

UNIVERSITÉ DE LIÈGE
Faculté des sciences
Sciences géographiques

Généralisation conceptuelle par modèle

Application aux données du 1/50 000^e de l'IGN

*Mémoire présenté par
Pierre HALLOT
pour l'obtention du titre de
licencié en sciences géographiques
option géomatique et géométrologie*

*Année académique
2004 – 2005*

Je tiens ici à remercier mon promoteur Monsieur Donnay pour son suivi, son soutien et ses conseils avisés lors de la réalisation de ce mémoire.

Je remercie également Monsieur Erpicum et Monsieur Petit pour leurs conseils et leur écoute.

Ce travail, réalisé en collaboration avec l’Institut géographique national, n’aurait jamais été possible sans l’aide et le soutien de Madame Anne Féchir et de Monsieur Jan De Waele tous deux responsables du projet GenMap. Qu’ils en soient vivement remerciés.

Merci à tous les chercheurs du laboratoire SURFACE et plus particulièrement à François Laplance et Christophe Chaidron.

Enfin je remercie mes parents, ma famille et mes proches pour leurs relectures, leur aide, leur soutien, et surtout leur patience.

Table des matières

Remerciements.....	2
Table des matières	3
1 Introduction.....	6
1.1 La généralisation.....	7
1.1.1 Cartes de base et cartes dérivées	7
1.1.2 Généralisation automatique	8
1.1.3 Généralisation conceptuelle par modèle.....	9
1.2 Appariement des données géographiques.....	11
1.3 État de l'art.....	14
1.3.1 Les agences cartographiques nationales étrangères.....	14
1.3.2 Le projet GenMap	16
1.4 Objectif du mémoire	17
1.4.1 Crédit d'un modèle conceptuel des données du 1/50 000 ^e	17
1.4.2 Appariement des modèles du 1/50 000 ^e et de SGISR	17
1.5 Étapes de réalisation	18
2 Description des données disponibles à l'IGN.....	20
2.1 Projet SGISR (Seamless Geographic Information System of Reference).....	20
2.1.1 Contexte	20
2.1.2 Objectif et but.....	21
2.1.3 Catalogue de données SGISR	22
2.1.4 Description du modèle conceptuel	23
2.2 Données géographiques du 1/50 000^e 2^e édition.....	24
2.2.1 TOC (Table Of Contents).....	24
2.2.2 Top50V-Gis – Liste des objets à modéliser	25
3 Élaboration d'un modèle conceptuel pour les données du 1/50 000^e	29
3.1 Prescriptions à l'élaboration du modèle	30
3.1.1 Choix de conserver une même sémantique	30
3.1.2 « Ressemblance » à SGISR	30
3.2 Modélisation des domaines du 1/50 000^e	31
3.2.1 High-Voltage Network	31
3.2.2 Administrative division	34
3.2.3 Landuse et landscape.....	35
3.2.4 Rail transport network	36
3.2.5 Road transport	37
3.2.6 Watercourse.....	38
3.2.7 Network	41
3.2.8 Structure	41
3.2.9 Complexe	43
3.2.10 Altimetry	43
3.2.11 Autre	43
3.3 Modélisation des relations entre domaines.....	44
3.4 Énumérations	45
3.5 Modèle conceptuel des données du 1/50 000^e	46
3.6 Remarques suite à la validation de l'IGN	46

4 Appariement des MCD SGISR et 1/50 000.....	47
4.1 Introduction.....	47
4.2 Assertion de correspondances Inter-schémas (ACI).....	48
4.2.1 Définition générale	48
4.2.2 Conflits détectés et représentés grâce aux ACI	50
4.2.3 Intégration et fonctions.....	56
4.3 Exploitations des ACI pour la déclaration des contraintes de sélection	56
4.3.1 Critères de sélections sémantiques	56
4.3.2 Critères de sélections géométriques	57
4.3.3 Changements de dimensions	58
4.4 Définition des ACI pour l'ensemble des objets du modèle SGISR et 1/50 000.....	59
4.4.1 High-Voltage Network	59
4.4.2 Administrative division	63
4.4.3 Landcover	66
4.4.4 Rail transport network	69
4.4.5 Road transport	72
4.4.6 Hydrography	78
4.4.7 Network	84
4.4.8 Structure	87
4.4.9 Altimetry	92
4.5 Appariement des associations	92
4.6 Conclusions préliminaires sur l'appariement	96
5 Conclusions	97
6 Bibliographie.....	100
Liste des annexes.....	103

*« La carte n'est pas le monde :
elle n'est que le regard qu'un homme a un jour posé sur lui. »*

--- Rémi Caron

1 Introduction

À l'heure actuelle, les données géographiques sont sauvegardées dans des bases de données. Avec des impératifs de production élevés et des temps de mise à jour différents pour chaque représentation cartographique, les agences cartographiques nationales ont été obligées de produire une base de données différentes pour chaque échelle. Les coûts d'acquisitions et de mise à jour des données géographiques étant considérables, un constat s'est rapidement imposé : il serait plus efficace de définir un processus basé sur une seule filière de production de données topo-géographiques de référence, une seule collecte d'informations de mise à jour et un modèle de données permettant des applications à différentes échelles grâce notamment à des outils de généralisation.

L'Institut géographique national belge s'est lancé en 2000 dans le développement d'un système d'information géographique de référence qui doit pouvoir répondre à ce constat. Ce projet est appelé « Seamless Geographic Information System of Reference ». Les représentations cartographiques généralisées dont celles à l'échelle du 1/50 000^e devront être produites à partir de cette base de données de références.

Pour ce faire, deux possibilités s'offrent aux responsables des services en charge de la généralisation (responsables du projet GenMap). Premièrement, il est possible d'effectuer ce que l'on appelle une généralisation automatique des données de référence, cette généralisation automatique permettrait d'obtenir directement les données nécessaires à la représentation cartographique. Une autre solution serait de conserver la représentation actuelle du 1/50 000^e et de ne propager que les mises à jour d'une base de données à l'autre. Dans les deux cas, il est nécessaire de savoir premièrement vers quoi les données ou les mises à jour doivent être généralisées et ensuite comment celles-ci vont y arriver. Ces deux étapes correspondent à ce que l'on appelle une généralisation conceptuelle. Celle-ci qualifie la généralisation effectuée entre deux bases de données géographiques pour passer d'un niveau d'abstraction à un autre.

