

Les individus face au changement social: analyse de la structure sociale en France sur un siècle 1836-1936

Projet de recherche soutenu par l'Agence Nationale de la Recherche

SOC **FACE**

ANR
AGENCE
NATIONALE
DE LA
RECHERCHE



Un projet de recherche collaboratif entre

Historiens, économistes, démographes

Experts en intelligence artificielle



Archivistes



Les services d'archives départementales et municipales

De multiples objectifs et challenges :



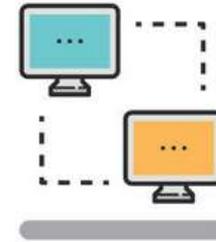
Technologie

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



Science

Etudes en économie historique et démographie : évolution du marché du travail, de l'urbanisme, de la structure sociale, des mobilités.



Valorisation

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

Pourquoi ?

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.

Objectifs

Collecter, traiter, retranscrire, organiser et analyser l'ensemble des listes nominatives du recensement de 1836 à 1936 (20 recensements).

Livrables

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.

Des données individuelles à l'échelle nationale

Comblent un vide entre enquêtes nationales à partir de statistiques agrégées et monographie plus locales afin de bâtir une histoire économique et sociale de la France à la fois individuelle et nationale.

Un outil pour la recherche

Base d'étude ou d'analyse pour de futures enquêtes historiques: matrice de la recherche en histoire quantitative contemporaine.

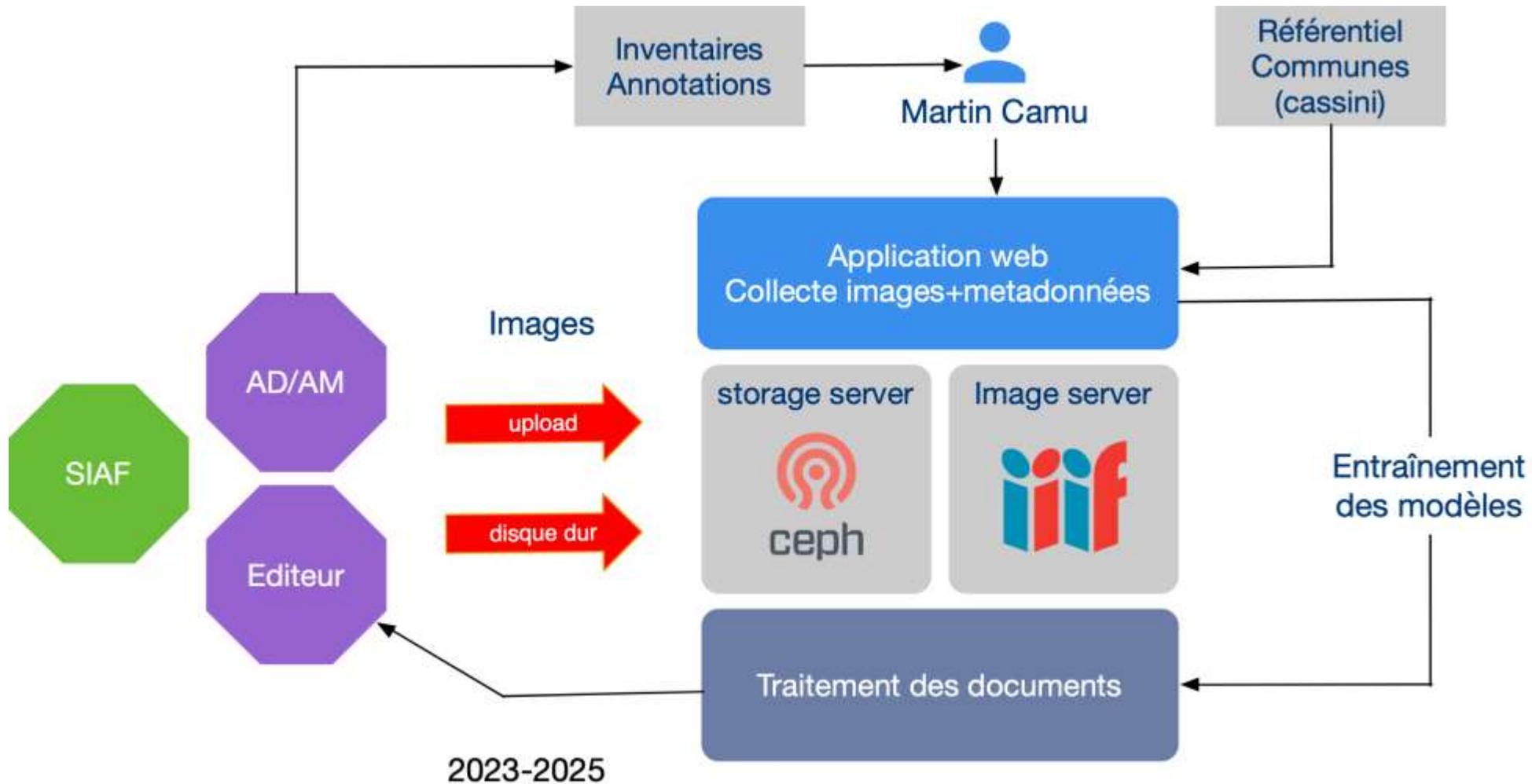
Liens possibles avec la période actuelle.

