

# 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

Christopher Kermorvant  
Lionel Kesztenbaum

anr<sup>©</sup>

Conseil supérieur des archives  
13/09/2022



*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**



## Mobiliser les recensements du passé

➤ États-Unis: 9 recensements complets (*full count*) de 1850 à 1940.

- Explosion des travaux grâce aux appariements (linking) entre recensements.
- Publications nombreuses et de grande qualité.
- Diversité des thématiques: mobilité (géographique ou sociale), santé et mortalité, choix politiques...



➤ Royaume-Uni: 1851 à 1911.

➤ Autres: Irlande, Canada, pays scandinaves...





# SOC FACE



Une liste nominative:  
Aubervilliers, 1921

312 — 4 —

DESIGNATION		NUMÉROS. par quartier, village, bâtiment ou n°.			NOMS	ANNÉE	LIEU	NATIONAL.	SITUATION	PROFESSION	
des quar- tiers, villages ou bâtimens	des n°s des villages ou bâtimens	des numéros	des maisons	des maisons	DE FAMILLE	de MARIAGE	de MARIAGE	NAT. LIEU	en chef de ménage		
		2			Coribout Lucie	1886	Reims	F	épouse	sans	
		3			Coribout Henri	1909	Paris	"	enfant	"	
		4			Coribout Alphonse	1912	"	"	"	"	
		5			Coribout Henri	1916	Aubervilliers	"	"	"	
		6			Coribout André	1920	"	"	"	"	
		7			Pierre Henriette	1881	Reims	"	bell. sœur	Magasinier	Commerciaux Valenciennes
		8			Pierre Liliame	1910	Bagnolle	"	filie	sans	
		9			Pierre Lucien	1918	Warghote	"	veuve	sans	
		1			Parmenier Juliette	1877	Weschâteau	"	Chef	maçon	
		2			Parmenier Lucie	1867	Lambach	"	épouse	sans	
		3			Parmenier Maximilien	1885	Paris	"	enfant	Confectionnier	Chulman
		4			Parmenier Emma	1900	St Denis	"	"	empl.	Ministère Pensions
		5			Parmenier Ludovic	1902	"	"	"	plûr	Châ
Rue des Fillettes		1			Caliez Georges	1878	Lille	"	Chef	Commerciaux	Magist
		2			Caliez Céline	1884	Stade	"	épouse	sans	
		1			Renaud Jean	1869	Longueuil	"	Chef	chef équipe	Magist

## 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 suivies

- ❖ 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.
- ❖ Lien avec la période contemporaine.

## 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.





## L'apport des appariements: étudier la mobilité

1886

31	36	Baillardin	Joseph	30	ouvrier de passage
		Houssard	Octave	30	propriétaire de ferme
		Baillardin	Joseph	5	en fte
		Baillardin	Joseph	3	en fte
36		Barnier	Henri	21	propriétaire de ferme
37		Boussard	Antoine	20	ouvrier de passage
38	36	Welaunoy	Armand	10	ouvrier de passage
39		Boussard	Eugène	32	propriétaire de ferme
40	37	Chapin	Julien	10	ouvrier de passage
41		Féat	Henri	20	ouvrier de passage
42		Guille	Joseph	21	ouvrier de passage
43	38	Grandin			
44		Boussard			
45	39	Boussard			
46		Boussard			

1896

1	1	Juste	Joseph	29	ouvrier	chef
2	2	Boussard	Henri	21	ouvrier	
3	3	Boussard	Joseph	22	ouvrier	
4	4	Boussard	Henri	21	ouvrier	chef
5	5	Boussard	Henri	21	ouvrier	jeune
6	6	Boussard	Henri	21	ouvrier	jeune
7	7	Boussard	Henri	21	ouvrier	jeune
8	8	Boussard	Henri	21	ouvrier	jeune
9	9	Boussard	Henri	21	ouvrier	jeune
10	10	Boussard	Henri	21	ouvrier	jeune