Le 1/50 000^e ne possède actuellement pas de modèle conceptuel de ses données, ce mémoire consistera en partie à en créer un. Ensuite, pour analyser la manière dont les données peuvent se propager d'une base à l'autre, nous effectuerons un appariement du modèle conceptuel des données de référence avec celui que nous aurons créé. Ce mémoire se découpe en quatre parties. La première définit les différentes généralisations ainsi que les concepts d'appariement. Ensuite vient une brève description des données sur lesquelles nous travaillons. Les deux dernières parties représentent le travail personnel. Elles contiennent la modélisation conceptuelle des données ainsi que l'appariement des modèles par une méthode originale.

1.1 La généralisation

La représentation de l'espace géographique passe par deux questions essentielles : « que veut-on représenter ? » et « comment veut-on le faire ? » [RUAS A. 2004, p. 19]. Il s'agit premièrement de choisir et de sélectionner l'information qui sera représentée pour le besoin exprimé : le « quoi ». Le « comment » définit les règles de représentations graphiques ou informatiques aussi bien pour la géométrie des objets que pour leur description sémantique.

Les diverses utilisations de l'information géographique passent par des échelles d'analyse différentes. Chaque échelle correspond à une abstraction différente du monde réel. Il est donc logique de ne pas retrouver les mêmes informations à chaque échelle. [MULLER J.C. et al. 1995, p.5] & [BUTTENFIELD B. & McMASTER R. 1991, p.60]

La réduction d'échelle ne peut-être effectuée comme une simple homothétie de tous les objets qui composent la carte. En effet, elle deviendrait alors inutilisable pour des raisons de lisibilité graphique et sémantique. Il est nécessaire d'effectuer une généralisation. [RUAS A. et al. 2002, p.31] La généralisation est un processus de synthèse de l'information. Elle consiste en une réduction de la quantité d'information, une mise en valeur de l'information la plus importante pour l'échelle d'analyse considérée en passant par le respect de l'information initiale. Elle ne va pas sans une certaine déformation ou déplacement des objets cartographiés. La généralisation implique une interprétation raisonnée et réclame un « sens géographique ».

1.1.1 Cartes de base et cartes dérivées

Actuellement, l'information géographique est stockée dans des bases de données dites *bases de données géographiques* ou dans des *systèmes d'informations géographiques*. A chaque échelle d'analyse ou d'abstraction correspond un schéma conceptuel différent. Les cycles de mise à jour plus courts des données à petites échelles ont obligé la plupart des agences cartographiques nationales à maintenir des bases de données valables pour différentes gammes d'échelles. [RUAS A. et al. 2002, p.50] Chaque base de données géographiques possède ses propres filières de mise à jour. La représentation à plus grande échelle est appelée carte de base. Les autres échelles plus petites sont dérivées à partir de celles-ci. A l'IGN, le 1/50 000^e dont la première édition est partiellement dérivée du 1/10 000^e (carte de base) possède également sa propre filière de complément terrain. [IGN 2005, *Considérations conceptuelles sur la nouvelle carte de base 1 : 10 000*] & [IGN 2005, *Le SIG et la carte au 1 : 10 000*] Le projet SGISR¹ a pour objectif de remédier à cette situation.

¹ « Seamless geographic information system of reference» dont le développement est en cours à l'IGN. Cfr. 2.1.

La figure 1 nous montre (à droite) la situation existante dans la plupart des agences cartographiques, ainsi que la situation idéale (à gauche) où toutes les représentations dérivent d'une seule base de données géographiques.

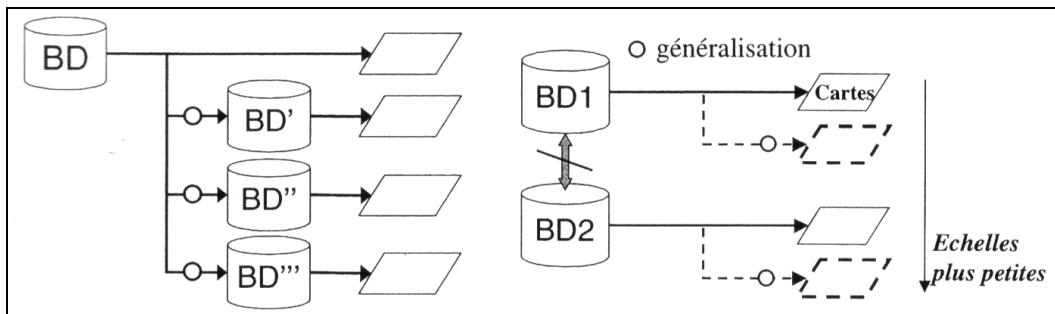


Figure 1 : A gauche, la situation où les bases de données à plus petite échelle sont dérivées à partir d'une base de référence ; à droite, une coexistence de bases de données indépendantes. [RUAS A. et al., 2002, fig. 2.3.]

1.1.2 Généralisation automatique

La généralisation automatique est le processus qui va permettre en partant de données numériques de référence de produire une représentation à plus petite échelle. Elle passe par la mise en évidence de ce qui est porteur de sens (pertinent pour cette échelle). On parle de généralisation automatique lorsqu'il y a mise en œuvre d'algorithmes qui effectuent des opérations de généralisation (élimination-sélection, agrégation, déplacement, simplification de forme, exagération de forme, dilatation, changement de dimensions).

Plusieurs étapes sont nécessaires pour la création d'une représentation à petite échelle à partir d'une base de données valables pour une gamme d'échelles plus grandes. Plusieurs auteurs [BUTTENFIELD B. & MCMASTER R. 1991, p.53], [BEAT P. & WEIBEL R. 1999], [WARE J. & JONES. C, 1998] & [SHEEREN D. 2005, p. 12] ont essayé d'automatiser complètement les processus de généralisation. Cependant, l'ensemble de ceux-ci est d'une complexité telle que peu de résultats globaux ont été obtenus en pratique. Actuellement, la recherche s'axe principalement sur des processus particuliers et des enchaînements à réaliser de manière interactive.

