



Lettre d'information n°41 - Supplément

avril 2026 - page 1

SocFace - un projet hors normes

Journée d'étude du 24 mars 2026 organisée aux Archives nationales de Pierrefitte par le Service interministériel des Archives de France (SIAF) et l'Institut national d'études démographiques (Ined) à l'occasion de la mise en ligne sur le site FranceArchives, une semaine auparavant, des premières données du projet **SocFace**. A noter que 370 personnes étaient inscrites à la journée du 24 mars, en présentiel ou en visio.

Rappel du projet

Le nom est constitué de **Soc** pour société et **Face** pour visage, soit visages de société ou visages d'une société. Projet initié en 2021 pour réaliser une base de données de tous les individus ayant vécu en France entre 1836 et 1936 à partir des **recensements** en utilisant l'IA, l'**HTR** (reconnaissance de caractères manuscrits). Objectifs : analyser le changement social dans la longue durée, améliorant ainsi considérablement la compréhension des structures économiques et sociales françaises, et permettre l'accès aux listes nominatives pour le grand public. <https://socface.site.ined.fr/>

Quelques (gros) chiffres

- 20 recensements à traiter : 1836-1936
 - 18 millions d'images de recensement produites par les services d'archives (départementales/municipales)
 - env. 400 millions de noms de personnes à extraire
 - env. 3,2 milliards de données (noms, prénoms, professions, lieux, etc.)
- Dépasse en nombre tous les projets jusque-là mis en œuvre par les services d'archives y compris Le Grand Mémorial qui traite 11 millions de noms.
- A noter que l'utilisation de l'IA a permis de diminuer de manière significative le coût de production des données, soit env. 42 fois moins cher* que pour Le Grand Mémorial. L'IA a également permis de réaliser le projet beaucoup plus rapidement.

Quelques autres projets utilisant l'IA sont à mentionner :

- **Recensements de Paris** (qui a donné lieu à l'exposition du Musée Carnavalet) - distinct de SocFac mais parfois avec les mêmes équipes. [A consulter ici](#).
- **Simara** (Saisie d'inventaires manuscrits assistée par reconnaissance automatique) aux [Archives nationales](#)
- **Transcription de manuscrits anciens** à partir du XVII^e siècle aux AD de la Côte-d'Or. [A lire ici](#).

* Ce qui a, par contre, un coût non négligeable, c'est la numérisation.

Données non exhaustives et des erreurs

Il manque quelques départements, des communes ou années (archives détruites), des individus (population comptée à part).

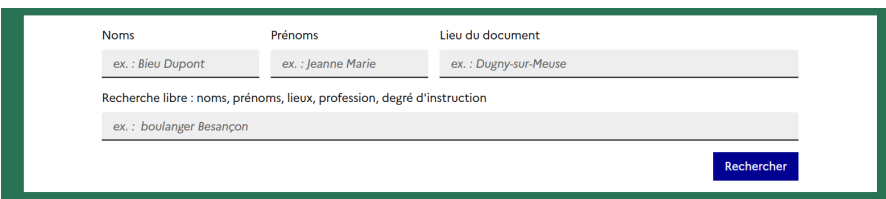
A cela, s'ajoutent quelques erreurs dues au traitement automatique : images corrompues, matching, classification, regroupements en ménage incorrects. A noter aussi le choix d'un seul modèle d'IA pour tous les documents ; en outre il n'a pas été prévu de correction manuelle comme cela a été fait pour Paris). Il est rappelé que les relevés ou indexations effectués par des humains comportent aussi des erreurs, peut-être plus difficiles à détecter que celles produites par l'HTR qui pourront être corrigées au fil du temps. s de la base.



Mise en ligne dès à présent

Malgré les imperfections mentionnées, il a été décidé **une mise en ligne** le 17 mars dernier sur "la base de noms" du portail **FranceArchives**. <https://francearchives.gouv.fr/>

Pour le moment, pas de recherche avancée. Possibilité d'utiliser le ? ou le * pour remplacer une lettre. Dans la recherche libre, on peut utiliser la barre de séparation (alt + 6).



