

STRUCTURATION DES RESSOURCES GEOGRAPHIQUES AUTOUR DES TOPONYMES ET D'UNE ONTOLOGIE FORESTIERE

Marius MASSALA*, Franck FAVETTA*, Robert LAURINI*

{prenom.nom}@liris.cnrs.fr

(*) Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, Domaine Scientifique de la Doua, LIRIS, Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne cedex

Mots clés: Toponyme, ontologie géographique, forêt, métadonnée

Keywords: gazetteer, geographic ontology, forest, metadata

Palabras clave : Topónimo, ontología geográfica, bosque

Résumé

L'exploitation et l'utilisation croissante de l'information géographique a conduit à une augmentation du volume informationnel. En effet, le volume des données mises à la disposition des utilisateurs, l'hétérogénéité des données elles-mêmes (multisources) combinée au développement rapide des nouvelles technologies de l'information et de la communication (NTIC) entraînent des difficultés en termes de recherche d'information. Face à cette problématique, les organismes producteurs et utilisateurs des données géographiques ont besoin d'architectures adaptées pour mieux gérer leur patrimoine informationnel géographique. L'exploitation des toponymes est une piste potentielle. La définition d'une architecture globale intégrant les toponymes, les métadonnées et les limites floues peut apporter des réponses efficaces en termes de recherche et diffusion de l'information géographique notamment dans un contexte de gestion environnementale.

1. Introduction

La dernière conférence sur le climat tenue à Copenhague (2009) a permis de mettre en lumière les enjeux liés à la préservation de l'environnement. Partant du constat que près de la moitié des forêts qui couvraient initialement notre planète ont été défrichées et 30 pour cent ont été fragmentées [15], la préservation

des 20 pour cent restant qui se situent essentiellement dans le bassin de l'Amazonie, en Afrique centrale, au Canada, en Asie du Sud-Est et en Russie devient plus que primordiale. C'est dans ce cadre que certains pays d'Afrique Centrale dont l'espace territorial est couvert par une immense forêt portant le nom de Bassin du Congo ont conjugué leurs efforts pour mettre en commun leurs énergies dans la procédure de sauvegarde de cet écosystème. Enfin de faire face aux engagements liés au développement durable, les pays de cette région se sont engagés à gérer leur patrimoine forestier et une bonne partie de données environnementales par des données de télédétection et les bases de données. A cet effet, l'Agence pour le Développement de l'Information Environnementale (ADIE) a été créée. Cette dernière regroupe les pays suivant le Cameroun, le Gabon, le Congo, le Tchad, la Centrafrique, la Guinée Equatoriale, et le Congo Démocratique. L'objectif de l'agence est de mutualiser l'information récoltée non seulement par elle-même, mais aussi par les Organisations Non Gouvernementales (ONG) spécialisées. Devant cette tâche, l'ADIE s'est retrouvée confrontée d'une part, au problème d'hétérogénéité des données et de référentiel géographique commun dans le cadre de la division du territoire, et d'autre part à la problématique des toponymes dans la localisation des ressources. En effet, les pays membre de cette agence n'ont pas tous la même langue officielle (français, anglais, espagnol, etc.). En plus de cela s'ajoute l'utilisation des langues vernaculaires dans la l'attribution des noms des entités géographique. Dans cet article, nous proposons une approche méthodologique qui permette d'établir le lien entre le répertoire des toponymes et les objets géographiques aux limites floues.

2. Illustration de la problématique des toponymes par l'exemple

La dénomination des objets géographiques dépend en général d'un ou plusieurs pays ou simplement d'autorités internationales comme c'est le cas pour les noms d'objets extra terrestres comme les constellations ou les nouvelles planètes. Les différents outils, et travaux scientifiques présentés dans l'état de l'art montrent la complexité de la gestion des toponymes dans le cadre de la recherche d'information. Le cas du *lac de Genève* est une parfaite illustration de ces objets géographiques dont le nom est lié au pays ou à la région (Lac de Lausanne en Suisse et lac Léman en France). Aussi, toute proposition de solution doit prendre en compte non seulement les variantes linguistiques des toponymes mais aussi les synonymes dialectaux et leur évolution historique. C'est le cas par exemple du *Mont Blanc* qui est appelé *Monte Bianco* en Italie ou du *Mont Sinai* qui peut varier en *Sinaa* ou *Sinail* selon le pays (Egypte ou Israël) d'usage. De même, certaine entité ont connu des changements d'appellation tout au long de leur histoire. C'est le cas du fleuve *Congo* par exemple situé en République Démocratique du Congo. Anciennement il fut appelé *Barbila*. Ce fleuve en amont portait le nom de *Lualala*. Puis au fil du temps, le nom *Zaire* s'est imposé et cela jusqu'en 1997. Aujourd'hui, le fleuve est appelé *Congo*. De plus, ce toponyme désigne également le fleuve, mais aussi deux pays frontaliers de ce même fleuve différenciés par leurs capitales et aussi un vaste couvert forestier qui s'étend sur six pays de la même région. Divers exemples peuvent être multipliés en matière de gestion des toponymes.