Si une généralisation automatique complète pouvait être mise en œuvre, il serait possible d'obtenir une « carte » à la « carte » en fonction des besoins de l'utilisateur, chaque représentation étant dérivée automatiquement des données de référence. [RUAS A. et al. 2002, p.42] De plus, pour les agences cartographiques nationales, la généralisation automatique permettrait de ne pas devoir systématiquement stocker l'ensemble des données de chaque représentation. Celle-ci pourrait être dérivée d'une échelle de référence à chaque fois qu'il en serait nécessaire.

La recherche de processus de généralisation automatique est freinée par le fait que les plus grands producteurs de SIG-logiciels ne souhaitent pas investir beaucoup dans le domaine de la généralisation automatique. Ils estiment qu'il est plus rentable pour eux de proposer une gamme d'outils de généralisation interactifs. [RUAS et al. 2002, pp. 88-89] La généralisation automatique intéresse un nombre restreint de chercheurs qui commencent à s'organiser pour partager les résultats de leur travail (ICA Workshop par exemple). Les processus de généralisation cartographique ont été divisés en deux groupes : la généralisation conceptuelle et la généralisation structurelle.

1.1.3 Généralisation conceptuelle par modèle

Un des principaux objectifs de la généralisation dite « de modèle » est une réduction contrôlée des données d'un point de vue spatial et thématique. Cette réduction de volume permet de diminuer l'espace nécessaire au stockage de l'information et d'accélérer les processus de traitements effectués sur les données.

A coté de cette réduction de taille, un des objectifs les plus importants de la généralisation conceptuelle est de généraliser des bases de données géographiques à différents niveaux d'abstraction. Cette opération équivaut à dériver un modèle conceptuel de données d'une représentation à partir d'un autre modèle conceptuel de données sources. La généralisation conceptuelle est essentiellement guidée par des questions d'analyse : Quelles sont les tendances ? Quelle est la moyenne spatiale ? A contrario, la généralisation structurelle est principalement axée sur des questions de communication : problèmes de lisibilité, distinction, ... [BUTTENFIELD B. & MCMASTER R. 1991, p.63] La généralisation conceptuelle par modèle est une étape préalable à la généralisation structurelle (nommée cartographique dans la plupart des articles quand elle intervient après une généralisation conceptuelle).

Lorsque l'on parle de base de données géographiques, il n'est pas évident de savoir exactement à quoi on fait référence. Une solution proposée par [BRASSEL K. & WEIBEL R. 1988] est de distinguer base de données géographiques et cartographiques :

- une base de données géographiques regroupe des données dont la position de chaque objet est **conforme** à la précision géométrique et à la granularité de saisie. Une base de données géographiques possède un espace de signification qui est constitué des concepts géographiques décrits dans la base de données. Ces concepts sont décrits par le schéma des données et par les spécifications.
- une base de données cartographiques regroupe des données géographiques dont la position de chaque objet est **proche** de la précision géométrique de saisie mais dont la taille et la position de chaque objet sont conformes à l'échelle mathématique de représentation. Une base de données cartographiques possède un espace de

signification qui est constitué des concepts géographiques décrits dans la base de données. Ces concepts sont décrits par le schéma des données et par les spécifications.

Notons qu'il ne faut pas, suite à ces définitions, croire qu'une base de données cartographiques n'est pas une base de données géographiques. Dans le langage courant néanmoins, ces deux types de bases de données sont souvent appelés *base de données géographiques*.

La figure 2 permet de situer où se trouvent les processus de généralisation conceptuelle par modèle et les processus de généralisation cartographique. Le passage entre deux niveaux de détails, entre deux bases de données géographiques, résulte d'une généralisation conceptuelle par modèle. Le passage d'une base de données géographiques à une base de données cartographiques s'effectue, lui, par une généralisation cartographique.

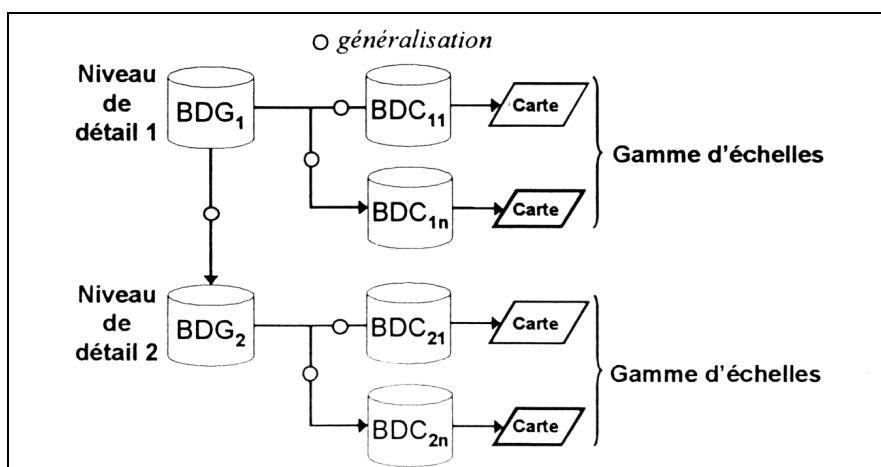


Figure 2 : Processus de généralisation conceptuelle et cartographique [RUAS A., 2004, fig. A.10.]

Les opérateurs de généralisation mis en place lors de ces deux processus ne sont pas les mêmes. Pour une généralisation par modèle, il s'agit essentiellement de sélections, d'éliminations, de regroupements et de changements de dimensions. Tandis que les déplacements, exagérations, typifications, équilibrages, schématisations,... destinés à une meilleure lisibilité relèvent de la généralisation cartographique.

