

# 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 table with multiple columns and rows of text. The text is dense and appears to be a list or ledger.



A small, vertically oriented handwritten table with multiple columns and rows of text. The text is dense and appears to be a list or ledger.



A larger, vertically oriented handwritten table with multiple columns and rows of text. The text is dense and appears to be a list or ledger.

DESIGNATION		NUMÉROS, PAR QUARTIER, VILLAGES, HAMEAUX ou VILLES			NOMS	ANNÉE	LIEU	NATIONALITÉ	SITUATION PAR RAPPORT	PROFESSION	Notes	
des quartiers, villages ou hameaux	des rues ou des villes	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	"	"	"	époux	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	1875	Combrigny	"	"	"	

Une liste nominative:

Aubervilliers, 1921

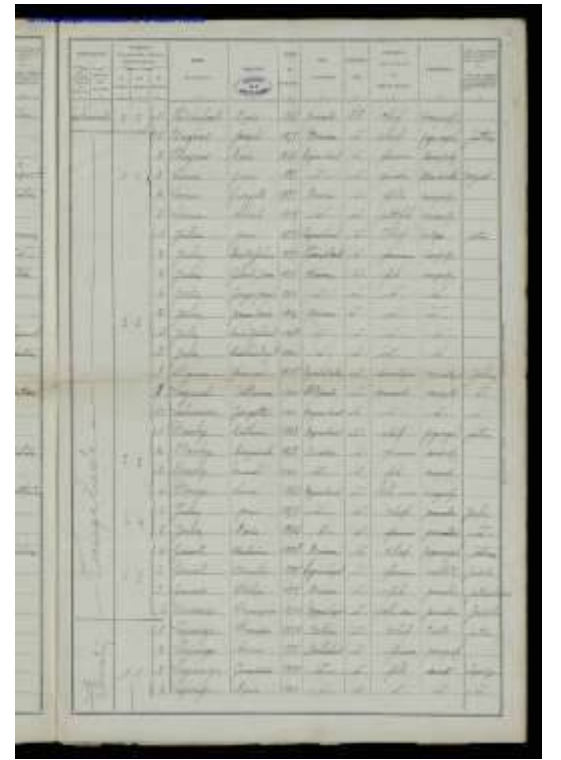
Rue  
des  
Elleltes

REGISTRATION		NUMEROUS, the number, being shown in it.			NAME	RESIDENCE	SEX	AGE	NATIONALITY	SITUATION	PROFESSION	Remarks	
1	2	3	4	5	6	7	8	9	10	11	12	13	
		11	2	Eribout	Lucie	186	Reims	F		épouse	ans		
			3	Eribout	Samy	1909	Reims				enfant		
			4	Eribout	Alphonse	1912							
			5	Eribout	René	1916	autisillon						
			6	Eribout	Audré	1920							
				7	Pierre	Jennette	1881	Reims			bell. love	Magasin	Boislaucourt Mancini
			8	Pierre	Liliane	1912	Baguette			Uice	ans		
			9	Pierre	Lucien	1918	Hoinglot			neveu	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	DATE DE NAISSANCE					AGE	OBSERVATIONS
				AN	MOIS	JOUR	HEURE	MINUTE		
20	4	Boyer	Boyer	18	10	15			73	
21	5	Boyer	Boyer	18	10	15			71	
22	6	Boyer	Boyer	18	10	15			69	
23	7	Boyer	Boyer	18	10	15			67	
24	8	Boyer	Boyer	18	10	15			65	
25	9	Boyer	Boyer	18	10	15			63	
26	10	Boyer	Boyer	18	10	15			61	
27	11	Boyer	Boyer	18	10	15			59	
28	12	Boyer	Boyer	18	10	15			57	
29	13	Boyer	Boyer	18	10	15			55	
30	14	Boyer	Boyer	18	10	15			53	
31	15	Boyer	Boyer	18	10	15			51	
32	16	Boyer	Boyer	18	10	15			49	
33	17	Boyer	Boyer	18	10	15			47	
34	18	Boyer	Boyer	18	10	15			45	
35	19	Boyer	Boyer	18	10	15			43	
36	20	Boyer	Boyer	18	10	15			41	
37	21	Boyer	Boyer	18	10	15			39	
38	22	Boyer	Boyer	18	10	15			37	
39	23	Boyer	Boyer	18	10	15			35	
40	24	Boyer	Boyer	18	10	15			33	
41	25	Boyer	Boyer	18	10	15			31	
42	26	Boyer	Boyer	18	10	15			29	
43	27	Boyer	Boyer	18	10	15			27	
44	28	Boyer	Boyer	18	10	15			25	
45	29	Boyer	Boyer	18	10	15			23	
46	30	Boyer	Boyer	18	10	15			21	
47	31	Boyer	Boyer	18	10	15			19	
48	32	Boyer	Boyer	18	10	15			17	
49	33	Boyer	Boyer	18	10	15			15	
50	34	Boyer	Boyer	18	10	15			13	
51	35	Boyer	Boyer	18	10	15			11	
52	36	Boyer	Boyer	18	10	15			9	
53	37	Boyer	Boyer	18	10	15			7	
54	38	Boyer	Boyer	18	10	15			5	
55	39	Boyer	Boyer	18	10	15			3	
56	40	Boyer	Boyer	18	10	15			1	

