



**HAL**  
open science

# Topomine : une application web d'exploration itérative de la toponymie française

Eric Mermet, Eric Grosso

► **To cite this version:**

Eric Mermet, Eric Grosso. Topomine : une application web d'exploration itérative de la toponymie française. 2023. halshs-03931476

**HAL Id: halshs-03931476**

**<https://shs.hal.science/halshs-03931476v1>**

Preprint submitted on 9 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

# Topomine : une application web d'exploration itérative de la toponymie française

Eric Mermet<sup>1</sup>, Eric Grosso<sup>2</sup>

<sup>1</sup> CNRS / EHESS, CAMS - ISC-PIF,  
@ [eric.mermet@ehess.fr](mailto:eric.mermet@ehess.fr),

www : [cams.ehess.fr/eric-mermet](http://cams.ehess.fr/eric-mermet)

<sup>2</sup> Mapping Edinburgh

**Résumé** Topomine est une application web permettant d'explorer et de visualiser des données relatives à la toponymie en France. La toponymie est porteuse de sens multiples sur le temps long et parfois cachés dans l'évolution des langues et dialectes français. Agrégeant plusieurs bases de données issues de différents producteurs, l'outil Topomine allie une recherche avancée par mots-clés qui permet de faire émerger et de révéler des logiques spatiales sous-jacentes à certains phénomènes d'études en histoire, en géographie, en études environnementales, en littérature, en études du genre ou encore en généalogie.

**Keywords:** Toponymie, Bases de données toponymiques, Cartographie, Exploration de données

**Abstract.** Topomine is a Web application to explore and visualise data related to toponymy in France. Toponymy carries multiple meanings over time, sometimes hidden in the evolution of French languages and dialects. By combining several databases from different producers, Topomine combines an advanced keyword search that makes it possible to reveal the spatial logic underlying certain phenomena in fields such as history, geography, environmental studies, literature, gender studies or genealogy.

**Keywords:** Toponymy, Toponymic databases, Cartography, Data mining

## 1 Introduction

Cet article présente l'application web Topomine développée dans le but de pouvoir interroger, visualiser et explorer des données relatives à la toponymie française. La toponymie est porteuse de sens multiples sur le temps long et parfois cachés dans l'évolution des langues et dialectes français : noms de familles, accidents topographiques, noms de plantes,

noms de métiers, etc. [Ros48]. Hydronymie, odonymie<sup>3</sup>, oronymie<sup>4</sup>, synphytotoponymie[ALGM14]<sup>5</sup>, choronymie<sup>6</sup>, microtoponymie, etc. sont autant de facettes de la toponymie qu’il est possible d’exploiter dans une recherche ciblée.

Dès les années 1930, il existe un regain pour l’analyse de la toponymie qui est matérialisé par plusieurs congrès internationaux [noa38] [Éb48] [Dau49] et qui illustrent déjà l’intérêt grandissant pour l’analyse fine de noms de lieux dans des disciplines telles que la géographie, l’archéologie et l’histoire [Dau60].

Plusieurs travaux récents ont déjà été entrepris au sujet de l’approche d’analyses toponymiques fines, informatisées ou non, en abordant par exemple la thématique du changement climatique au travers de la répartition ancienne du hêtre à travers ses traces toponymiques [TTM+15], ou celle de l’identification de sites gallo-romains alpins [Man15]. Dans le contexte d’autres travaux s’intéressant davantage à des aspects linguistiques, citons par exemple les analyses sur la nature des préfixes et des suffixes dans la toponymie [Vac99] [Vac04] ou encore l’analyse des génériques – c’est-à-dire la qualifications des voies –, dans l’odonymie [BPP21].

Il est à souligner que peu d’outils d’exploration de données toponymiques sont d’ores et déjà accessibles en ligne, à l’exception par exemple de l’application DicoTopo<sup>7</sup>, développée par l’École Nationale des Chartes, qui met à disposition une partie des dictionnaires topographiques de la France<sup>8</sup> établis à partir de 1859. De fait, toute nouvelle étude, analyse ou recherche toponymique nécessite généralement la mise en place d’outils et de méthodes ad hoc. Cela implique nécessairement des barrières de faisabilité, synonyme d’un important investissement en temps, qui peuvent avoir comme conséquence la non utilisation de l’apport toponymique alors que celle-ci pourrait apporter un éclairage ou un point de vue différent intéressant, contribuant ainsi à l’enrichissement des résultats de l’étude ou de la recherche effectuée.