La généralisation conceptuelle, définie dans [NYERGES R. 1985 in RUAS A. 2004, p. 37] comme étant une abstraction de l'information géographique, consiste en la définition d'un nouveau schéma conceptuel ainsi que son instantiation (création des objets matérialisant les classes préalablement définies). Dans le cadre de ce mémoire, où nous disposons d'une base de données initiale décrivant un niveau de détails assez élevés (SGISR), cela se traduit par :

- la définition d'un nouveau schéma conceptuel qui correspond à un autre niveau de détail ;
- le transfert des données ou l'appariement de la base initiale vers la nouvelle base en cours de construction.

Les opérations utiles à la réalisation du nouveau schéma à partir du premier et le transfert des données sont la réification (matérialisation d'un concept par un objet ou une classe), l'association, la généralisation et l'agrégation.

Dans la pratique, la généralisation conceptuelle va consister en la redéfinition de nouvelles classes (principalement la suppression de classes et d'attributs ou l'apparition de classes suite à la généralisation de plusieurs autres) et de nouvelles relations (traduisant l'apparition de nouveaux concepts). Il est également nécessaire de redéfinir les nouvelles règles de sélections sémantique et géométrique des objets présents dans une classe. Par exemple, toutes les entités dont la surface est inférieure à un seuil ne seront pas reprises, les segments inférieurs à une certaine longueur non plus, etc.

De nouveaux concepts peuvent également apparaître, comme par exemple lors de l'agrégation de plusieurs entités de la base de données de référence. C'est le cas notamment des bâtiments qui se regroupent en zones urbaines, ...

Lors d'une généralisation conceptuelle par modèle il est dès lors nécessaire de définir :

- un nouveau schéma conceptuel et ses contraintes :
 - o les nouvelles classes ;
 - o les nouvelles règles de sélection ;
 - o la résolution sémantique des nouveaux attributs.
- une nouvelle description géométrique :
 - o les nouvelles règles de représentation de la géométrie
 - o la nouvelle résolution ;
 - o la nouvelle précision géométrique ;
 - o la nouvelle granularité.

1.2 Appariement des données géographiques

Une fois le nouveau schéma créé, nous avons vu qu'il était nécessaire d'instancier la nouvelle base de données. Dans le cadre de ce mémoire, les données dérivées ne sont pas créées de toute pièce, mais elles existent déjà. Elles seront chargées dans le nouveau modèle conceptuel qui sera lui-même apparié à la base de référence. Il va donc être nécessaire d'effectuer un appariement entre les données des deux bases de données géographiques (base de référence et base généralisée).

L'appariement de bases de données est l'opération qui consiste à exprimer une relation explicite entre les objets des bases qui représentent les mêmes phénomènes du monde réel. [DEVOGELE T. 1997, p.33] Il est nécessaire de faire la distinction entre l'appariement et l'intégration. L'intégration est l'opération qui consiste à fusionner deux bases de données

différentes [BATINI C. et al. 1996, p.327]. L'appariement est quant à lui une étape préalable à l'intégration, mais il est aussi utilisé dans de nombreux autres cas comme par exemple le regroupement de données juxtaposées, la propagation des mises à jour dans une base de données client, le recalage de données sur un référentiel, le contrôle de qualité et la superposition de couches pour fusionner les géométries. [BADARD T. 2000]

L'appariement de deux bases de données géographiques se fait premièrement sur base de leur schéma et ensuite sur base de leurs instances (fig. 3). Dans le cadre de notre travail, la base de données de référence sur laquelle nous travaillons n'est pas encore instanciée, il ne sera donc pas possible d'effectuer un appariement sur les données. Nous distinguons quand même tous les types d'appariement, bien que nous nous focaliserons essentiellement sur l'appariement sémantique (effectué au niveau des modèles conceptuels des données).

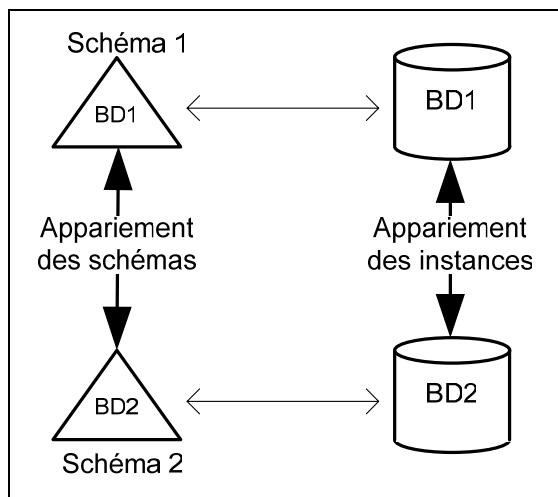


Figure 3 : Appariement de deux bases de données.

Pour déclarer les correspondances au niveau des schémas des bases de données, il est possible d'utiliser le langage des « **Assertions de Correspondance Interschémas** » développé pour des bases de données classiques par [SPACCAPITRA S. et al. 1992, PARENT C. & SPACCAPITRA S. 1996] et adapté aux données géographiques par [DEVOGELE T. 1997]. Ce langage formel permet d'expliciter, en vue d'une intégration, les correspondances qui existent entre deux schémas de bases de données. Il est enrichi pour permettre de travailler sur des données géographiques. Nous donnerons une description plus détaillée de ce langage dans le chapitre 4 destiné à l'appariement.

Une fois que les classes en correspondance sont identifiées, il est plus aisément de détecter quels éléments (les instances) sont en correspondance. L'appariement des données semble assez aisément pour l'homme quand il superpose deux jeux de données. Comme on peut le voir sur la figure 4, on peut « intuitivement » retrouver les correspondances sur base de la notion de « ressemblance ».

