

Alliant précision scientifique et démarche collaborative, la base de données géographiques mondiale GeoNames s'est imposée comme une ressource indispensable pour les chercheurs comme pour les bibliothécaires.

GeoNames : pierre angulaire des données de référence géographiques

GeoNames.org¹ est un répertoire mondial de données géographiques compilant 12 millions de caractéristiques et 25 millions de noms. Régi par une licence d'attribution ouverte, ce dépôt non seulement facilite le téléchargement et l'édition des données, mais contribue également à la création collaborative de connaissances à l'échelle mondiale. GeoNames.org se distingue par le fait qu'il s'appuie sur des sources officielles, notamment les agences nationales de cartographie. Par sa rigueur et son exactitude scientifiques, ses solides mesures pour garantir la qualité des données, et la description complète des champs, GeoNames constitue un outil inestimable et une ressource essentielle pour les chercheurs et les bibliothécaires à la recherche de données fiables.

Attributs des entités du répertoire : au cœur de GeoNames.org se trouve une compilation minutieuse d'attributs. Ceux-ci comprennent des éléments fondamentaux tels que le nom, la classe d'entités, le code d'entités, la latitude, la longitude, le pays, la hiérarchie administrative (région, département, arrondissement, commune), la population, l'altitude et le fuseau horaire. La représentation de l'empreinte géographique sous la forme d'un point, qui désigne souvent l'église ou le siège du gouvernement pour les lieux peuplés, ajoute de la précision à l'information.

Modèle d'attribution : GeoNames.org adopte un modèle basé sur l'attribution, s'éloignant de l'approche strictement collaborative. Ce modèle permet une utilisation plus libre des données, facilitant une intégration transparente avec les ensembles de données des utilisateurs. L'accent mis sur l'attribu-

tion encourage un environnement collaboratif dans lequel le référentiel devient un espace partagé pour l'amélioration des connaissances.

Sources officielles et collaboration : dans un souci d'exactitude, GeoNames.org s'appuie largement sur des sources officielles, notamment les agences nationales de cartographie et les bureaux de statistiques. La disponibilité croissante des données officielles sous licence ouverte ouvre la voie à un processus d'intégration continue, s'alignant sur l'objectif primordial de la plateforme qui consiste à utiliser des sources faisant autorité. Alors que les États-Unis ont lancé le mouvement de diffusion des données par les sources gouvernementales sous licences libres, la plupart des autres pays suivent et diffusent leurs données. La situation est plus difficile dans les pays en développement où des organisations d'aide et de secours collectent et publient des données qui sont souvent absentes des sources gouvernementales officielles.

Identifiants pérennes et variantes de noms : GeoNames.org utilise un identifiant pérenne, Geonameld, qui assure la stabilité des références. Le nom principal est présenté en anglais ou dans des variantes internationales, accompagné de noms alternatifs liés. Cette caractéristique enrichit l'ensemble des données, en tenant compte des diverses particularités linguistiques et de la nomenclature historique.

Prise en charge des noms alternatifs : le répertoire adopte une approche flexible en prenant en charge les noms alternatifs. Ces noms alternatifs sont assortis d'attributs tels que le code de langue ISO, le statut de préférence, les variations de longueur et la signification historique ou vernaculaire. Pour les noms historiques, l'inclusion des champs « depuis » et « jusqu'à » permet de suivre l'évolution temporelle de la nomenclature. Pour l'affichage aux utilisateurs finaux, il est recommandé d'utiliser le nom alternatif marqué comme *isShort* dans la langue concernée. Le champ du code de langue pour les noms alternatifs accueille également des pseudo-codes comme *link* pour une url vers un site web (le plus souvent l'article Wikipedia correspondant). D'autres pseudo-codes sont utilisés, comme *post* pour les codes postaux, *icao* et *iata* pour les codes d'aéroport.

Crédit Adobe stock, par Keitma



Identifiant végétal, généré à l'aide de l'IA

[1] <https://www.geonames.org>

Garantie de qualité et processus de confirmation :

Geonames.org accorde une grande importance au maintien de l'intégrité des données en les soumettant à un processus de garantie qualité approfondi, comprenant plus de 1 700 vérifications. Ces vérifications systématiques permettent d'identifier rapidement toute incohérence. En outre, les modifications pertinentes sont soumises à un processus de confirmation.

