



Collaborative research between:

Historians, economists, demographers

Archivists

Experts in machine learning



With the contribution of local French Archives, départementales and municipales

Lionel Kesztenbaum



Archivdag VLA
Dudelange, Luxembourg
06/10/2022

SocFace:

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

Collecting, processing, transcribing, and organizing all French individual census lists from 1836 to 1936 (20 censuses, metropolitan France only).

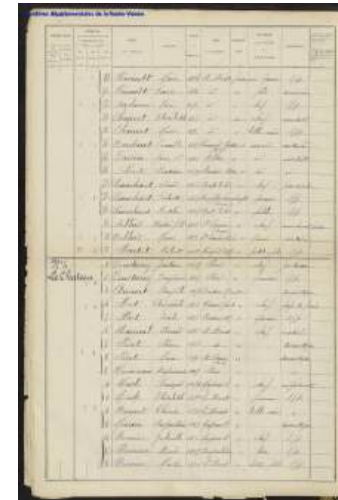
SocFace aims at producing a complete microdatabase of all individuals who lived in France between 1836 and 1936 and at using it to study social change in the long run.



Number	Name	Age	Sex	Profession	Marital Status
1
2
3



Number	Name	Age	Sex	Profession	Marital Status
1
2
3



Number	Name	Age	Sex	Profession	Marital Status
1
2
3

Why Socface?

- Microdata are increasingly important for research in quantitative social sciences: economics, history, sociology, demography...
- Explosion in quality of automated writing recognition (especially manual writing) and treatment of images.

**Mass production of microdata
at the national level**

**Develop new methods to
extract individual-level data
from a very large set of archival
document images**

**Disseminate individual
information produced to the
general public, genealogists,
and researchers**

Roadmap

- Preamble: listes nominatives du recensement.
- Processing images and extracting information.
- A social and economic microhistory at the national level.
- Dissemination and valuation of the data and the archives.

Liste nominative:
 Saint-Vincent-sur-Jabron,
 Alpes-de-Haute-Provence
 1936

RESIGNATION	NUMEROS				NOMS	ANNÉE	LIEU	NATIONALITÉ	SITUATION	PROFESSION	Fait par le chef d'entreprise, ou le directeur, ou le propriétaire - possesseur.	
	DES ANS	DES ANS	DES ANS	DES ANS								
1	2	3	4	5	DE FAMILLE	DE	DE	LITE	ou			
1	2	3	4	5		SAISON	NAISSANCE		chef de ménage			
Mouton	1-2	115			Andrie	Louis	1888	La Mee	français	Chf.	Cultivateur	Pétrole
		116			Andrie	Thérèse	1891	Bastans	id.	épouse	gfp.	
		117			Andrie	Louis	1922	B. Vincent	id.	fil.	gfp.	
		118			Andrie	Louis	1885	St-Martin	id.	veuf	tailleur	Pétrole
		119			Jordan	Roulette	1898	Hospitall	id.	veuf	gfp.	
St-Antoine	21-23	119			Dumont	Edmond	1872	St-Martin	id.	Chf.	Cultivateur	Pétrole
		120			Dumont	Maria	1874	Vojeux	id.	épouse	gfp.	
		121			Dumont	Josette	1899	St-Martin	id.	fil.	Cultivateur	
		122			Girardot	Gabriel	1830	id.	id.	petit-fil.	gfp.	
St-Martin	43-44	123			Michel	Gulielm	1884	Arthon	id.	Chf.	Cultivateur	Pétrole
	44-45	124			Michel	Elie	1891	St-Martin	id.	Chf.	Cultivateur	Pétrole
St-Pierre	45-46	125			M. Soubat	Simon	1886	id.	id.	Chf.	gfp.	Pétrole
Arson	47-48	126			Eymette	Emile	1907	Vojeux	id.	Chf.	Cultivateur	Pétrole
		127			Eymette	Emile	1909	St-Martin	id.	épouse	gfp.	
		128			Eymette	Rosine	1924	id.	id.	fil.	gfp.	
	49-50	129			Soubat	Jean	1877	id.	id.	Chf.	Cultivateur	Pétrole
		130			Soubat	Simon	1878	id.	id.	épouse	gfp.	
		131			Soubat	Thérèse	1925	id.	id.	fil.	gfp.	
		132			Soubat	Gabriel	1913	id.	id.	fil.	Cultivateur	
		133			Nicolas	Justine	1922	id.	id.	petit-fil.	gfp.	
49-50	134			Artyès	Marianne	1890	id.	id.	Chf.	Cultivateur	Pétrole	
Vauville	50-50	135			Maurin	Louis	1865	Arthon	id.	Chf.	Cultivateur	Pétrole
		136			Maurin	Maria	1878	Arthon	id.	épouse	gfp.	
		137			Maurin	Elie	1901	Arthon	id.	fil.	Cultivateur	
		138			Maurin	Elie	1905	Arthon	id.	fil.	Cultivateur	
		139			Maurin	Maria	1861	Arthon	id.	veuf	gfp.	
Thompigny	50-51	140			Bouchet	Arthur	1889	St-Martin	id.	Chf.	Cultivateur	Pétrole
		141			Bouchet	Rosine	1909	Arthon	id.	épouse	gfp.	
		142			Bouchet	Emile	1876	St-Martin	id.	veuf	gfp.	
		143			Bouchet	Louis	1931	id.	id.	fil.	gfp.	