---

3. Étude des toponymes des voies de communication

4. Étude des toponymes des reliefs

5. Étude des toponymes basés sur le nom des plantes

6. Étude des toponymes ayant une caractéristique géographique physique ou une particularité environnementale

7. <https://dicotopo.cths.fr/>

8. Voir par exemple pour le département du Gard : <https://gallica.bnf.fr/ark:/12148/bpt6k39295t>

Aujourd'hui, les bases de données relatives à la toponymie sont éparées et maintenues par différents producteurs de données thématiques, nationaux ou locaux : l'IGN<sup>9</sup>, le Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique, le SANDRE<sup>10</sup>, l'École des Chartes, l'EHESS<sup>11</sup>. La décentralisation et l'hétérogénéité de ces bases de données ne permet actuellement pas leur interrogation croisée de manière aisée, et de fait, leur comparaison et leur visualisation, en particulier cartographique. Bien entendu, cette interrogation est possible au prix d'une intégration parfois laborieuse et donc chronophage, puisque requérant des compétences multiples en matière de SIG, de programmation et de gestion de bases de données [GGGS16].

C'est dans l'idée de pallier à ces multiples problématiques que l'application Topomine a été initialement pensée et conçue.

Le fait que Topomine agisse en tant qu'agrégateur de données relatives à la toponymie française dans un contexte applicatif, que nous visons à terme comme étant cohérent et uniforme, constitue une première valeur ajoutée de l'application. Cette agrégation a uniquement été rendue possible grâce à la libération des données sous forme de données ouvertes (données sous licences libres) et des principes FAIR<sup>12</sup> au cours de la dernière décennie. Les bases de données mises à disposition à travers l'application Topomine peuvent ainsi être considérées comme "une mine de données" explorable qu'il est possible de "miner" pour y trouver l'information toponymique recherchée, d'où l'étymologie de Topomine.

La principale et seconde valeur ajoutée de Topomine réside dans la possibilité de facilement effectuer des recherches toponymiques avancées à partir d'un large éventail de bases de données toponymiques en langue française, et dont les résultats obtenus sont dynamiquement représentés cartographiquement. Ces représentations permettent d'appréhender une quantité importante de toponymes issus de différentes sources, et ainsi de révéler plus facilement de possibles logiques sémantiques et spatiales sous-jacentes, existantes entre plusieurs lieux considérés ou encore rela-

---

9. Institut national de l'information géographique et forestière

10. Service d'Administration Nationale des Données et Référentiels sur l'Eau

11. École des Hautes Études en Sciences Sociales par le biais du Laboratoire de Démographie Historique et du Centre de Recherche Historique

12. Les données dites FAIR se doivent d'être "Findable" – Faciles à trouver –, "Accessibles", "Interopérables", et "Réutilisables". Topomine est donc par définition une application favorisant ces principes.

tives à un phénomène d'étude donné. L'ensemble de ces fonctionnalités est librement ouvert à tous les utilisateurs, sans restriction d'accès. Elles sont de plus aisément accessibles, Topomine étant proposé comme une application web sur étagère qui ne requiert aucune installation logicielle de la part des utilisateurs, si ce n'est celle d'un simple navigateur web.

Notons que deux études ont déjà bénéficié en amont de l'utilisation de Topomine. La première porte sur l'analyse de l'approvisionnement de Paris en poisson par les chasse-marée du XIII<sup>ème</sup> siècle au XVIII<sup>ème</sup> siècle [Rob18]. La seconde, quant à elle, offre un nouvel éclairage sur la toponymie bretonne [Del21]. Il est également à souligner que d'autres études et recherches sont actuellement en cours de publication <sup>13</sup>.

Sans offrir les mêmes fonctionnalités ni la même finesse historique que celle proposée par exemple dans DicoTopo, Topomine entend permettre une interrogation sémantique et spatiale reposant sur d'importantes et riches bases de données, majoritairement contemporaines. De plus, de par le nombre actuel d'objets que Topomine permet d'interroger et qui dépasse actuellement les 13.6 millions, les résultats offerts bénéficient tant d'une homogénéité sur l'ensemble du territoire que d'une relative exhaustivité.

Nous présentons dans un premier temps les bases de données disponibles dans l'application Topomine ainsi que leurs caractéristiques, puis, dans un second temps, nous introduisons les possibilités offertes par Topomine, en particulier les différentes fonctionnalités de recherche, et enfin, nous concluons en montrant les fonctionnalités avancées de l'application (notamment celles offertes par l'API), tant actuelles que celles à venir.

## 2 Bases de données intégrées à Topomine

Aujourd'hui, la principale difficulté dans la construction d'une application telle que Topomine réside dans l'agrégation de différentes bases de données, celles-ci étant le plus souvent hétérogènes entre elles. Cette hétérogénéité s'explique globalement par des productions indépendantes de données, réalisées parallèlement par différentes institutions ayant généralement des missions et des thématiques abordées différentes. Dans les cas

---

13. Voir par exemple la communication suivante d'Anaïs Beaubois : [https://www.youtube.com/watch?v=U\\_8cdhCOMVA](https://www.youtube.com/watch?v=U_8cdhCOMVA)

où ces thématiques demeureraient sensiblement similaires, les choix historiques, individuels et particuliers des modèles de données peuvent quasiment à eux-seuls expliquer les raisons de ces hétérogénéités constatées.

