



Un projet de recherche collaboratif entre

Historiens, économistes, démographes

Archivistes

Experts en intelligence artificielle



Et les services d'archives départementales et municipales

Christopher Kermorvant
Lionel Kesztenbaum

Financement sur **3 ans ½**



SocFace:

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).

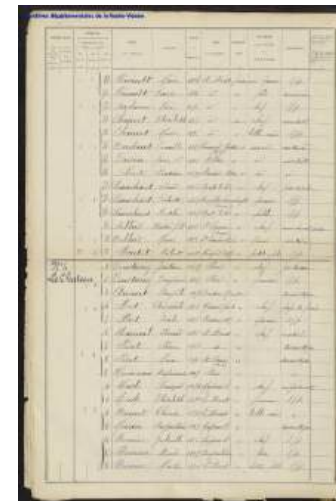
SocFace produira une base de données complète des individus ayant vécu en France entre 1836 et 1936 et l'utilisera pour analyser le changement social dans la longue durée.



A page from a 19th-century French census document, showing a grid of handwritten entries for individuals. The page is filled with dense, cursive handwriting, organized into columns and rows, typical of a population register.



A page from a 19th-century French census document, showing a grid of handwritten entries for individuals. The page is filled with dense, cursive handwriting, organized into columns and rows, typical of a population register.



A page from a 19th-century French census document, showing a grid of handwritten entries for individuals. The page is filled with dense, cursive handwriting, organized into columns and rows, typical of a population register.

De multiples objectifs et enjeux



Science

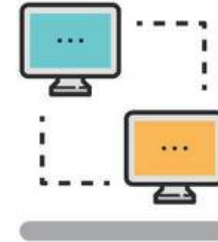
Etudes en économie, en histoire et en démographie : évolution du marché du travail, des inégalités, de la structure sociale, des mobilités.



Technologie

Reconnaissance automatique d'écriture manuscrite, analyse de tableau, traitement de plusieurs million d'images, accès à de multiples sources.

(TLR initial = 3)

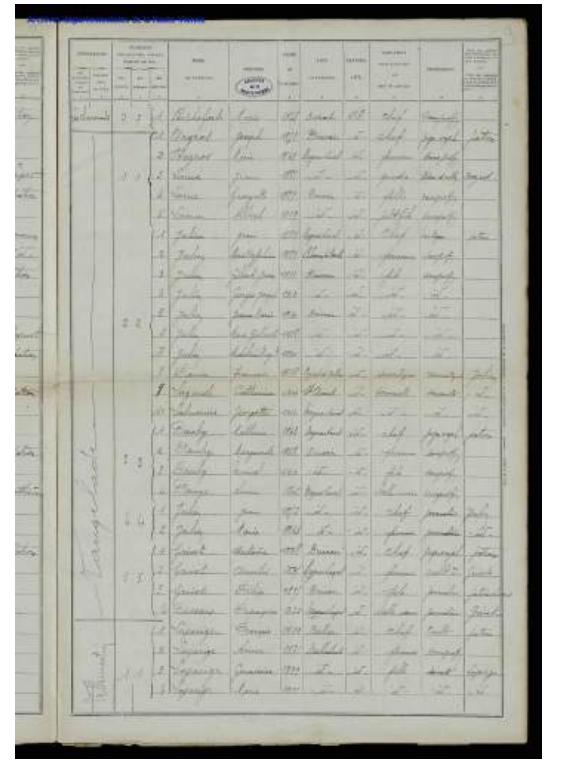


Valorisation

Mise à disposition des données extraites en accès libre, versement aux propriétaires des fonds, valorisation par les portails.

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.



The image shows a page from an old document, possibly a census or administrative list. It features a grid with multiple columns and rows. The text is written in a cursive script, likely from the 18th or 19th century. The grid contains names and numbers, and there are some markings and annotations throughout the page. The document is aged and has a slightly yellowed appearance.

DESIGNATION		NUMÉROS, PAR QUARTIER, VILLAGES, hameaux ou rcs.			NOMS	PRÉNOMS	ANNÉE	LIEU	NATIONALITÉ	SITUATION	PROFESSION	REMARQUES
des quartiers, villages ou hameaux.	des rcs ou des villages.	des numéros.	des messes.	des jardins.	DE FAMILLE.		de naissance.		LITÉ.	PAR RAPPORT au chef de ménage.		12.
				2	Eribout	Lucie	1886	Reims	F	épouse	ans	
				3	Eribout	Senni	1909	Claris	"	enfant	"	
				4	Eribout	Alphonse	1912	"	"	"	"	
				5	Eribout	Reni	1916	Aubervilliers	"	"	"	
				6	Eribout	Audé	1920	"	"	"	"	
				7	Pierre	Henriette	1881	Reims	"	belle sœur	Magasinier	Boulonnais Valenciennes
				8	Pierre	Liliane	1912	Bagnollet	"	Nièce	sans	
				9	Pierre	Lucien	1918	Wizy-lez-Reims	"	neveu	sans	
				10	Parmentier	Justave	1877	Wafchateau	"	Chef	maçon	
				11	Parmentier	Lucie	1867	Lambach	"	épouse	sans	
				12	Parmentier	Maximilien	1885	Paris	"	enfant	Confectionnier	Chulman
				13	Parmentier	Emma	1900	W. Reims	"	"	empl.	Minist. Reims
				14	Parmentier	Ludovic	1902	"	"	"	plu	Calix
				15	Caliez	Georges	1878	Lill	"	Chef	combustible	Magist
				16	Caliez	Céline	1884	Tachy	"	épouse	sans	
				17	Renault	Jean	1869	Combray	"	Chef	chef équipe	Magist
				18	Renault	Oliver	1869	Combray	"	"	"	