DÉSIGNATION		NUMÉROS PAR QUARTIER, VILLAGE, hameau ou rue,			NOMS DE FAMILLE.	PRÉNOMS.	ANNÉE de NAIS- SANCE.	LIEU de NAISSANCE.	NATIONA- LITÉ.	SITUATION PAR RAPPORT au chef de ménage.	PROFESSION.	13
des QUAR- TIERS, villages ou hameaux.	DES RUES dans les villes.	des mai- sons.	des mé- nages.	des indi- vidus.	6	7	8	9	10	11	12	13
Moulin		41-42-		114	André	Luis	1888	La Mées	français	Chef	Cultivateur	Patron
				115	André	Félicienne	1893	Saint-Vincent	id	Épouse	s/p.	
				116	André	Luis	1922	Saint-Vincent	id	fil.	s/p.	
				117	André	Luis	1865	Ch ^{neuf} M ^{neuf}	id	beau-père	Cardonnier	Patron
				118	Genevieve	Paulette	1934	Hospitalet	id	petite fille	s/p.	
St Antonin		42-43-		119	Durand	Ferdinand	1872	Saint-Vincent	id	Chef.	Cultivateur	Patron
				120	Durand	Murie	1874	Nages	id	épouse	s/p.	
				121	Durand	Gustave	1915	Saint-Vincent	id	fil.	Cultivateur	
				122	Genevieve	Gabrielle	1930	id	id	petite fille	s/p.	
Pis Neuf.		43-44-		123	Marcel	Gabriel	1884	Arbonne	id	Chef	Cartonnier	Patron
				124	Michel	Elic	1891	Saint-Vincent	id	Chef	Cultivateur	Patron
Pis Patin		45-46-		125	V ^{ie} Embert.	Léonie	1884	id	id	Chef.	s/p.	Patron
				126	Genevieve	Luis	1907	Nages	id	Chef	Cartonnier	Patron

Pour les patrons,
chefs d'entreprise,
ouvriers à domicile,
inscrire : patron.

Pour les employés
et ouvriers, indi-
quer le nom du
patron ou de l'en-
treprise qui les
emploie.

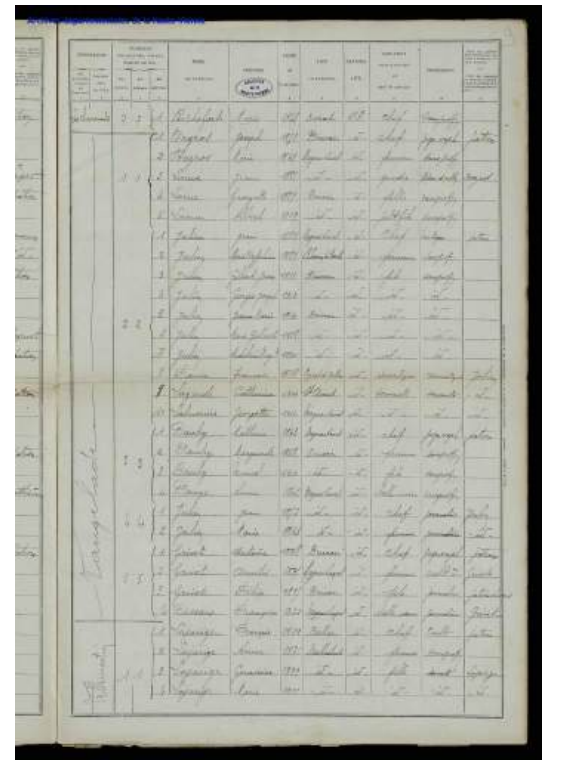
Liste nominative:

Saint-Vincent-sur-Jabron,
Alpes-de-Haute-Provence

1936

Why using *listes nominatives*?