Toutefois, notons que ces hétérogénéités constituent également un atout majeur quant à l’exploration et l’interrogation de ces bases. En effet, certaines similarités sémantiques et/ou proximités spatiales contenues intrinsèquement dans ces bases offrent d’avantageuses possibilités d’analyse, car permettant d’accéder à des résultats d’une plus grande richesse, révélant ainsi les nuances liées, par exemple, aux légères différences de graphies ou de variations de positions géométriques. Pour cette raison, ces bases de données peuvent ainsi être considérées comme étant complémentaires et c’est pourquoi le choix est fait dans Topomine de conserver l’intégralité des objets présents initialement dans chacune des bases.

Topomine permet ainsi d’accéder simplement et en quelques clics à la majorité des bases toponymiques en langue française. Ces dernières sont intégrées dans une base de données spatiales PostgreSQL<sup>14</sup> / PostGIS<sup>15</sup>. Cette solution offre l’avantage de centraliser l’ensemble des données au sein d’une unique base de données de manière homogène, d’effectuer des mises à jour régulières dès l’instant où les jeux de données utilisés sont eux-mêmes mis à jour, et enfin d’ajouter si possible à l’avenir de nouveaux jeux de données.

À ce jour<sup>16</sup>, Topomine intègre 5 différentes bases de données :

- un assemblage des noms de lieux présents dans plusieurs thèmes de l’ensemble des départements de la BD TOPO v3 de l’IGN<sup>17</sup>, à l’exception pour l’instant des CROM-TOM,
- un assemblage des voies nommées issues également de la même BD TOPO v3 de l’IGN,
- la base de données Carthage du SANDRE<sup>18</sup> pour l’hydronymie,

---

14. <https://www.postgresql.org/>

15. <https://postgis.net/>

16. Janvier 2023

17. <https://geoservices.ign.fr/documentation/donnees/vecteur/bdtopo>

18. <https://www.sandre.eaufrance.fr/concept/base-de-donnees-sur-la-cartographie-thematique-des-agences-de-leau>

- la base de données FANTOIR<sup>19</sup> qui ne contient initialement pas d'éléments de géométrie, mais un code de commune INSEE, que nous avons joint avec la base de données GeoFla<sup>20</sup>,
- la base de données de l'EHESS dite base de données démographique Cassini : Des chefs-lieux de Cassini aux communes de France (1756-1999)[CVDM17].

L'ensemble des objets constitue une base interrogeable de plus de 13.6 millions d'objets, dont le détail est fourni dans le tableau suivant :

Base de données	Nombre d'objets	Dernière mise à jour
Toponymes / BD NYME	2.397.281	Juillet 2021
Odonymes	2.784.805	Juillet 2021
Hydronymes	179.905	Février 2019
FANTOIR (Communes)	36.836	Novembre 2022
FANTOIR (Voies et Lieux-dits)	8.197.957	Novembre 2022
Cassini	43.792	Avril 2021

Détaillons à présent les caractéristiques des différentes les bases de données en vue de leur intégration dans Topomine ainsi que certains traitements ayant été réalisés afin de les enrichir et de les homogénéiser.

### **Les bases de données noms de lieux ponctuels et voies nommées issues de la version 3 de la BD TOPO de l'IGN**

Avant l'apparition de la version 3 de la BD TOPO (BD TOPO v3) en mars 2019, il existait à l'IGN une base de données thématique dédiée à la toponymie française. Cette base de données, appelée BD NYME, regroupait en plusieurs classes déployée pour la France entière : les communes, les lieux-dits habités et non-habités, les hydronymes, les oronymes, les voies de communication et les voies ferrés ainsi qu'une classe d'inclassables dénommée divers. Le nombre d'objets dépassait alors les 1.5 millions.

Depuis l'apparition de la BD TOPO v3 de l'IGN, les thèmes relatifs à la toponymie y ont été largement intégrés et le produit BD NYME dont la dernière mise à jour datait de 2016 a disparu du catalogue. Ce

19. <https://www.data.gouv.fr/fr/datasets/fichier-fantoir-des-voies-et-lieux-dits/>

20. <https://geoservices.ign.fr/geofla>

nouveau produit IGN intègre des thèmes toponymiques additionnels<sup>21</sup>, en lien notamment avec la modélisation INSPIRE : toponymie relative au bâti, toponymie de l'hydrographie, un grand thème de lieux nommés (détails orographiques, lieux-dits non habités, toponymie lieux nommés, zone d'habitation), toponymie des services et activités, toponymie des transports et enfin une classe d'objets relative à la toponymie des zones réglementées. À ces thèmes s'ajoute une classe des voies nommées, relative donc à l'odonymie. Toutefois, le téléchargement de la BD TOPO v3 fut d'abord limité à une archive compressée par département et surtout les conditions de licence d'utilisation ne permettait pas une mise à disposition libre.

En janvier 2021, l'ouverture d'une grande partie des bases de données de l'IGN en OpenData<sup>22</sup> dont la BD TOPO a permis une utilisation massive des données topographiques françaises. Bien qu'à ces débuts, l'accès aux données se limitait à des archives individuelles basées sur un découpage départementale, il fut très rapidement possible d'automatiser le téléchargement et l'assemblage des thèmes pour la France entière. C'est ce que nous avons mis en place pour tous les thèmes relatifs à la toponymie<sup>23</sup>.

