
Des SIG aux ontologies géographiques

Nadine Cullot* – Christine Parent – Stefano Spaccapietra*** – Christelle Vangenot*****

* Laboratoire LE2I, Université de Bourgogne, France

Nadine.Cullot@u-bourgogne.fr

** HEC-INFORGE, Université de Lausanne, Suisse

Christine.Parent@unil.ch

*** Laboratoire de Bases de Données, Ecole Polytechnique de Lausanne, Suisse

{Stefano.Spaccapietra, Christelle.Vangenot}@epfl.ch

RÉSUMÉ. L'évolution des technologies et d'Internet suscite un besoin croissant de partage d'informations sur le Web qui passe par la compréhension des informations mises à disposition et une automatisation des processus de navigation, d'extraction ou de mises à jour de ces informations. L'information géographique n'échappe pas à ces nouveaux besoins. Une réponse prometteuse à ces objectifs est le développement d'ontologies. Les ontologies permettent la spécification de connaissances agréées par une communauté de personnes et partageables sur le Web. Ce partage nécessite la représentation de la sémantique des informations afin de les rendre compréhensibles à une communauté d'utilisateurs relativement à un domaine ou à une activité. Plusieurs modèles ont été développés pour la modélisation des ontologies qui trouvent leurs fondements dans des domaines comme l'intelligence artificielle, la représentation des connaissances ou les bases de données. Les logiques de description et les outils d'inférences afférents connaissent un nouvel essor pour la modélisation des ontologies. Les modèles conceptuels de bases de données étendus sont aussi une voie prometteuse. L'information géographique est actuellement le parent pauvre de ce foisonnement et les ontologies spatiales restent à construire. Cet article s'attache tout d'abord à décrire les besoins spécifiques des ontologies en terme de modélisation et d'interrogation pour l'information géographique. Puis une étude comparative des modèles basés sur la logique de description et des modèles conceptuels de bases de données s'attache à montrer les forces et les lacunes de chacune de ces approches et les aspects requêtes sont présentés. Enfin le besoin de caractéristiques temporelles et multi-représentation est discuté.

ABSTRACT.

MOTS-CLÉS : SIG, Ontologies, Interopérabilité

KEYWORDS: GIS, Ontologies, Interoperability

1. Introduction

Le développement des technologies Internet ouvre de nouveaux horizons dans le domaine du partage d'informations. L'information géographique n'échappe pas à cette tendance et les besoins de modèles, de méthodes et d'outils adaptés pour représenter, manipuler et partager l'information géographique sur le Web deviennent cruciaux. Ce partage d'informations passe par leur compréhension et une automatisation des processus de navigation, d'extraction et de mise à jour de ces informations. Les ontologies sont une voie prometteuse pour décrire et partager des informations agréées par une communauté de personnes, accessibles sur le Web. Cependant différents types d'ontologies peuvent être considérées selon leurs contenus et leurs utilisations.

Les ontologies de type thesaurus. Ces ontologies, aussi appelées taxonomiques, visent à définir des termes pour fournir un vocabulaire de référence pour harmoniser les noms désignant des données dans différentes applications. Les termes sont organisés en hiérarchies de généralisation/specialisation et des relations de synonymie, composition ... peuvent être spécifiées. Wordnet est sans doute la plus représentative des ontologies de type thesaurus. Elle fournit une ensemble de définitions de termes, structurées en arborescence et des relations entre les termes (synonymes, composition,...)(//www.cogsci.princeton.edu/wn/). On peut considérer ce type d'ontologie comme la première génération d'ontologie. Elles sont essentiellement utilisées de façon exploratoire pour apporter une compréhension commune d'un vocabulaire. Mais les besoins de modélisation et de partage d'informations plus complexes ont conduit à une évolution des ontologies vers des modèles plus riches.

Les ontologies descriptives. Ces ontologies vont au-delà de la définition de taxonomies et visent à modéliser un domaine ou une activité. Elles sont plus proches de la définition de schémas conceptuels de bases de données et s'attachent à modéliser les informations à l'aide de concepts et de relations sémantiquement riches. L'ontologie apporte alors une connaissance sur un domaine ou une activité qui peut être utilisée par un concepteur pour modéliser des applications particulières. Ces ontologies peuvent être utilisées comme un outil de compréhension des structures de données décrivant les concepts du domaine (dans ce cas elles peuvent ne pas posséder d'instances), ou bien, si elles possèdent des instances, elles peuvent être utilisées comme des bases de données, notamment en étant accessibles au public via le Web. De nombreux environnements ont été proposés pour modéliser, interroger et gérer des ontologies. Une brève description de trois d'entre eux, représentatifs des tendances en cours, est donnée ci-après. Il s'agit de KAON [MOT 02], DOGMA [JAR 02], DAML+OIL [HOR 02]. Mais, ces travaux n'offrent pas de concepts spécifiquement dédiés à la modélisation de l'information géographique. Un challenge important est donc l'évolution des ontologies descriptives traditionnelles vers des ontologies permettant réellement de prendre en compte et gérer l'information géographique. Nous appelerons ces ontologies, les ontologies géographiques. Une alternative à l'extension des travaux ci-dessus vers la prise en compte de l'information géographique est d'étudier dans quelle mesure un modèle conceptuel de données spatio-temporelles, avec des outils visuels associés pour la conception du schéma et l'expression des requêtes, peut assurer à ce

type d'ontologie des services de création, de consultation et d'évolution. La modélisation conceptuelle est une technique qui a fait ses preuves, et nous sommes convaincus qu'elle peut rendre d'excellents services dans le domaine des ontologies. Pour cette étude, nous proposons le modèle MADS [SPA 99] avec son environnement d'outils de type CASE [MUR]. Ce modèle est représentatif de l'état de l'art en modélisation spatio-temporelle.