3. Etat de l'art

Sur le plan scientifique, un grand nombre des travaux se sont orientés vers la mise en relation du nom du lieu soit avec son footprint (empreinte géométrique d'une entité géographique) actuelle ou alors avec les ressources attachées à ce toponyme. C'est dans cette logique que s'inscrivent les outils géoname ou flicker. Ensuite, on a vu apparaître d'autres outils plus élaborés comme la bibliothèque d'Alexandrie (Alexandria Digital Library). La modélisation de ce répertoire a permis d'intégrer les différentes appellations d'un toponyme et de les associer au footprint correspondant. L'exploitation de l'ADL est couplée à un thésaurus qui permet de définir la typologie des entités. Progressivement le besoin de passer du simple album annoté vers des systèmes d'information logique plus élaborés s'est imposé.

Les auteurs [9, 10, 11] ont montré que les nouveaux répertoires des toponymes doivent désormais intégrer des relations topologiques qualitatives et quantitatives. Ils affirment qu'il était désormais possible de définir des relations topologiques tournées vers les différents types de trait : par exemple une île doit être dans l'eau intérieure, une gare doit être à côté d'une ligne de chemin de fer, une mangrove ne peut être qu'à côté de la mer etc. L'ensemble des relations permises peut aussi contenir des restrictions juridiques par exemple, une concession d'exploitation forestière et un parc ne peuvent pas se superposer entre eux, ou l'un ne peut contenir l'autre. La spécification de telles relations aiderait à résoudre les problèmes que les approches géométriques sont « aveugles » généralement à cause des annotations fausses des utilisateurs [11]. Le fait de faire respecter des règles pour les relations topologiques permises est une approche prometteuse qui permet d'éliminer les ressources annotées qui violent les règles. Abordant dans le même sens, [9, 10, 11] proposent de développer une méta-ontologie de type de trait qui fournit des définitions formelles aux types de trait fondamentaux trouvés dans les index géographiques. Ainsi, si un index géographique ne contient pas le jeu de données approprié ou n'est pas responsable d'une région géographique spécifique, la question est réexpédiée à un autre index géographique. Par exemple, quelques index géographiques peuvent contenir seulement des données validées par une autorité, pendant qu'un autre index géographique peut conserver de nouveaux traits contribués par des volontaires qui n'ont pas été évalués jusqu'ici. L'infrastructure d'index qu'ils proposent soutient le développement d'une interface indépendante. De leur côté, [7] proposent un prototype qui permet d'encoder et de visualiser les phénomènes environnementaux sur une carte. Leur outil repose sur une base de métadonnées destinée à décrire la sémantique des données (essentiellement les cartes). La base des métadonnées utilise un thésaurus du domaine agricole pour le choix des termes appropriés. Dans le domaine des systèmes logiques, ce sont les projets SPIRIT et GEOLIS qui ont retenu notre attention. SPIRIT propose un moteur de recherche web avec des capacités d'interrogation spatiale. Pour cela, une géo-ontologie a été construite qui permet de décrire la terminologie des lieux et une structure de l'espace géographique. Elle traduit explicitement l'imbrication géographique des lieux, par exemple : Rennes est une ville en Bretagne, la Bretagne est une région de France, la France est un pays d'Europe. L'intérêt de ce type d'ontologie est également d'associer à chaque lieu les différents noms qui le désignent, par exemple : Bretagne, Breizhoneg, Brittany [2]. Chaque lieu est associé un ensemble de noms (les différentes dénominations possibles de ce lieu), un descriptif du lieu (exemple: ville, région, pays) ainsi qu'un ensemble d'emprises spatiales décrivant la géométrie du lieu à différents niveaux de précision. L'ensemble des lieux est ordonné de façon hiérarchique dans la géo-ontologie selon une relation d'inclusion de lieux. Les relations d'inclusion sont complétées par des relations de superposition et d'adjacence entre lieux. La géo-ontologie intervient, entre autre, dans l'interprétation des requêtes où elle permet :

- ✓ Désambiguïsation de requêtes. Par exemple, lorsque le terme « *Paris* » apparaît dans une requête, le système SPIRIT peut suggérer à l'utilisateur « Paris, France » ou « Paris, Texas ».
- ✓ Généralisation d'un nom de lieu à l'ensemble de ses dénominations connues (par exemple la traduction dans différentes langues).

