

VigiCité

Outil citoyen de transparence démocratique

White paper méthodologique

Croisement de données publiques françaises pour la mesure objective de la transparence des responsables publics et des marchés publics.

Version : 1.0 — mai 2026

Licence : AGPL-3.0-only

URL : <https://vigicite.org>

Code source : disponible sur demande (Tor / clearnet)

Statut : prépublication, ouverte au peer review citoyen

Ce document est publié sans auteur nommé. Le projet est opéré de façon anonyme pour des raisons légales et éthiques liées à la nature des données publiques croisées. La méthodologie présentée ici est intégralement reproductible : toutes les sources de données sont libres d'accès, le code est open source.

Résumé

VigiCité est un outil open source qui croise les données publiques françaises (déclarations HATVP, marchés publics DECP, registres d'entreprises, votes parlementaires, comptes de campagne, registre des lobbys, Journal Officiel) pour produire des fiches factuelles sur 6 696 responsables publics. Quatre dimensions de score sont calculées sur chaque élu : cohérence patrimoniale, concentration des marchés, densité de réseau, transparence déclarative. Un score composite agrège ces dimensions. Chaque score est accompagné d'un intervalle de confiance et d'une liste exhaustive de sources reproductibles. L'outil n'émet aucun jugement : il rend visibles des faits que les citoyens vérifient eux-mêmes.

Mots-clés : transparence démocratique, open data, marchés publics, HATVP, civic tech, données publiques, conflits d'intérêts, France.

1. Introduction

1.1 Contexte

Depuis la loi du 11 octobre 2013 relative à la transparence de la vie publique, la France dispose d'un dispositif déclaratif unique en Europe : tout responsable public (élu, ministre, haut fonctionnaire) doit publier ses déclarations d'intérêts et de situation patrimoniale auprès de la Haute Autorité pour la Transparence de la Vie Publique (HATVP). Parallèlement, depuis 2018, les marchés publics supérieurs à 25 000 euros HT sont publiés en données ouvertes (Données Essentielles de la Commande Publique, DECP). Combinées au registre national des entreprises et au répertoire national des associations, ces sources permettent en théorie de croiser intérêts déclarés et flux financiers publics.

En pratique, ces données restent silotées, dispersées et peu exploitées par le grand public. VigiCité comble ce manque en agrégeant, croisant et rendant lisibles ces sources pour tout citoyen.

1.2 Principe directeur

VigiCité ne juge pas, il rend visible. Le mot « corruption » n'apparaît dans aucun output. Toutes les conclusions sont laissées à l'appréciation du lecteur. Chaque chiffre est accompagné de sa source URL, et chaque élu dispose d'un droit de réponse intégré à sa fiche.

1.3 Symétrie démocratique

L'État utilise le croisement de données pour surveiller les citoyens (fichiers, scoring social, profilage fiscal). VigiCité applique le principe de symétrie : ce que l'État fait avec nos données, les citoyens peuvent le faire avec celles des responsables publics, *dès lors qu'elles sont publiques et légalement accessibles.*

2. Sources de données

Toutes les sources sont en libre accès, légalement publiables, et rafraîchies automatiquement. Aucune donnée privée, scrapée, ou obtenue par moyens détournés n'est utilisée.

Source	Provider	Volume	Fréquence
Élus HATVP	hatvp.fr CSV	6 696	Quotidienne
Déclarations HATVP	hatvp.fr CSV	14 643	Quotidienne
Participations financières	hatvp.fr XML	36 203	Quotidienne
Marchés publics DECP	data.gouv.fr	116 607	Quotidienne
Entreprises (SIREN)	recherche-entreprises.api.gouv.fr	26 842	À la demande

Associations RNA	media.interieur.gouv.fr	1 946 202	Hebdomadaire
Lobbys	hatvp.fr registre	3 950	Quotidienne
Budgets collectivités	data.ofgl.fr	69 934	Mensuelle
Votes Assemblée Nationale	data.assemblee-nationale.fr	952 231	Hebdomadaire
Comptes de campagne	CNCCFP via data.gouv	6 292	Annuelle
Journal Officiel	echanges.dila.gouv.fr	30 338	Quotidienne

Tableau 1 — Sources utilisées par VigiCité (mai 2026). Volumes indicatifs.