1906

1	1	Boussard	Henri	111	ouvrier	chef	ouvrier	jeune
2	2	Boussard	Henri	112	ouvrier	chef	ouvrier	jeune
3	3	Boussard	Henri	113	ouvrier	chef	ouvrier	jeune
4	4	Boussard	Henri	114	ouvrier	chef	ouvrier	jeune
5	5	Boussard	Henri	115	ouvrier	chef	ouvrier	jeune
6	6	Boussard	Henri	116	ouvrier	chef	ouvrier	jeune
7	7	Boussard	Henri	117	ouvrier	chef	ouvrier	jeune
8	8	Boussard	Henri	118	ouvrier	chef	ouvrier	jeune
9	9	Boussard	Henri	119	ouvrier	chef	ouvrier	jeune
10	10	Boussard	Henri	120	ouvrier	chef	ouvrier	jeune





## Technologie: reconnaissance d'écriture

*OCR : écriture imprimée*

*HTR : écriture cursive*

ÉTAT NOMINATIF

Segmenter

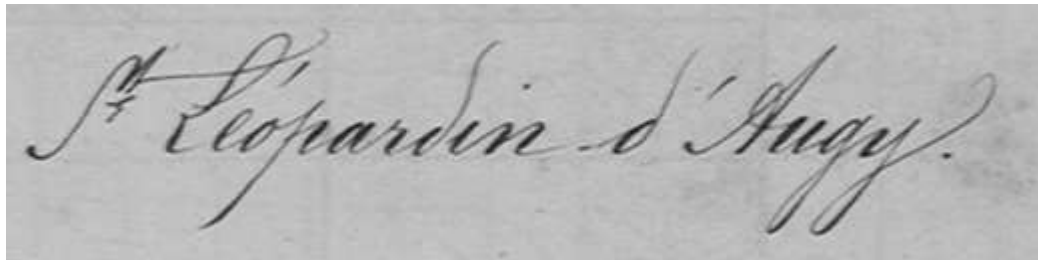


É T A N O M I N F  
T A I  
T



Identifier

ETAT NOMINATIF

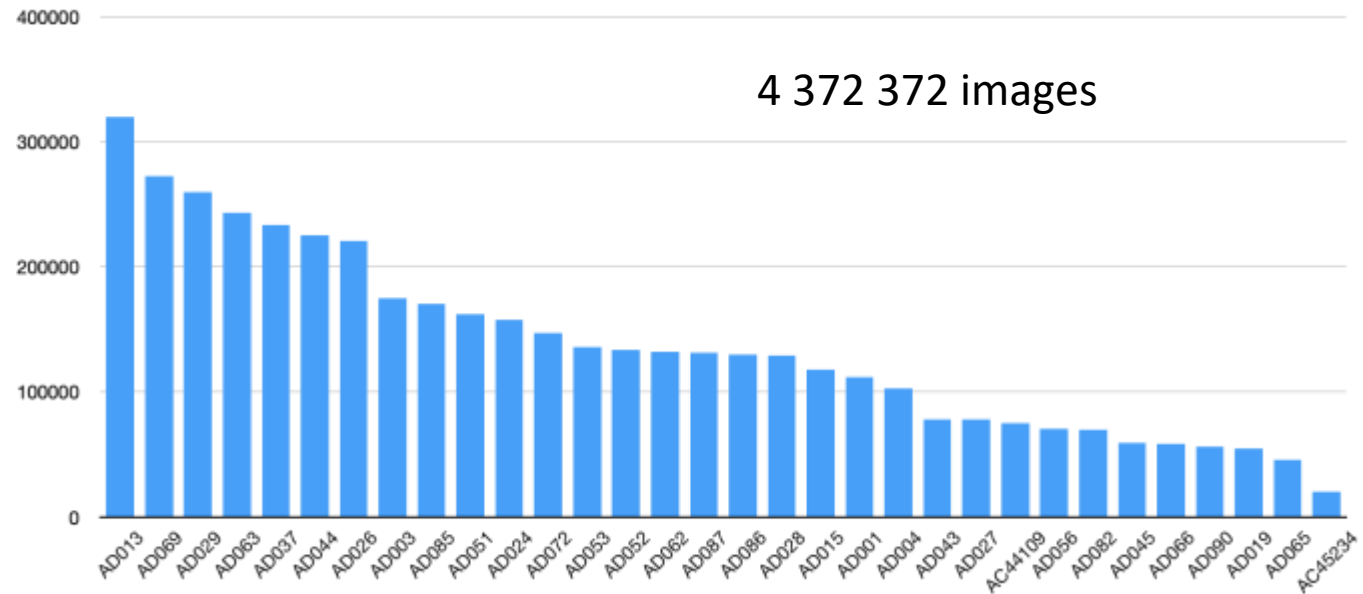
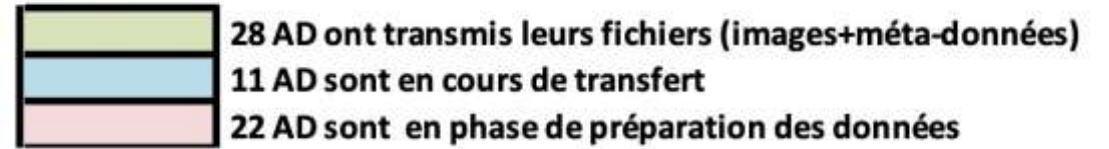
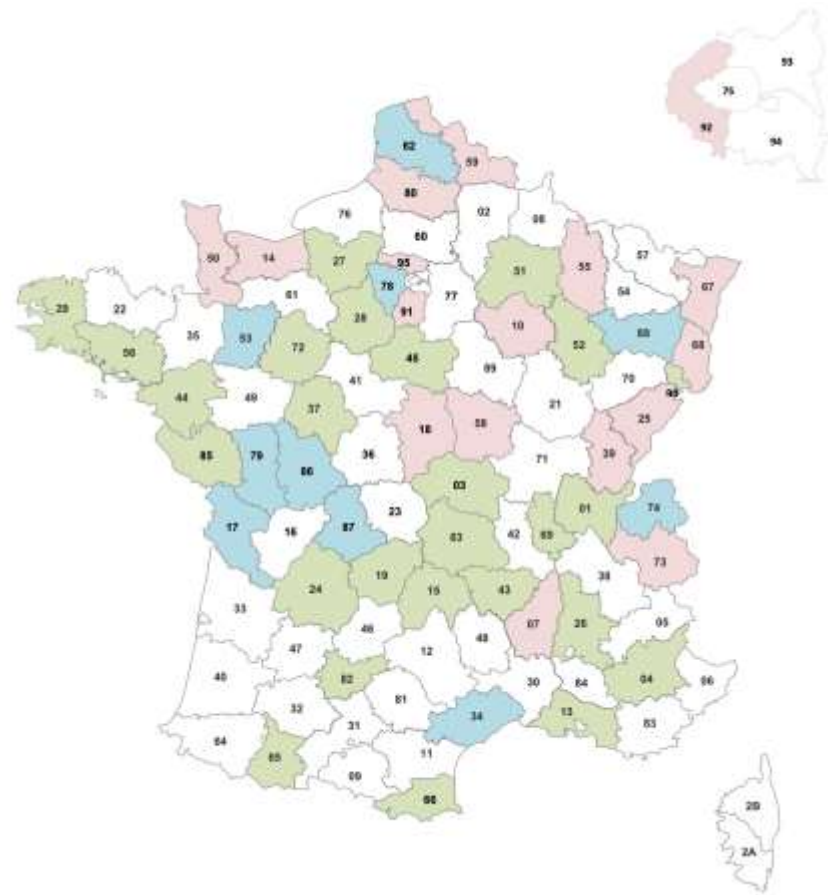


Pour identifier les lettres, il faut les *segmenter*.  
Pour segmenter les lettres, il faut les *identifier*.