D'une manière générale, les systèmes SPIRIT, GEOLIS et l'Alexandrie Digital Library (ADL) sont certainement parmi les plus aboutis en matière de répertoire de toponyme que nous avons consulté. Certains d'entre eux ont intégré quelques relations topologiques qualitatives binaires (inclusion, intersection, adjacence, etc.) et même quantitatives. Le tableau 1 résume quelques aspects fonctionnels de certains outils.

| | Relations topologiques | ontologie | Fonctionnalités et atouts | Limites |
|-----------------|--|------------------|---|---|
| ADL | <ul style="list-style-type: none"> - Superposition - Intersection - Inclusion | Oui (XML) | <ul style="list-style-type: none"> - Affichage des des ressources. - Catalogage des métadonnées | <ul style="list-style-type: none"> - Pas de désambiguïsation des noms - Non prise en compte des évolutions du footprint - Non intégration des limites floues dans le footprint - Absence de synonymie entre toponyme - Complexité dans la description des métadonnées |
| GEONAMES | Hierarchie et voisinage | Non | Affichage vectoriel des objets | <ul style="list-style-type: none"> - Non prise en compte des évolutions du footprint - Pas d'affichage des ressources associées au footprint |
| GEOLIS | Intégrées dans la description des objets | Non | <ul style="list-style-type: none"> - Affichage des ressources sous forme cartographique - Stockage des données au format GML - utilisation du service PofFS pour la navigation sur les données | <ul style="list-style-type: none"> - Non prise en compte des évolutions du footprint - Pas d'affichage des ressources associées au footprint - Stockage limité au fichier GML et ne prend pas en compte les autres formats de fichier - Affichage limité aux fichiers vectoriel |
| SPIRIT | <ul style="list-style-type: none"> - Inclusion - Adjacence - superposition | Oui (OWL) | <ul style="list-style-type: none"> - désambiguïsation des lieux réussie - taxonomie des lieux | <ul style="list-style-type: none"> - pas d'évolution du footprint des lieux dans le temps |

Tableau 1 : Comparaison fonctionnelle entre différents gazetteer.

A travers ce rapide panorama, il ressort que les outils ci-dessus ne possèdent pas de gazetteer capable de ressortir la géométrie d'un objet en fonction des appellations que celui-ci a connu, encore moins les ressources associées aux différents toponymes. Face à cette problématique, nous nous proposons de développer un outil complet basé sur les technologies libres. Cette proposition émane d'une réflexion conjointe menée entre informaticiens et thématiciens faisant appel à la géomatique, à l'environnement et à la foresterie. L'outil que nous comptons développer devra être capable de répondre aux interrogations suivantes :

- ✓ Quelle est la meilleure façon de faire le lien entre un toponyme, un objet, ces différentes appellations en langue officielle ou vernaculaire, sa typologie et l'évolution temporelle de sa géométrie?
- ✓ Comment retrouver les ressources qui appartiennent (ou ont appartenu) à deux objets géographiques contigus dont les limites sont floues?
- ✓ Comment assurer la cohérence topologique en intégrant la notion de continuité et de discontinuité dans la construction des relations entre les concepts de l'ontologie?

4. Proposition d'un service de toponymes

Jusqu'à ces dernières années, les missions du chercheur et des organismes de recherche n'intégraient que rarement la gestion du patrimoine scientifique, en tout cas pas son optimisation [1]. A cet effet, dans le cadre de la gestion des données géographiques de l'ADIE. Il s'agit d'un service de type " **observatoire** ". Il se caractérise par l'acquisition répétitive de données sur des sites de référence, selon des protocoles précis et en général pour une longue période avec, bien sûr, un accès direct ou filtré aux données (droits d'accès, profils personnalisés,...). Ce service sera construit autour d'un répertoire des noms des forêts et localités des pays d'Afrique centrale et une ontologie forestière dans laquelle les relations topologiques qualitatives spécifiques à cette forêt y figureront.

4.1 Modèle

La conception que nous avons adoptée part de l'expérience de dénomination des noms des lieux des pays d'Afrique centrale. Celle-ci repose sur quatre éléments essentiels à savoir : le toponyme, l'entité géographique qui lui est associée, la ressource et les classes de l'ontologie. Le modèle des toponymes et de description des ressources s'efforce de prendre en compte toutes les caractéristiques ci-dessus énumérées. Il se présente sous la forme d'un modèle Entité - Relation volontairement simplifié et ne traite que des problèmes toponymiques qui se rapportent aux caractères latins.