Les ontologies géographiques. Elles recouvrent 1) les ontologies de l'espace plus spécifiquement dédiées à la description des concepts qui caractérisent l'espace comme le point, la ligne, etc... Ces ontologies sont typiquement élaborées par de grands organismes de normalisation. l'OpenGIS (www.opengis.net) propose par le biais de GML (Geography Markup Language) un langage pour structurer et échanger des informations géographiques. En particulier, GML apporte un quasi-standard sur la définition des géométries des objets. 2) les ontologies de domaines géographiques comme une ontologie modélisant les concepts des données hydrauliques, ou bien une ontologie décrivant les concepts des données des réseaux électriques, ... Ce sont des ontologies "métier", développées par une communauté d'utilisateur du domaine concerné. 3) les ontologies spatialisées (ou spatio-temporelles), qui sont des ontologies dont les concepts sont localisés dans l'espace. Une composante temporelle est souvent nécessaire en complément pour la modélisation de l'information géographique, car les applications géographiques manient aussi très souvent des données temporelles, voire spatio-temporelles. MADS offre des facilités pour la modélisation des informations spatio-temporelles, ce point sera discuté dans le paragraphe 5. Les ontologies de l'espace et de domaines géographiques peuvent être de type thesaurus si elles sont limitées à la description de vocabulaire ou bien descriptives si elles incluent une description plus complexe ou sémantiquement plus riche de l'espace ou du domaine considéré. Les ontologies spatio-temporelles sont typiquement descriptives et peuvent s'appliquer à tout contenu nécessitant une modélisation localisée dans l'espace et le temps. C'est ce type d'ontologie géographique que nous considérerons dans la suite de cet article, illustré sur un exemple simplifié d'ontologie modélisant un domaine skiable.

Les ontologies géographiques ont des besoins spécifiques qui sont liés à aux besoins de :

- définir la spatialité à l'aide de types de données spatiaux variés (ligne, point, surface simple, ...), de types d'objets spatiaux (i.e. d'objets ayant des attributs spatiaux), de relations spatiales comme les relations topologiques et/ou des champs continus (raster).
- définir intentionnellement des concepts spatiaux à l'aide d'axiomes contenant des prédictats spatiaux.
- raisonner sur la spatialité des instances i.e. inférer à partir des relations spatiales décrites l'ensemble des relations spatiales valides.

Comme toutes les ontologies, les ontologies géographiques peuvent être utilisées pour l'exploration, mais aussi l'extraction d'informations et au-delà pour l'interopération de SIG. Des travaux récents tels que [HAK 01] montrent l'usage des ontologies pour résoudre l'hétérogénéité sémantique des SIG. Fonseca et al. proposent une architec-

ture pour un système de type ODGIS (Ontology-Driven GIS) [FON 02] pour l'intégration de GIS.

Les différents modèles proposés dans la littérature pour la modélisation des ontologies trouvent leurs fondements dans des domaines tels que l'intelligence artificielle, la représentation des connaissances ou les bases de données. Les travaux issus de l'IA mettent davantage l'accent sur l'aspect raisonnement et techniques d'inférence. Les logiques de description connaissent un essor tout particulier pour le développement de modèles ontologiques [MCG 01]. Par exemple, le système RACER (www.fh-wedel.de/mo/racer/) [HAA 01] implémente la logique de description *SHIQ* [HOR 00] avec les outils d'inférence associés. Le langage OWL (Web Ontology Language : www.w3.org/TR/owl-features/) est une révision de DAML+OIL [HOR 02]. OWL DL introduit la possibilité de définir des axiomes et est basé sur la logique de description. Les logiques de description ne proposent pas à l'origine de facilités pour la modélisation des informations spatiales mais de nombreux travaux ont été développés pour étendre les logiques de description pour prendre en compte les informations spatiales et/ou temporelles. A. G. Cohn et S.M. Hazarika [COH 01] ont proposé une vue d'ensemble des techniques de représentation et des raisonnements sur les informations spatiales qualitatives et notamment la définition et le raisonnement de relations topologiques. V. Haarslev et al. [HAA 98] ont proposé une extension des logiques de description pour le raisonnement spatio-terminologique. Elle intègre le raisonnement sur les descriptions terminologiques et les relations topologiques. Cette approche est détaillée au paragraphe 3.

Le système KAON (<http://kaon.semanticweb.org/>) que l'on peut considérer comme une approche de représentation des connaissances, définit un langage pour la modélisation des ontologies qui inclut la définition de la structure d'une ontologie avec des concepts, des attributs, des relations binaires orientées qui peuvent être symétriques, transitives et avoir des inverses. Des contraintes de cardinalités peuvent être définies et les concepts et les relations peuvent être organisés en hiérarchies de généralisation/specialisation. Les concepts et les relations peuvent être instanciés et des facilités sont offertes pour permettre à un objet d'être à la fois concept et instance d'un concept.