**Paradoxe de Sayre**



## Collecte des images



Projection France: 15M d'images, 30To

## Créer des données d'entraînement

NOM DES VILLAGES U OU ZEE		NOMS DU FAMILLE	ANNÉE	LIEU	NATIONAL	SITUATION	PROFESSION
du	au	PRÉNOMS	en	en	LETT.	PAR RAPPORT AU CHEF DE MÉNAGE	en QUEL DE MÉNAGE
1	2	3	4	5	6	7	8
10		Arcueil Jean Louis	1891	Paris	FR		
11		Corcos Charles	1870	Montauban	FR		
12		Corot Lucien	1915	Paris	FR		
13		Corot Jeanne	1891	Paris	FR		
14		De Villis Domatien	1881	Paris	FR		
15		De Villis Domatien	1911	Paris	FR		
16		Salaman Victor	1896	Paris	FR		
17		Salaman Louis	1896	Paris	FR		
18		Salaman Louis	1897	Paris	FR		
19		Arzel Victor	1890	Paris	FR		
20		Arzel Louis	1911	Paris	FR		
21		Arzel Yves	1911	Paris	FR		
22		Arzel François	1887	Paris	FR		
23		Arzel Charles	1887	Paris	FR		
24		Arzel Charles	1897	Paris	FR		
25		Arzel Yves	1897	Paris	FR		
26		Arzel Victor	1879	Paris	FR		
27		Arzel Charles	1913	Paris	FR		
28		Arzel Marie Anne	1886	Paris	FR		
29		Arzel Jean	1871	Paris	FR		

M. Arzel

Victoire

Repassuse

Ch. de m.