3. Moteur de croisement

Cinq types de relations sont détectés et stockés dans une table unifiée *relations* (*source_type*, *source_id*, *target_type*, *target_id*, *relation_type*, *source_url*). Toutes les relations conservent l'URL de leur source pour traçabilité.

3.1 Marché → entreprise

Lien direct par SIREN attributaire. Aucune ambiguïté possible. Couvre 100 % des marchés DECP dont l'attributaire dispose d'un SIREN valide (filtrage des entrées vides).

3.2 Élu → entreprise (département)

Relation agrégée : une entreprise qui obtient au moins deux marchés publics dans le département d'un élu est liée à cet élu. Cette relation est *contextuelle, pas accusatoire* : elle indique une coïncidence géographique entre mandat et activité économique.

3.3 Élu → entreprise (HATVP)

Lien fort : l'élu déclare lui-même détenir des parts ou un mandat social dans l'entreprise. Le matching s'effectue par similarité fuzzy (*rapidfuzz*, seuil 88 %) entre le nom déclaré et la raison sociale SIRENE, validé par le SIREN reconstruit. C'est le signal le plus directement intéressant : déclaration officielle du conflit d'intérêts potentiel par l'élu lui-même.

3.4 Marché → association

Lien par SIRET ou SIREN reconstruit (9 premiers chiffres) entre la table associations (RNA) et l'attributaire DECP. Détecte les associations recevant des marchés publics.

3.5 Élu → association (similarité nom)

Approche pragmatique en attendant l'enrichissement RNA des dirigeants (non gratuit). Une similarité partial-ratio supérieure ou égale à 92 % entre *prénom + nom* de l'élu et le titre d'une association du même département crée une relation marquée comme *presidence_probable*. Les faux positifs sont possibles (homonymie) ; le citoyen vérifie.

4. Scoring multi-dimensionnel

Quatre dimensions sont calculées pour chaque élu, normalisées sur [0, 100] où 100 est l'optimum théorique. Un cinquième score, composite, agrège les quatre dimensions par moyenne pondérée.

4.1 Cohérence patrimoine

Mesure l'écart entre patrimoine déclaré HATVP et les indicateurs extérieurs disponibles (participations déclarées, mandats sociaux). Le score décroît quand des éléments sont déclarés incomplètement ou incohérents entre les déclarations successives.

```
score = 100 × max(0, 1 - Σ ecart_normalises / N_elements)
```

4.2 Concentration des marchés (HHI)

Indice de Herfindahl-Hirschman (HHI) appliqué à la distribution des marchés publics par attributaire dans le département. HHI faible = diversité d'attributaires (sain). HHI élevé = concentration sur un petit nombre d'entreprises (à interpréter).

```
HHI = Σ (part_i)2 × 10000, normalise : score = 100 × (1 - HHI / 10000)
```

4.3 Densité du réseau

Centralité de l'élu dans le graphe biparti élus / entreprises / marchés (NetworkX). Score décroissant avec le nombre et la force des liens. Un élu sans relation détectée a un score de 100 ; un élu présent dans 50 + croisements descend rapidement.

4.4 Transparence déclarative

Complétude des déclarations HATVP : présence d'une déclaration d'intérêts, d'une déclaration patrimoniale, mention de l'élu au Journal Officiel pour ses prises de fonction, présence ou absence dans le registre des lobbys.

4.5 Score composite

Moyenne pondérée :

```
composite = 0.30 × coherence + 0.20 × concentration + 0.25 × reseau + 0.25 ×  
transparence
```

Ces pondérations sont volontairement transparentes et discutables. Le code source les centralise dans *src/scoring/thresholds.py* et tout fork peut les recalibrer.