Les systèmes tels que DOGMA [JAR 02][MEE 01] sont issus des bases de données (www.starlab.vub.ac.be/research/dogma.htm). Se basant sur la constatation que, dans un schéma, les contraintes (d'identifiant, de cardinalité ou d'intégrité) sont sujettes à changer d'une application à l'autre, alors que la structure de base de données est, elle, plus stable, les auteurs proposent un modèle en deux parties : une partie (la structure des données) indépendante de toute application particulière, et plusieurs jeux alternatifs de contraintes, chaque jeu constituant une personnalisation de l'ontologie générale pour une application donnée. Une notion de contexte est aussi introduite. L'approche MADS que nous proposons, décrite au paragraphe suivant, s'inscrit aussi dans une approche bases de données.

La définition des modèles et des outils pour les ontologies répondent à des besoins spécifiques qui sont décrits dans le paragraphe 2. Une description du modèle MADS, d'un modèle basé sur l'extension au domaine spatial de la logique de description et

d'éléments de comparaison entre ces deux approches sont présentés au paragraphe 3. Le paragraphe 4 discute des aspects requête. Enfin, les caractéristiques temporelles et représentation multiple sont abordées en section 5.

2. Spécificités des ontologies

Les besoins spécifiques des ontologies peuvent être identifiés selon quatre points : 1) la modélisation des informations et de leurs sémantiques qui nécessite des modèles suffisamment riches et pour l'information géographique, des concepts de description des caractéristiques spatiales à la fois sous forme discrète et/ou continue. 2) le raisonnement pour être capable d'inférer/classifier des informations et vérifier la cohérence des descriptions. 3) les instances pour gestion des données si l'ontologie en comporte et 4) les requêtes pour interroger l'ontologie aussi bien au niveau des instances qu'au niveau du schéma.

Modélisation et raisonnement. Nous avons décrit dans l'introduction différents modèles utilisés pour la description des ontologies. Les justifications de ces approches sont liées aux besoins spécifiques des ontologies en terme d'évolution. Une ontologie est par définition une représentation d'un consensus entre une communauté de personnes qui peuvent être amenées à compléter ou modifier cette ontologie même si elle comporte déjà des instances. Au contraire des bases de données, dont le schéma est défini avant la définition des instances et qui impose un certain nombre de critères de cohérence et de règles d'intégrité incontournables lors de l'insertion, la suppression et la mise à jour des données, la modélisation d'une ontologie doit permettre une certaine flexibilité dans sa définition et son évolution, ce qui signifie notamment que les informations (schéma ou instances) peuvent être incomplètes et que les mises à jour peuvent être faites sans avoir une connaissance complète des informations de l'ontologie. Cela nécessite donc la mise en œuvre d'outils qui permettent de vérifier la cohérence du schéma mais aussi des instances de façon dynamique. Ces spécificités expliquent l'engouement pour les modèles logiques et notamment les logiques de description qui offrent en plus des aspects purement modèle, des techniques d'inférence et de vérification de la validité et de la cohérence d'un ensemble d'axiomes. Par ailleurs, les modèles conceptuels de bases de données profitent d'une expérience largement reconnue dans le domaine de la modélisation du monde réel. Ils sont donc naturellement de bons candidats pour les ontologies mais doivent être étendus/enrichis afin d'offrir la possibilité de définir de nouvelles entités à l'aide d'axiomes et des outils d'inférence pour vérifier la cohérence des informations et la classifier (mécanisme de subsumption).

Instances et requêtes. Les Systèmes de Gestion de Bases de Données apportent toutes les fonctionnalités nécessaires à la gestion des données d'une base. Les ontologies qui gèrent des instances doivent s'appuyer sur ces outils mais la frontière entre schéma et instances est moins clairement définie pour une ontologie, au sens l'évolution du schéma et des instances est soumise à des contraintes plus souples et que l'exploration de l'ontologie (requêtes) se fait aussi bien au niveau des instances qu'au

niveau du schéma. Les langages de requêtes doivent donc être suffisamment riches pour permettre ce type d'interrogation sur le schéma lui-même.

Le paragraphe 3 propose une comparaison des modèles BD et des modèles DL à travers l'étude du modèle MADS et d'un modèle DL étendu au spatial qui a été proposé par Haarslev et al [HAA 98].

3. Modèles ontologiques géographiques : BD versus LD

Ce paragraphe s'attache tout d'abord à présenter brièvement les caractéristiques de MADS en les illustrant à l'aide d'un extrait d'une ontologie modélisant un domaine skiable (figure 1). L'environnement MADS fournit un éditeur de schémas qui permet la construction et la visualisation sous forme graphique des schémas et un éditeur de requêtes pour l'interrogation.

MADS. Le modèle MADS est un modèle conceptuel spatio-temporel qui permet de définir des types d'objets complexes, leurs attributs et leurs domaines, des relations entre les types d'objets, des relations entre les objets comme l'agrégation, les relations topologiques qui contraignent la géométrie des objets liés (disjonction, adjacence, croisement, chevauchement, inclusion, égalité) ou certaines relations de causes et effets pour la modélisation de l'évolution ou l'extension d'objets et la génération d'objets, des contraintes d'intégrité (contraintes de cardinalités, de clés, ...). MADS possède des types abstraits de données spatio et temporel pour la modélisation des phénomènes spatio-temporels. Les types d'objet et de relation peuvent être organisés en hiérarchies de généralisation/specialisation. MADS offre aussi des liens d'héritage spécialisés appelés raffinement, redéfinition et surcharge particulièrement utiles pour la modélisation de la géométrie des objets. Les attributs peuvent être simples ou complexes, mono ou multi valués, ordonnés ou non. Dans figure 1, *DomaineSkiable*, *Piste*, *TronçonPiste* ... sont des types d'objets. *DomaineSkiable* possède un attribut complexe *météo* variable dans l'espace et le temps, noté $f(S,T)$, qui est une fonction associant à chaque point du domaine skiable et à chaque instant du domaine temporel couvert par la base de données, une valeur complexe décrivant la météo à ce point et à cet instant là. Le type d'objet *Piste* possède une géométrie complexe dérivée par union des géométries des tronçons de pistes *TronçonPiste*. Le type d'objet *TronçonPiste* possède une géométrie de type ligne orientée ou surface et deux attributs variables dans le temps pour la qualité de la neige et sa hauteur. Les relations *Comprend*, *Composé* sont des relations de type *agrégation*. Les relations *Départ*, *Arrivée* sont de type topologique adjacent. La relation *Proche* est de type contrainte spatiale définie par l'utilisateur (distance des géométries < 200m). Une description plus complète de MADS peut être trouvée dans [MUR].