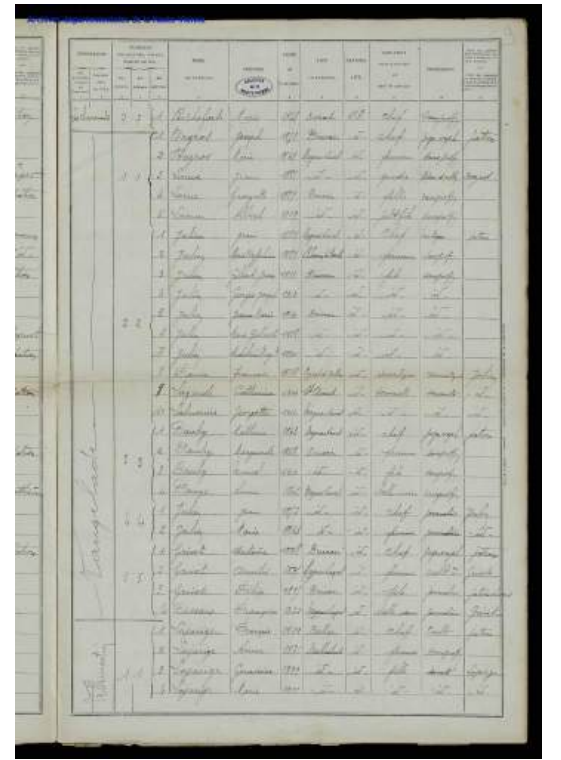
Une liste nominative:

Aubervilliers, 1921

Rue
des
Fillettes

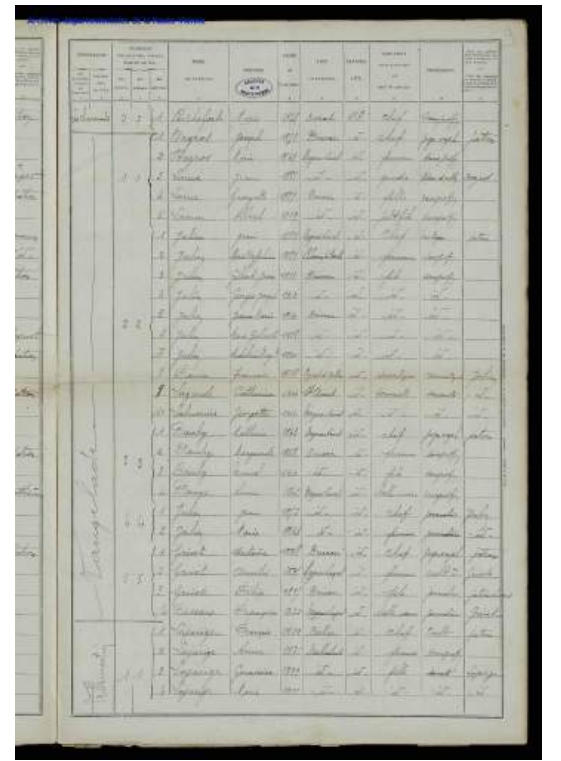
Pourquoi les listes nominatives?

- 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.



Les listes nominatives: défis

- Documents décentralisés et nombreux scripteurs.
- Volume inconnu mais important (entre 5 et 15 M d'images).
- Conditions de conservation/numérisation variables.
- Lecture en 2 dimensions.
- Evaluation de l'erreur.
- Structurer, consolider, croiser l'information.
 - Utiliser la structure des listes et les informations agrégées pour contrôler la transcription.
 - Utiliser la stabilité temporelle pour améliorer les données.



The image shows a page from a handwritten ledger or list. The page is filled with rows of text, organized into columns. The handwriting is in a cursive script, typical of the 19th or early 20th century. The columns appear to contain names, dates, and other numerical or categorical data. The paper is aged and slightly yellowed. The text is written in dark ink, and the overall layout is that of a formal record-keeping document.