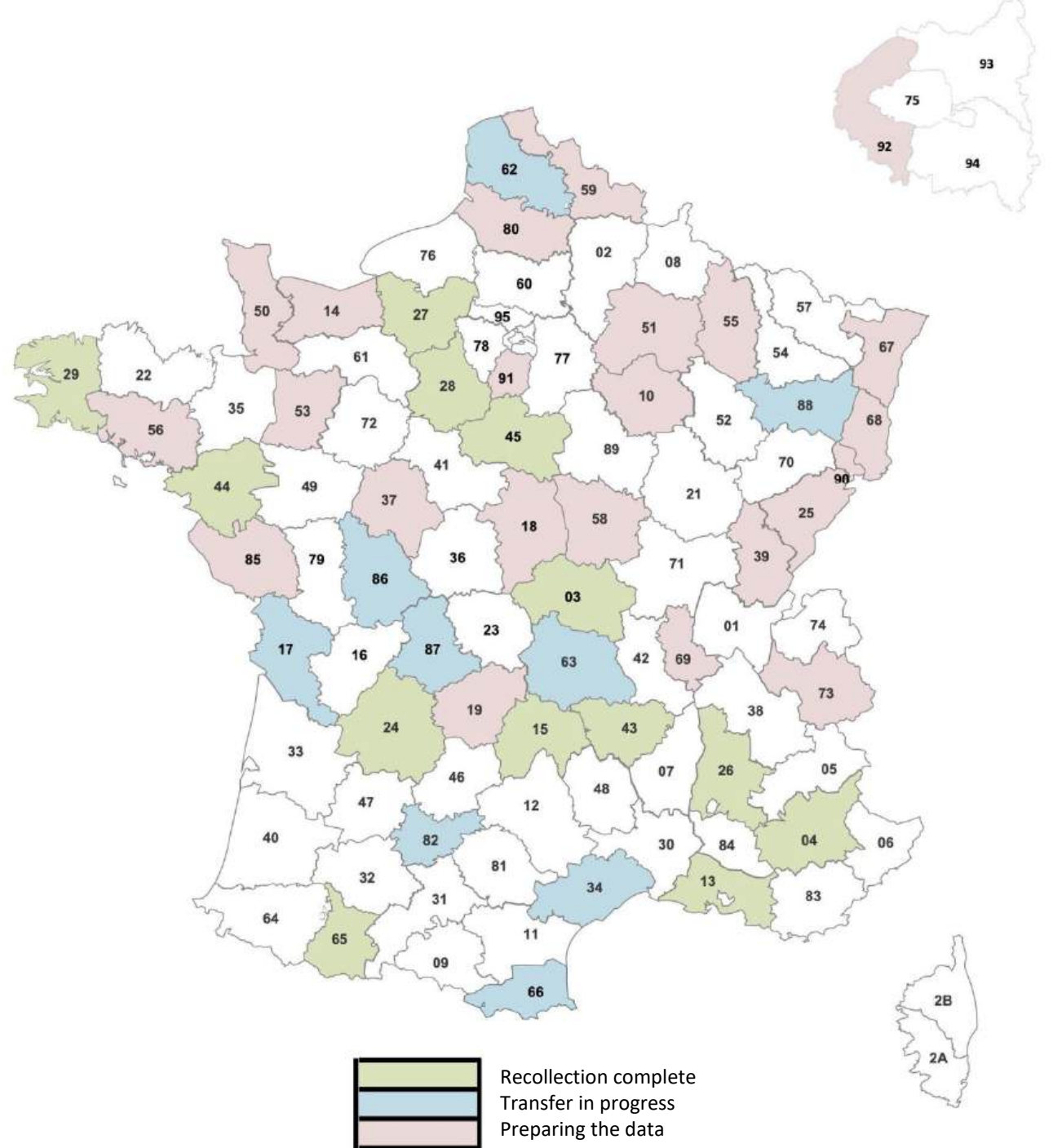
- A standard, numerous and quite simple source.
- A source (relatively) stable over time.
- Already digitalized by many archival depositories.
- A national, uniform source.
- Allows to build a database of France as a whole (almost...).
 - An ideal source for scaling up HTR.
 - A source that matters only at a (very) large scale.