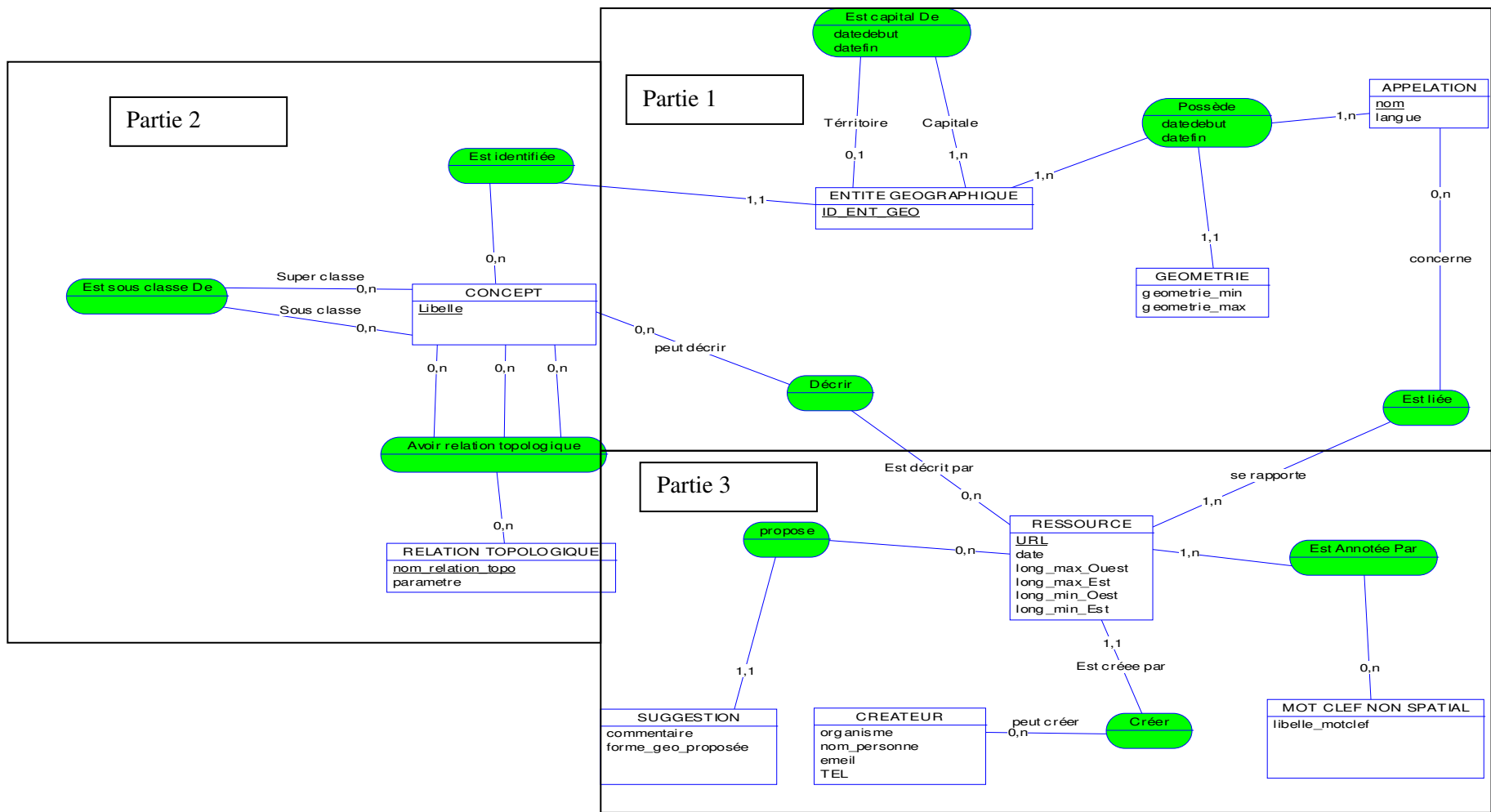


Figure 1 : modèle de description des toponymes, des ressources et des concepts de l'ontologie.

La première partie du modèle concerne la toponymie des noms. Cette partie s’efforce d’apporter une solution dans la résolution de tous les conflits que nous avons précédemment soulevés à travers les différents exemples. Le modèle est structuré autour de l’entité « *Appellation* » qui identifie les toponymes et sert de point de départ pour définir toute la lexicographie du répertoire. Aussi, cette entité permet d’identifier pour chaque toponyme les caractéristiques géométriques (rectangle englobant et géométrie minimale et maximale) des entités comme l’illustre la norme ISO 19115.