MADS est particulièrement bien adapté à la modélisation des informations spatiales comme le montre l'exemple du domaine skiable (figure 1). De plus, MADS possède quelques concepts de dérivation prédéfinis, comme les relations topologiques dont les instances sont automatiquement dérivées des géométries des instances des

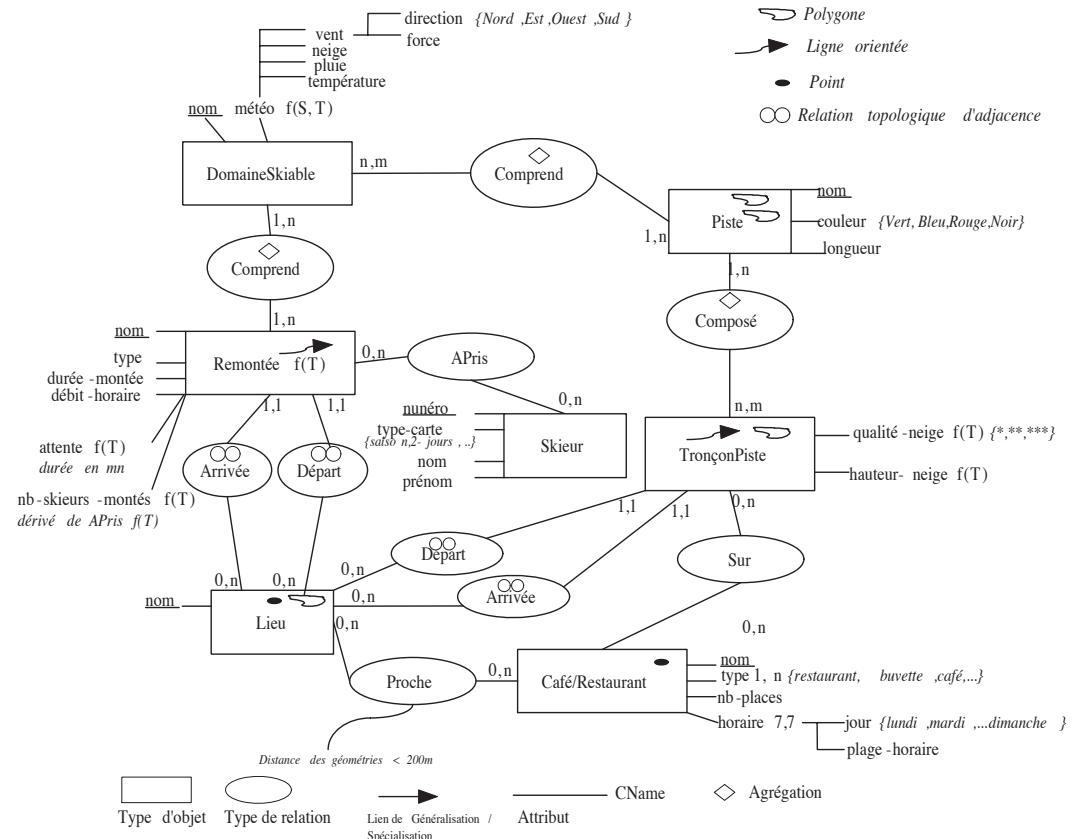


Figure 1. Extrait de l'ontologie "Domaine skiable" en MADS

types d'objets liés. Mais contrairement aux modèles basés sur les logiques de description étendues au spatial, MADS ne permet pas de raisonner sur la spatialité des instances. Ce dernier point est un des objectifs des modèles basés sur les LD étendues au domaine spatial. Nous allons illustrer ce type d'approche à l'aide des travaux qui ont été proposés par Haarslev et al [HAA 98] en reprenant l'exemple de l'extrait de l'ontologie du domaine skiable précédemment décrit en MADS. Puis nous donnerons des éléments comparatifs de ces deux approches.

Logique de description. Dans les systèmes basés sur la logique de description, la base de connaissances comporte une TBox qui contient la description des concepts et des rôles (schéma) et la ABox qui contient les instances. Les concepts peuvent être primitifs ou définis par des axiomes comportant des opérateurs d'union, intersection et négation. Ils peuvent être organisés en hiérarchie de généralisation/specialisation (inclusion). Les rôles sont des relations binaires entre les concepts. Des contraintes de

valeurs (cardinalités) peuvent être données. La définition de domaines concrets permet d'introduire des types de données comme les entiers, les rationnels ou comme nous le verrons ci-dessous les polygones pour la gestion des géométries. Des mécanismes de raisonnement (subsumption) permettent de classifier et de contrôler la validité des informations décrites.