Recollection of the data in the *Archives départementales*

≈ 2,5M images, 5To

(13 ADs completed, 8 in
progress, out of 95)



Challenges and objectives



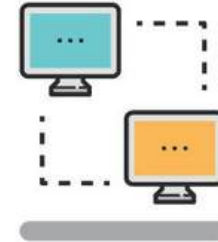
Social Science

Studies in economics, history, and demography: cause and consequences of long-term evolution of labor market, inequality, social structure and mobility.



Technology

Handwritten text transcription, table layout analysis, extracting individual-level data from millions of images.



Diffusion

Disseminating in Open Access the data extracted, through the Archives, to the general public and the researchers

Challenges and objectives



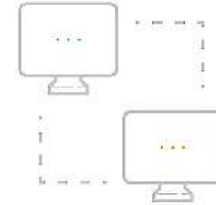
Social Science

Studies in economics, history, and demography: cause and consequences of long-term evolution of labor market, inequality, social structure and mobility.



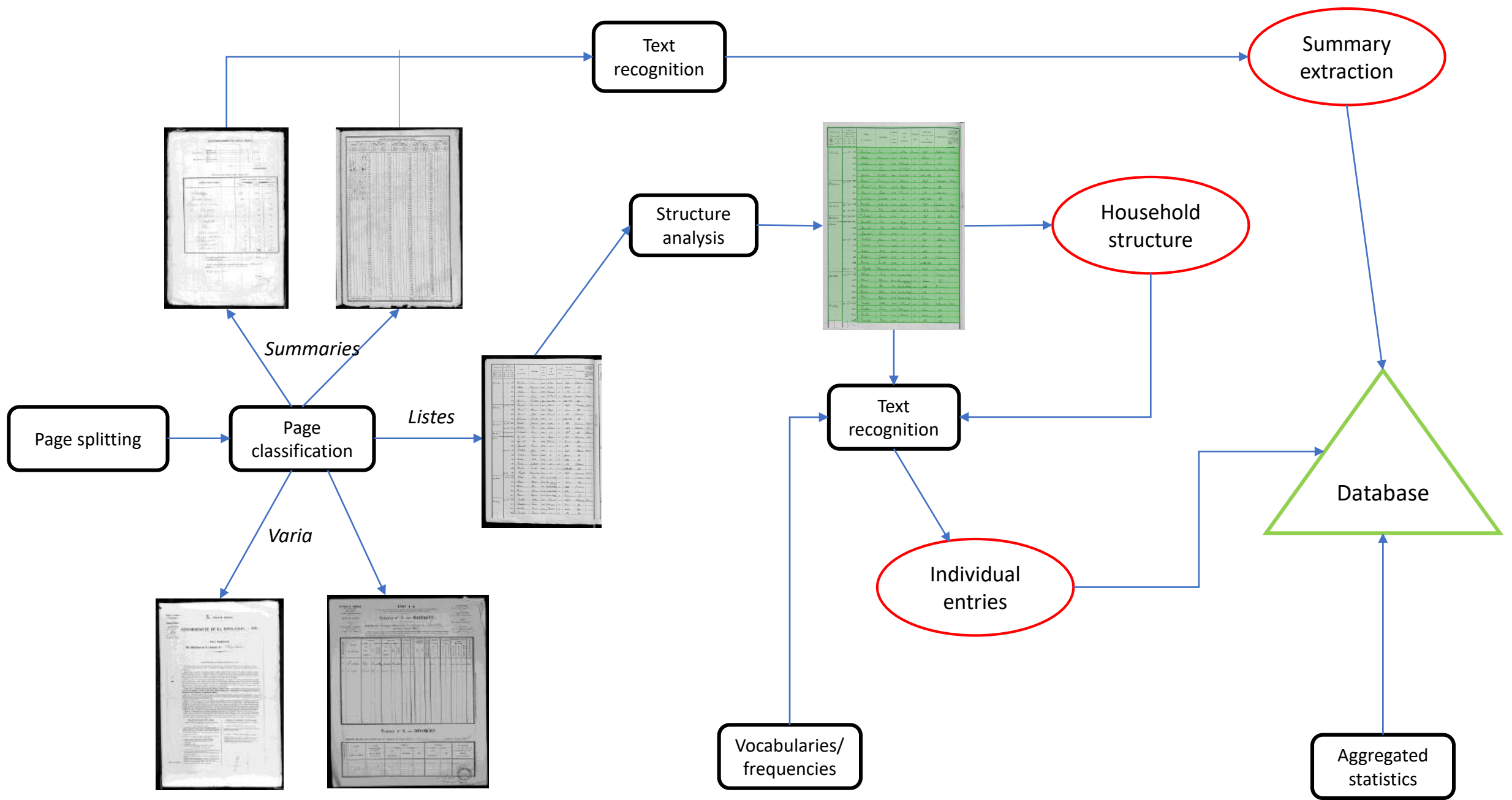
Technology

Handwritten text transcription, table layout analysis, extracting individual-level data from millions of images.



Diffusion

Disseminating in Open Access the data extracted, through the Archives, to the general public and the researchers



Structure Analysis

DESIGNATION	BUREAU					NOMS	PRENOMS	ANNEE de naissance	LEGE de naissance	NATIONALITE	SITUATION PAR RAPPORT au chef de ménage	PROFESSION	Pour mariage, état, profession, nationalité, etc.
	1	2	3	4	5								
Mandou	41-42-43	41	42	43	44	Andrie	Em	1888	Léves	Em	Chf	Cultivateur	Père
		45	Andrie	Blanc	1891	Em	Em	Chf	4/2				