1836

Cases à cocher

N°	NOM	PROFESION	LIEU DE NAISSANCE	DATE DE NAISSANCE					AGE	OBSERVATIONS
				AN	MOIS	JOUR	HEURE	MINUTE		
30	Boyer	Boyer	Boyer	18	10	15			73	
31	Boyer	Boyer	Boyer	18	10	15			71	
32	Boyer	Boyer	Boyer	18	10	15			69	
33	Boyer	Boyer	Boyer	18	10	15			67	
34	Boyer	Boyer	Boyer	18	10	15			65	
35	Boyer	Boyer	Boyer	18	10	15			63	
36	Boyer	Boyer	Boyer	18	10	15			61	
37	Boyer	Boyer	Boyer	18	10	15			59	
38	Boyer	Boyer	Boyer	18	10	15			57	
39	Boyer	Boyer	Boyer	18	10	15			55	
40	Boyer	Boyer	Boyer	18	10	15			53	
41	Boyer	Boyer	Boyer	18	10	15			51	
42	Boyer	Boyer	Boyer	18	10	15			49	
43	Boyer	Boyer	Boyer	18	10	15			47	
44	Boyer	Boyer	Boyer	18	10	15			45	
45	Boyer	Boyer	Boyer	18	10	15			43	
46	Boyer	Boyer	Boyer	18	10	15			41	
47	Boyer	Boyer	Boyer	18	10	15			39	
48	Boyer	Boyer	Boyer	18	10	15			37	
49	Boyer	Boyer	Boyer	18	10	15			35	
50	Boyer	Boyer	Boyer	18	10	15			33	
51	Boyer	Boyer	Boyer	18	10	15			31	
52	Boyer	Boyer	Boyer	18	10	15			29	
53	Boyer	Boyer	Boyer	18	10	15			27	
54	Boyer	Boyer	Boyer	18	10	15			25	
55	Boyer	Boyer	Boyer	18	10	15			23	
56	Boyer	Boyer	Boyer	18	10	15			21	
57	Boyer	Boyer	Boyer	18	10	15			19	
58	Boyer	Boyer	Boyer	18	10	15			17	
59	Boyer	Boyer	Boyer	18	10	15			15	
60	Boyer	Boyer	Boyer	18	10	15			13	
61	Boyer	Boyer	Boyer	18	10	15			11	
62	Boyer	Boyer	Boyer	18	10	15			9	
63	Boyer	Boyer	Boyer	18	10	15			7	
64	Boyer	Boyer	Boyer	18	10	15			5	
65	Boyer	Boyer	Boyer	18	10	15			3	
66	Boyer	Boyer	Boyer	18	10	15			1	

1886

âge

N°	NOM	PROFESION	LIEU DE NAISSANCE	DATE DE NAISSANCE					AGE	OBSERVATIONS
				AN	MOIS	JOUR	HEURE	MINUTE		
61	Boyer	Boyer	Boyer	18	10	15			73	
62	Boyer	Boyer	Boyer	18	10	15			71	
63	Boyer	Boyer	Boyer	18	10	15			69	
64	Boyer	Boyer	Boyer	18	10	15			67	
65	Boyer	Boyer	Boyer	18	10	15			65	
66	Boyer	Boyer	Boyer	18	10	15			63	
67	Boyer	Boyer	Boyer	18	10	15			61	
68	Boyer	Boyer	Boyer	18	10	15			59	
69	Boyer	Boyer	Boyer	18	10	15			57	
70	Boyer	Boyer	Boyer	18	10	15			55	
71	Boyer	Boyer	Boyer	18	10	15			53	
72	Boyer	Boyer	Boyer	18	10	15			51	
73	Boyer	Boyer	Boyer	18	10	15			49	
74	Boyer	Boyer	Boyer	18	10	15			47	
75	Boyer	Boyer	Boyer	18	10	15			45	
76	Boyer	Boyer	Boyer	18	10	15			43	
77	Boyer	Boyer	Boyer	18	10	15			41	
78	Boyer	Boyer	Boyer	18	10	15			39	
79	Boyer	Boyer	Boyer	18	10	15			37	
80	Boyer	Boyer	Boyer	18	10	15			35	
81	Boyer	Boyer	Boyer	18	10	15			33	
82	Boyer	Boyer	Boyer	18	10	15			31	
83	Boyer	Boyer	Boyer	18	10	15			29	
84	Boyer	Boyer	Boyer	18	10	15			27	
85	Boyer	Boyer	Boyer	18	10	15			25	
86	Boyer	Boyer	Boyer	18	10	15			23	
87	Boyer	Boyer	Boyer	18	10	15			21	
88	Boyer	Boyer	Boyer	18	10	15			19	
89	Boyer	Boyer	Boyer	18	10	15			17	
90	Boyer	Boyer	Boyer	18	10	15			15	
91	Boyer	Boyer	Boyer	18	10	15			13	
92	Boyer	Boyer	Boyer	18	10	15			11	
93	Boyer	Boyer	Boyer	18	10	15			9	
94	Boyer	Boyer	Boyer	18	10	15			7	
95	Boyer	Boyer	Boyer	18	10	15			5	
96	Boyer	Boyer	Boyer	18	10	15			3	
97	Boyer	Boyer	Boyer	18	10	15			1	

