



Un projet de recherche collaboratif entre

Historiens, économistes, démographes

Archivistes

Experts en intelligence artificielle



Et les services d'archives départementales et municipales

Lionel Kesztenbaum



Présentation Ligeo
25/01/2022

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.



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**

DESIGNATION		NUMÉROS, PAR QUARTIER, VILLAGES, HAMEAUX ou VILLES			NOMS	PRÉNOMS	ANNEE	LIEU	NATIONAL.	SITUATION	PROFESSION	REMARQUES
des quartiers, villages ou hameaux	des rues des villes	des numéros	des ménages	des habités	DE FAMILLE		de naissance	de naissance	LITÉ	PAR RAPPORT au chef de ménage		
1	2	3	4	5	6	7	8	9	10	11	12	13
				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	Boulonnais Valenciennes
				8	Pierre	Liliane	1912	Bagnollet	"	filie	sans	
				9	Pierre	Lucien	1918	Woisylot	"	neveu	sans	
				1	Parmenier	Julave	1877	Mesfiteau	"	chef	maçon	
				2	Parmenier	Lucie	1867	Lambach	"	épouse	sans	
				3	Parmenier	Maximilien	1888	Paris	"	enfant	Confectionnier	Chulman
				4	Parmenier	Emma	1900	W. Denis	"	"	empl.	Minist. Pensions
				5	Parmenier	Ludovic	1902	"	"	"	plum.	Calix
				1	Caliez	Georges	1878	Lille	"	chef	Condamné	Magist.
				2	Caliez	Céline	1884	Taches	"	épouse	sans	
				1	Renauld	Jean	1869	Combrigny	"	chef	chef équipe	Magist.
				2	Renauld	Olivia	1875	Combrigny	"	"	"	

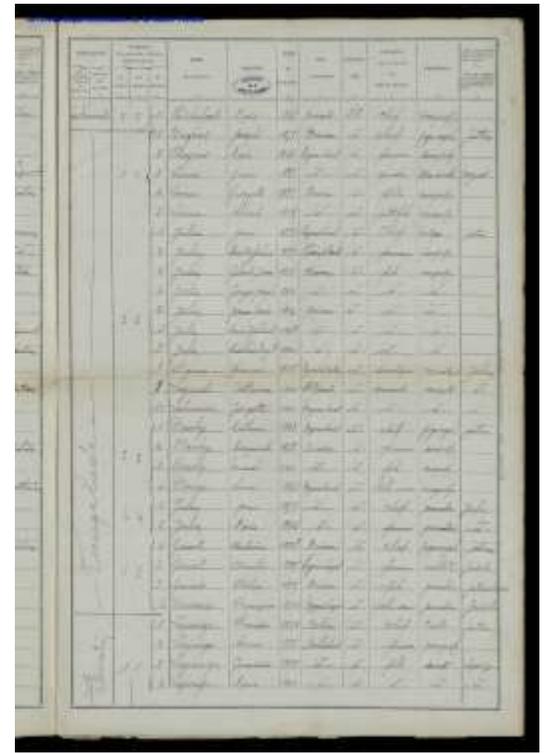
Une liste nominative:

