

#### Un projet de recherche collaboratif entre

Historiens, économistes, démographes

**Archivistes** 

Experts en intelligence artificielle







Archives départementales et municipales



Christopher Kermorvant Lionel Kesztenbaum



Webinaire SIAF-archives 17/10/2023



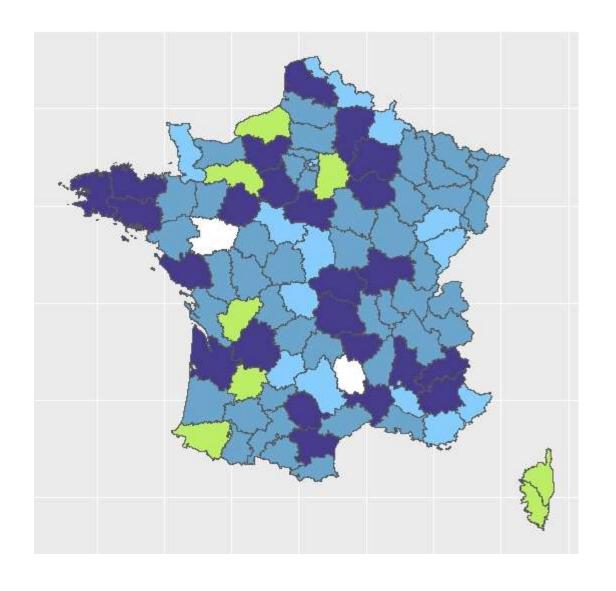
#### SocFace: bilan à mi-parcours

- > Deux ans écoulés, deux ans à venir...
- > Collecte des images auprès des services d'archives.
- Préparation du traitement des images: premiers modèles, échantillon d'entrainement, identification des obstacles, etc.
- > Organisation de la préparation des données extraites.
- > Arrivée des premières données (AD045 + FRAC45234).
- > Premiers tests sur les appariements.



# Etat de la collecte au 16/10/2023

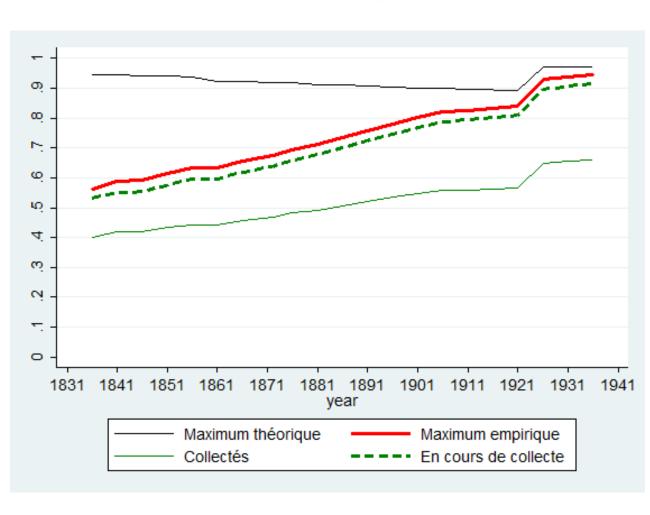
	AD	(AIVI
En cours de collecte	27	(0)
Images collecées	59	(3)
Images dans Arkindex	25	(1)
Ne participe pas (encore)	8	(6)
Pas d'images	2	



#### Population couverte

- En théorie: 20 recensements x 36 000 communes sur 100 ans
  - \* Entre 10 et 20 millions d'images (doubles).
  - Entre 400 et700 millions d'entrées (lignes).
- > Réalité (à ce jour):
  - Environ 7 millions d'images.
  - \* Environ 300 millions de lignes.
- > Réalité (à venir, estimations grossières):
  - \* Autour de 12 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)