Modèle premium et mises à jour en temps voulu :

la plateforme comprend un modèle premium proposant des mises à jour par téléchargements quotidiens gratuits. Les abonnés Premium bénéficient d'un abonnement mensuel leur donnant accès à un ensemble de données améliorées soumises aux processus de garantie de qualité, alors que les extractions quotidiennes gratuites sont des « *work in progress* ». Cette approche à deux niveaux permet de trouver un équilibre entre l'accessibilité et la fiabilité des données.

Divisions administratives hiérarchiques :

Geonames.org fournit une structure hiérarchique pour les divisions administratives, enrichissant automatiquement les entrées des utilisateurs avec les relations administratives des frontières. Cette structure s'étend aux champs relatifs aux pays, y compris un attribut spécifique (cc2) pour les pays alternatifs, ce qui est particulièrement pertinent pour les éléments tels que les montagnes ou les lacs bordant plusieurs pays. Les frontières des pays et des divisions administratives jouent un rôle essentiel dans le processus d'enrichissement en améliorant la compréhension contextuelle des caractéristiques géographiques.

Relations « parents-enfants » et caractéristiques créées par les utilisateurs :

une caractéristique distinctive de Geonames.org est la définition d'une relation « parent-enfant » entre les caractéristiques, établissant une hiérarchie secondaire. Pour les entités créées par les utilisateurs, la plate-forme complète le fuseau horaire et l'élévation approximative à partir des modèles DEM (*Digital Elevation Models*, ou coordonnées en trois dimensions) garantissant ainsi l'exhaustivité, même en l'absence de données d'élévation explicites.

Par essence, Geonames.org apparaît non seulement comme un dépôt de données géographiques mais aussi comme un outil polyvalent, méticuleusement conçu pour répondre aux besoins des bibliothécaires. Grâce à un mélange judicieux de précision scientifique, de collaboration et de fonctionnalités centrées sur l'utilisateur, il s'impose comme une ressource indispensable dans le paysage dynamique des études géographiques.

MARC WICK

Fondateur de GeoNames
marc@geonames.org

• • • PLEIADES : UNE BASE DE DONNÉES GÉOGRAPHIQUES POUR LES MONDES ANCIENS¹

Développée dans le cadre d'un projet conjoint de l'Institute for the Study of the Ancient World de l'Université de New York et du Ancient World Mapping Center de l'Université de Caroline du Nord (Institut pour l'étude du monde antique et Centre de cartographie du monde antique), Pleiades est, depuis sa mise en ligne en 2007, la plus importante base de données géographiques des mondes anciens.

SES FONCTIONNALITÉS

Pleiades opère une distinction entre les lieux, les toponymes et les localisations. Le lieu est l'unité de base correspondant à une entrée dans la base de données et peut posséder plusieurs noms et localisations.

La base de données est riche de plus de 40 000 lieux, 37 000 noms et 43 000 localisations. Chaque lieu est décrit dans une page de présentation possédant un URI et indiquant sa localisation sous la forme de degrés décimaux, les noms associés dans différentes langues et à différentes époques, les autres lieux en relation – c'est-à-dire incluant ou faisant partie du lieu décrit – des mots-clés, des références bibliographiques distinguant les sources anciennes et contemporaines, des liens vers d'autres bases de données comme GeoNames, des marqueurs permettant de lier des photographies déposées sur Flickr et des liens de téléchargement. Ces présentations font l'objet d'une évaluation par les pairs.

Pleiades permet de télécharger les données relatives à un lieu aux formats Atom, JSON, KML, RDF+XML et Turtle depuis sa page de présentation ou l'ensemble des données aux formats CSV, JSON, KML et RDF depuis la page *Downloads*.

SES LIMITES

Produit par la collaboration de différents projets œuvrant notamment dans le domaine des études classiques, Pleiades donne accès à de nombreuses informations géographiques, mais principalement concentrées sur le monde méditerranéen gréco-romain. Néanmoins, le projet est toujours en cours et permet à chacun de participer à l'enrichissement des données y compris dans un périmètre géographique plus large.

NICOLAS SOUCHON²

Assistant-égyptologue à l'Institut français d'archéologie orientale
nsouchon@ifao.egnet.net

[1] <https://pleiades.stoa.org>

[2] EPHE, PSL, AOROC UMR 8546, Paris/IFAQ, Le Caire