Diffusion et valorisation

Mise à disposition des données brutes dans une base nationale (sur FranceArchives) et sur les sites des Archives Départementales.

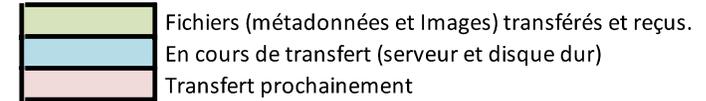
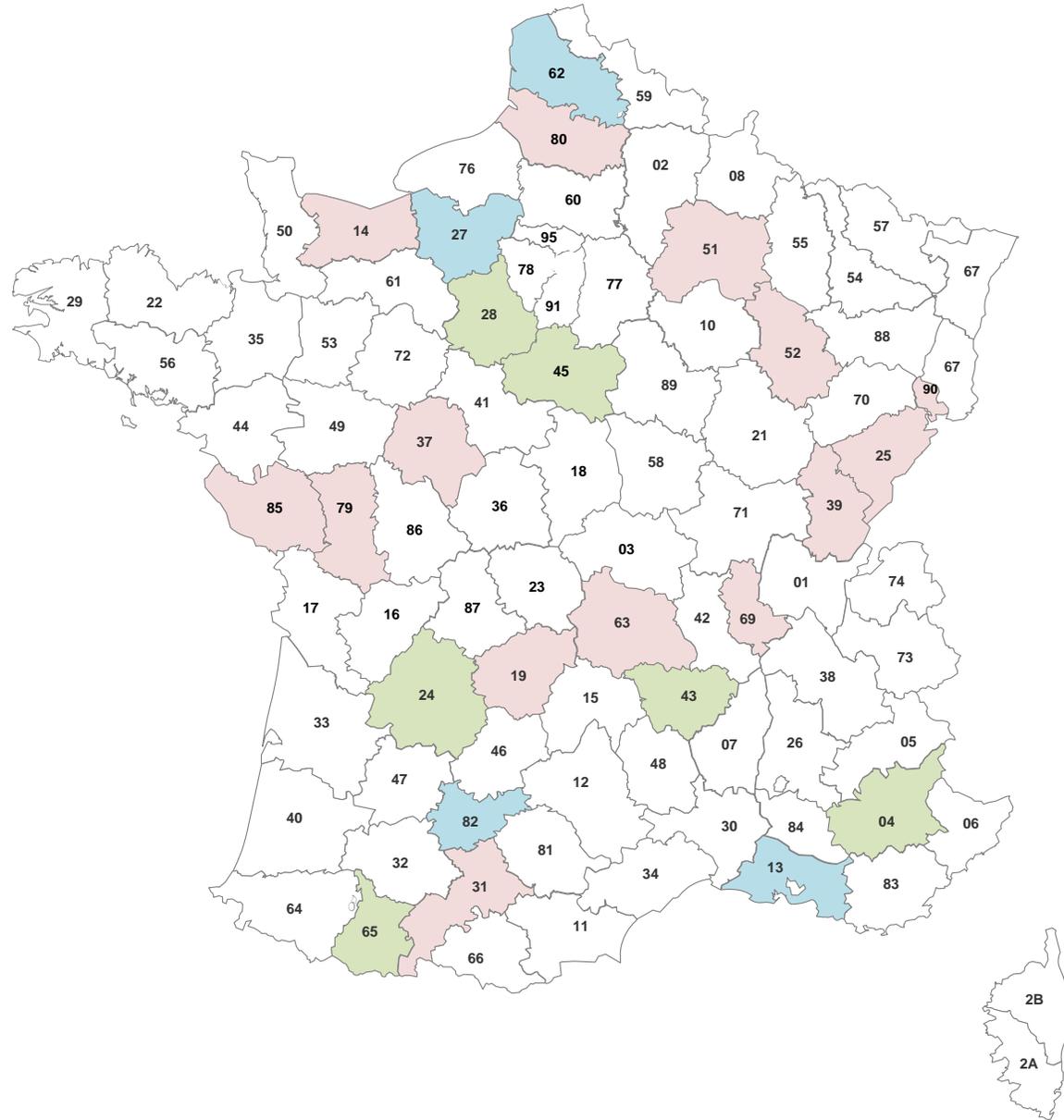
Mise à disposition des données de la recherche.