Dès lors, l'assemblage de ces objets nous permet de réaliser deux ensembles :

- une couche d'entités ponctuelles qui s'apparentent à l'ancienne BD NYME tout en l'étendant. Elle agrège toutes les classes de toponymie que nous avons citées. Cet ensemble constitue 2.397.281 objets.
- une couche d'entités ponctuelles relatives aux voies nommées. Celle-ci représente 2.784.805 objets.

### **La base de données BD Topage du SANDRE**

Anciennement dénommée BD Carthage, la base de données élaborée par le SANDRE concerne la construction d'un référentiel qui décrit les entités hydrographiques présentes sur le territoire français. L'objectif de cette base de données réside ainsi dans la constitution d'un référentiel national permettant de localiser des données relatives à l'eau. La BD Topage fait

---

21. Voir à ce propos le document descriptif de contenus : [https://geoservices.ign.fr/sites/default/files/2022-11/DC\\_BDTopo\\_3-2.pdf](https://geoservices.ign.fr/sites/default/files/2022-11/DC_BDTopo_3-2.pdf)

22. Sous licence Etalab

23. Aujourd'hui, il est maintenant possible de télécharger la BD TOPO v3 en version SQL France entière, via 8 fichiers



converger les objets de la BD Carthage avec les objets de la BD TOPO de l'IGN.

Bien qu'un thème présent dans la BD TOPO v3 concerne directement l'hydronymie, nous avons constaté que les hydronymes présents dans la BD Topage complétaient assez bien certains manques de la BD TOPO v3.

Dans la BD Topage, les objets peuvent être référencés avec un ou plusieurs toponymes<sup>24</sup> incluant les possibles variations suivantes : bassin versant, nœud hydrographique, plan d'eau, cours d'eau.

Nous avons donc exploité ces aspects de la BD Topage pour enrichir la base de données de l'application Topomine de 179.905 objets.

### **La base de données Cassini de l'EHESS**

Le laboratoire de démographie historique du Centre de Recherche Historique de l'EHESS travaille depuis plus de 20 ans à l'établissement de bases de données relatives à l'analyse de la démographie en France [Cla00] [MST03].

Les données du site Cassini de l'EHESS<sup>25</sup> couplé à un projet de saisie collaborative et d'appariement automatique ont mené à l'établissement d'une base de données de la démographie des communes en France. Dans cette donnée riche, nous trouvons une donnée toponymique qui donne une profondeur historique qui peut remonter pour certaines communes jusqu'au milieu du XVIIIème siècle en passant parfois par des changements de noms datant de l'an 3 et des noms datant de 1801.

Cette base de données comportent 43.792 objets utilisables.

### **La base de données FANTOIR**

Le fichier FANTOIR<sup>26</sup> (FANTOIR étant l'acronyme de Fichier ANnuaire TOpographique Initialisé Réduit) liste par commune :

- les voies (rues, avenues, ...),
- les lieux-dits (utilisés surtout dans les zones rurales),

---

24. Voir à cet effet le document descriptif : [https://www.sandre.eaufrance.fr/ftp/documents/fr/pre/eth/1/sandre\\_pres\\_eth\\_1.pdf](https://www.sandre.eaufrance.fr/ftp/documents/fr/pre/eth/1/sandre_pres_eth_1.pdf)

25. Site Cassini de l'EHESS : <https://cassini.ehess.fr>

26. <https://www.data.gouv.fr/fr/datasets/fichier-fantoir-des-voies-et-lieux-dits/>

- les ensembles immobiliers (voiries situées dans les copropriétés, les lotissements),
- les pseudo-voies (par exemple canaux ou stations de métro).

Ce fichier produit par le Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique depuis Juin 2013, est mis à disposition sous Licence Ouverte / Open Licence version 2.0<sup>27</sup>, et est régulièrement mis à jour, a minima plusieurs fois chaque année.

Le fichier FANTOIR contient l'ensemble des références topographiques qu'elles soient annulées ou actives. Ce fichier, ne contenant aucune donnée cartographique ou géométrique, a fait l'objet d'un travail préparatoire d'enrichissement spatial. Cet enrichissement est réalisé par jointure en s'appuyant sur les codes INSEE<sup>28</sup> entre le fichier FANTOIR et la couche administrative communale issue des données de la BD TOPO v2 datant de 2009, l'objectif étant double, à savoir maximiser la précision spatiale en s'appuyant sur les données du référentiel topographique IGN, et apparier au mieux les codes INSEE à l'aide de données produites avant que ne se produisent les nombreuses fusions communales, en particulier à partir de 2012. Ainsi, la couche administrative communale issue des données de la BD TOPO v2 de 2009 permet de bénéficier de 36.702 unités communales contre les 34.945 communes françaises au 1er janvier 2023. Les géométries polygonales des communes BD TOPO V2 ont été transformées en objets ponctuels<sup>29</sup>. Enfin, les anciennes communes dont les codes INSEE n'ont pu être apparierés ont fait l'objet d'un appariement manuel à l'aide des données provenant de multiples source, à savoir FANTOIR, INSEE, Wikipédia/Wikidata et OpenStreetMap.