1936

Année de naissance

# Des « normes » locales

DÉSIGNATION	NUMÉROS			NOMS	PRÉNOMS	ÂGE	NATIONALITÉ	PROFESSION	POSITION	OBSERVATIONS	
	DES QUARTIERS, VILLES, BOULEVARD, RUE	DES	DES								
DES	DES	DES	DES	DES	DES	DES	DES	DES	DES	DES	
				371	Jeanjean	Philippine	48	français	sans profession	se traite	
				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, BOULEVARD, RUE	DES	DES								
DES	DES	DES	DES	DES	DES	DES	DES	DES	DES	DES	
				39	Lambert	Anton	21	f	ouvrier	enfant	
					Lambert	Antoine	20	- id -	- id -	- id -	
				41	Blondel	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*

# 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 des villages ou hameaux	des maisons	des maisons	des maisons	DE FAMILLE	PRÉNOMS	de NAISSANCE	de NAISSANCE	LIBRE	ou chef de ménage		
1	2	3	4	5	6	7	8	9	10	11	12	13
Rue des Éillettes	115	11	2	Erribout	Lucie	1866	Reims	F	épouse	ans		
			3	Erribout	Léoni	1909	Paris	"	enfant	"		
			4	Erribout	Alphonse	1912	"	"	"	"		
			5	Erribout	Reni	1916	subisvillain	"	"	"		
			6	Erribout	Audré	1920	"	"	"	"		
		4	Pierre	Léonide	1881	Reims	"	bell. loeur	Magasin	Parabonville		
		8	Pierre	Liliane	1918	Baguville	"	Uicia	ans			Valenciennes
		3	Pierre	Lucien	1918	Woignoles	"	revenu	ans			
		1	Parmenier	Jules	1877	Wassigny	"	chef	maçon			
		2	Parmenier	Lucie	1867	Lambach	"	époux	ans			
3	Parmenier	Maximilien	1885	Paris	"	enfant	Confectionner	Chulman				
4	Parmenier	Emma	1900	Wassigny	"	"	empl.	Merisier	Paris			
5	Parmenier	Ludovic	1902	"	"	"	"	chef	Calix			
1	Caliez	Jorges	1878	Lille	"	chef	Confectionner	Mag				
2	Caliez	Céline	1884	Stache	"	époux	ans					
1	Renaud	Jean	1869	Combray	"	chef	chef équipe	Mag				