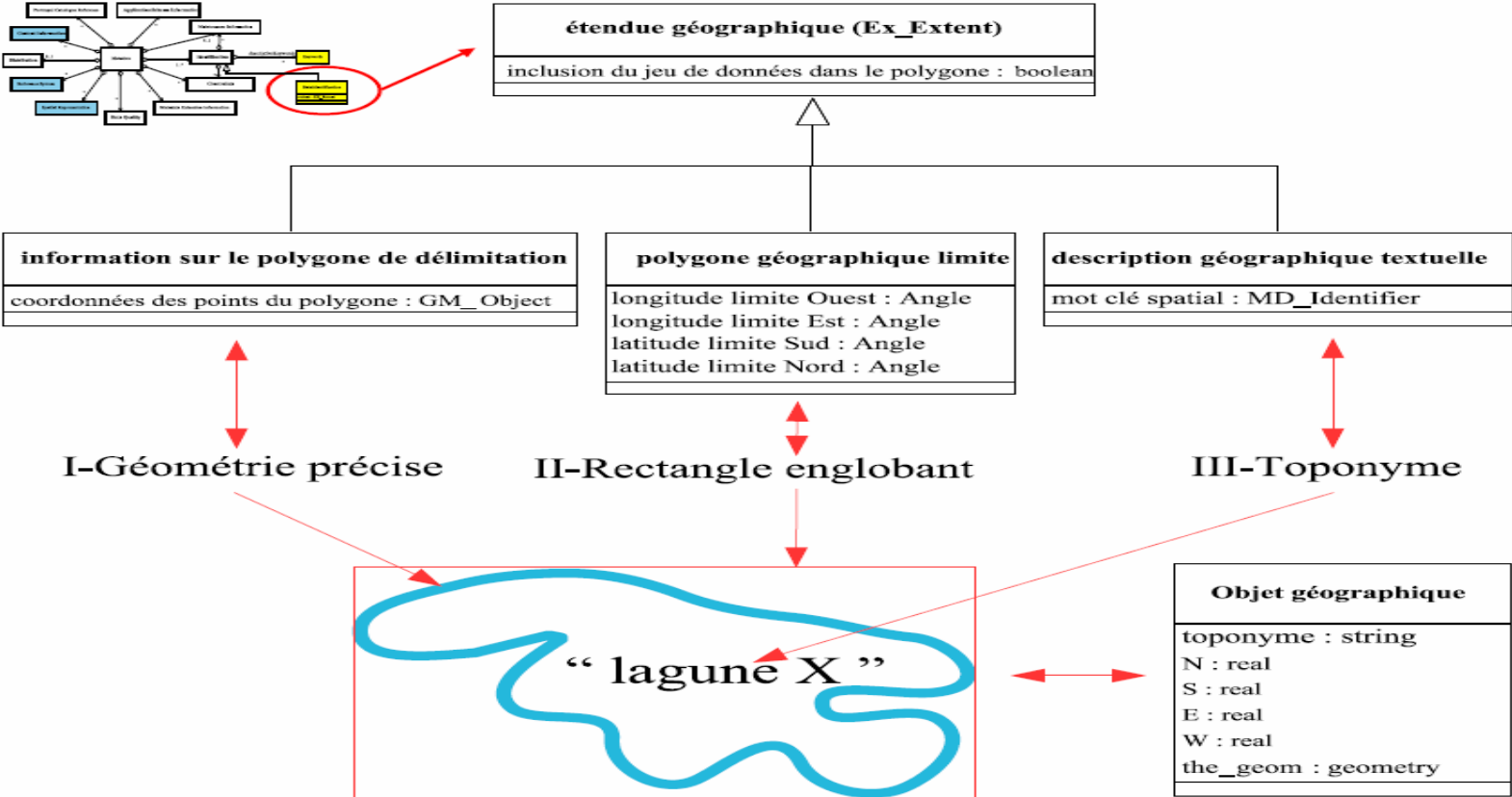


Figure 2 : Illustration des trois types de descripteurs géographiques de l'ISO 19115[3]

L'encadré portant le numéro un (1) qui se rapporte aux toponymes de la figure 1, est une description des entités géographiques. Ces derniers sont en relation avec l'un des concepts qui peuplent l'ontologie (partie deux). Cette relation a pour objectif de déterminer la typologie des objets géographiques (fleuve, ville, pays, etc.). La détermination des concepts s'est faite sur la base du découpage administratif et des caractéristiques des forêts des pays du golfe du Congo. Dans le chapitre qui suit nous nous attellerons à expliquer les raisons de ces choix. Enfin, la troisième partie (ressources) du modèle, permet de rattacher les ressources informationnelles à la partie toponymique et à l'ontologie.