		46	Andrie	Em	1907	Blanc	Em	Chf	4/2				
		47	Andrie	Em	1915	Blanc	Em	Chf	4/2				
		48	Jardin	Em	1920	Blanc	Em	Chf	4/2				
		49	Jardin	Em	1925	Blanc	Em	Chf	4/2				
Bastien	44-45-46	44	45	46	47	Dumont	Em	1898	Blanc	Em	Chf	Cultivateur	Père
		48	Dumont	Em	1915	Blanc	Em	Chf	4/2				
		49	Dumont	Em	1920	Blanc	Em	Chf	4/2				
		50	Jardin	Em	1925	Blanc	Em	Chf	4/2				
Bis Nant	43-44-45	43	44	45	46	Michel	Em	1891	Blanc	Em	Chf	Cultivateur	Père
		47	Michel	Em	1901	Blanc	Em	Chf	4/2				
Bis Bastien	45-46-47	45	46	47	48	P. Bastien	Em	1891	Blanc	Em	Chf	4/2	Père
		49	Em	1901	Blanc	Em	Chf	4/2					
Ammi	48-49-50	48	49	50	51	Eymette	Em	1895	Blanc	Em	Chf	Cultivateur	Père
		52	Eymette	Em	1905	Blanc	Em	Chf	4/2				
		53	Eymette	Em	1915	Blanc	Em	Chf	4/2				
		54	Em	1925	Blanc	Em	Chf	4/2					
		55	Em	1935	Blanc	Em	Chf	4/2					
		56	Em	1945	Blanc	Em	Chf	4/2					
		57	Em	1955	Blanc	Em	Chf	4/2					
		58	Em	1965	Blanc	Em	Chf	4/2					
		59	Em	1975	Blanc	Em	Chf	4/2					
		60	Em	1985	Blanc	Em	Chf	4/2					
Vauville	49-50-51	49	50	51	52	Manni	Em	1895	Blanc	Em	Chf	Cultivateur	Père
		53	Manni	Em	1905	Blanc	Em	Chf	4/2				
		54	Manni	Em	1915	Blanc	Em	Chf	4/2				
		55	Manni	Em	1925	Blanc	Em	Chf	4/2				
		56	Manni	Em	1935	Blanc	Em	Chf	4/2				
Champag	50-51-52	50	51	52	53	Baudet	Em	1895	Blanc	Em	Chf	Cultivateur	Père
		54	Baudet	Em	1905	Blanc	Em	Chf	4/2				
		55	Baudet	Em	1915	Blanc	Em	Chf	4/2				
		56	Baudet	Em	1925	Blanc	Em	Chf	4/2				
		57	Baudet	Em	1935	Blanc	Em	Chf	4/2				