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

# *COLLECTE*

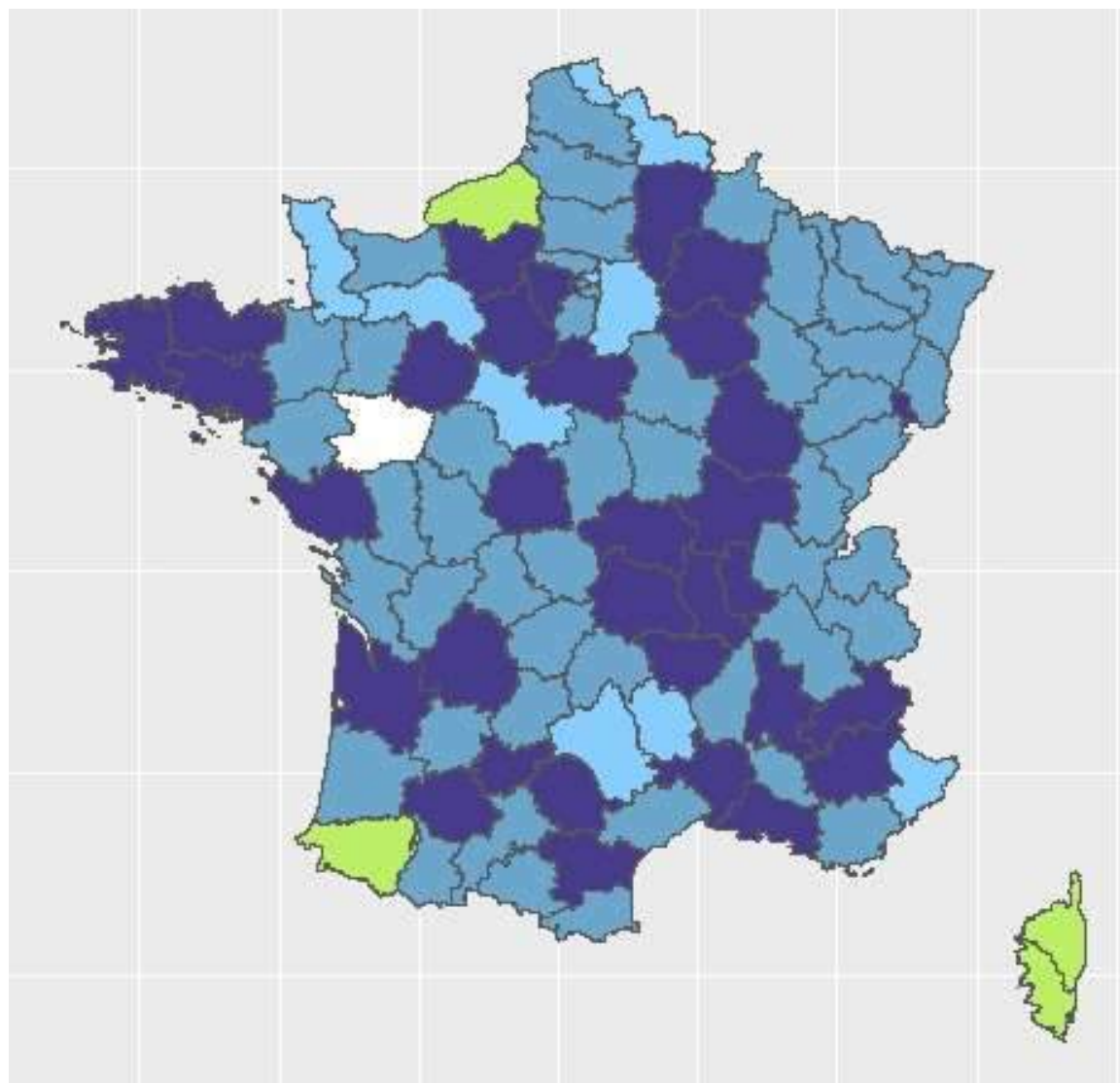
Obtenir et organiser beaucoup d'images





## 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).

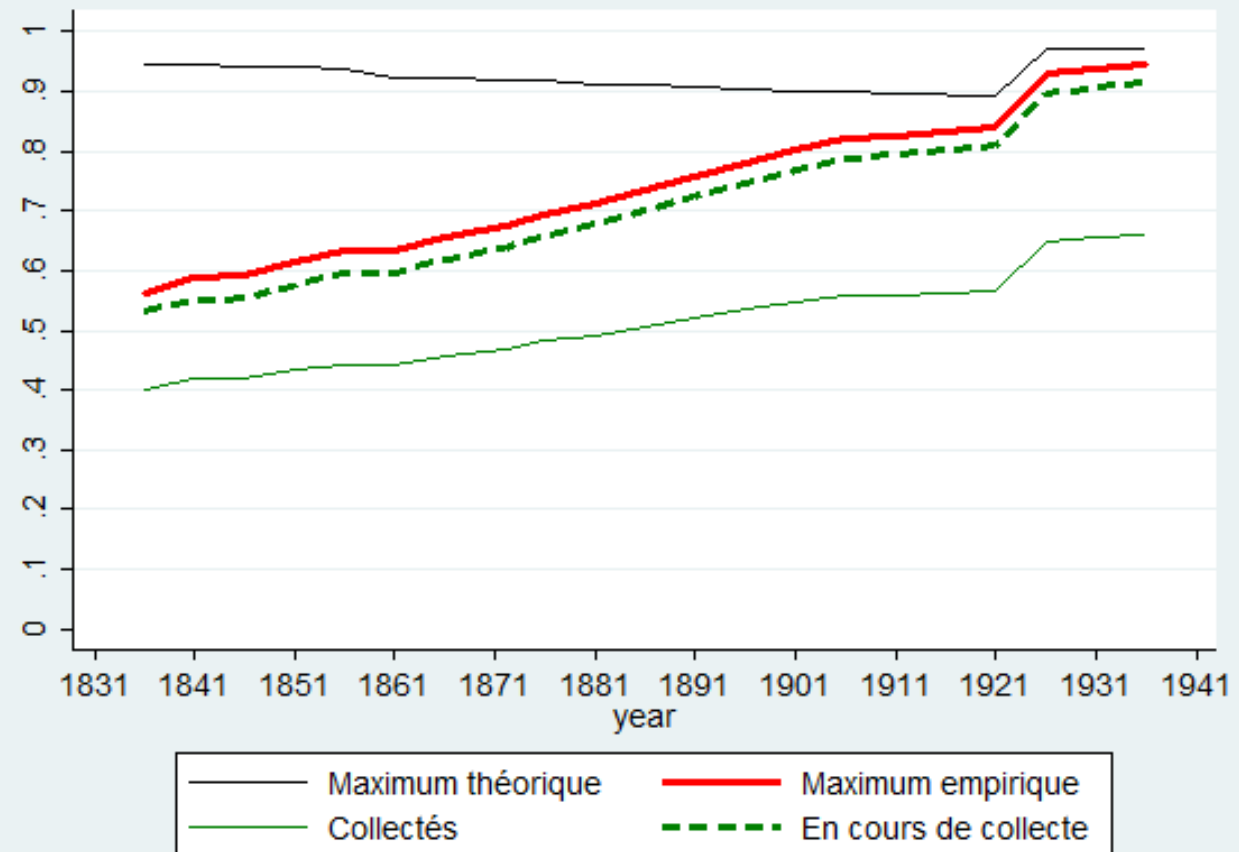
➤ Réalité (à ce jour):

