

Notice d'information sur la protection des données dans le cadre de la recherche scientifique

FEMDB *French Electoral Manifestos Database*

Version 1.0 – Janvier 2026

Les informations recueillies vous concernant vont faire l'objet d'un traitement dans le cadre du projet intitulé du projet piloté par Frédéric Gonthier, enseignant-chercheur à Sciences Po Grenoble – UGA et au laboratoire Pacte (UMR 5194).

Le traitement a été porté au registre tenu par la DPD du CNRS (numéro de traitement : 2-25214) et peut être mis en œuvre. Le responsable de traitement est Laurence Dumoulin, agissant en tant que directrice du laboratoire Pacte (UMR 5194), dont l'adresse est Pacte/IEP - BP 48 - 38040 Grenoble Cedex 9 - France

I. Description du projet et des objectifs

La **FEMDB – French Electoral Manifestos Database** a pour objectif de constituer et de mettre à disposition une base de données inédite regroupant les **professions de foi électorales** produites par les candidats et listes lors des scrutins politiques français depuis 2015. Ces documents programmatiques, élaborés dans le cadre institutionnel de la compétition électorale, permettent d'étudier l'évolution des discours politiques, des registres de revendication et des stratégies de positionnement partisan. Les douze élections suivantes sont couvertes :

Élection	Dates du scrutin	Remarques
Régionales 2015	6 et 13 décembre 2015	Deux tours
Présidentielles 2017	23 avril et 7 mai 2017	Deux tours
Législatives 2017	11 et 18 juin 2017	Deux tours
Européennes 2019	26 mai 2019	Scrutin à un seul tour
Municipales 2020	15 mars et 28 juin 2020	Premier tour avant confinement, second tour reporté à juin pour cause COVID
Législatives 2020 partielles	Variable selon les circonscriptions (principalement septembre 2020)	Scrutins complémentaires dans plusieurs circonscriptions annulées en 2017

Élection	Dates du scrutin	Remarques
Départementales 2021	20 et 27 juin 2021	Initialement prévues en mars, reportées en juin à cause du COVID
Régionales 2021	20 et 27 juin 2021	Tenues en même temps que les départementales
Présidentielles 2022	10 et 24 avril 2022	Deux tours
Législatives 2022	12 et 19 juin 2022	Deux tours
Européennes 2024	9 juin 2024	Scrutin à un seul tour
Législatives 2024	30 juin et 7 juillet 2024	Élections législatives anticipées. Professions de foi des Français.es de l'étranger non incluses

La FEMDB s'inscrit dans une démarche interdisciplinaire en science politique, humanités numériques et traitement automatique des langues. Elle fournit une ressource structurée permettant notamment :

- d'examiner les **programmes électoraux** à grande échelle ;
- d'analyser **l'évolution des priorités politiques et des clivages partisans** dans le temps ;
- de **croiser les professions de foi avec des métadonnées électorales fiables** pour explorer les dynamiques partisans, territoriales et sociodémographiques.

La base légale du traitement repose sur **l'exécution d'une mission d'intérêt public à des fins de recherche scientifique** (article 6, 1.e du RGPD).

II. Nature et origine des données

1. Sources des données

Les professions de foi électorales sont collectées à partir de dépôts documentaires sur les sources ouvertes suivantes :

- Pour les Présidentielles de 2017 : <https://archives.aubervilliers.fr/Professions-de-foi-des-candidats-aux-presidentielles>
- Pour toutes les autres élections, [GitHub](https://programme-candidats.interieur.gouv.fr/) du collectif Regards Citoyens pour la collecte des professions de foi publiées sur <https://programme-candidats.interieur.gouv.fr/>.

Les métadonnées sont collectées à partir de la source ouverte suivante :

- Données officielles des élections agrégées publiées par le **Ministère de l'Intérieur** (<https://www.data.gouv.fr/datasets/donnees-des-elections-agregees/>)

2. Types de données traitées

Les variables de la FEMDB couvrent deux dimensions :

a) Données descriptives sur les candidats et listes

- Nom, prénom, sexe, année de naissance de la tête de liste
- Profession déclarée de la tête de liste
- Nuance politique officielle de la liste (Ministère de l'Intérieur)
- Investiture par un parti ou une coalition
- Codes géographiques (département, circonscription)

b) Données textuelles et issues des documents

- Texte intégral des professions de foi (fichiers PDF numérisés et OCRisés si scannés)
- Méthode d'extraction (parseable / OCR)
- Texte corrigé des artefacts OCR
- Métadonnées sur le scrutin (type, année, tour, département et circonscription si pertinent)

Ces données proviennent **exclusivement** de documents publics et de fichiers électoraux officiels. Aucune donnée sensible supplémentaire n'est collectée.

III. Pipeline de traitement des données

Afin d'assurer la qualité et la traçabilité de la base, un pipeline automatisé a été mis en œuvre :

1. **Analyse des fichiers** : distinction entre PDF numérisés directement lisibles et PDF scannés.
2. **OCRisation** des fichiers scannés via ABBYY FineReader (Huma-Num).
3. **Appariement automatisé** entre professions de foi et fichiers de candidatures du Ministère, grâce à des techniques d'appariement flou (RapidFuzz).
4. **Correction des artefacts OCR** via un modèle de langage avancé (DeepSeek-R1, 32B paramètres), exécuté localement ou sur le cluster GRICAD.
5. **Apurement et corrections non automatisés complémentaires** via la liste des occurrences
6. **Structuration tabulaire** des données : une ligne = une profession de foi, avec métadonnées harmonisées.

L'ensemble des scripts seront mis à disposition pour garantir la **reproductibilité** des traitements.

IV. Participation

Votre participation à ce projet est entièrement libre et volontaire.

V. Sécurité, conservation et partage

- **Sécurité** : les données sont stockées sur des serveurs sécurisés de l'UGA, avec accès restreint aux membres autorisés du projet.
- **Conservation** : les données sont conservées **20 ans** en base active, puis archivées.
- **Transfert** : aucun transfert hors de l'Union européenne.
- **Partage scientifique** : la base sera mise à disposition sous licence ouverte (CC-BY-SA), dans le respect des droits liés aux documents électoraux.

VI. Droits des personnes concernées

Conformément au RGPD, vous pouvez :

- Accéder aux données vous concernant, en obtenir copie ;
- Vous opposer au traitement de ces données ;
- Demander leur rectification en cas d'erreur ;
- Demander la limitation du traitement.

Le droit à l'effacement et à la portabilité peut être limité, car le traitement poursuit une mission d'intérêt public (articles 17 et 20 du RGPD).

Vous pouvez exercer ces droits en vous adressant à :

Frédéric Gonthier | Pacte – UMR 5194 ; Pacte / IEP – BP 48, 38040 Grenoble Cedex 9, France

Vous pouvez également contacter la Déléguée à la Protection des Données (DPD) du CNRS à l'adresse suivante :

CNRS – Service protection des données, 2 rue Jean Zay, 54519 Vandoeuvre-lès-Nancy
dpd.demandes@cnrs.fr

Si vous estimez que vos droits Informatique et Libertés ne sont pas respectés, vous avez la possibilité d'introduire une réclamation en ligne auprès de la **CNIL, 3 Place de Fontenoy, TSA 80715 – 75334 Paris Cedex 07 (<https://www.cnil.fr/>)**.