Aubervilliers, 1921

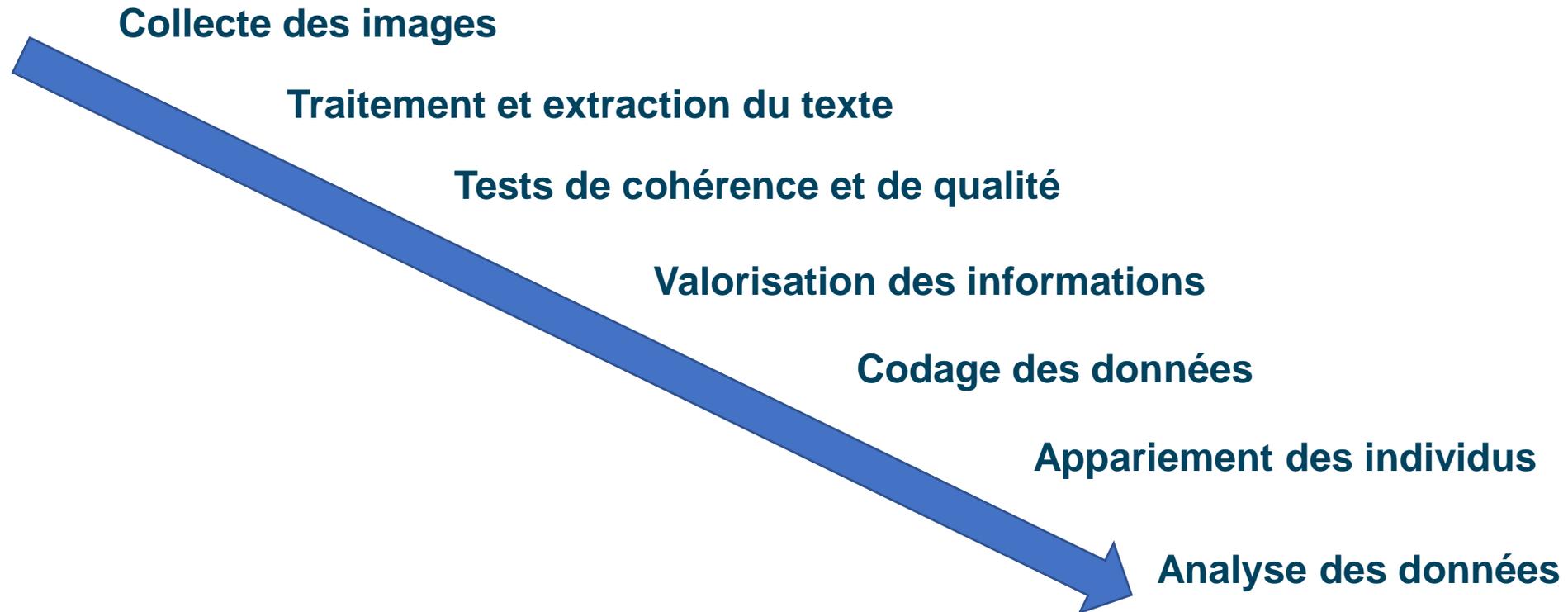
Rue
des
Ellelles

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.



Etapes et objectifs



Etude pilote: l'ensemble des étapes sur une sélection d'archives – Objectif: un an

Puis montée en charge progressive à partir d'octobre 2022

Incorporation des services d'archives peu à peu et retours au fur et à mesure

De multiples objectifs et enjeux



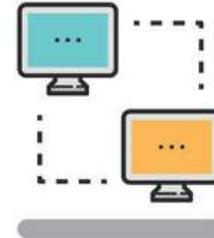
Science

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.



Valorisation

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

De multiples objectifs et enjeux



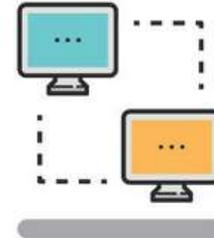
Science

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.



Valorisation

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

Listes nominatives de recensement

Défis du projet:

- Documents décentralisés
- Volume inconnu mais important (entre 5 et 15 M d'images ?)
- Conditions de conservation/numérisation variables
- Lecture en 2 dimensions
- Evaluer l'erreur
- Structurer, consolider, croiser l'information

Lecture en 2 dimensions

312 — 4 —

DÉSIGNATION		NUMÉROS, PAR QUARTIER, VILLAGES, HAMEAUX ou TERRES			NOMS	PRÉNOMS	ANNÉE	LIEU	NATIONALITÉ	SITUATION	PROFESSION	REMARQUES
des villages ou hameaux	des quartiers ou villages	des maisons	des maisons	des maisons	DE FAMILLE		de naissance	de naissance	LITTE.	ou chef de ménage		
1	2	3	4	5	6	7	8	9	10	11	12	13
Rue des Éillettes	115	11	2	2	Loribout	Lucie	1886	Reims	F	épouse	ans	
			3	3	Loribout	Léoni	1909	Paris	"	enfant	"	
			4	4	Loribout	Alphonse	1912	"	"	"	"	
			5	5	Loribout	Reni	1916	subisillies	"	"	"	
			6	6	Loribout	Audré	1920	"	"	"	"	
		4	4	Pierre	Léonide	1881	Reims	"	bell. soeur	Magasin	Parabonnes Municipales	
		8	8	Pierre	Liliane	1918	Baguette	"	filie	ans		
		3	3	Pierre	Lucien	1918	Woignoles	"	veuve	ans		
		1	1	Parmenier	Juliane	1877	Meschesnois	"	chef	maçon		
		2	2	Parmenier	Lucie	1867	Lambach	"	époux	ans		
3	3	Parmenier	Maximilien	1885	Paris	"	enfant	Confectionn.	Chulman			
4	4	Parmenier	Emma	1900	St Denis	"	"	empl.	Mémoire Paris			
5	5	Parmenier	Ludovic	1902	"	"	"	"	chef	Calix		
1	1	Caliez	Jorges	1878	Lille	"	chef	Confectionn.	Mag			
2	2	Caliez	Celine	1884	Stache	"	époux	ans				
1	1	Renault	Jean	1869	Combray	"	chef	chef équipe	M			