DESIGNATION	BUREAU					NOMS	PRENOMS	ANNEE de naissance	LEGE de naissance	NATIONALITE	SITUATION PAR RAPPORT au chef de ménage	PROFESSION	Pour mariage, état, profession, nationalité, etc.
	1	2	3	4	5								
Carpentier	41-42-43	41	42	43	44	Carpentier	Em	1895	Blanc	Em	Chf	Cultivateur	Père
		45	Carpentier	Em	1905	Blanc	Em	Chf	4/2				
		46	Carpentier	Em	1915	Blanc	Em	Chf	4/2				
		47	Carpentier	Em	1925	Blanc	Em	Chf	4/2				
		48	Carpentier	Em	1935	Blanc	Em	Chf	4/2				
		49	Carpentier	Em	1945	Blanc	Em	Chf	4/2				
		50	Carpentier	Em	1955	Blanc	Em	Chf	4/2				
		51	Carpentier	Em	1965	Blanc	Em	Chf	4/2				
		52	Carpentier	Em	1975	Blanc	Em	Chf	4/2				
		53	Carpentier	Em	1985	Blanc	Em	Chf	4/2				

Liste nominative:
 Saint-Vincent-sur-Jabron,
 Alpes-de-Haute-Provence
 1936

DESIGNATION		NUMEROS PAR QUARTIER, VILLAGE hameau ou rue.			NOMS DE FAMILLE.	PRÉNOMS.	ANNÉE de NAIS- SANCE.	LIEU de NAISSANCE.	NATIONA- LITÉ.	SITUATION PAR RAPPORT au chef de ménage.	PROFESSION.	13
des QUAR- TIERS, villages ou hameaux.	DES RUES dans les villes.	des ma- sons.	des mé- nages.	des indi- vidus.								
1	2	3	4	5	6	7	8	9	10	11	12	13
Mondin		41-42-		114	André	Luis	1884	La Mées	française	Chef	Cultivateur	Patron
				115	André	Helvénus	1883	Sisters	id	Epouse	g/p.	
				116	André	Luis	1922	S'Vincent	id	Fils.	g/p.	
				117	André	Luis	1865	Ch. ^{neuf} M ^{me}	id	beau-père	Coiffeur	Patron
				118	Gervand	Pauletta	1934	Hospitales	id	petite fille	g/p.	
St Antonin		42-43-		119	Durand	Ferdinand	1872	S'Vincent	id	Chef.	Cultivateur	Patron
				120	Durand	Muris	1874	Nogers	id	épouse	g/p.	
				121	Durand	Gustav	1915	S'Vincent	id	fils	Cultivateur	
				122	Gervand	Gabrielle	1930	id	id	petite fille	g/p.	
Pis Neuf.		43-44-		123	Martel	Gulrich	1884	Chabon	id	Chef	Coiffeur	Patron
				124	Michel	Elic	1891	S'Vincent	id	chef	Cultivateur	Patron
Pis Patin		45-46-		125	V. Sembat.	Lionie	1883	id	id	Chef.	g/p.	Patron
				126	Eysenbacher	Luis	1904	Nogers	id	Chef.	Cultivateur	Patron

Pour les patrons,
 chefs d'entreprise,
 ouvriers à domicile,
 inscrire : patron

Pour les employés
 et ouvriers, indi-
 quer le nom du
 patron ou de l'en-
 treprise qui les
 emploie.

Testing for coherence and cohesion

Two pages of a population census form, showing a grid of columns and rows for recording data. The pages are slightly tilted and overlapping.

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 épars (2 ^e section du cadre A).....	.
Population municipale. (TOTAL du cadre A).....	40.067
Population comptée à part, conformément à l'article 3 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.....	40.832 40.211
absents le jour du recensement.....	621
Dont.....	1006
nes	12398
au cours	14462
des	9460
années	2879
1920 ou 1921.....	621
1901 à 1919 (1 à 19 ans).....	
1881 à 1900 (20 à 39 ans).....	
1861 à 1880 (40 à 59 ans).....	
1860 et antérieurement (60 ans et plus).....	
	<i>Age inconnu</i>

40.832

BÉCAPITULATION.