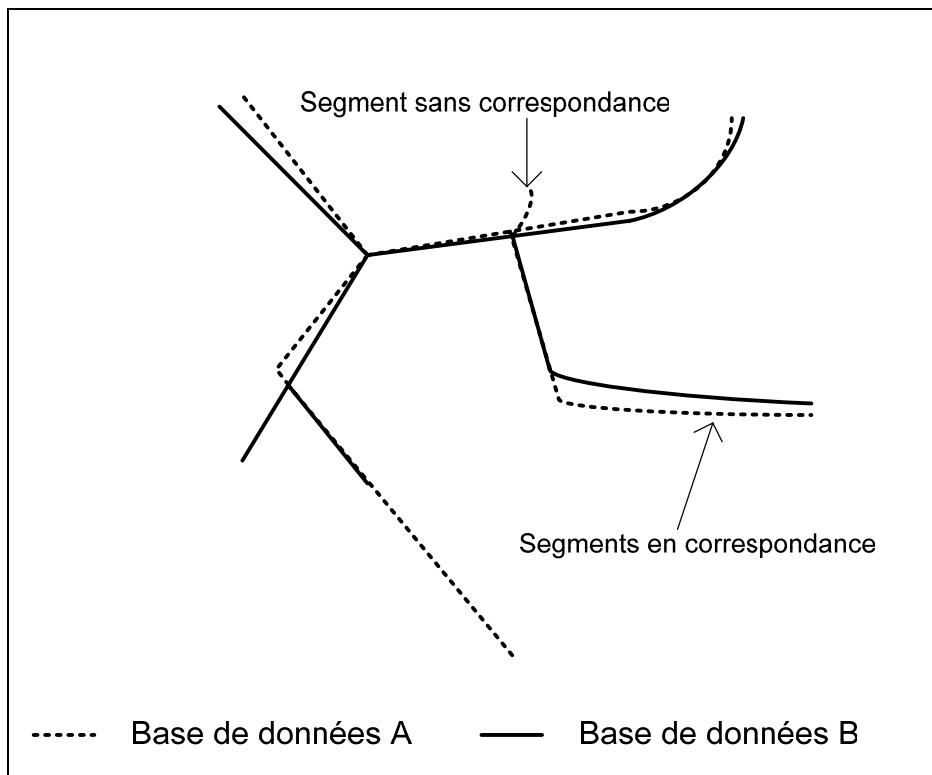


Figure 4 : Appariement de deux jeux de données superposés.

Pour rendre l'appariement automatisable, la difficulté est de traduire en algorithmes la notion de ressemblance que nous effectuons visuellement assez facilement. Cette difficulté fait que l'on se tourne généralement vers un processus semi-automatique. Chaque outil développé est spécifique à un problème donné, ils s'appuient sur différentes caractéristiques des objets. On en distingue trois types :

- l'**appariement sémantique** est similaire au mécanisme d'intégration de données classiques (les objets sont appariés sur base d'un identifiant commun) ;
- l'**appariement topologique** utilise les relations topologiques entre les différents objets pour apparier les données ;
- l'**appariement géométrique** consiste à apparier les données géographiques sur base de leur localisation et de leur forme.

Le laboratoire du COGIT (Conception Objet et Généralisation de l'Information Topographique) de l'IGN France a développé de nombreux outils permettant un appariement quasi automatique des instances. Il a notamment été utilisé pour l'étude de l'extraction des mises à jour entre deux bases de données géographiques [BADARD T. 2000] et l'étude de la cohérence de l'intégration de bases de données [SHEEREN D. 2005].

1.3 État de l'art

Nous allons dans la suite du présent chapitre passer en revue quelques démarches suivies dans plusieurs agences cartographiques nationales pour effectuer la généralisation des diverses représentations cartographiques. Le cas que nous allons traiter en Belgique est assez particulier dans le sens où l'IGN conçoit actuellement un « nouveau » système d'information géographique. Nous verrons plus loin que cela nous permet d'effectuer une nouvelle démarche un peu différente de celle des autres pays.

1.3.1 Les agences cartographiques nationales étrangères

L'IGN France possède une carte de base à l'échelle de 1/20 000. Elle a également comme mission de réaliser un jeu de quatre bases de données constituant un « référentiel spatial à grande échelle » (RGE). Ce référentiel contient quatre composants qui correspondent à des produits de l'IGN. Il s'agit d'une base de données images (BD Ortho®), d'une base de données topographiques (BD Topo®), d'une base de données foncières (BD Parcellaire®) et enfin d'une base de données d'adressage (BD Adresse®). La BD Topo®, d'une précision de l'ordre du mètre, est produite par départements et répartie en dix domaines indépendants : voies de communications routières, voies ferrées, ... [IGN France 2005, *Descriptif Technique de la BD Carto®*] & [IGN France 2005, *Descriptif Technique de la BD Topo® Pays v1.2*]

D'autres bases de données sont également constituées à des échelles conceptuelles plus petites, comme par exemple la BD Carto®. Plusieurs thèses [DEVOGELE T. 1997], [SHEEREN D. 2005] & [BADARD T. 2000] développées au laboratoire COGIT proposent des moyens d'intégration de bases de données géographiques en vue de représentations multiples,... Ceux-ci sont nécessaires pour des domaines divers comme la cartographie électronique multi-échelle, la propagation des mises à jour, etc. Un processus de généralisation complet a par exemple été effectué entre les bases de données BD Carto® à l'échelle conceptuelle du 1/50 000^e vers une nouvelle base Top100 à l'échelle du 1/100 000^e. [NEUFFER D. et al. 2004, pp.19-21] Cependant toutes les autres bases de données restent indépendantes et sont produites de manière séparée. [SHEEREN D. 2005, p. 13]

L'agence cartographique britannique, l'Ordnance Survey a développé un système d'information géographique sur l'ensemble de son territoire. Celui-ci est appelé OS MasterMap. Il inclut des données topographiques de chaque objet géographique : bâtiments, routes, cabines téléphoniques, ... Chaque objet possède un identifiant unique appelé TOID (Topographic Objetc IDentifier). La base de données est destinée à une représentation à une échelle du 1/1 250^e pour les zones urbaines et du 1/2 500^e pour les zones rurales. [Ordnance Survey 2005, *OS MasterMap Userguide v2.1*] Lors du développement de l'OS MasterMap a été introduit la possibilité de dériver automatiquement des représentations

à plus petite échelle se basant sur la philosophie « capture once, use many times » (digitaliser une fois, utiliser plusieurs fois). [REVELL P. 2004] Des outils de généralisation automatique sont en cours de développement notamment pour créer une représentation à l'échelle du 1/50 000^e automatiquement. La plate-forme qui a été utilisée pour développer ces outils de généralisation est « Clarity ». Celle-ci est basée sur la technologie Agent. [DUCHENE C. 2003]

Comme dans le cas de la France, on peut constater qu'il y a une certaine volonté de créer des bases de données de référence possédant une filière de mise à jour unique et de dériver les autres représentations à partir de celles-ci.