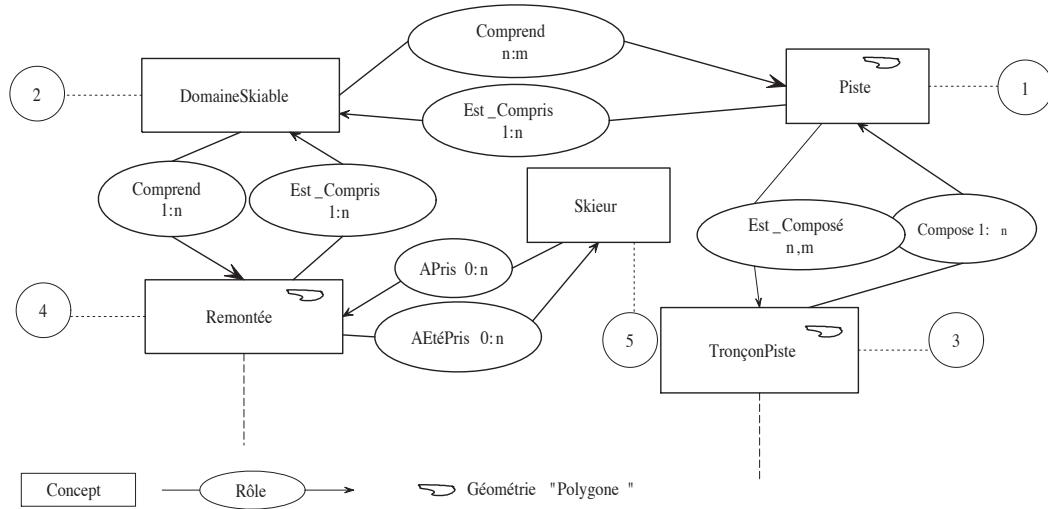


Figure 2. Extrait de l'ontologie "Domaine skiable" en LD (partie 1).

Les figures 2 et 3 proposent de façon visuelle la modélisation d'une partie de l'exemple du domaine skiable en logique de description sans les relations spatiales définies sur les instances. Les géométries des objets sont de type polygone.

Les entités *DomaineSkiable*, *Piste*, *TronçonPiste* ... sont des concepts atomiques. Des rôles binaires relient ces concepts (*Comprend*, *EstCompris*, ...). Des contraintes de cardinalités peuvent être définies sur les rôles. Les attributs sont décrits à l'aide de rôles fonctionnels appelés caractéristiques. Seuls les caractéristiques du concept *Piste* sont détaillées sur la figure 3.

Extension spatiale des LD. V. Haarslev et al ont proposé une extension de la logique de description $\mathcal{ALD}(\mathcal{D})$ appelée $\mathcal{ALDRP}(\mathcal{D})$ [HAA 98] pour raisonner sur les relations qualitatives entre les régions spatiales et sur les propriétés des données quantitatives. L'extension proposée gère les objets spatiaux et leurs relations à l'aide de prédictats (\mathcal{RP} signifie Rôles basés sur des Prédictats). Un domaine concret *Polygone* est introduit pour gérer la géométrie des objets.

Les relations topologiques entre les objets sont décrites à l'aide de huit prédictats élémentaires schématisés dans le figure 4 (égalité, disjonction, adjacence, recouvrement strict, inclusion tangentielle, inclusion stricte et les relations inverses de ces deux derniers).

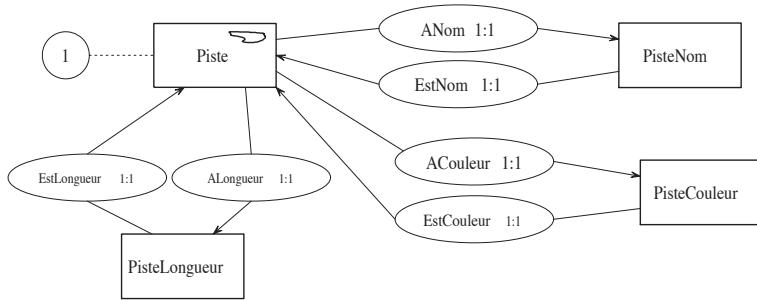


Figure 3. Extrait de l'ontologie "Domaine skiable" en LD (partie 2).

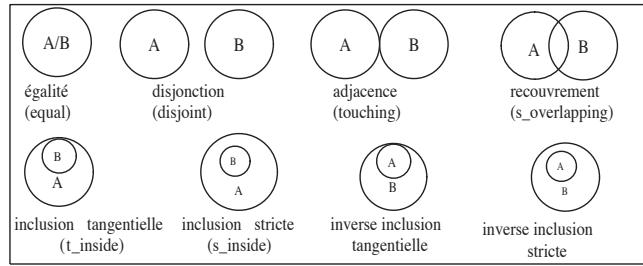


Figure 4. Relations élémentaires entre 2 régions A et B.

nières relations. Une définition formelle de ces relations est donnée dans [HAA 98]. Une relation composite est une disjonction de relations élémentaires.

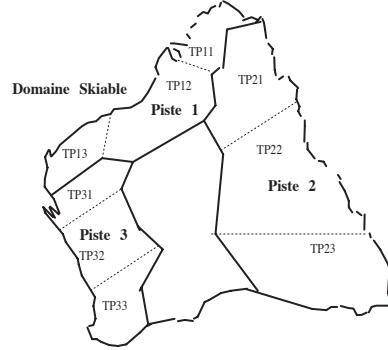


Figure 5. Exemple simplifié d'un domaine skiable.