Lecture en 2 dimensions

312 — 4 —

DÉSIGNATION		NUMÉROS. PAR QUARTIER, VILLAGES, HAMEAUX ou FER.			NOMS		ANNÉE	LIEU	NATIONAL.	SITUATION	PROFESSION.	
des carrés, cours, villages ou hameaux	des carrés, des villages ou hameaux	des maisons	des maisons	des maisons	DE FAMILLE.	PRÉNOMS.	de NAISSANCE.	de NAISSANCE.	LITTÉ.	de MARIAGE.		
1	2	3	4	5	6	7	8	9	10	11	12	13
Rue des Éclottes	115	11	2	Erribout	Lucie	1886	Reims	F	épouse	ans		
			3	Erribout	Léoni	1909	Paris	"	enfant	"		
			4	Erribout	Alphons	1912	"	"	"	"		
			5	Erribout	Reni	1916	subisilluis	"	"	"		
			6	Erribout	Audie	1920	"	"	"	"		
		4	Pierre	Léonide	1881	Reims	"	belle sœur	Magasin	Parabonnie		
		8	Pierre	Liliane	1918	Baguette	"	filie	ans			Alencien
		3	Pierre	Lucien	1918	Winghote	"	revenu	ans			
		1	Parmenier	Justave	1877	Wesbatau	"	chef	maçon			
		2	Parmenier	Lucie	1867	Lambach	"	épouse	ans			
		3	Parmenier	Maximilien	1885	Paris	"	enfant	Confectionn.	Chulman		
		4	Parmenier	Emma	1900	W. Denis	"	"	empl.	Minist. P. et M.		
		5	Parmenier	Ludovic	1902	"	"	"	"	chef	Calix	
		1	Caliez	Jorges	1878	Lill	"	chef	Confectionn.	Mag.		
		2	Caliez	Clélie	1884	Éclottes	"	épouse	ans			
1	Renault	Jean	1869	Combrigny	"	chef	chef équipe	Mag.				

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

Tests de cohérence et de cohésion

C. — RÉCAPITULATION GÉNÉRALE de la population de la commune.

Population agglomérée au chef-lieu (1 ^{re} section du cadre A ci-contre).....	40.067
Population éparsé (2 ^e section du cadre A).....	765
Population municipale. (TOTAL du cadre A).....	40.832
Population comptée à part, conformément à l'article 2 du décret du 5 octobre 1920. (Total du cadre B ci-dessus).....	765
TOTAL GÉNÉRAL de la population de la commune.	40.832
présents le jour du recensement	40.211
absents le jour du recensement	621
Dont	
1920 ou 1921	1006
nes	12.398
au cours	14.462
des	9.460
années	2.879
1861 à 1880 (40 à 59 ans).....	621
1860 et antérieurement (60 ans et plus).....	10.899
Âge inconnu	

— 8 —

RÉCAPITULATION.

A. — Tableau récapitulatif de la population inscrite sur la liste administrative et répartition de cette population en population agglomérée et population éparsé.

CATEGORIES DE POPULATION	NOMBRE	
	hommes	total
Population agglomérée au chef-lieu	39.302	40.067
Population éparsé	530	765
Total	39.832	40.832

B. — Tableau récapitulatif de la population comptée à part en exécution de l'article 2 du décret du 5 octobre 1920.

CATEGORIES DE POPULATION	NOMBRE	
	hommes	total
Refugiés	1006	1006
Population agglomérée au chef-lieu	12.398	13.404
Population éparsé	9.460	10.406
Total	21.864	23.810

C. — Récapitulatif général de la population de la commune.

Population agglomérée au chef-lieu (1 ^{re} section du cadre A ci-contre).....	40.067
Population éparsé (2 ^e section du cadre A).....	765
TOTAL GÉNÉRAL de la population de la commune.	40.832
présents le jour du recensement	40.211
absents le jour du recensement	621
Dont	
1920 ou 1921	1006
nes	12.398
au cours	14.462
des	9.460
années	2.879
1861 à 1880 (40 à 59 ans).....	621
1860 et antérieurement (60 ans et plus).....	10.899
Âge inconnu	

Valorisation: retour aux archives (et au-delà...)

- Mise à disposition des données brutes aux archives
 - ❖ Sur une base nationale (sur FranceArchives).
 - ❖ Sur les bases des Archives Départementales.
 - ❖ Idéalement avec le lien vers les photos.
- Mise à disposition des données de la recherche
- Développement de nouvelles technologies pour la reconnaissance en masse
 - ❖ TLR final = 6/7.
- Ouverture sur d'autres sources: un modèle pour la mise à disposition des archives?

Merci pour votre attention

<https://socface.site.ined.fr/>

Christopher Kermorvant
kermorvant@tekliia.com

Lionel kesztenbaum
lionel.kesztenbaum@ined.fr