4.2 Ontologie

L'espace géographique que nous étudions a été subdivisé en quatre grandes catégories à savoir les espaces administratifs, les espaces réglementés, l'occupation du sol et les espaces en eau. En ce qui spatial administratif: Dans lequel on retrouve le découpage administratif des pays. Ces quatre types sont rangés en deux catégories. On a d'une part les concepts qui présentent une géométrie fixe (sharp) et ceux dont la géométrie est variable (fuzzy). Nous considérons qu'un polygone a des limites fixes lorsque les limites de celui-ci sont clairement définies et qu'il n'y a pas de zone de transition (appelée zone floues) pour passer d'un polygone à un autre. Au nombre de polygones aux limites fixes, nous avons mis toutes les classes du concept *espace administratif* à l'exception de villes qui selon notre approche ne sont pas clairement délimitées. Ensuite il y a les *zones à réglementation*. Dans la première catégorie nous avons les zones administratives comme les régions, les départements et les communes, puis on a les espaces à réglementation à infimes densité humaine. Ce par exemple les concessions forestières, les parcs, les réserves et jardins (botanique, zoologique...) de toutes natures. Les concepts rangés dans la catégorie fuzzy sont ceux dont les limites sont variables et qui possède des zones de transition aux limites imprécises. Par exemple la limite géographique exacte entre une mangrove et une forêt inondée n'est pas très claire à déterminée. Si sur le plan de la représentation cartographique cela semble claire, dans la réalité par contre celle-ci n'est pas triviale comme le montre la figure ci après.

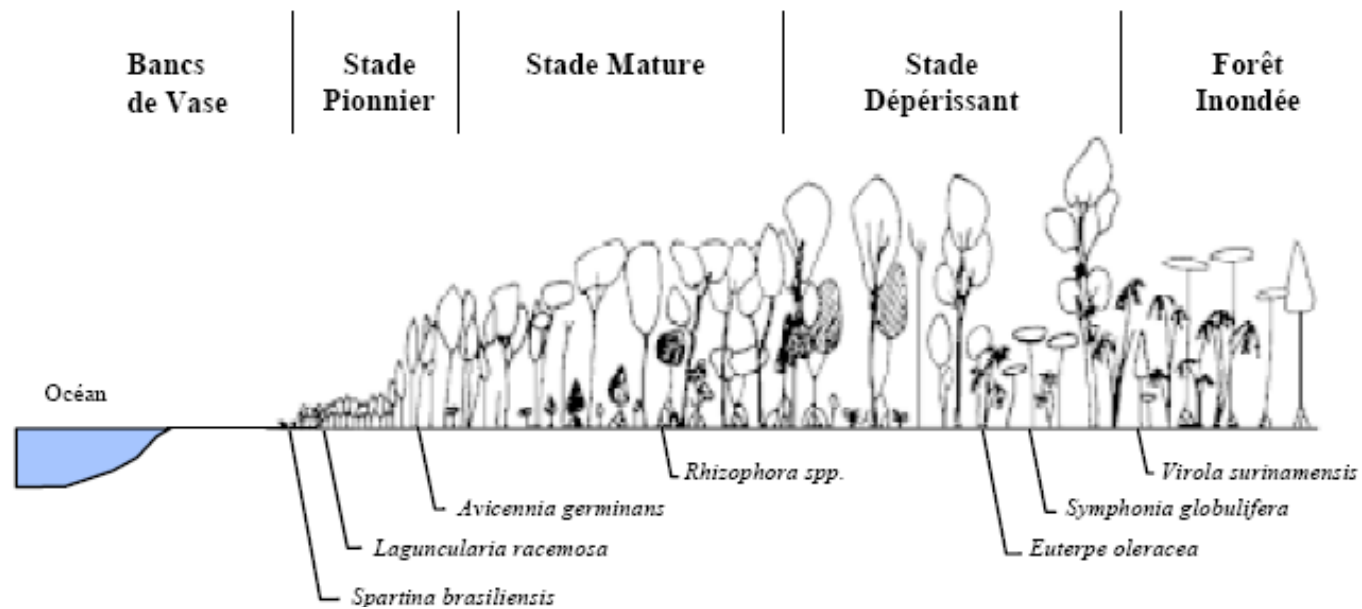


Figure 3 : Stade de développement d'une mangrove guyanaise Française d'après [13]

Aux nombres des polygones ayant des limites floues nous avons recensé trois catégories de concepts : l'occupation du sol (forêt primaire et secondaire, la savane, les steppes, etc.), les villes et les espaces en eaux. En ce qui concerne les deux derniers, nous avons intégré dans la tranche des polygones aux limites floues pour divers raisons. S'agissant des espaces en eaux (lac ou cours d'eau), leurs lits sont souvent sujets à variation selon les types de saison. Ainsi nous

avons intégré les paramètres de lit majeur en période de crue et de lit mineur lors des saisons dites sèches (sans pluies). Pour ce qui concerne les villes, nous sommes partis du constat qu'à l'intérieur et/ou à la périphérie des grandes villes, il existe des communes plus petites. Si ces dernières (les communes) sont administrativement autonomes, par contre sur le plan géographique elles ont tendance à être confondues aux grandes agglomérations. C'est le cas par exemple de la ville de Lyon et de la commune de Villeurbanne en France. La deuxième tendant à être confondue avec la première bien que celle-ci semble avoir à première vue des limites fixes. Pour ce qui concerne les relations topologiques entre les concepts, nous avons définie trois types d'opérateurs binaires pour modéliser l'espace. Ce sont les opérateurs *Tuches*, *Meet*, *Cover*, *Cover By*, *Intersection*. Ces opérateurs vont servir à définir des associations qualitatives spatiales entre les différentes régions de l'ontologie.