1890

idem

idem

Nom  
Veuve Arzel

Prénoms  
Victoire

Profession  
Repassuse

Situation par rapport au chef de ménage  
Chef de ménage

Employeur  
idem

Age  
idem

Date de naissance  
1890

Etat civil (choix parmi : Garçon / Homme / Veuf / Fille / Femme / Veuve)  
Veuve

Nationalité  
idem

Lieu de naissance  
idem

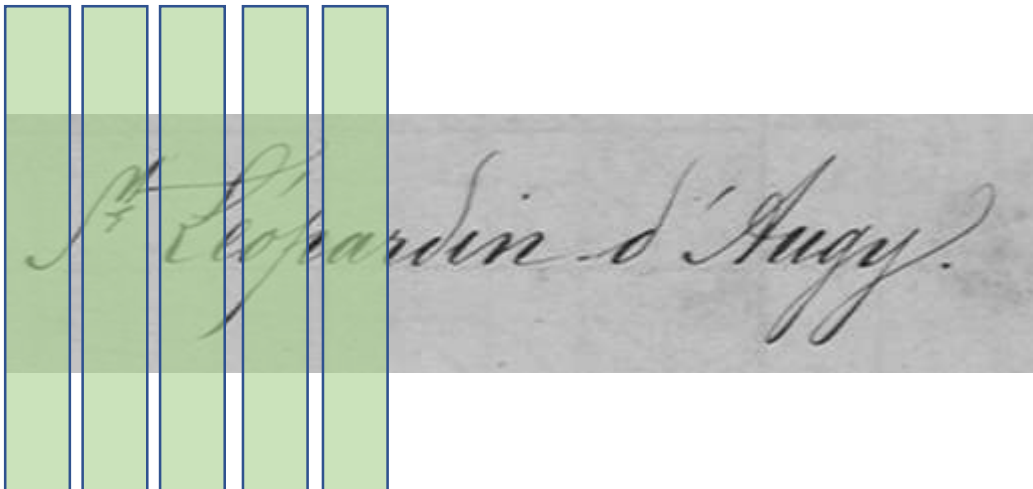
Observations  
idem

Skip task

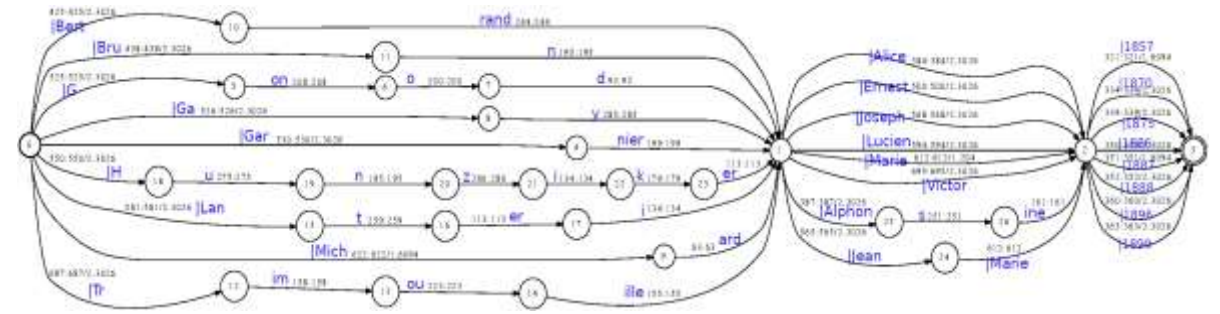
Annuler



## Entraîner des modèles: reconnaissance de texte avec contrainte syntaxique



Modèle optique : fait des hypothèses de lettres



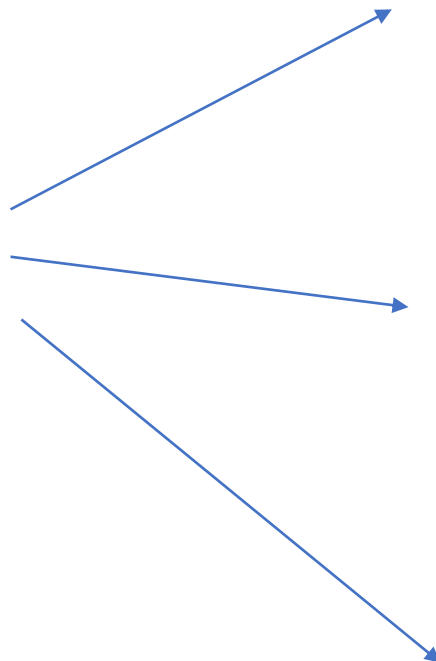
Modèle linguistique: grammaire de la langue pour trouver l'hypothèse la plus probable



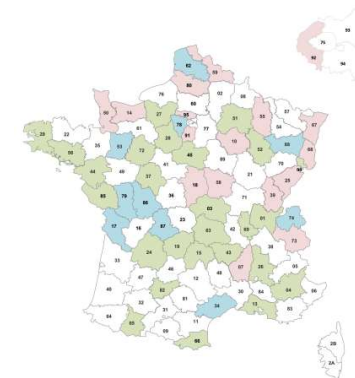
Traiter des millions d'images

NUMERO	NOM	DATE	LIEU	PROF	REMARQUES
41	Rosette	Valentin 1862	Bellefleur	Femme	fil
42	Rosette	Vigier 1878	St-Pierre	fil	site
43	Rosette	Blanc 1891	Bellefleur	petit fil	liger
44	Rosette	Laurent 1881			liger
45	Rosette	Joseph 1888			sieste
46	Rosette	Marie 1901			petit fil
47	Rosette	Julie 1897	Bellefleur		Cultivat
48	Rosette	Rosale 1878	burvau	époux	
49	Rosette	Joseph 1877	Bellefleur		fil
50	Rosette	Gabriel 1879			fil
51	Rosette	Julienne 1890			fil
52	Rosette	Julien 1901	Sepvion	petit fil	
53	Rosette	Julie 1877	Bellefleur		Cultivat
54	Rosette	Joseph 1877	Bellefleur		fil
55	Rosette	Joseph 1877	Bellefleur		fil
56	Rosette	Marie 1878	Segne	petit fil	sieste

