremattre de flature	1	contremattre de flature	0	5.59E-06
contremaute	1	contremaute	0	5.59E-06
contremautr	1	contremautr	0	5.59E-06
contreme d'usine	1	contreme d'usine	0	5.59E-06
contremeitre	2	contremeitre	0	1.12E-05
contremeitre mateur	1	contremeitre mateur	0	5.59E-06
contremeneur de mar	1	contremeneur de mar	0	5.59E-06
contrementier	6	contrementier	0	3.35E-05
contrementier malayer	1	contrementier malayer	0	5.59E-06
contremestre	1	contremestre	0	5.59E-06

Dissemination: retour aux archives, et au-delà

- Raw database to be distributed by the Archives
 - ❖ On a national database (*FranceArchives*), with a search engine.
 - ❖ Sur les bases des Archives Départementales.
 - ❖ Direct links with the image.
- Encoded database to be distributed for research
 - ❖ A database where various information are organized and encoded (occupation, place of birth, etc.).
 - ❖ A database with probabilistic linkage between individuals.
- Opening on other sources: a model for disseminating French national archives?

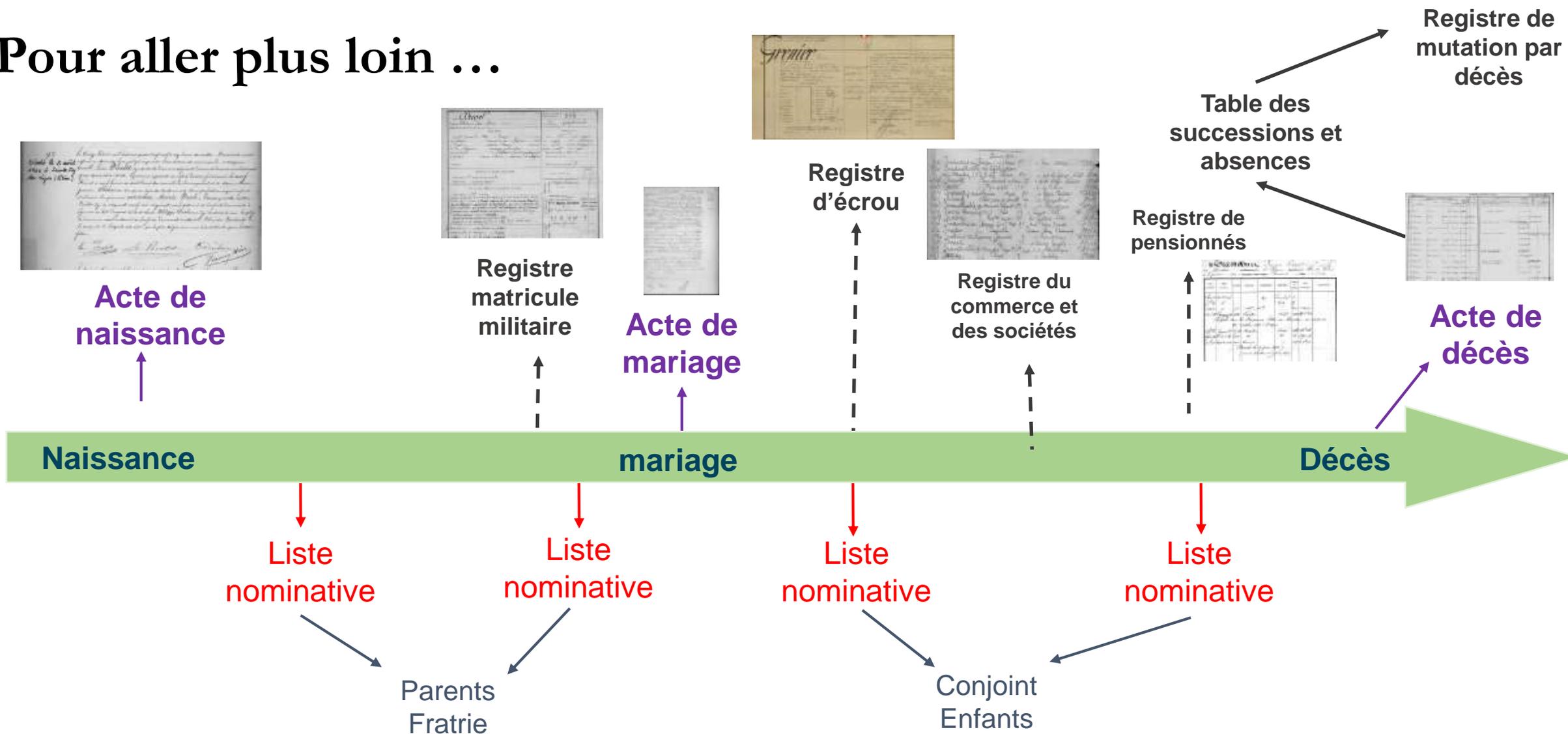


Pour aller plus loin: développer les interactions au-delà de Socface

- Développer un modèle général d'appariement
 - ❖ Adaptable à la problématique (deux points à 5 ans d'intervalle, trajectoire complète d'un individu, un ménage à plusieurs moments du temps, etc.).
 - ❖ Qui mesure les biais possibles.
- Etude de la mobilité (géographique et sociale)
 - ❖ Ajout d'autres sources pour affiner la qualité les informations.
- Une suite immédiate
 - ❖ Volontaires welcome!



Pour aller plus loin ...



Pour aller plus loin: développer les interactions au-delà de Socface

- Développer un modèle général d'appariement
 - ❖ Adaptable à la problématique (deux points à 5 ans d'intervalle, trajectoire complète d'un individu, un ménage à plusieurs moments du temps, etc.).
 - ❖ Qui mesure les biais possibles.
- Etude de la mobilité (géographique et sociale)
 - ❖ Ajout d'autres sources pour affiner la qualité les informations.
- Une suite immédiate
 - ❖ Volontaires welcome!

SOC FACE



Merci de votre attention



INSTITUT NATIONAL
D'ÉTUDES DÉMOGRAPHIQUES



PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS



T E K L I A

FranceArchives
PORTAIL NATIONAL DES ARCHIVES