A. — TABLEAU RÉCAPITULATIF de la population inscrite sur la liste nominative et répartition de cette population en population agglomérée et population épars.

MUNICIPALITÉS, VILLAGES, HAMEAUX, MOULINS, etc.	VILLE				ÉTRANGERS
	de sexe masculin	de sexe féminin	de sexe total	de sexe total	
1 ^{re} Section, inscrite sur la liste nominative (Population agglomérée)					
Commune de Châtenoy-le-Français	287	402	689	287	392
	1.485	1.067	2.552	1.485	1.067
2 ^e Section, villages, hameaux, moulins et habitations en dehors de l'agglomération de Châtenoy, inscrite sur la liste nominative (Population épars)					
	1.581	1.067	2.648	1.581	1.067
TOTAL de la population agglomérée et épars	2.068	1.529	3.597	2.068	1.529

B. — TABLEAU RÉCAPITULATIF des populations comptées à part en exécution de l'article 3 du décret du 5 octobre 1920.

EXTÉRIEURES DE POPULATION	RENDRE		
	de sexe masculin	de sexe féminin	de sexe total
Refugiés.....	1006	1006	2012
Militaires et marins inscrits sur la liste nominative des réfugiés.....	589	589	1178
Autres réfugiés.....			
Militaires des Forces armées alliées.....			
Militaires des Forces armées alliées en France.....			
Militaires des Forces armées alliées en France.....			
Autres militaires des Forces armées alliées.....			
Depuis la déclaration de guerre.....			
Autres militaires des Forces armées alliées.....			
Autres militaires des Forces armées alliées.....			
Autres militaires des Forces armées alliées.....			
Autres militaires des Forces armées alliées.....			
TOTAL de la population comptée à part	1595	1595	3190

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 épars (2 ^e section du cadre A).....	.
Population municipale. (TOTAL du cadre A).....	40.067
Population comptée à part, conformément à l'article 3 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.....	40.832 40.211
absents le jour du recensement.....	621
Dont.....	1006
nes	12398
au cours	14462
des	9460
années	2879
1920 ou 1921.....	621
1901 à 1919 (1 à 19 ans).....	
1881 à 1900 (20 à 39 ans).....	
1861 à 1880 (40 à 59 ans).....	
1860 et antérieurement (60 ans et plus).....	
	<i>Age inconnu</i>

40.832

Challenges and objectives



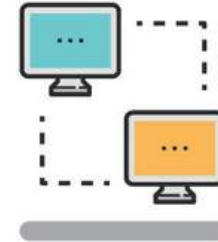
Social Science

Studies in economics, history, and demography: cause and consequences of long-term evolution of labor market, inequality, social structure and mobility.



Technology

Handwritten text transcription, table layout analysis, extracting individual-level data from millions of images.



Diffusion

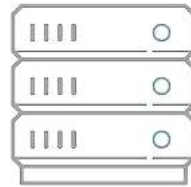
Disseminating in Open Access the data extracted, through the Archives, to the general public and the researchers

Challenges and objectives



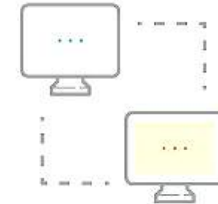
Social Science

Studies in economics, history, and demography: cause and consequences of long-term evolution of labor market, inequality, social structure and mobility.



Technology

Handwritten text transcription, table layout analysis, extracting individual-level data from millions of images.



Diffusion

Disseminating in Open Access the data extracted, through the Archives, to the general public and the researchers

Microdata at the national level

- An economic and social history at both individual and national levels
 - ❖ To fill the gap between national studies from aggregated statistics and local monographies.
 - ❖ A very important moment in the history of France: industrialization, urbanization, deep social transformations.
- Different ways to organize research: one cross-section (a year, with spatial variation); repeated cross-section (up until 20 years); linked individuals (trajectories between censuses).
- And in the future?
 - ❖ The basis for future historical studies as a platform for contemporary quantitative history.
 - ❖ Link with other sources (civil registers, military registers, ...) and databases (e.g., genealogical records).
 - ❖ Connection with the contemporary period.

Analyzing social change

➤ Structural change in the long run

- ❖ Transformation of the labor market: spatial variations, gender inequality, ...
- ❖ Evolution of transport systems: effects on the spatial distribution of the population...