4.6 Intervalles de confiance

Chaque score est accompagné d'une marge de confiance dérivée du nombre de sources disponibles, du nombre de matches fuzzy, et des lacunes documentées. Un score de 60 avec marge ± 8 ne se lit pas comme un score de 60 avec marge ± 2 .

5. Limites et biais documentés

5.1 Biais de complétude HATVP

Les élus locaux (maires de communes < 20 000 habitants, conseillers municipaux) ne sont pas tenus de déclarer leurs intérêts auprès de la HATVP. Notre couverture est donc plus dense pour les responsables nationaux et régionaux.

5.2 Biais DECP

Les marchés inférieurs à 25 000 euros HT ne sont pas publiés. Certains montants DECP sont des valeurs sentinelles ; un filtrage est appliqué pour exclure les montants $> 1 \times 10^{10}$ dans les agrégations publiques. Le matching SIREN n'est pas toujours fait par les acheteurs publics : les entrées vides représentent ~ 8 % du flux.

5.3 Biais d'homonymie

Le matching fuzzy nom-prénom génère des faux positifs lorsque plusieurs personnes portent le même nom. Le seuil (88 % participations, 92 % associations) est calibré pour minimiser les faux positifs au prix d'une sous-détection.

5.4 Biais temporel

Les déclarations HATVP, les marchés DECP et les budgets OFGL n'ont pas le même rythme de publication. Une fiche élu reflète l'état des sources à la date de dernier rafraîchissement, indiqué en pied de page de chaque fiche.

6. Reproductibilité

Toute la pipeline VigiCité est reproductible sur n'importe quelle machine respectant les prérequis : Python 3.12, libs *pyproject.toml*, accès Internet vers les sources officielles, 4 Go disque, 1.4 Go RAM.

Le code source est publié sous AGPL-3.0. Les forks sont encouragés (Belgique, Suisse, Québec : structures HATVP-équivalentes). Un dataset hebdomadaire CSV/JSON anonymisé (hash SHA-256 des identifiants) est mis à disposition pour les chercheurs.

7. Droit de réponse et correction

Chaque fiche élu intègre un formulaire de droit de réponse. L'élu, ou un mandataire désigné, peut demander la correction d'un fait erroné, l'ajout d'une précision contextuelle, ou la suppression d'une donnée qu'il estime à tort attribuée. Les corrections sont appliquées sous 7 jours après vérification documentaire.

Le droit de réponse n'est pas un mécanisme de censure : si un fait est sourcé par une donnée publique officielle, il sera maintenu (éventuellement annoté par la réponse de l'élu).

8. Conclusion et appel

VigiCité existe parce que les outils de transparence démocratique doivent émerger de la société civile, pas seulement des institutions qu'ils observent. Le projet est ouvert : associations, journalistes, chercheurs, civic tech sont invités à utiliser, critiquer, forker, améliorer.

Trois pistes d'évolution sont à l'étude : enrichissement RNA dirigeants (API DJEPVA), intégration des comptes annuels d'entreprises (Greffes), forks territoriaux Belgique/Suisse/Québec.

9. Références

- [1] Loi n° 2013-907 du 11 octobre 2013 relative à la transparence de la vie publique.
- [2] HATVP, *Référentiel open data*, <https://www.hatvp.fr/livraison/opendata/>
- [3] Décret n° 2018-1075 du 3 décembre 2018 portant publication des données essentielles de la commande publique.
- [4] Code de bonnes pratiques pour la statistique européenne, Eurostat.
- [5] OECD, *Open, Useful, Reusable Government Data (OURdata) Index*, 2023.
- [6] Anticor, *Rapport annuel sur la transparence publique*.
- [7] Transparency International France, *Indice de perception de la corruption*.
- [8] Regards Citoyens, *Méthodologie nosdeputes.fr*, <https://github.com/regardscitoyens>

Citation suggérée : VigiCité, *White paper méthodologique 1.0*, mai 2026, <https://vigicite.org>, licence AGPL-3.0.