Le fichier FANTOIR utilisé, à savoir celui de Novembre 2022, contient 8.234.901 objets.

---

27. <https://www.etalab.gouv.fr/licence-ouverte-open-licence/>

28. L'Institut national de la statistique et des études économiques publie un code officiel géographique définit pour chaque commune de France.

29. L'opération spatiale utilisée s'appuie sur la méthode géométrique "ST\_PointOnSurface" de PostGIS qui transforme chaque polygone communal en un point contenu au sein de ce polygone

### 3 Présentation de l'application

Ces bases de données éparses et massives peuvent alors ici trouver un cadre commun d'interrogation en ligne. L'application<sup>30</sup> permet une exploration itérative par accumulation de couches de recherche.

L'architecture de l'application repose sur une interface web dite une application cliente qui se charge de l'interaction avec une base de données PostgreSQL qui stocke l'ensemble des bases dans une schéma spécifique à l'application. Afin de permettre une autre utilisation que celle offerte par l'application web Topomine, mais aussi pour rassembler et décloisonner ces données éparses, nous avons mis en place une API<sup>31</sup> qui ouvre des possibilités intéressantes pour qui souhaitent intégrer les résultats des recherches toponymiques dans d'autres applications ou encore pour les exploiter dans des carnets Jupyter<sup>32</sup> sans avoir à passer obligatoirement par l'application web Topomine.

Malgré tout, les fonctionnalités de recherche et d'affichage que nous proposons sont relativement simples et directes. L'interface présente en haut une barre de recherche centrale sur laquelle l'utilisateur peut entrer un texte de recherche. L'action de la touche entrée ou du bouton rechercher va déclencher une recherche et si des résultats sont présents, une couche de données va s'afficher sur la visualisation cartographique et sur la zone de gestion des couches dans la partie gauche de l'application. La recherche est configurable à droite de la barre de recherche de manière à rechercher parmi toutes les bases de données, ou de sélectionner une ou plusieurs des bases disponibles.

Si cette utilisation est simple, la vraie valeur ajoutée de Topomine tient dans l'aspect offert par les possibilités d'interrogation de la base de données. En effet, pour rechercher et visualiser des objets toponymiques, dont le nombre peut être important, nous pouvons avoir recours à une interrogation avancée qui intègre des métacaractères (ou jokers) ou encore

---

30. L'application et sa documentation sont disponibles : <https://paris-timemachine.huma-num.fr/app/topomine/> pour l'application et <https://paris-timemachine.huma-num.fr/app/topomine/doc/> pour sa documentation

31. API : Application Programming Interface pour interface de programmation d'application, disponible à l'adresse <https://paris-timemachine.huma-num.fr/api/topomine/docs>

32. Voir par exemple le démonstrateur Callisto : <https://hnlab.huma-num.fr/blog/2021/05/26/callisto-un-demonstrateur-jupyter/>

l'utilisation d'expressions régulières.

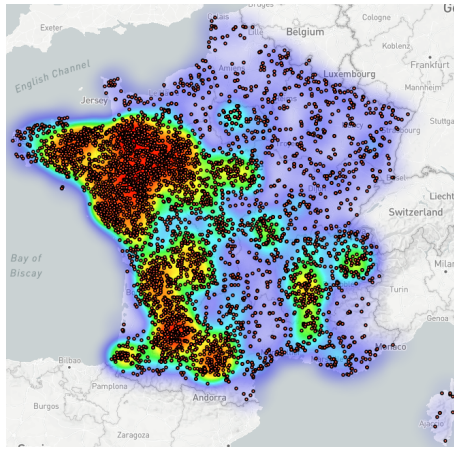
Les **métacaractères** ou **jokers** sont des caractères spécifiques permettant d'effectuer des recherches à partir d'un mot, d'un terme ou d'une expression incomplète. Ils autorisent ainsi de combler une partie manquante contenue au sein d'une chaîne de caractères, et donc par extension, favorisent une interrogation plus souple de corpus linguistiques. Par défaut, Topomine propose deux métacaractères que les utilisateurs peuvent utiliser pour effectuer des recherches avancées.

Le premier métacaractère est le caractère tiret bas (aussi appelé underscore en anglais), symbolisé par `_`. Celui-ci permet de combler un caractère manquant. Par exemple la recherche du terme `“gr_s”` recherche les formes `gras`, `grbs`, `grcs`, etc. et dans notre cas les toponymes `gras`, `gris`, `gros`, sont révélés, cartographiés et assemblés au sein d'une même couche résultante.