La Pologne qui possède une carte de base à l'échelle du 1/10 000^e est actuellement en cours d'élaboration d'un processus lui permettant d'en dériver sa représentation à l'échelle du 1/50 000^e. L'agence cartographique utilise l'environnement de travail DynaGEN© d'Intergraph©. [CHYBICKA I. et al 2004]

Il est intéressant de remarquer que même certains des logiciels effectuant de la généralisation de données proposent également d'effectuer une généralisation de modèle avant l'utilisation des algorithmes de généralisation structurelle. [NEUFFER D. et al. 2004]

En Allemagne, chaque « Land » est chargé d'effectuer lui-même l'acquisition des données en fonction d'un modèle et d'un catalogue commun. Cependant, celui-ci étant interprété de diverses manières, il est nécessaire d'harmoniser les différentes vues avant la généralisation en un modèle à une échelle conceptuelle du 1/50 000^e. La généralisation du modèle de référence (Base-DLM) vers le 1/50 000^e (DLM50) fait l'objet du projet ATKIS qui inclut les étapes suivantes : homogénéisation des données, sélection des classes, sélection d'objets individuels en fonction de valeurs d'attributs, changement de géométries, ... Ces étapes correspondent à celles de la généralisation conceptuelle par modèle. [MULLER J.-C. et al. 1995]

En 2000, le KMS (Agence cartographique du Danemark) a commandé à Laser-Scan© le développement d'un schéma de généralisation basé sur la technologie AGENT. La généralisation porte essentiellement sur les bâtiments. Un long processus de préparation des données est préalablement nécessaire. [NEUFFER D. et al. 2004, pp.19-21] & [KMS 2004, *An Update on Activities at KMS*].

Les quelques exemples que nous venons de décrire rapidement nous montrent qu'actuellement la plupart des agences cartographiques élaborent des processus permettant la généralisation la plus automatique possible de leurs bases de données vers des échelles conceptuelles plus petites. La majeure partie des ces projets intègrent une phase de généralisation conceptuelle avant la généralisation structurelle.

1.3.2 Le projet GenMap

L'IGN Belgique dispose actuellement de trois bases de données. La première à l'échelle conceptuelle du 1/10 000^e qui contient un inventaire de base du territoire et qui est intéressante pour des analyses détaillées (TOP10V-Gis). La deuxième à l'échelle du 1/50 000^e qui est une synthèse de la première et destinée à des applications régionales et militaires (TOP50V-Gis). (Rappelons que l'échelle du 1/50 000^e est l'échelle de l'OTAN). Et enfin, la troisième qui est à l'échelle du 1/250 000^e donne un aperçu des grandes structures. [IGN 2005, *Structure et codage des données TOP10V-GIS et TOP50V-GIS*]

Les données de la première édition du 1/50 000^e ont été obtenues par généralisation soit à partir de la carte numérique au 1/10 000^e, là où elle était disponible, soit à partir de la carte analogique au 1/25 000^e. De plus, des compléments terrains particuliers ont été effectués pour cette échelle. La mise à jour des données au 1/50 000^e a débuté en 2001. Cette deuxième édition sera terminée en 2007.

En 2001, l'IGN a décidé dans le cadre d'un plan stratégique quinquennal, de mettre en place un système d'information géographique appelé « Seamless Geographic Information System of Reference » (SGISR). Cette base de données topo-géographiques doit être le support futur des représentations aux échelles du 1/10 000^e et 1/50 000^e. Il n'y aura plus de filière de mise à jour propre aux données du 1/50 000^e. Deux options seront alors possibles pour produire des cartes à moyenne et petites échelles à jour : la généralisation automatique ou la propagation des mises à jour (fig. 5). Comme nous l'avons montré précédemment, la généralisation automatique complète d'une représentation cartographique n'étant pas encore possible, c'est l'option de la propagation des mises à jour qui a été choisie en raison des impératifs de production.

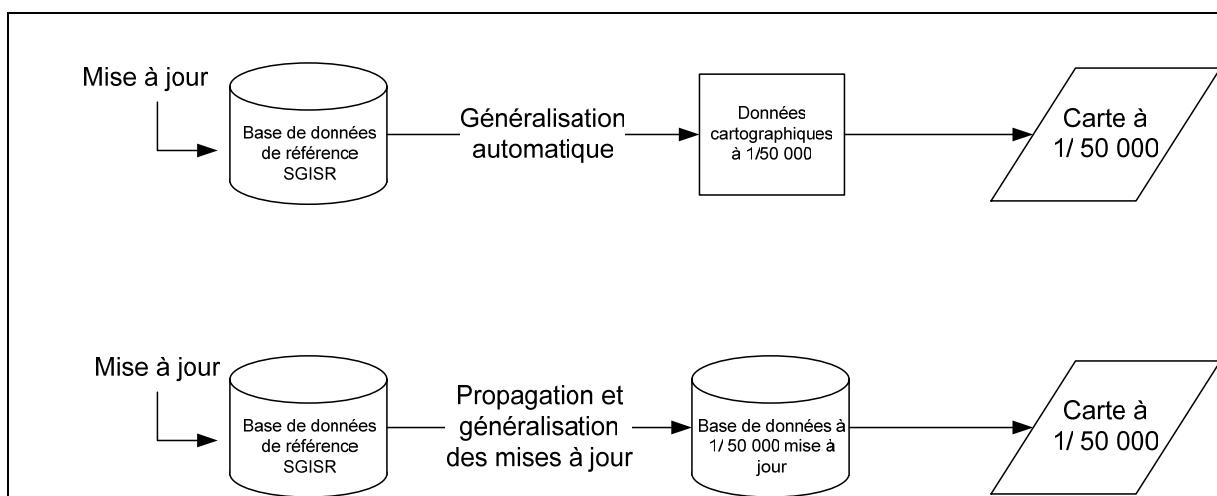


Figure 5 : Production de carte par généralisation automatique au dessus et par propagation des mises à jour en dessous.