5. Exemple de requête

Pour illustrer nos travaux, nous allons prendre deux exemples de requêtes. Le premier concerne l'évolution de la géométrie d'une région, le deuxième l'affichage des ressources concernant cette même région avec d'autres paramètres spécifiés par l'utilisateur. La figure 2 présente le comportement du système lorsqu'un utilisateur soumet la requête suivante « quelle furent les formes géométriques de la Yougoslavie respectivement de 1960 à 1995 ? et de 2007 à 2009 ? ». Dans une telle requête, quatre éléments nous apparaissent essentiels : un toponyme (la Yougoslavie), une période (1960 à 1995 puis 2007 à 2009), un type (c'est un pays), et enfin une géométrie. Le résultat d'une telle requête doit être une image vectorielle qui montre pour chacune des périodes saisies, la géométrie du toponyme sélectionné. Dans l'étape (1), l'utilisateur construit sa requête. Celle-ci est soumise à la base des toponymes. Une évaluation de cette dernière est effectuée par le système pour voir si au cours de cette période il existe une géométrie correspondante et s'il n'existe pas une ambiguïté (exemple : le nom désigne deux pays différents) sur ce toponyme (étape 2). Suite à cette évaluation, deux situations peuvent se présenter, soit la géométrie existe et il n'y a pas d'ambiguïté, dans ce cas une demande est faite au SGBD spatial (étape 3) qui récupère auprès du serveur de fichier le footprint (étape 4) et l'envoie au serveur cartographique (en vue de sa reconstitution (étape 5). Ce dernier renvoie ensuite le footprint associé à l'utilisateur pour visualisation (étape 6). Si l'utilisateur avait également spécifié l'affichage des ressources, celles-ci seront affichées dans une gille sous forme de liste. Par contre, si pour la période sélectionnée, la géométrie n'existe pas (ou plus) à l'exemple de la période 2007 à 2009 pour l'ancienne Yougoslavie, ou s'il y a ambiguïté sur le toponyme, il peut avoir un silence ou alors le système propose de désambiguïser le toponyme à l'aide d'une liste proposée à l'utilisateur (exemple : Washington : Etat, ou Washington : capitale de l'Etat de Washington, etc.). La figure 3 montre le cheminement de la requête dans le système.