Search interface showing fields for Noms (ex: Bieu Dupont), Prénoms (ex: Jeanne Marie), Lieu du document (ex: Dugny-sur-Meuse), and Recherche libre (ex: boulanger Besançon). A Rechercher button is visible.

NB. Dans un premier temps, les données de 55 AD, 7 Archives municipales et des Anom ont été mises en ligne. La suite s'effectuera progressivement d'ici l'été 2026.

Double diffusion et circuit de mise à jour des données

Les données seront ensuite rétrocédées aux **services d'archives** qui les mettront en ligne sur leurs **propres sites**. A noter que quelques ajustements techniques seront nécessaires pour les liens vers les URL des visionneuses. Adaptation aussi des éditeurs de solutions logicielles.

Les services d'archives auront alors la possibilité d'apporter des corrections, peut-être en sollicitant des bénévoles dans le cadre d'un projet collaboratif. Chaque service d'archives sera autonome.

Le pourcentage d'erreurs détectées varie entre 2 et 30% selon le type d'informations, le plus grand nombre d'erreurs concernant les professions et les lieux de naissance.

Il y aura ensuite un retour vers FranceArchives pour mettre à jour la base nationale ; il faudra aussi prévoir un circuit régulier de mise à jour.

Cadre juridique

Il est rappelé les dispositions législatives quant à la communicabilité en lien avec la protection des données à caractère personnel.

Il est rappelé aussi une disposition de 1887 qui aurait pu être dramatique si elle avait été appliquée partout : "*Documents à supprimer - Les papiers qui ne sont pas de nature à être conservés indéfiniment (...) Population - les tableaux dressés par les mairies pour le recensement quinquennal de la population et les bulletins individuels inutiles lorsque les opérations de recensement subséquent sont terminées, c'est-à-dire environ après six ans.*"

Bien que la disposition ait été très peu respectées, on peut néanmoins déplorer quelques pertes.

FranceArchives

Portail national des Archives

Recherche dans la base de noms →



Recensements de population

Données sur les recensements de la population française produites dans le cadre du projet de recherche SocFace.

[Rechercher dans les recensements de population](#)

Un levier pour les politiques territoriales

Le responsable des AD de l'Aisne a expliqué comment le projet SocFace est une opportunité pour l'Aisne. En effet dans ce département, touché par les deux guerres mondiales, la liste de recensements préfectorale a disparu. Il ne reste que des collections communales dispersées dans les quelques 800 communes du département. SocFace est un argument pour accélérer les dépôts communaux (pas toujours bien compris par les communes) et la numérisation. Dans ce département où il n'y a pas d'antenne universitaire en histoire, les AD s'appuient sur les associations historiques ou généalogiques.

Constatation : les nouvelles technologies et les sites marchands obligent les associations généalogiques à se repenser et à plus collaborer avec les AD.

Table ronde 1 : utilisateurs et réutilisateurs face à SocFace

Pour la **chercheuse du CNRS**, cela va changer toute une série de travaux. Alors qu'il y a de moins en moins de chercheurs, cela va faciliter le travail et donner de l'appétence pour d'autres travaux, comme la possibilité de cartographier des professions et peut-être relancer des vocations d'historiens démographes.

Pour la **Fédération Française de Généalogie**, il y a un potentiel énorme. C'est une porte d'entrée vers un parcours de vie. Cela peut faire revenir les généalogistes vers les centres d'archives. Concernant les associations, elles sont à une grande période de basculement. Les grandes idées de relevés arrivent au bout. SocFace nous dit : il faut faire autrement. L'idée c'est la réutilisation, comment devenir des relais pour les associations. L'IA est-elle un bien ou un mal ? SocFace va aider à démontrer l'intérêt de l'IA. Cela peut aider à faire revenir les jeunes.

La FFG va aider au maximum à faire connaître SocFace, en faisant redescendre l'information vers les associations.