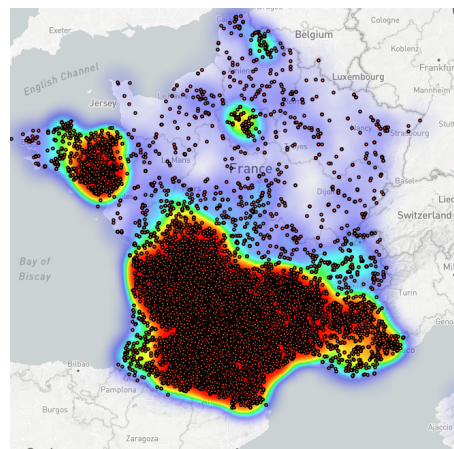
Le second métacaractère est le caractère astérisque, symbolisé par `*`. Ce dernier permet de remplacer un ou plusieurs caractères. La requête `“gr*s”` renvoie ainsi des résultats plus hétérogènes : les mêmes `gras`, `gris` et `gros` (un ensemble de caractères pouvant aussi être un unique caractère), mais aussi par exemple et non exhaustivement `grais`, `grandvilliers`, `graviras`, ou encore des toponymes avec des espaces : `grande lyes`, `grands champs`, etc., voire même des tirets comme par exemple `gros-bois`. En résumé, tous les toponymes qui commencent par `gr` et qui se terminent par un `s` sont considérés comme des résultats valides.

Dès lors, il est possible d'interroger les bases à partir de préfixes ou de suffixes. Par exemple, tous les toponymes ayant pour préfixe `belle-` sont accessibles à l'aide de la requête `“belle*”`, et l'ensemble des toponymes finissant par le suffixe `-ac` le sont à l'aide la requête `“*ac”`. Les résultats cartographiques de ces requêtes sont respectivement représentées dans les figures 1 et 2.

Une utilisation de Topomine plus avancée est possible grâce à l'emploi d'**expressions régulières** - ou `regex`. Celles-ci fournissent un moyen simple de repérer des schémas dans une chaîne de caractères[Fou05] en utilisant d'autres métacaractères et/ou des opérateurs qui, une fois assem-



**Fig. 1.** Cartographie des toponymes avec un préfixe en belle-



**Fig. 2.** Cartographie des toponymes avec un suffixe en -ac

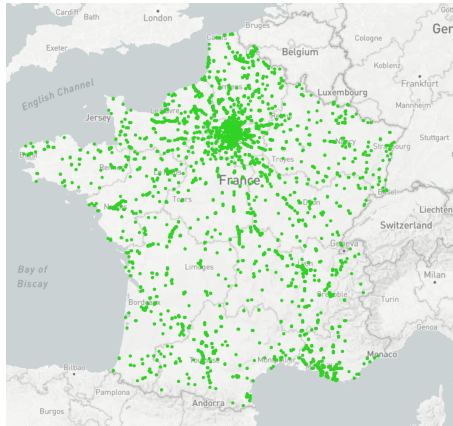
blés<sup>33</sup>, permettent d'aller encore plus loin dans l'interrogation des bases de données toponymiques. Cette fonctionnalité s'active en sélectionnant l'option "Expression Régulière" disponible dans la recherche avancée de Topomine.

Grâce à ces expressions, il devient par exemple possible :

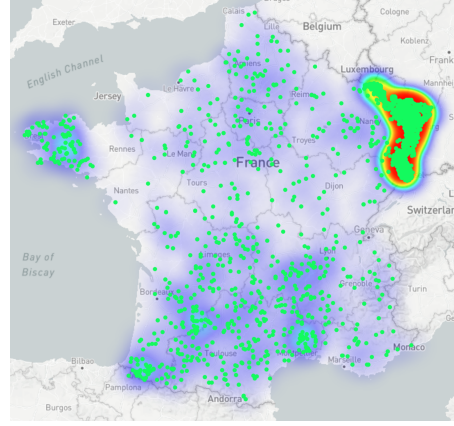
- de rechercher parmi les toponymes un préfixe (l'opérateur `^` – accent circonflexe – indique le début d'une chaîne de caractère) ou un suffixe (l'opérateur `$` –dollar- indique la fin d'une chaîne de caractère. Ainsi les requêtes `^belle` et `ac$` renverront respectivement les toponymes de préfixe en belle- et de suffixe -ac,
- d'utiliser des opérateurs booléens (le caractère `|` – barre verticale, appelée encore "pipe" – pour l'opérateur OR<sup>34</sup>), ainsi la requête `paris|marseille` renverra tous les toponymes contenant paris ou marseille. Dans les expressions régulières, contrairement à la recherche simple, la recherche n'est pas stricte. Sur une recherche portant sur la base "Voies nommées", les résultats renvoyés seront par exemple : tous les odonymes relatifs à Paris comme "rue de paris", "avenue de paris", "route de paris à genève", etc. mais

33. Différents guides sont disponibles décrivant les caractères et opérateurs de expressions régulières, voir par exemple : [https://developer.mozilla.org/fr/docs/Web/JavaScript/Guide/Regular\\_Expressions/Cheatsheet](https://developer.mozilla.org/fr/docs/Web/JavaScript/Guide/Regular_Expressions/Cheatsheet)

34. Opérateur OU en logique booléenne



**Fig. 3.** Cartographie des routes de Paris et de Marseille



**Fig. 4.** Cartographie des toponymes d'une longueur de 12 caractères avec contraintes

aussi ceux relatifs à Marseille comme "route de nîmes à marseille", "rue jean marseille", etc., voir la figure 3.