La figure 5 illustre un domaine skiable (simplifié) avec 3 pistes dont les géométries sont considérées comme des polygones. Des relations topologiques similaires à celles exprimées sur la figure 1 MADS peuvent être décrites à l'aide des prédictats comme :

- l'inclusion des géométries des tronçons de piste dans une piste
 $g_inside(TP11, Piste1)$
où g_inside est une relation définie par $g_inside \equiv t_inside \vee s_inside \vee equal$
- l'adjacence de 2 tronçons de piste
 $touching(TP11, TP12)$.

De nouveaux concepts peuvent être décrits en utilisant l'opérateur de construction de concepts $\exists f.P$. On peut par exemple définir le concept d'une zone de piste facile (*zone_piste_facile*) par l'inclusion de la géométrie des instances de la zone de piste facile dans la géométrie de la piste 1. On suppose que tout objet possède un attribut *has_area* qui décrit sa représentation spatiale polygone.

$$\text{zone_piste_facile} \doteq \exists has_area.g_inside_{Piste1}$$

Le mécanisme de subsumption permet de classifier les concepts ainsi définis.

Plus de détails sont donnés dans [HAA 98].

Comparaison. Plusieurs points de comparaison sont à souligner dans ces deux approches qui concernent 1) la structure des objets 2) les relations spatiales 3) les concepts définis, les vues et la classification des concepts.

Structure des objets. Les modèles conceptuels tels que MADS offrent la possibilité de définir des objets complexes et des relations sémantiquement riches qui donnent une meilleure représentation du monde réel qu'une modélisation à l'aide d'objets simples et de relations binaires qui conduit à un éclatement des entités du monde réel et diminue largement la lisibilité de la description. Les figures 1, 2 et 3 en sont une illustration. En ce qui concerne la géométrie des entités modélisées, MADS offre un panel riche de type de données spatiaux et permet les géométries dérivées comme la géométrie d'une piste qui est l'union des géométries de ses tronçons. L'exemple d'extension de DL choisi, ne propose qu'un seul type de données, le polygone. Ce choix s'explique par la complexité des algorithmes d'inférence sur des types spatiaux variés. La restriction aux seuls polygones permet d'utiliser des algorithmes efficaces mais restreint fortement la richesse de description du modèle. Le modèle DL offre la possibilité de définir intentionnellement des concepts qui sont automatiquement classifiés dans la hiérarchie des concepts grâce au mécanisme de subsumption. Il permet aussi de définir de nouveaux rôles à l'aide d'un opérateur $(\exists f).(f).P$ [HAA 98].

Relations spatiales. MADS permet de définir des relations topologiques au niveau du schéma, ces contraintes étant appliquées/propagées lors de la création des instances. Dans le modèle DL étendu, les relations spatiales sont décrites au niveau des instances à l'aide des prédicts topologiques introduits avec le domaine concret polygone, ce qui rend plus complexe la gestion de instances et ne permet pas d'avoir au niveau du schéma une visibilité des relations topologiques décrites.

Classification des concepts. Un des points forts des approches DL étendu est la possibilité offerte de définir des concepts à l'aide d'axiomes et d'offrir des outils pour les classifier. Dans le modèle DL étendu, ce mécanisme d'inférence a été étendu afin de prendre en compte les relations topologiques définies. Ainsi en utilisant la table

des compositions sur les relations topologiques, des inférences sur ces relations sont réalisées. Plus de détails peuvent être trouvés dans [WES 02]. Les modèles conceptuels n'offrent en général pas la possibilité de définir de nouvelles entités par le biais d'axiomes, ce qui correspond davantage à l'approche BD de non-redondance des informations. Cependant le mécanisme de vues construites à partir d'une requête rejoint par certains aspects la définition de concepts définis. En MADS, la construction d'une vue induit son placement dans la hiérarchie des types d'objets, elle n'est cependant pas considérée comme un nouveau type d'objet. Certains modèles conceptuels proposent un mécanisme de partitionnement qui permet de classifier automatiquement la population d'un type d'objet entre ses sous-types en fonction de prédictats. Par exemple, les prédictats (*Personne.age <20*) partitionnent le type *Personne* en 2 sous-types, *Jeune* et *Adulte*. Une différence fondamentale entre le mode de raisonnement appliqué en DL et celui en général adopté en BD est l'utilisation de l'hypothèse du monde clos pour les BD et l'hypothèse du monde ouvert pour les DL. Ceci reflète la différence entre les deux approches : les BD supposent que toutes les informations sont décrites et les DL supposent que les informations décrites peuvent être incomplètes.

4. Interrogation des ontologies

L'interrogation d'une ontologie, diffère de l'interrogation d'une base de données traditionnelle selon plusieurs points fondamentaux. Une ontologie est en général une source de données complexe, souvent incomplète et évolutive dans le temps. Le simple utilisateur qui souhaite accéder à l'ontologie effectue des recherches exploratoires voire incrémentales pour obtenir les informations qu'il souhaite. Il n'a pas en général la connaissance globale du schéma de l'ontologie et doit donc disposer d'un langage qui lui permette d'interroger à la fois le schéma de l'ontologie pour en découvrir son contenu et ses instances de façon homogène. Les administrateurs de l'ontologie peuvent être, ou bien une communauté de personnes ciblées ou bien un ensemble variable de personnes comme un groupe de travail. Ils peuvent être amenés à étendre et faire évoluer le schéma de l'ontologie et son contenu sans une connaissance préalable complète de cette ontologie et doivent donc disposer de moyens d'interrogation du schéma de l'ontologie à la fois pour l'explorer mais aussi pour s'assurer de sa cohérence. Le raisonnement à la fois sur le schéma et sur les instances pour classifier les concepts et les instances et vérifier leur cohérence est un point crucial pour la maintenance de l'ontologie. Les développeurs ont besoin d'accéder à l'ontologie pour en extraire des informations pour les intégrer ou les utiliser dans des applications. Au-delà de la simple interrogation, ils ont besoin de services de type API pour explorer et extraire des données de l'ontologie. L'interrogation d'une ontologie se caractérise donc par deux aspects importants 1) la possibilité d'interroger ou explorer le schéma aussi bien que le contenu de l'ontologie et 2) le raisonnement sur les concepts et les instances de l'ontologie.