- ❖ 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

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

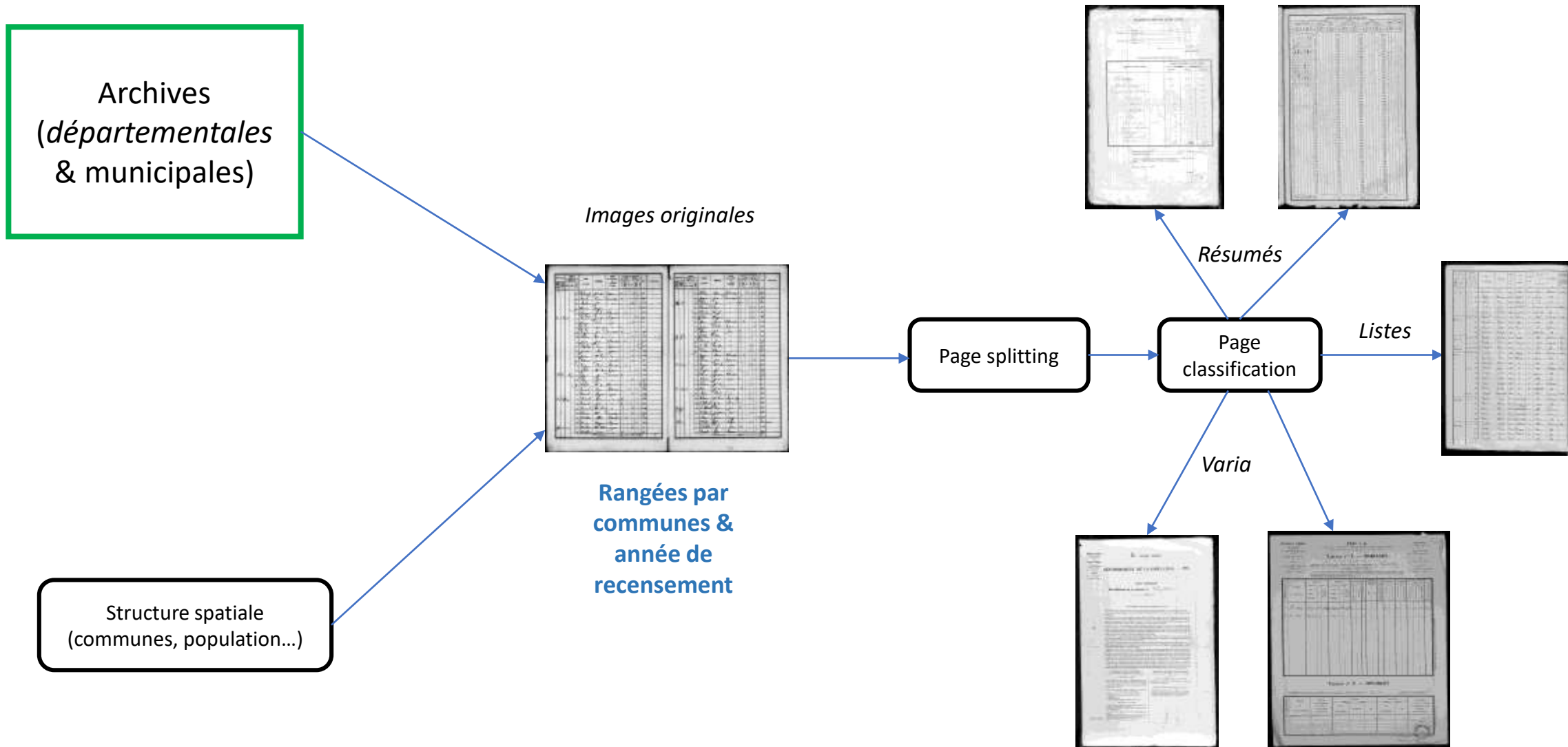
A small, vertical thumbnail image of a handwritten document, likely a list or ledger, with multiple columns and rows of text.