Diffusion



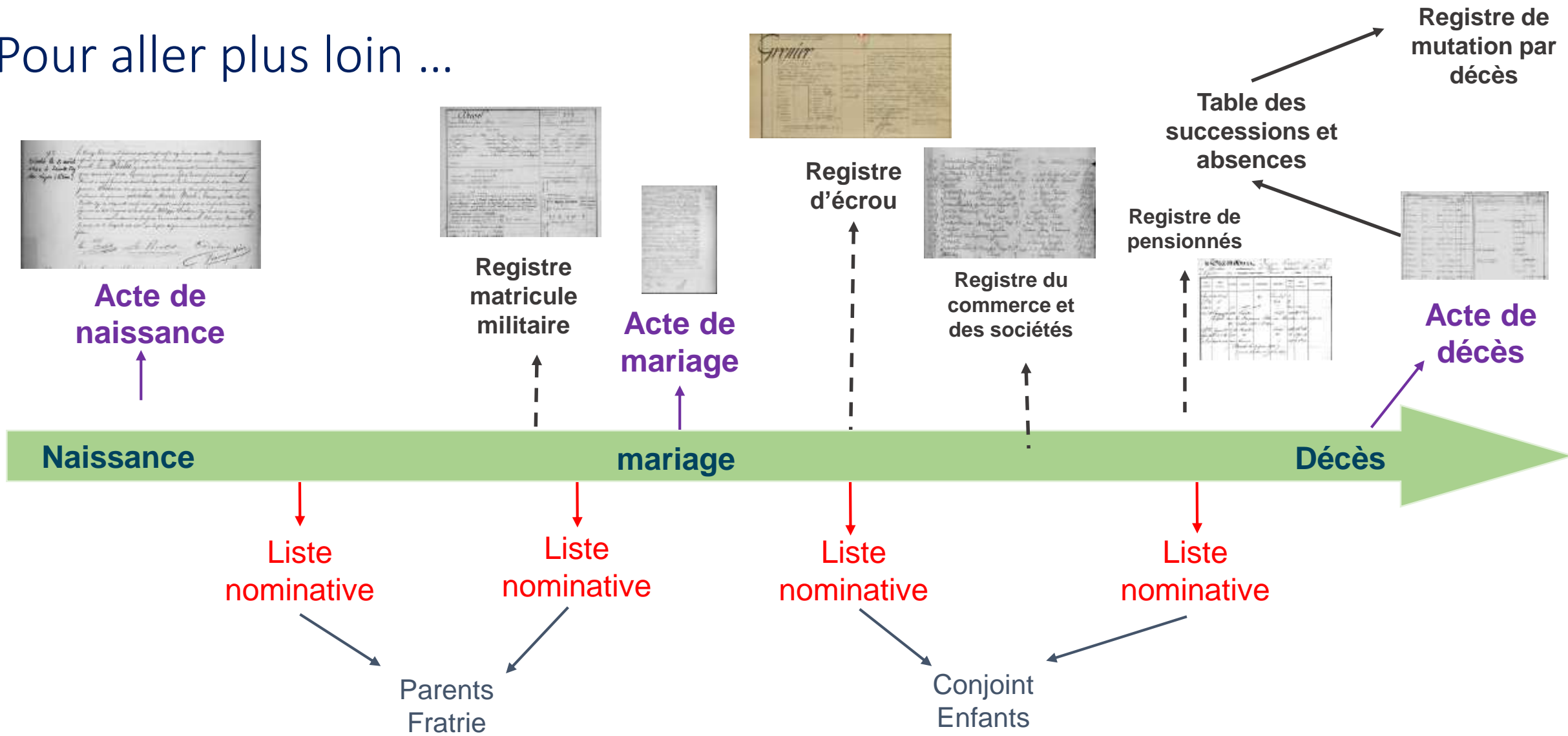
France Archives



Services d'archives départementales/municipales

Base de données de recherche  
Codage et appariements

## Pour aller plus loin ...



SOC FACE



Merci de votre attention



INSTITUT NATIONAL  
D'ÉTUDES DÉMOGRAPHIQUES



PARIS SCHOOL OF ECONOMICS  
ÉCOLE D'ÉCONOMIE DE PARIS

