

# Le thème de la biodiversité dans les contributions du *Grand Débat national*

**Mots clés :** biodiversité, contributions écrites, informations localisées, modèle de langue, textométrie, géomatique

## Contexte

En janvier 2019, le Gouvernement français a mis en place le Grand Débat national<sup>1</sup> (GDN) afin de recueillir les doléances de l'ensemble de la population, comme une des réponses au mouvement des Gilets jaunes. Deux modalités de contributions ont été proposées : des cahiers papier ouverts en mairie d'expression libre sans consignes de thème, nommés les Cahiers citoyens, et une plateforme en ligne contenant des questions de quatre thématiques. Les Cahiers ont été numérisés puis transcrits sous la supervision de la Bibliothèque nationale de France (BnF) et les réponses aux questions de la plateforme numérique ont été sauvegardées telles qu'écrites numériquement.

Le stage s'insère dans les travaux du groupe de recherche Geosem (EHESS et IGN) sur l'analyse sémantique et géographique du GDN. Des stages ont déjà été réalisés en 2023, 2024 et 2025 sur les thèmes de l'énergie [4], notamment éolienne [5], et de la densification périurbaine [3]. Le thème de ce stage est la biodiversité, c'est-à-dire les espèces animales et végétales, et les milieux de vie dans lesquels elles évoluent.

## Sujet

L'enjeu du stage est d'identifier les préoccupations et les opinions concernant la biodiversité tels que formulés dans les contributions du GDN, et de les interroger dans un contexte de transition écologique et de recherche de conciliations entre développement économique et préservation des écosystèmes ([1], [2]). Les missions prévues sont les suivantes :

- constitution d'un sous-corpus biodiversité : il s'agira d'extraire dans les contributions les phrases évoquant la biodiversité. Des ressources pour construire un vocabulaire pourront être utilisées, telles que [6] ;
- analyse des aspects constitutifs du thème de la biodiversité : l'hypothèse, à confirmer ou infirmer, est que les phrases du sous-corpus sont réparties selon des thématiques telles que les changements des populations et des milieux de vie, les mesures de préservation et les réglementations, les interactions avec des activités économiques et des projets d'aménagement du territoire, etc. ;
- caractérisation des regards que portent les contributeurs sur le thème : éventuelles prépondérances de préoccupations, polarités d'opinions, émotions ;
- étude de la géographie des informations extraites : par la localisation des contributions à la commune (code INSEE dans les Cahiers citoyens et code postal dans la plateforme numérique) et par les entités nommées spatiales dans les contributions.

Le stage utilisera des méthodes de traitement automatique des langues (modèles de langue et encodages de mots et de phrases correspondants, outils de classification), et de géomatique (méthodes d'analyse spatiale à partir de données complémentaires de description des territoires). Les productions attendues sont les suivantes :

- une méthode d'extraction des phrases sur le thème de la biodiversité et le sous-corpus correspondant ;

---

<sup>1</sup> Site officiel : <https://granddebat.fr/>

- des traitements permettant d’identifier les différents aspects constitutifs du thème et les regards portés ;
- des méthodes d’analyse spatiale, et les données et ressources permettant de les mettre en œuvre ;
- un rapport de stage ou un document technique détaillant les réalisations du stage.

Le code produit devra être rendu disponible et commenté dans le but d’en faciliter le partage et la réutilisation.

### **Références**

- [1] Archana Bele et Ujwala Chakradeo. “Public perception of biodiversity: a literature review of its role in urban green spaces”. In: *Journal of Landscape Ecology* 14.2 (2021), p. 1-28, doi: 10.2478/jlecol-2021-0008
- [2] Victoria Campbell-Arvai. “Engaging urban nature: improving our understanding of public perceptions of the role of biodiversity in cities”. In: *Urban Ecosystems* 22 (2019), p. 409-423, doi: 10.1007/s11252-018-0821-3
- [3] Baptiste Descamps. “Étude de la perception de la densification périurbaine dans des consultations citoyennes”. Mémoire de master, St-Mandé : Master 2 Science des données pour la décision publique, École nationale de la statistique et de l’analyse de l’information, juillet 2024.
- [4] Inès Lebrun Moro. “Le thème de l’énergie dans le Grand Débat national”. Mémoire de master, St-Mandé : Master 2 Science des données pour la décision publique, École nationale de la statistique et de l’analyse de l’information, juillet 2025.
- [5] Matilde Monnier. “L’analyse spatiale des Cahiers Citoyens appliquée au thème de l’écologie”. Mémoire de master, St-Mandé : Master 2 Métiers de la coopération et du développement en Amérique Latine, Université Sorbonne Nouvelle, juillet 2023.
- [6] Patrick Triplet. “Dictionnaire encyclopédique de la diversité biologique et de la conservation de la nature”. In : 11<sup>ème</sup> édition (2025), 1424 p.

### **Formation requise**

Ce stage s'adresse aux étudiant.e.s de master 2 en informatique/science des données ou en traitement automatique des langues (TAL) avec une formation suffisante pour l’utilisation autonome d’un langage de programmation (de préférence Python et R) et d’outils de TAL (outils fondés sur l’apprentissage, modèles de langue, classifieurs, si possible outils statistiques de lexicométrie).

### **Lieu du stage**

Le stage se déroulera au LASTIG sur le site de l'école Géodata Paris à Champs-sur-Marne, 6-8 avenue Blaise Pascal (accès : RER A - *Noisy-Champs*).

Le LASTIG (laboratoire en sciences et technologies de l'information géographique pour la ville intelligente et les territoires durables) est une unité mixte de recherche sous la tutelle de l'Université Gustave Eiffel et de l'IGN (Institut national de l'information géographique et forestière). Il mène des recherches finalisées en sciences et techniques de l’information géographique couvrant l’ensemble du cycle de vie de la donnée géographique, de son acquisition à sa visualisation, en passant par sa modélisation, son intégration et son analyse.

### **Durée**

La durée du stage prévue est de 5 mois, avec un début possible à partir de mars 2026.

## **Gratification**

La gratification du stage est calculée à partir du taux horaire qui suit la législation en vigueur : <https://www.service-public.gouv.fr/particuliers/vosdroits/F32131>.

## **Encadrement du stage**

Catherine Domingues, chercheure HDR en TAL, catherine.domingues@ign.fr\\

Laurence Jolivet, chercheure en géomatique, laurence.jolivet@ign.fr\\

## **Pour candidater**

Des entretiens seront organisés à partir de novembre 2025. Le dossier de candidature est à envoyer aux encadrantes et devra contenir les documents suivants : CV, derniers relevés de notes (M1, et premier semestre de M2 si possible), description des enseignements suivis (un lien vers le site internet de la formation est le bienvenu), dernier mémoire ou rapport de stage, lettre de motivation.