➤ Shocks

- ❖ Short-, medium- and long-term consequences.
- ❖ e.g., phylloxera crisis; World War One, Courrières (1906).

➤ Spatial organization of economic activities

- ❖ Linking individual information with spatial database.
- ❖ Consequences of the transition from agriculture to industry at the very local level.

Linking individuals: building trajectories to study mobility

➤ Life-cycle trajectories

- ❖ Following individuals all their life, through time and space.
- ❖ As children and in adulthood.
- ❖ Transition between the two.

➤ Study geographic and social mobility...

➤ ... and how it changed over time.

Linking individuals between censuses

1886

92	Baillandier	Edouard	44	♂	Notaire	chef de ménage
93	Houssard	Octavie	30	♀	propriétaire	sa femme
94	Baillandier	Edouard	5	♂	"	leur fils
95	Baillandier	Marie	3	♀	"	leur fille
96	Besnard	Welfphine	72	♀	propriétaire	meie du chef
97	Bourcelot	Emili	20	♀	cuisinière	Domestique

1896

1	Julles	Joseph	52	♂	ouvrier	chef
2	Renaud	Gene	43	♂	receveur	"
3	Buchard	Jeanne	47	♀	Domestique	F
4	Abadie	Gene Charles	40	♂	percepteur	chef
1	Delamay	Amable	50	♂	propriétaire	époux
2	Beaudet	Eugene	43	♂	"	époux
1	Baillandier	Edouard	53	♂	notaire	époux
2	Houssard	Octavie	40	♀	propriétaire	épouse
3	Robert	Gene	19	♂	Domestique	F
1	Leroy	Victor	30	♂	marchand	époux
2	Mile	Gene	36	♀	"	épouse
3	Leroy	Victor	11	♂	"	fils
4	Leroy	Gene	7	♀	"	id
5	Mile	Eugene	27	♀	Domestique	serv. pièce

1906

1	Baillandier	Edouard	1842	Notaire	♂	chef	titulaire	patron
2	Besnard	Octavie	1836	Propriétaire	♀	épouse	"	"
3	Baillandier	Edouard	1880	Notaire	♂	fils	"	"
4	Baillandier	Marie	1883	Notaire	♀	fil	"	"
5	Leroy	Gene	1881	Notaire	♀	Domestique	serv. pièce	Baillandier
6	Baillandier	Edouard	1880	Notaire	♂	id	id	id
1	Leroy	Gene	1870	Notaire	♀	chef	titulaire	patron
2	Baillandier	Gene	1871	Notaire	♀	épouse	titulaire	patron
1	Baillandier	Gene	1880	Notaire	♀	chef	"	patron
2	Baillandier	Gene	1881	Notaire	♀	petit-fils	titulaire	patron
1	Baillandier	Edouard	1882	Notaire	♂	chef	titulaire	patron
2	Baillandier	Gene	1881	Notaire	♀	épouse	titulaire	patron
3	Baillandier	Gene	1882	Notaire	♀	id	id	id

Challenges and objectives



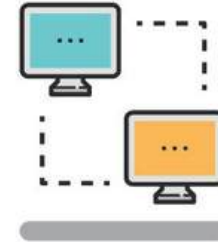
Social Science

Studies in economics, history, and demography: cause and consequences of long-term evolution of labor market, inequality, social structure and mobility.



Technology

Handwritten text transcription, table layout analysis, extracting individual-level data from millions of images.



Diffusion

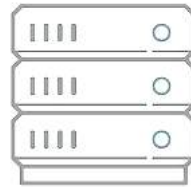
Disseminating in Open Access the data extracted, through the Archives, to the general public and the researchers

Challenges and objectives



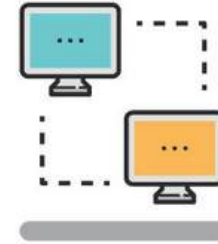
Social Science

Studies in economics, history, and demography: cause and consequences of long-term evolution of labor market, inequality, social structure and mobility.



Technology

Handwritten text transcription, table layout analysis, extracting individual-level data from millions of images.



Diffusion

Disseminating in Open Access the data extracted, through the Archives, to the general public and the researchers

Dissemination: back to Archival deposits, and beyond

- 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?

Thank you!

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

<http://socface.org>

[**contact@socface.org**](mailto:contact@socface.org)