Etape 1: collecte des images+meta-données



Etape 1: progression de la collecte

5 AD completes (images+méta-données)
6 AD en cours de transfert
14 AD en préparation



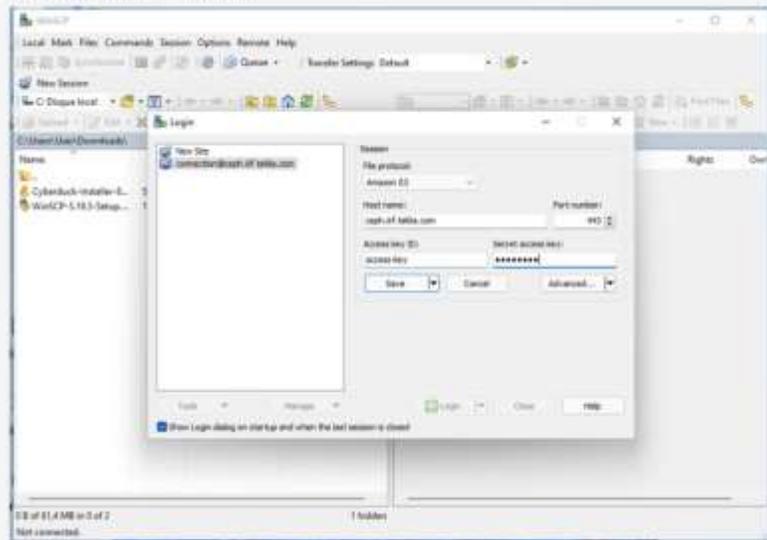
Etape 1: comment participer

Téléverser les images

Pour les systèmes Windows, il est préférable d'utiliser WinSCP qui est également une application graphique pour gérer des documents distants sur le bucket S3.

- 1. Installer WinSCP : <https://winscp.net/eng/index.php>
- 2. Démarrer WinSCP
- 3. Créez une nouvelle connexion et configurez-la

1. sélectionner le type amazon S3
2. entrez l'adresse du serveur : ceph.liif.teldia.com
3. entrez votre identifiant de clé d'accès
4. entrer votre clé d'accès secrète.



Envoyer un disque dur par la poste



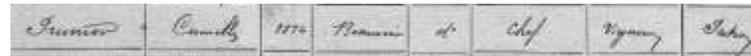
INED - Martin Camu
9, cours des Humanités
CS50004
93322 Aubervilliers Cedex

Etape 2: Entraînement des modèles IA

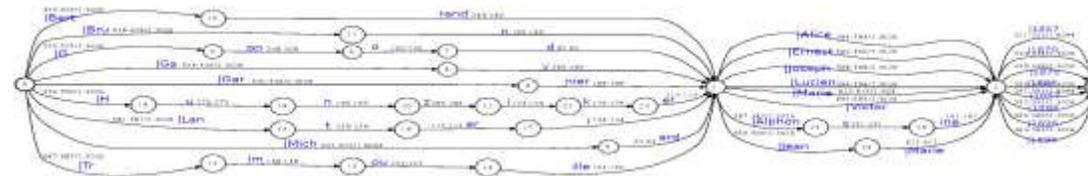
Détection des lignes de texte



Reconnaissance de texte avec contrainte syntaxique



Modèle optique



Modèle linguistique

Validation de cohérence Avec contraintes