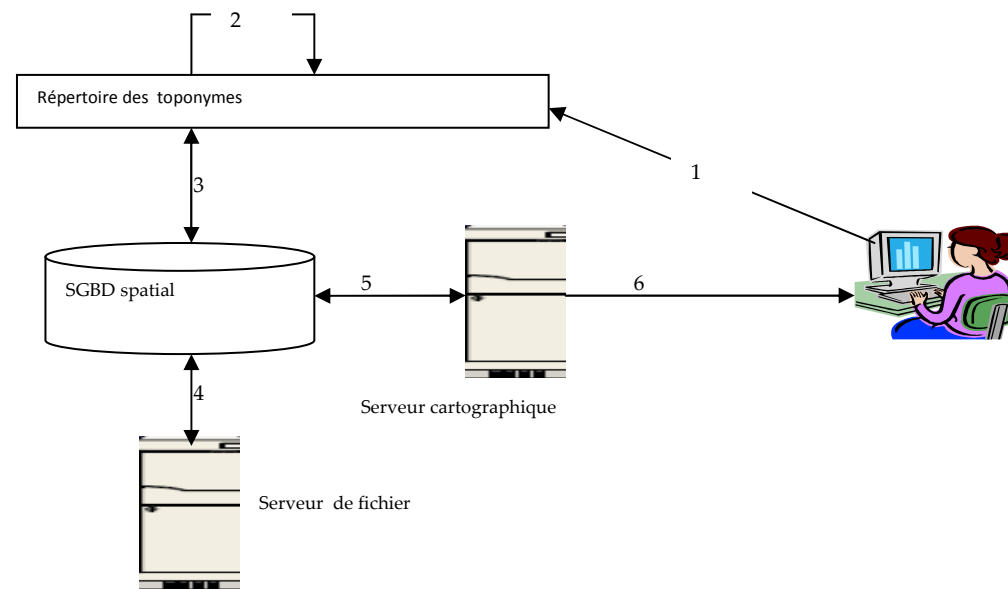


Figure 4 : Présentation des différentes étapes de traitement de la requête

6. Conclusion et perspectives

Le besoin d'accès à l'information a conduit au développement des solutions de plus en plus performantes. Parmi ces solutions se trouvent celles dont les fondements reposent sur l'utilisation des répertoires des toponymes. A la différence de la première catégorie des répertoires de toponymes qui étaient essentiellement bâtis sur les noms officiels contrôlés par les instances toponymiques, la nouvelle génération quant à elle intègre en plus du côté vernaculaire des toponymes, la gestion des données relevant des objets géographiques plus petit comme par exemple les études d'impacts menées à l'intérieurs de zones (villages, campements, pistes, etc.) dont les noms ne sont pas toujours connus sur le plan national et/ou international. Aussi, en nous servant des opérateurs

topologiques définis par Max Egenhofer[...], nous allons intégrer la problématique des limites floues dans la description des objets. C'est dans ce cadre que s'inscrivent nos travaux de recherche. Tous les modèles que nous avons présentés constituent des premières versions. Actuellement nous nous attelons à affiner et implémenter le premier d'entre eux : le dictionnaire des toponymes.

Bibliographie

- [1] BARDE Julien., LIBOUREL Thérèse., Maurel Pierre., *Ontologies et métadonnées pour le partage d'information géographique*, European Journal of GIS and Spatial Analysis - Revue internationale de géomatique, vol 14, N° 2-2004, pp 199-217.
- [2] BEDEL Olivier, *GEOLIS : Un Système d'information logique pour l'organisation et la recherche de données géolocalisées*, Thèse de Doctorat, Université de Renne 1, Janvier 2009.
- [3] DESCONNETS JC, LIBOUREL Thérèse, S. Clerc., *Cataloguer pour diffuser les ressources environnementales*, Actes de la conférence INFORSID 2007, Perros-Guirec, 2007, pp 344-361.
- [4] DESCONNETS JC, MOYROUD Nicolas, LIBOUREL T., *Méthodologie de mise en place d'observatoires virtuels via les métadonnées*, Actes de la XXI ème conférence INFORSID'2003, Nancy 3-6 juin 2003, p 253-269.
- [5] DE LONGUEVILLE Bertrand, OSTLÄNDER Nicole, KESKITALO Carina., *Addressing vagueness in Volunteered Geographic Information (VGI) –A case study*, Article under Review for the International Journal of Spatial Data Infrastructures Research, Special Issue GSDI-11, submitted 2009-03-01.
- [6] DESJARDINS Robert., *La télédétection : Perspective analytique*, Collection : savoir Plus Université, 2003.
- [7] GUTIERREZ Claudia, SERVIGNE Sylvie., *Métadonnées spatiales temps réel*, ingénierie des systèmes d'information, métadonnées et nouveaux SI, volume 12, N° 2-2007, pp 96-119.
- [8] INSTITUT GEOGRAPHIQUE NATIONAL, *BD NYME*, Version 2, description de contenu, Février 2009.
- [9] JAILLARD Pierre., *Entre guerres et conflits : la planète sous tension*, Festival international de géographie, Saint Dié-des-Vosges, Novembre 2008, SIG la lettre N° 101, Novembre 2008.
- [10] JANOWICZ K, KEBLER C., *The role of ontology in improving gazetteer interaction*, Int. Journal of Geographical Information Science, volume 22, N° 10-2008, pp 1129–1157.
- [11] JANOWICZ, K., SCHWARZ, M., WILKES M.: *Implementation and Evaluation of a Semantics-based User Interface for Web Gazetteers*. At the Visual Interfaces to the Social and the Semantic Web (VISSW 2009) Workshop in conjunction with the International Conference on Intelligent User Interfaces (IUI 2009). CEUR-WS Proceedings Vol. 443. Sanibel Island, Florida. February 2009.
- [12] KEBLER C, JANOWICZ K., BISHR M, *An Agenda For The Next Generation Gazetteer: Geographic Information Contribution and Retrieval* , 17th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2009, November 4-6, 2009, Seattle, Washington, USA, Proceedings, pp. 91-100.
- [13] PROISY Christophe, *Apport des données radars à synthèse d'ouverture pour l'étude de la dynamique des écosystèmes forestie*, thèse de Doctorat, Université Paul Sabatier Toulouse III, 1999.
- [14] WORLD RESSOURCES INSTITUTE, *Un premier regard sur l'exploitation forestière au Gabon*, 2000.
- [15] WORLD RESSOURCES INSTITUT, *Atlas forestier interactif du Cameroun*, version 2, document de synthèse, 2007.