# ***TRAITEMENT***

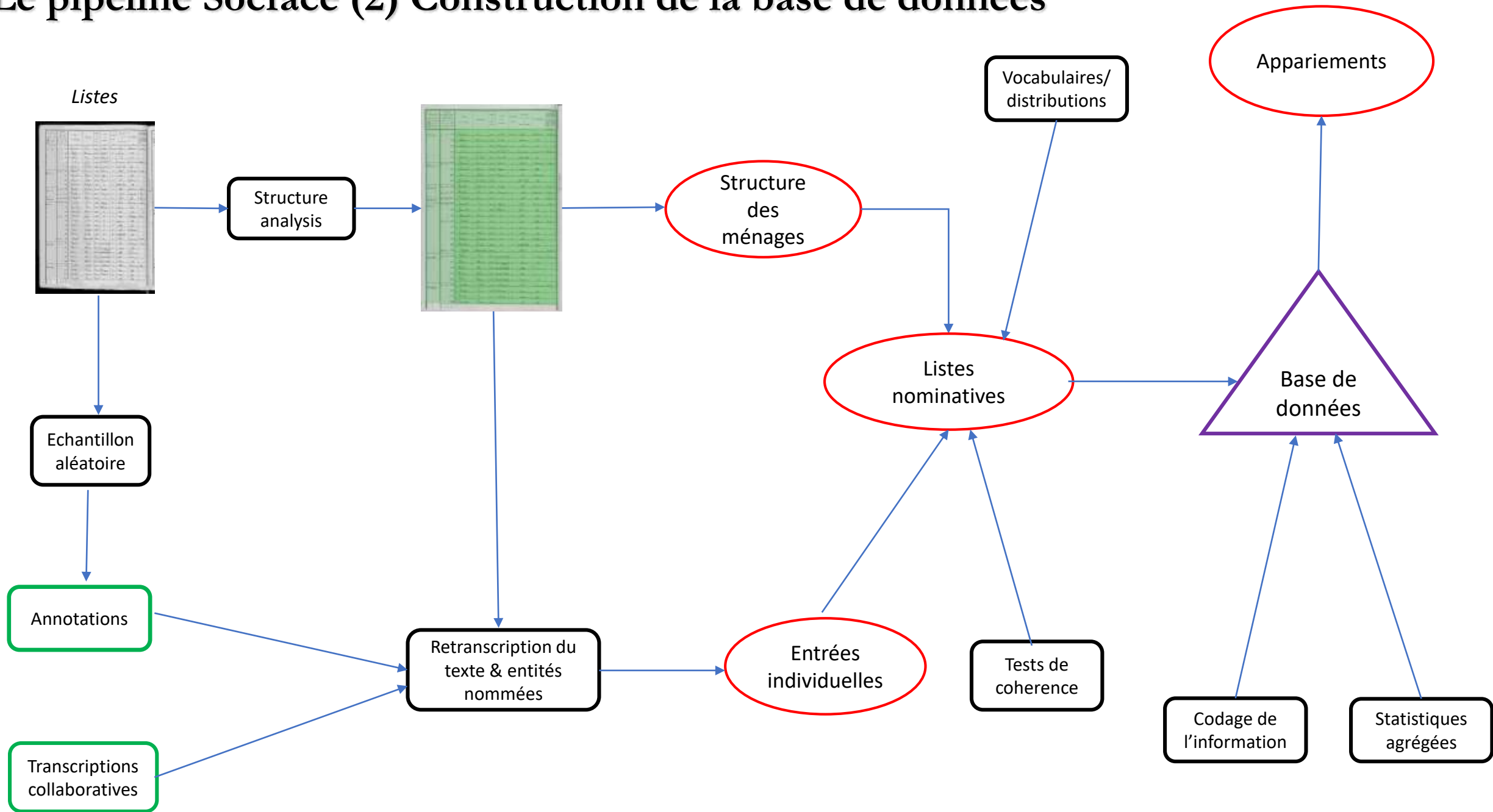
**Extraire l'information des listes**

A vertical thumbnail image of a handwritten document, similar to the one in the top left, showing a list with multiple columns.A vertical thumbnail image of a handwritten document, showing a list with multiple columns and rows of text.