La généralisation automatique permettrait d'obtenir une représentation sans devoir sauvegarder le résultat pour des traitements. Il n'y aurait pas de BDG à l'échelle du 1/50 000^e, les données de celles-ci étant directement généralisées de la BDG de référence à chaque fois qu'il le serait nécessaire. La propagation des mises à jour nécessite une BDG à l'échelle du 1/50 000^e car seules les mises à jours sont propagées et généralisées d'une à l'autre.

1.4 Objectif du mémoire

L’Institut Géographique National belge (IGN) dispose actuellement d’un jeu de données à une échelle conceptuelle du 1/50 000^e. Comme nous l’avons vu dans l’état de l’art, la conception du projet SGISR impose que toutes les représentations cartographiques dérivent de cette nouvelle base de données topo-géographiques de référence. Dans un premier temps, les mises à jour introduites dans la base de données de référence seront propagées dans les bases de données dérivées.

L’objectif de ce mémoire va donc consister à effectuer une généralisation conceptuelle par modèle des données de référence pour arriver à l’échelle conceptuelle du 1/50 000^e. En d’autres termes, nous allons montrer pour chaque objet de référence comment il dérive vers le 1/50 000^e. Ceci permettra de lier les modèles conceptuels des données aux différentes échelles pour dans un premier temps propager les mises à jour et peut-être plus tard généraliser automatiquement. Cet objectif se divise en deux étapes.

1.4.1 Crédit d’un modèle conceptuel des données du 1/50 000^e

La généralisation conceptuelle nécessite de créer un nouveau modèle et de définir les nouvelles contraintes sémantiques et géométriques de représentation. Dans notre cas, la représentation existant déjà, il va donc être nécessaire de créer un modèle qui conserve la même sémantique que la carte actuelle. Cela passe par un apprentissage approfondi de l’ensemble des objets présents à cette échelle et des contraintes de représentations.

1.4.2 Appariement des modèles du 1/50 000^e et de SGISR

Le second objectif est d’étudier les deux modèles SGISR et 1/50 000 pour vérifier s’il est possible de dériver totalement les objets de l’un vers l’autre. Ceci constitue une étape de la généralisation conceptuelle avant l’instanciation de la nouvelle base de données. Dans notre cas, la BDG SGISR n’est pas encore instanciée. Il nous faudra dès lors trouver le moyen de vérifier la possibilité de généralisation uniquement à l’aide du modèle, les algorithmes d’appariements travaillant sur les instances ne pouvant être mis en œuvre.

De plus, il est nécessaire de trouver le moyen de reconstruire tous les objets qui ne seraient pas directement dérivables. Il est également demandé d'exprimer d'une manière explicite l'ensemble des critères de sélection sémantique et géométrique pour les objets de la nouvelle base. On pourra donc montrer de quelle manière un objet évolue de SGISR vers le 1/50 000^e.

La figure 6 situe l'objectif du travail. La BDG SGISR et la BDC Top50V-Gis existent déjà. La démarche s'effectuera dans deux sens. Premièrement, la création du modèle conceptuel des données du 1/50 000^e suit la flèche 1, ensuite l'appariement des deux modèles et l'énonciation des règles de représentation suivent la flèche 2.

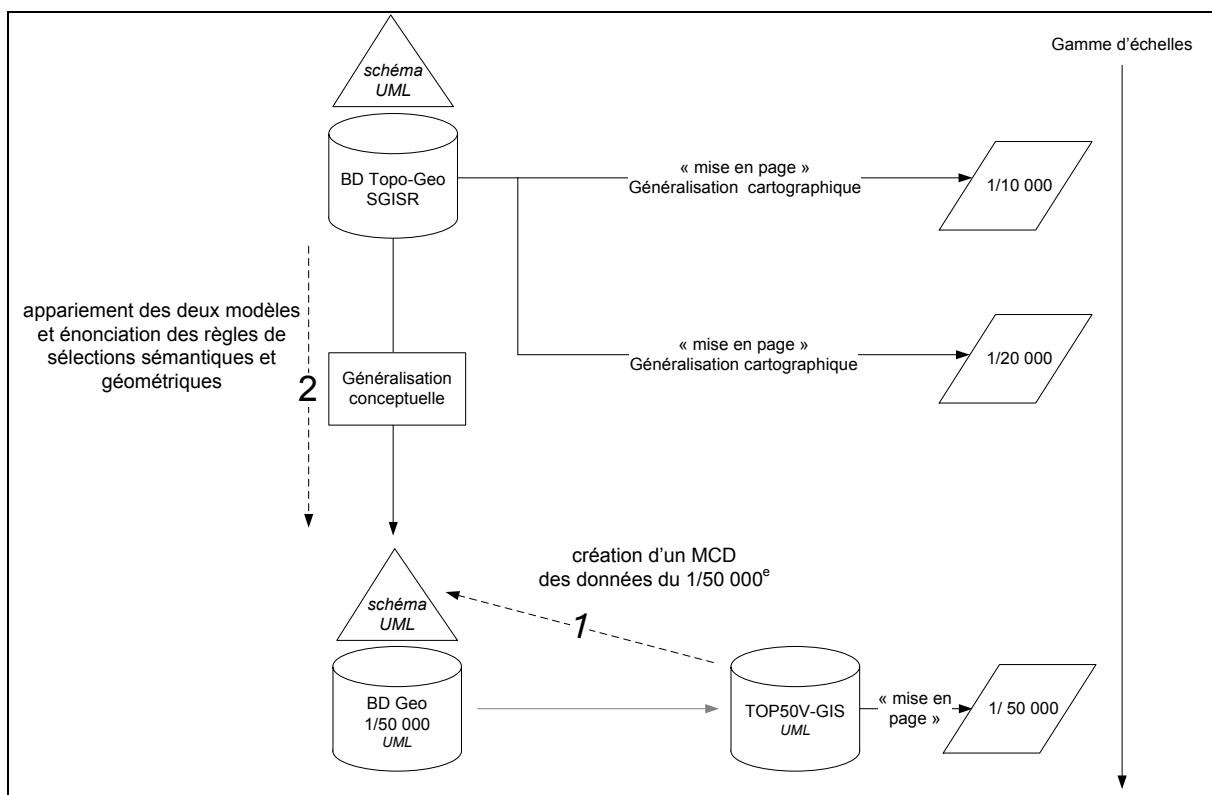


Figure 6 : Objectif et méthodologie du travail.