Exploration du schéma. Elle doit permettre d'obtenir des informations sur les concepts existants, les relations liées à un concept, les propriétés des relations (sy-

métrique, transitive, inverse) si elles existent, et tous les éléments du schéma. Certains systèmes comme KAON [MOT 02] ou RACER [HAA 01] permettent ce type d'interrogation sous forme de fonctions d'évaluation ou de requêtes. RACER dispose de requêtes prédéfinies comme *describe-concept*, *describe-role*, *reflexive ?*, *symmetric ?*, *transitive ?*, *feature ?*, *role-domain*, *role-range*, etc ... KAON supporte aussi ce type d'interrogation par le biais de requêtes prédéfinies *Properties_From*, *Properties_To*, *Domain_Concept*, *Range_Concept*, etc ... Les extensions de LD au spatial sont directement issues de systèmes de LD comme RACER et sont donc susceptibles de disposer des mêmes fonctionnalités d'interrogation du schéma mais enrichies par des fonctionnalités d'interrogation des relations ou attributs spatiaux. En MADS [MUR], l'interrogation du schéma peut se faire en définissant le méta-modèle de MADS et en interrogeant cette meta-base à l'aide d'un des langages génériques de MADs (visuel ou algébrique). Cependant une démarche plus naturelle pour l'utilisateur pour interroger le schéma de l'ontologie en MADS, est de le visualiser et de l'explorer à l'aide de l'éditeur de schéma MADS.

Exploration des instances. Les systèmes issus de BD tel que MADS s'appuient sur des langages de requête à la SQL qui offrent une large panel de fonctionnalités pour interroger les instances de la base et permettent aussi de créer de nouveaux ensembles d'objets. Les systèmes issus des LD ou LD étendues disposent en général de fonctionnalités prédéfinies pour interroger les instances décrites dans les ABox. RACER [HAA 01] dispose de fonctionnalités telles que *retrive-concepts-instances*, *retrive-individual-filters*, etc ...

Raisonnement. Certaines requêtes aussi bien sur le schéma que sur les instances peuvent nécessité la mise en oeuvre de mécanismes de raisonnement comme par exemple 1)la classificaton d'un concept dans une hiérarchie, connaître les concepts qui subsument un concept et vice et versa. Ce type de requêtes s'expriment en RACER par des fonctionnalités comme *conceptsubsumes-p*, *role-subsumes-p*,..., 2) les tests de satisfabilité d'un concept (*concept-satisfiable-p*) ou 3) les tests de cohérence des TBox et ABox(*check-tbox-coherence*, *check-abox-coherence*). Les systèmes de LD étendus intègrent en plus des mécanismes de raisonnement sur les terminologies des mécanismes de raisonnement sur les relations spatiales décrites comme les relations topologiques. KAON [MOT 02] supporte uniquement un certain nombre de requêtes telles que connaître les super ou sub concepts d'un concepts, d'un role, etc ... MADS par le biais de définitions de vues permet la définition de nouveaux objets qui sont classifiés dans la hierarchie des objets exitants.

5. Spatio-temporel et multi-représentation

Les ontologies ont vocation à décrire le monde réel et à rendre les informations modélisées accessibles et partageables. Les ontologies géographiques intègrent la dimension spatiale qui est souvent liée à une dimension temporelle pour rendre compte d'informations évolutives dans le temps et dans l'espace. Tout comme une entité peut évoluer dans le temps, elle peut aussi être sujette à différentes representations selon

le contexte dans lequel elle est considérée. L'intégration dans les ontologies géographiques des trois composantes : espace, temps et multi-représentation est un challenge important pour assurer aux ontologies une réelle pertinence. Des nombreux travaux ont été réalisés pour prendre en compte la composante temporelle aussi bien dans le domaine des bases de données que dans les logiques de description. Les approches peuvent différer par les références de temps choisis (instants, intervalles, ...) et par les mécanismes choisis pour introduire le temps (opérateurs temporels, changements d'états, ...). Artale et Franconi proposent dans [ART 03] différentes extensions temporelles de la logique de description. MADS est un modèle spatio-temporel, la caractéristique temporelle est supportée par un type abstrait de données temporel qui permet de définir des attributs estampillés dans le temps pour conserver l'évolution de leurs valeurs. Une hiérarchie de types de données temporels est aussi proposée. Le cycle de vie des objets peut aussi être pris en compte (création, suspension de validité, réactivation, suppression). Un concept peut être défini comme spatio-temporel pour modéliser son évolution dans le temps et ses déformations. La composante représentation multiple est aussi supportée par MADS qui offre la possibilité d'estampiller tout constructeur du modèle (objets/attributs/relations) avec deux informations indiquant un point de vue et une résolution. La notion de point de vue permet d'identifier les informations pertinentes pour dans un certain contexte et pas dans d'autres. La notion de résolution permet de connaître le niveau de détails au-dessous duquel l'objet n'est pas pertinent. Plus de détails sur les aspects temporels et représentation multiple de MADS sont donnés dans [SPA 99], [MUR].