# Le pipeline Socface (1) collecte et analyse des images



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



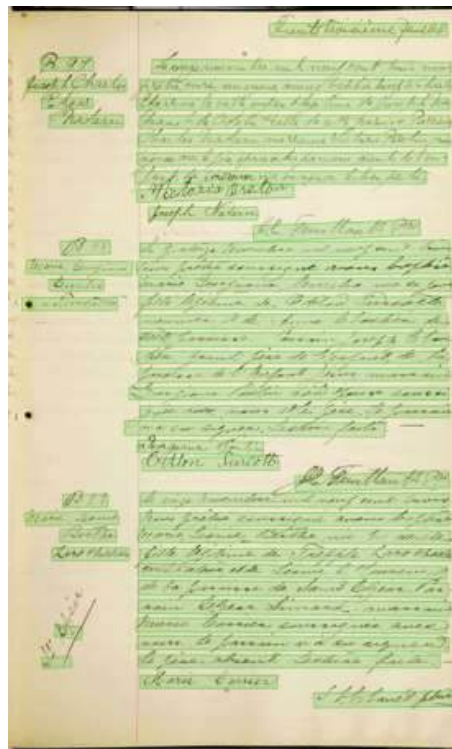
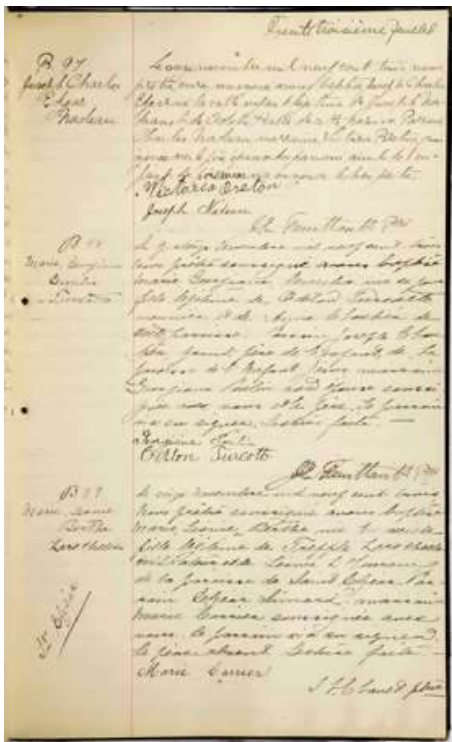
# Chaîne de traitement automatique

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 . Lechoire 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 . Lechoire 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	ÂGE	SEX	ÉTAT	OCCUPATION	LIEU DE NAISSANCE	NOM DE LA FAMILLE	
							PREMIER	DEUXIÈME
4	Juliette	8	F	enfant		l. bille		
5	Albertine	3	F	enfant				
6	Jules	19	M	domestique	chauger			
7	Jeanne	22	F					
8	Charles	32	M	chef				
9	Hermance	27	F	épouse				
10	René	6	M	enfant				
11	Juliette	3	F	enfant				
1	Jean	50	M	chef				
2	Marie	46	F	épouse				
3	Juliette	19	F	sa fille				
4	Georgette	12	F					
5	Maurice	10	M					
6	Andrée	1	F					
7	Philomène	47	F	meunier	chef			

surname Robert, firstname Juliette, occupation néant, link l. bille, 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. bille, 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 handwritten table with multiple columns and rows of text, likely a list of names and associated data.



A small, vertically oriented handwritten table with multiple columns and rows of text, likely a list of names and associated data.



A larger, vertically oriented handwritten table with multiple columns and rows of text, likely a list of names and associated data.



## 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	Baillardin	Henri	30	ouvrier	de passage
		Houssard	Octave	30	ouvrier	de passage
		Baillardin	Henri	30	ouvrier	de passage
		Baillardin	Henri	30	ouvrier	de passage
32	36	Baillardin	Henri	30	ouvrier	de passage
		Baillardin	Henri	30	ouvrier	de passage
33	36	Baillardin	Henri	30	ouvrier	de passage
		Baillardin	Henri	30	ouvrier	de passage
34	36	Baillardin	Henri	30	ouvrier	de passage
		Baillardin	Henri	30	ouvrier	de passage

1896

1	1	Julien	Joseph	30	ouvrier	de passage
2	1	Henri	Henri	30	ouvrier	de passage
3	1	Henri	Henri	30	ouvrier	de passage
4	1	Henri	Henri	30	ouvrier	de passage
5	1	Henri	Henri	30	ouvrier	de passage
6	1	Henri	Henri	30	ouvrier	de passage
7	1	Henri	Henri	30	ouvrier	de passage
8	1	Henri	Henri	30	ouvrier	de passage
9	1	Henri	Henri	30	ouvrier	de passage
10	1	Henri	Henri	30	ouvrier	de passage