- de rechercher spécifiquement des termes avec des variations de voyelles avec l'expression : "gr[aiou]s". Les voyelles entre crochets sont considérés comme possibilités de remplacement. De la même façon, nous retrouverons tous contenant les toponymes gras, gris, gros seront renvoyés de manière non stricte,
- d'effectuer une recherche stricte, la requête "<sup>^</sup>paris\$" ne renverra que les toponymes ne contenant strictement que paris,
- de combiner des termes et des opérateurs pour effectuer une recherche plus avancées comme de tous les toponymes commençant par n'importe quelle lettre de l'alphabet, finissant par une voyelle qui n'est pas un y et d'une longueur de 12 caractères. La requête suivante "<sup>^</sup>[a-z]12[<sup>^</sup>aeiou]\$" permet cette interrogation et est représentée sur la figure 4.

## 4 Conclusion

L'application web Topomine ouvre des possibilités d'interaction et d'interrogation de plusieurs bases de données toponymiques en langue française (à savoir une BD NYME reconstruite, une base de données odonymiques, une base de données hydronymiques, la base de données FANTOIR enrichie spatialement, et la base de données historiques et démographiques de Cassini) mises à disposition de manières simplifiées et

centralisée. Topomine offre de puissants moyens de recherche de chaînes de caractères pour interagir avec ces bases dans le but de faire émerger et de révéler visuellement des logiques spatiales d'un ensemble de toponymes. Topomine est librement accessible à l'ensemble des utilisateurs, et ce sans barrière d'accès, que ce soit au niveau technique car ne nécessitant qu'un navigateur web, ou au niveau des droits d'utilisation. Ainsi, en favorisant pleinement les principes FAIR, Topomine permet de consolider la mise à disposition et l'accessibilité des bases de données toponymiques en langue française actuellement accessibles, leur interopérabilité et leur réutilisation, y compris dans d'autres applications.

Dans un avenir proche (durant l'année 2023), de nouvelles fonctionnalités, certaines étant déjà en maturation, d'autres étant à initier, vont être intégrées à Topomine :

- maintenance et mise à jour automatisée des bases de données avec amélioration des contenus et de leur homogénéité,
- prise en compte des données des CROM-TOM dans une application Topomine dédiée (sachant que certaines bases de données utilisées ne couvrent malheureusement pas ces territoires),
- prise en compte des différents découpages administratifs (régions, départements, etc.) dans la recherche et/ou dans la représentation cartographique des résultats,
- intégration possible de nouvelles bases de données comme par exemple celle de DicoTopo via son API dédiée,
- recherche toponymique par approches phonétiques<sup>35</sup>,
- sauvegarde d'une session de recherches afin de pouvoir la rejouer rapidement dans un contexte de présentation,
- prise en charge de l'intégration d'un corpus de termes fournis par l'utilisateur afin d'effectuer facilement des recherches multiples,
- création si possible d'autres applications Topomine s'appuyant sur des bases de données toponymiques en langues étrangères et pour des pays autres que la France.

De plus, afin de faciliter l'utilisation de l'API qui est au cœur de l'application web Topomine, et donc sa réutilisation dans d'autres con-

---

35. Recherche basée sur l'algorithme Soundex ou encore Metaphone. Pour plus d'information, voir <https://www.postgresql.org/docs/current/fuzzystmatch.html#id-1.11.7.26.6>

textes, de nouveaux clients seront prochainement mis à la disposition des utilisateurs et des développeurs :

- création de clients JavaScript et Python,
- création de clients dans d'autres langages de programmation – il est en effet à noter qu'il est théoriquement possible de générer automatiquement des clients dans de nombreux autres langages de programmation à partir de la description proposée de l'API Topomine qui, de par son respect de la spécification OpenAPI (version 3.0.2)<sup>36</sup>, offre l'avantage de pouvoir être interfacée avec de nombreux outils logiciels existants –,
- intégration des nouvelles fonctionnalités offertes par l'API Topomine au sein de l'extension Topomine pour QGIS<sup>37</sup> – extension conçue originellement et offrant l'avantage aux utilisateurs d'utiliser leurs propres bases de données toponymiques s'ils le souhaitent.

## 5 Acknowledgements

L'application Topomine est développée dans le cadre du consortium Paris Time Machine<sup>38</sup>. Son hébergement et sa diffusion sont permis grâce à l'appui institutionnel de la Très Grande Infrastructure de Recherche Huma-Num du CNRS<sup>39</sup>.