6. Conclusion

Les ontologies géographiques ont un avenir prometteur car elles sont un tremplin pour le partage automatisé d'informations dans des domaines où les données sont complexes à modéliser et onéreuses. Elles doivent jouer un rôle important de référence pour l'interopérabilité des systèmes. Leur développement est conditionné par des critères de réelle utilisation. Nous nous sommes efforcés dans cet article de présenter les différentes approches développées pour la modélisation d'ontologies générales et géographiques et de souligner les similitudes et les différences de ces modèles. Nous avons montré que les modèles conceptuels de base de données spatio-temporelles sont d'excellents candidats pour la modélisation des ontologies géographiques mais qu'ils doivent être enrichis pour permettre la définition intentionnelle de concepts. Ce dernier point est le point fort des modèles issus des logiques de description même si certains systèmes dédiés aux ontologies ne disposent pas de tous les mécanismes d'inférence des LD. L'intégration des composantes temporelles et représentation multiples nous semble un challenge important à relever. Nous avons mis aussi l'accent sur les besoins des ontologies qui diffèrent des besoins traditionnels de bases de données et des systèmes déductifs pour la modélisation et pour l'interrogation. Nous avons actuellement défini et implementé le modèle MADS qui supporte la modélisation spatio-temporelle et la gestion de contextes. Une étape supplémentaire vers les ontologies géographiques est la prise en compte des informations imprécises ou incomplètes. Le développement

d'un serveur d'ontologies s'appuyant sur les services actuellement fournis par MADS développés dans le projet IST MurMur [MUR] est à l'étude.

7. Bibliographie

- [ART 03] ARTALE A., FRANCONI E., « Temporal Description Logics », GABBAY D., FISHER M., VILA L., Eds., *Handbook of Time and Temporal Reasoning in Artificial Intelligence*, World Scientific Publishing (forthcoming), 2003, p. 349-372.
- [COH 01] COHN A., HAZARIKA S., « Qualitative Spatial Representation and Reasoning : An Overview », *Fundamenta Informaticae*, vol. 46, n° 1-2, 2001, p. 1-29.
- [FON 02] FONSECA F., EGENHOFER M., AGOURIS, CÂMARA G., « Using Ontologies in Integrated Geographic Information Systems », *Transactions in GIS Vol 6 N° 3*, 2002.
- [HAA 98] HAARSLEV V., LUTZ C., MÖLLER R., « Foundations of Spatiotemporal Reasoning with Description Logics », A.G. COHN L.K. SCHUBERT S., Ed., *Principles of Knowledge Representation and Reasoning : Proceedings of the Sixth International Conference (KR'98)*, 1998, p. 112-123.
- [HAA 01] HAARSLEV V., MÖLLER R., « RACER System Description », GORÉ R., LEITSCH A., NIPKOW T., Eds., *Proceedings of International Joint Conference on Automated Reasoning (IJCAR'2001)*, Springer-Verlag, 2001, p. 701-705.
- [HAK 01] HAKIMPOUR F., TIMPF S., « Using Ontologies for and Resolution of Semantic Heterogeneity in GIS », *4 th AGILE Conference on Geographical Information Science*, 2001.
- [HOR 00] HORROCKS I., SATTLER U., TOBIES S., « Reasoning with Individuals for the Description Logic SHIQ », MACALLESTER D., Ed., *Proc. of the 17th Int. Conf. on Automated Deduction (CADE-17)*, LNAI 1831, Springer-Verlag, 2000, p. 482-496.
- [HOR 02] HORROCKS I., « DAML+OIL : A reason-able Web Ontology Language », JENSEN C., ET AL., Eds., *EDBT 2002*, Springer-Verlag, LNCS 2287, 2002, p. 2-13.
- [JAR 02] JARRAR M., MEERSMAN R., « Formal Ontology Engineering in the DOGMA Approach », MEERSMAN R., TARI Z., ET AL., Eds., *CooPIS/DOA/ODBASE*, Springer-Verlag, LNCS 2519, 2002, p. 1238-1254.
- [MCG 01] MCGUINNESS D., « Description Logics Emerge from Ivory Towers », *Proceedings of the International Workshop on Description Logics*, Stanford, CA, 2001.
- [MEE 01] MEERSMAN R., « Ontologies and Databases : More than a Fleeting Resemblance », *OES/SEO Workshop Rome*, 2001.
- [MOT 02] MOTIK B., MAEDCHE A., VOLZ R., « A Conceptual Modeling Approach for Semantic-Driven Enterprise Applications », MEERSMAN R., TARI Z., ET AL., Eds., *CooPIS/DOA/ODBASE*, Springer-Verlag, LNCS 2519, 2002, p. 1082-1099.
- [MUR] MURMUR, « MurMur Consortium - MurMur Project : Multi-representations and Multiple resolution in geographic databases », Final Report. <http://lbdwww.epfl.ch/e/Murmur>.
- [SPA 99] SPACCAPIETRA S., PARENT C., ZIMANYI E., « Spatio-Temporal Conceptual Models : Data Structures + Space + Time », *7th ACM Symposium on Advances in Geographic Information Systems (ACM GIS'99)*, 1999, p. 26-33.
- [WES 02] WESSEL M., « On Spatial Reasoning with Description Logics - Position Paper », *Proceedings of the International Workshop in Description Logics 2002 (DL2002)*, Toulouse, France, 2002.