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 épars (2 ^e section du cadre A).....	765
Population municipale. (TOTAL du cadre A).....	40.067
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	40.832
absents le jour du recensement.....	621
Dont	1006
1920 ou 1921.....	12398
1901 à 1919 (1 à 19 ans).....	14462
1881 à 1900 (20 à 39 ans).....	9460
1861 à 1880 (40 à 59 ans).....	2879
1860 et antérieurement (60 ans et plus).....	621
àge inconnu.....	10.832

RÉCAPITULATION.

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

CATEGORIES - SEXES, MARIAGES	MARIAGES			
	de	de	de	de
de	de	de	de	de
1 ^{re} MARIAGES, autres que ceux des				
Catholiques	287	1092	1007	3911
Protestants				1249
Autres				
2 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
3 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
4 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
5 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
6 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
7 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
8 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
9 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
10 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
11 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
12 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
13 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
14 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
15 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
16 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
17 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
18 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
19 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
20 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
21 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
22 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
23 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
24 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
25 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
26 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
27 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
28 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
29 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
30 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
31 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
32 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
33 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
34 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
35 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
36 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
37 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
38 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
39 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
40 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
41 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
42 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
43 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
44 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
45 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
46 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
47 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
48 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
49 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
50 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
51 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
52 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
53 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
54 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
55 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
56 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
57 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
58 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
59 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
60 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
61 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
62 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
63 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
64 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
65 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
66 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
67 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
68 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
69 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
70 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
71 ^e MARIAGES				
Catholiques				
Protestants				
Autres				
72 ^e MARIAGES				
Catholiques				
Pro				

De multiples objectifs et enjeux



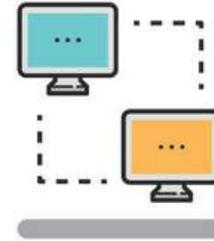
Science

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.



Valorisation

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

De multiples objectifs et enjeux



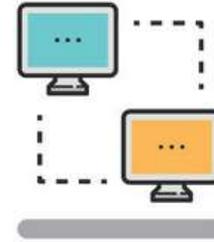
Science

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.



Valorisation

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

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 monographie 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

- Trajectoire des individus au cours de leur vie
 - ❖ Dans l'enfance et dans l'âge adulte.
 - ❖ Transition entre les deux.
- Etudier la mobilité géographique et sociale.
- Et son évolution au cours du temps.

L'apport des appariements: étudier la mobilité

1886

La Place.

31	36	92	Baillandier	Gérard	44	♂	Noblesse	chef de ménage
		93	Houssard	Octavie	30	♀	propriétaire	sa femme
		94	Baillandier	Gérard	5	♂	"	son fils
		95	Baillandier	Maurice	3	♂	"	son fils
		96	Bouvard	Welféine	72	♀	propriétaire	sa femme
		97	Bouzelet	Emile	20	♂	ouvrier	ouvrière
32	36	98	Nelamoy	Amable	40	♂	Coiffeur	chef de ménage
		99	Beardos	Eugénie	32	♀	M ^{re} au commerce	sa femme
33	37	100	Chapin	Julien	43	♂	rentier	chef de ménage
		101	Ferat	Mariette	60	♀	rentière	sa femme
		102	Julie	Joseph	42	♂	ouvrier	chef de ménage
34	38	103	Grandin	Nicolas	22	♂	Noblesse	ouvrier
39		104	Bichard	Jeannette				
40		105	Brigod	Marie				
41		106	Bouvard	Renée				

1896

La Place.

1	Julie	Joseph	72	♂	rentier	chef
2	Renard	Henri	42	♂	rentier	"
3	Buchard	Jeannette	47	♀	propriétaire	"
4	Chodron	Henri Charles	40	♂	rentier	chef
5	Nelamoy	Renée	50	♀	propriétaire	épouse
6	Grandin	Eugène	43	♂	"	épouse
7	Baillandier	Edmond	53	♂	rentier	épouse
8	Houssard	Octavie	40	♀	propriétaire	épouse
9	Kobus	Renée	37	♀	propriétaire	"
10	Vary	Julie	30	♀	marchande	épouse
11	Mile	Renée	35	♀	"	épouse
12	Vary	Julie	11	♀	"	fils
13	Vary	Renée	7	♀	"	id.
14	Mile	Eugène	27	♂	propriétaire	son fils

1906

La Place.