## References

- ALGM14. Marlène Albert-Llorca, Raphaële Garreta, and Jean-Paul Métaillé. Les plantes de montagne: regards et débats sur un patrimoine. 2014.
- BPP21. Dominique Badariotti, Fabien Pfaender, and Jean-Luc Pinol. Trente millions d'adresses : boulevards et impasses d'une recherche sur la complexité lexicale et spatiale des odonymes. 2021. Publisher: CNRS-UMR Géographie-cités 8504.
- Cla00. Sylvain Clavaud. Jean-Pierre Bardet et Claude Motte, (dir.), *Paroisses et communes de France, Creuse*. Dictionnaire d'histoire administrative et démographique, Paris, CNRS Editions, 2000. 14(1):289–290, 2000. Publisher: Persée - Portail des revues scientifiques en SHS.
- CVDM17. Pascal Cristofoli, Marie-Christine Vouloir, Bertrand Duméniou, and Claude Motte. Des chefs-lieux de Cassini aux communes de France (1756-1999) : données géo-historiques associant la position géographique des lieux

---

36. Description OpenAPI disponible à l'adresse <https://paris-timemachine.huma-num.fr/api/topomine/openapi.json>

37. <https://plugins.qgis.org/plugins/Topomine/>

38. <https://paris-timemachine.huma-num.fr/>

39. <https://www.huma-num.fr/>



- représentés sur la Carte Générale de la France dite de Cassini (1756-1789) à l’historique des communes de France (An III-1999)., 2017. Laboratoire de démographie et d’histoire sociale.
- Dau49. Albert Dauzat. Le troisième congrès international de toponymie et d’anthroponymie. 1(3):153–153, 1949. Publisher: Persée - Portail des revues scientifiques en SHS.
- Dau60. Albert Dauzat. *La toponymie française: Buts et méthodes, questions de peuplement, les bases pré-indo-européennes, noms de rivières, toponymie gallo-romaine. Un dépouillement régional : Auvergne et Velay.* FeniXX, 1960. Google-Books-ID: Ax6XDwAAQBAJ.
- Del21. Par Bertrand Deléon. Un nouvel éclairage sur la toponymie bretonne à l’aide d’un outil en ligne. 2021. Draft / article non publié.
- Fou05. Vincent Fourmond. *Les expressions régulières par l’exemple.* H & K, 2005. Google-Books-ID: lfW1DwAAQBAJ.
- GGGS16. Alain Guerreau, Marie-José Gasse-Grandjean, and Laure Saligny. FANTOIR géolocalisé : mode d’emploi simplifié, 2016. ISSN: 1623-5770 Issue: Hors-série n9 Number: Hors-série n9 Publisher: Centre d’études médiévales Saint-Germain d’Auxerre.
- Man15. Francis Manzano. Jalons pour une étude géotoponymique de quelques distributions macrotoponymiques dans l’espace gallo-roman, in *L’onomastique gallo-romane alpine* [[http://www.centre-etudes-francoprovencales.eu/cef/allegati/actes-2007\\_914.pdf](http://www.centre-etudes-francoprovencales.eu/cef/allegati/actes-2007_914.pdf)]. page 26, 2015.
- MST03. Claude Motte, Isabelle Séguy, and Christine Théré. *Communes d’hier, communes d’aujourd’hui. Les communes de la France métropolitaine, 1801-2001. Dictionnaire d’histoire administrative avec 1 Cédérom - Claude Motte, Isabelle Séguy, Christine Théré.* Classique de l’économie et de la population - Etudes et enquêtes historiques. INED paris edition, 2003.
- noa38. Actes du premier congrès International de Toponymie et d’Anthroponymie : actes et mémoires (Paris, 25-29 juillet 1938), 1938.
- Rob18. Sandrine Robert. Les chasse-marée et la route du poisson. Réduire la distance pour approvisionner Paris du XIIIe siècle au XVIIIe siècle. 2018. Publisher: Les Éditions de l’EHESS.
- Ros48. Charles Rostaing. Actes et mémoires du premier congrès de toponymie et d’anthroponymie, paris, institut de phonétique de la sorbonne, 1939. 60(237):326–330, 1948. Publisher: Persée - Portail des revues scientifiques en SHS.
- TTM<sup>+</sup>15. Michel Tamine, Dimitri Tarze, Sébastien Mustière, Vincent Badeau, and Jean-Luc Dupouey. Changement climatique et toponymie - Étude de la répartition ancienne du hêtre à travers ses traces toponymiques. 25(2):229–244, 2015. Number: 2 Publisher: Lavoisier SAS.
- Vac99. Françoise Vachey. Les suffixes toponymiques français : atlas et base de données, 1999.
- Vac04. Françoise Vachey. Le statut du suffixe dans les toponymes ? Entre synchronie et diachronie... 11(1):349–361, 2004. Publisher: Persée - Portail des revues scientifiques en SHS.

Éb48. Jean Ébersolt. Congrès international de toponymie et d'anthroponymie, paris, 15-18 juillet 1947. 3(3):374-375, 1948. Publisher: Persée - Portail des revues scientifiques en SHS.