Le **professeur d'histoire de Reims** voit dans SocFace un intérêt quant à son territoire marqué par l'ère industrielle et les mouvements de population dus à la Première Guerre Mondiale. Cela peut aussi constituer une éducation aux médias. Cela va permettre de rapprocher les élèves des informations ; ils feront eux-mêmes les recherches. SocFace va lever le problème de l'écriture manuscrite.

Le responsable des **AD 24** n'a pas encore pu tester puisque les données de ce département ne sont pas encore en ligne. De manière générale aux AD, les demandes des usagers portent sur des individus. La réponse doit être individuelle. Il va falloir être pédagogique pour expliquer les erreurs possibles. Cela va donner la possibilité de situer un individu dans son environnement. Les AD vont devoir mettre en avant l'existence de cette base. Comme il a été précédemment responsable des AD 45, il s'interroge sur les données de recensement qui avaient été indexées. Teklia, prestataire de service informatique confirme qu'il n'a pas été possible de reprendre les indexations réalisées par des humains dans le projet SocFace.

NB. Parmi les questions posées à l'issue de cette table ronde, on peut noter celle concernant les adresses. Teklia précise que malheureusement il n'a pas été possible de les intégrer dans le projet, celles-ci figurant de façon trop différentes sur les documents (horizontalement, verticalement).

Table ronde 2 : transformer l'essai - quelles perspectives pour l'IA dans le traitement et la mise à disposition des grandes typologies d'archives sérielles ?

Est-ce que l'on va reproduire le projet ? Questions à se poser :

- Quelles disponibilités d'archives (déjà) numérisées ?
- Existe-t-il des projets en partenariat avec d'autres structures de recherches ?

L'Ined a identifié quelques types d'archives qui pourraient constituer un nouveau partenariat :

- Les TSA
- Les fiches matricules
- Différentes sources pas forcément numérisées

On s'orienterait vers des projets plus locaux car en dehors des thèmes évoqués, il existe de grandes disparités entre les départements.

On a pu identifier des utilisations de l'IA pour générer des descriptifs de documents.

Il pourrait aussi s'agir de la reconstitution automatique de la généalogie des parcelles à partir du cadastre. C'est 150 ans d'histoire.

Les sites d'archives ont nécessité d'optimiser leurs budgets. D'où la question à se poser : quel est le prochain projet où l'IA permettrait d'utiliser le plein potentiel des ressources ? Quelques exemples :

- AD 23 - délibérations du conseil municipal avec recherches en plein texte
- Projet commun AD 88, 86 et 23 - reconstitution des tables de l'enregistrement pour le moment inexistantes (le modèle serait ensuite réutilisable pour d'autres sites d'AD)

Conclusion

L'IA va "pousser les archivistes dans leurs retranchements".

Ils vont accepter de publier des choses qui peuvent comporter des erreurs.

Intervenants, par ordre d'intervention, lors de la première partie de la journée :

- Marie-Françoise Limon-Bonnet - directrice des Archives nationales
- Bruno Ricard - chef du SIAF
- Christopher Kermorvant - chef de projet Teklia (prestataire technique)
- Mathieu Stoll - SIAF
- Jean-Charles Bédague - SIAF
- Lionel Kesztenbaum - Ined
- Manonmani Restif - SIAF - cheffe de projet FranceArchives
- Raphaël Baumard - AD 02
- Aude Guého - SIAF

Intervenants, par ordre d'intervention, lors des tables rondes de l'après-midi :

- Olivier Guionneau - secrétaire général FFG
- Henri Pinoteau - AD 24
- Nathalie Sage Pranchère - CNRS
- Laurent Guillaume - professeur collège Reims
- Jean-François Moufflet - Archives nationales
- Gaël Chenard - AD 23
- Christopher Kermorvant - chef de projet Teklia (prestataire technique)
- Lionel Kesztenbaum - INED
- Edouard Vasseur - Ecole nationale des Chartres