1906

1	1	Baillardin	Henri	30	ouvrier	de passage
2	1	Baillardin	Henri	30	ouvrier	de passage
3	1	Baillardin	Henri	30	ouvrier	de passage
4	1	Baillardin	Henri	30	ouvrier	de passage
5	1	Baillardin	Henri	30	ouvrier	de passage
6	1	Baillardin	Henri	30	ouvrier	de passage
7	1	Baillardin	Henri	30	ouvrier	de passage
8	1	Baillardin	Henri	30	ouvrier	de passage
9	1	Baillardin	Henri	30	ouvrier	de passage
10	1	Baillardin	Henri	30	ouvrier	de passage

# 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

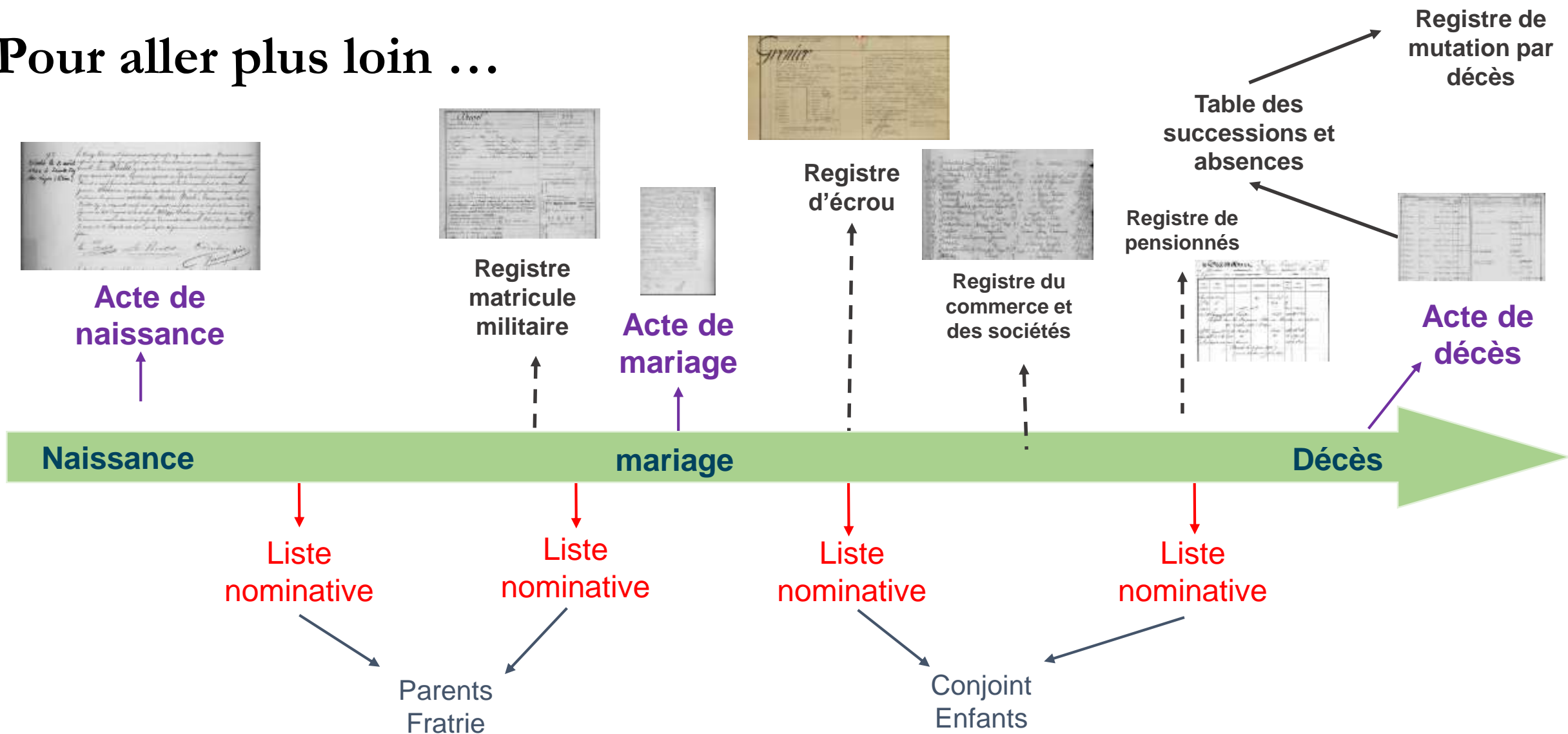
- Raw database to be distributed by the Archives
  - ❖ On a national database (*FranceArchives*), with a search engine.
  - ❖ Sur les bases des Archives Départementales.
  - ❖ Direct links with the image.
- Encoded database to be distributed for research
  - ❖ A database where various information are 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



TEKLIA

FranceArchives  
PORTAIL NATIONAL DES ARCHIVES