1	Baillandier	Gérard	1882	Noblesse	♂	chef	ouvrier	fils
2	Bouvard	Blaise	1885	rentier	♂	épouse	"	"
3	Baillandier	Edmond	1880	Noblesse	♂	fils	"	"
4	Vary	Julie	1883	Noblesse	♂	fils	"	"
5	Vary	Julie	1881	Noblesse	♂	rentier	rentier	fils
6	Mile	Renée	1880	Noblesse	♀	id.	id.	id.
7	Bouvard	Julie	1870	Noblesse	♀	chef	rentier	fils
8	Chodron	Julie	1871	Noblesse	♀	épouse	rentier	"
9	Bouvard	Henri	1880	Noblesse	♂	chef	"	fils
10	Bouvard	Henri	1882	Noblesse	♂	rentier	rentier	épouse
11	Bouvard	Henri	1883	Noblesse	♂	chef	rentier	fils
12	Bouvard	Henri	1884	Noblesse	♂	épouse	rentier	"
13	Bouvard	Henri	1885	Noblesse	♂	fils	"	"
14	Bouvard	Henri	1886	Noblesse	♂	chef	rentier	fils
15	Bouvard	Henri	1887	Noblesse	♂	épouse	rentier	"
16	Bouvard	Henri	1888	Noblesse	♂	fils	rentier	Bouvard
17	Bouvard	Henri	1889	Noblesse	♂	rentier	rentier	"
18	Bouvard	Henri	1890	Noblesse	♂	chef	rentier	Bouvard
19	Bouvard	Henri	1891	Noblesse	♂	épouse	rentier	"
20	Bouvard	Henri	1892	Noblesse	♂	fils	rentier	Bouvard
21	Bouvard	Henri	1893	Noblesse	♂	rentier	rentier	"
22	Bouvard	Henri	1894	Noblesse	♂	chef	rentier	Bouvard
23	Bouvard	Henri	1895	Noblesse	♂	épouse	rentier	"
24	Bouvard	Henri	1896	Noblesse	♂	fils	rentier	Bouvard
25	Bouvard	Henri	1897	Noblesse	♂	rentier	rentier	"
26	Bouvard	Henri	1898	Noblesse	♂	chef	rentier	Bouvard
27	Bouvard	Henri	1899	Noblesse	♂	épouse	rentier	"
28	Bouvard	Henri	1900	Noblesse	♂	fils	rentier	Bouvard

De multiples objectifs et enjeux



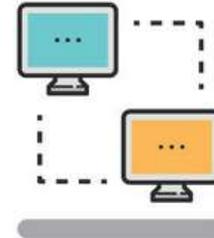
Science

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.



Valorisation

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

De multiples objectifs et enjeux



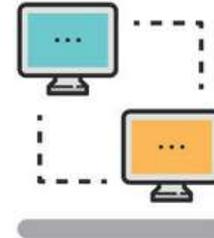
Science

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.



Valorisation

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

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
 - ❖ Base avec les codages des informations brutes.
 - ❖ Base des appariements.

- Ouverture sur d'autres sources: un modèle pour la mise à disposition des archives nationales?

Merci beaucoup!!!

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

<http://socface.org>

lionel.kesztenbaum@ined.fr

Spécificités du projet Socface

➤ Traitement automatisé de vastes ensembles d'images

- ❖ Plusieurs millions d'images, de provenances variées (différents éditeurs, formats, etc.).
- ❖ Ouvre des perspectives considérables pour la recherche en intelligence artificielle.

➤ Une coproduction entre historiens et informaticiens

- ❖ Utiliser la structure des listes pour contrôler l'information (ordre des individus, ménages, etc.).
- ❖ Utiliser les informations agrégées (externes ou non) pour contrôler le traitement.
- ❖ Codage des informations obtenues pour les rendre accessibles aux chercheurs (lieux, professions...).

➤ Appariement des individus pour les suivre au cours du temps

- ❖ Première réalisation à l'échelle de la France entière, sur 100 ans.
- ❖ Ouvre des perspectives considérables pour la recherche en sciences sociales.