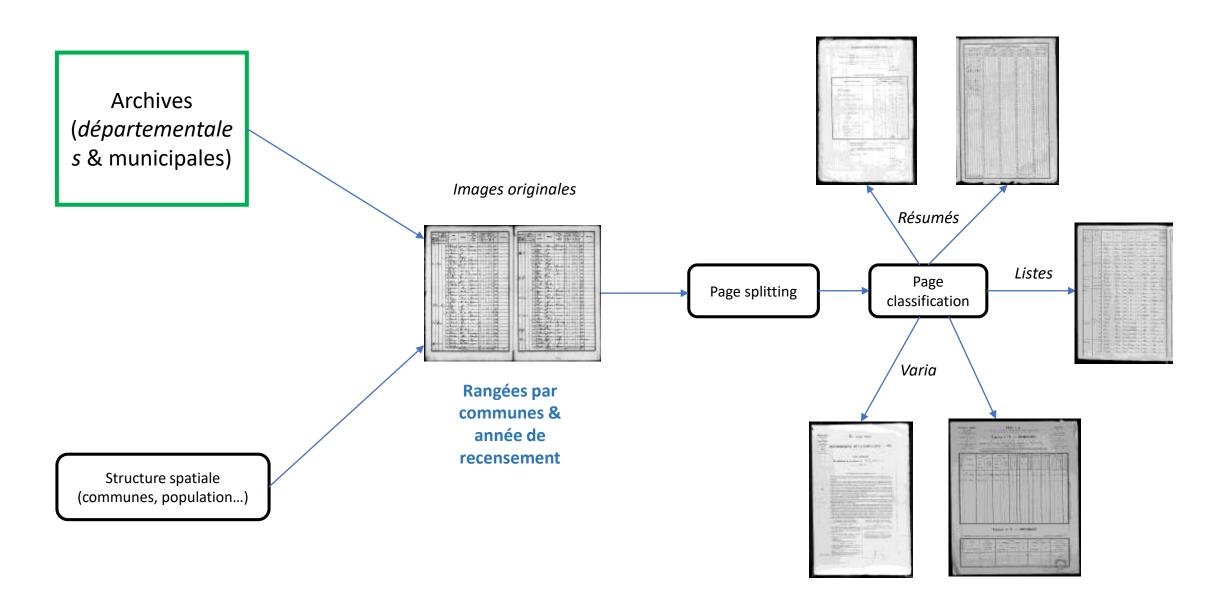




#### SocFace: bilan à mi-parcours

- > Deux ans écoulés, deux ans à venir...
- > Collecte des images auprès des services d'archives.
- Préparation du traitement des images: premiers modèles, échantillon d'entrainement, identification des obstacles, etc.
- > Organisation de la préparation des données extraites.
- > Arrivée des premières données (AD045 + FRAC45234).
- > Premiers tests sur les appariements.

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



#### Le pipeline Socface (2) Construction de la base de données **Appariements** Vocabulaires/ Listes distributions Structure Structure des analysis ménages Listes Base de nominatives données Echantillon aléatoire **Annotations** Retranscription du Entrées Tests de texte & entités individuellescoherence nommées Codage de Statistiques l'information agrégées Transcriptions collaboratives

## Un exemple avec retranscription et reconnaissance des entités nommées

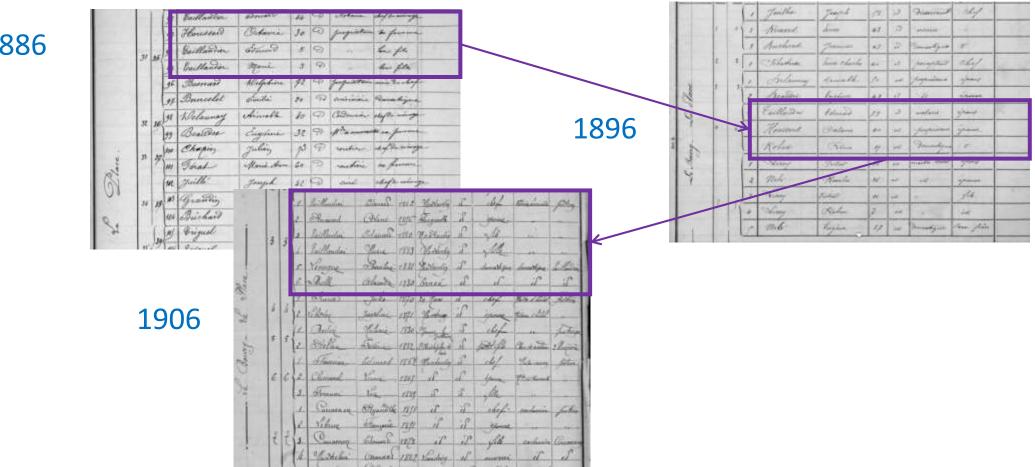
```
surname Robert firstname Juliette occupation néant linkl, bille age 8 nationality française
surname Robert firstname Albertine occupation idem linkidem age 3 nationality idem
surname Johan firstname Jules occupation domestique link chauger age 19 nationality idem
surname Belouet firstname Jeanne occupation idem linkidem age 22 nationality idem
surname Beignet firstname Charles occupation les linkchef age 32 nationality idem
surname Bezault firstname Hermance occupation idem linképouse 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 | leer link chef | age 50 | nationality idem
surname Gronni firstname Marie occupation idem linképouse age 46 nationality idem
surname Fortin firstname Juliette occupation néant links a fille age 19 nationality idem
surname Portin firstname Georgette occupation idem linkidem age 12 nationality idem
surname Portin firstname Maurice occupation idem linkidem age 10 nationality idem
surname Vappereau firstname Andrée occupation idem linkidem age 1 nationality idem
surname Grangeau firstname Philomen occupation meunier link chef age 47 nationality idem
```

DÉSIGNATION	PAR QU	NUMÉROS	VELLAUR.					1	and an annual control of the control	
day mas aura gran- gran, dans les ou	des	das	des	NOMS	PRÉNOMS	AOM	LITÉ.	PROPERMON,	STRUCTION DARK LE MÉDANE	OBSERVATIONS
on imenux. f #	3	4	5		7			10		12
			4	Robus	Julia	1	#.	Newt.	1 /11	
			5	Gobert "	Allyto	3	-	-	2	
1.50			6	Johan	- July	11		Donntyn	Thanger	
			1	Belows	James	22		2	-60	
The same			9	Beignet	Charles	32	*	14	Chil.	
100			1	Bezault	Hermance	27		-0	epour	
Va .			10	Bugnet	Pini	6	*		1. ft	
			11	Bugnet	Julatte	3	*	n	4 file	DEEL
			14	Tollin	Jean,	60		7h	-Chif.	
			2	Growni	Marie	116		2	époise	
100			3	Fortini -	Juliette.	19		Neme	A fille	-
PER			4	Poston 1	Georgette	12	h	- W	R	
		- 5	1	Gotton	Maurie	10		æ	D	
			6	Vapperen	Annie	1		2	Nourne	
To find			7	Grangen	Philimon	47		Meunir	Clif	Marie Land



### L'apport des appariements: étudier la mobilité

1886





#### Merci de votre attention