1.5 Étapes de réalisation

Nous commencerons par décrire l'ensemble des données disponibles à l'IGN nécessaires à notre travail. Il s'agit du catalogue des données de SGISR, du TOC (Table Of Contents, cfr. 2.2.1) et de la description du contenu des données numériques au 1/50 000^e.

Pour être sûr d'avoir une liste exhaustive de l'ensemble des données nécessaires à la représentation, nous modéliserons les classes du TOP50V-Gis (SIG géo-relationnel actuel). Ensuite, nous créerons en collaboration avec l'IGN un modèle conceptuel des données du 1/50 000^e. La modélisation s'effectuera dans le même langage que celui utilisé pour SGISR,

c'est-à-dire l'UML (Unified Modeling Language). [ROQUES P. 2004] & [NAIBURG E. & MAKSIMCHUCK R. 2002]

Le modèle ne sera pas directement élaboré pour l'ensemble des données, mais bien pour chaque domaine (découpage thématique de l'ensemble des objets composant la carte selon un thème). Il s'agira ensuite de retrouver l'ensemble des relations qui existent entre ceux-ci.

L'étape suivante consistera à trouver un moyen de représenter les correspondances entre les deux schémas et d'exprimer les règles de sélections sémantique et graphique. Pour les objets qui ne seront pas en correspondance, il sera nécessaire d'essayer de trouver une solution pour continuer la représentation au 1/50 000^e.

Enfin, la dernière étape, qui s'effectuera également tout au long du travail, consistera à faire vérifier la démarche par l'IGN pour respecter ses souhaits et contraintes.

2 Description des données disponibles à l'IGN

Dans le présent chapitre, nous allons passer en revue les données nécessaires à la réalisation de notre objectif et en donner une description. Nous parlerons principalement du SGISR ainsi que du 1/50 000^e sous format numérique.

2.1 Projet SGISR (*Seamless Geographic Information System of Reference*)

2.1.1 Contexte

Le projet SGISR s'inscrit dans un plan stratégique quinquennal de l'IGN intitulé « Stratégie pour Aujourd'hui et pour Demain ». [IGN 2002, *Schéma directeur du projet SGISR – version 1.1*] Ce plan d'action comporte 5 projets regroupés en deux catégories :

- 1 projet de production :
 - Le projet **10-2005** qui consiste à terminer pour fin 2005 la production des données vectorielles TOP10V à l'échelle conceptuelle du 1/10 000^e.
- 4 projets de développement :
 - **SGISR** – Seamless Geographic Information System of Reference
Nous décrirons ce projet plus en détail dans la suite du chapitre.
 - **DEM** – Digital Elevation Model
Qui a pour but d'élaborer un modèle numérique de terrain continu sur tout le territoire national et utilisable pour des applications intégrées aux informations de base de l'IGN.
 - **AGN** – Active Geodetic Network
Ce projet vise à préparer le remplacement progressif du réseau géodésique et du réseau de nivellation existant par un réseau de stations permanentes de référence pour établir un réseau géodésique et un réseau de nivellation actif interfacés avec le réseau EUREF.
 - **GID** – Geographic Information Distribution
L'objectif de ce projet est d'améliorer la distribution des informations à caractère géographique produites par l'IGN.

2.1.2 Objectif et but

Un des résultats escomptés dans le cadre du projet est la mise en place d'un SIG permettant la gestion des données topo-géographiques de référence de l'IGN.

Durant les dernières années, des études ont porté sur l'amélioration de la structure de distribution des données des SIG TOP10V et TOP50V (données vectorielles des deux SIG topographiques de l'IGN aux échelles conceptuelles du 1/10 000^e et du 1/50 000^e). Cela a abouti à la création d'une nouvelle structure SIG commune pour la distribution des données TOP10V et TOP50V. Il s'agit de la structure [IGN 2005, *Structure et codage des données TOP10V-GIS et TOP50V-GIS*].

Dans ce contexte d'évolution des méthodes de production et de structuration des données, des concepts et des outils associés aux SIG-logiciels, une idée maîtresse s'est rapidement imposée : « Dans l'avenir, la collecte des informations de mise à jour doit être unique, complète, centralisée et indépendante de l'échelle de production des données ». [IGN 2000, *Schéma directeur du projet SGISR – version 1.1*]

Cette réflexion a abouti à la remise en cause de la situation actuelle où deux SIG topographiques (TOP10V et TOP50V) existent parallèlement. Elle a conduit aux deux constatations suivantes :

- Il serait plus efficace de définir un processus basé sur une seule filière de production de données topo-géographiques de référence, une seule collecte d'informations de mise à jour et un modèle de données permettant des applications (civiles et militaires) aux échelles du 1/10 000^e, 1/20 000^e et 1/50 000^e grâce notamment à des outils de généralisation ;
- Il serait dès lors probablement indispensable de gérer ces données topo-géographiques de référence de façon continue dans l'espace plutôt que par découpage cartographique.

Les données topo-géographiques de référence intègrent les données vectorielles TOP10V et TOP50V, les données toponymiques, les données altimétriques et les données d'adressage.

C'est ce qui est exprimé par la notion de « système d'information géographique continu de référence », le « Seamless Geographic Information System of Reference » : **SGISR**.

Ce projet a donc pour objectif d'élaborer et de mettre en place :

- Une filière complète de collecte, de traitement et d'intégration de données de mise à jour dans les données topo-géographiques de référence ;

- Un système d'informations géographiques centralisé et continu spatialement sur le territoire national, pour le stockage, la gestion, l'édition et la diffusion des données topo-géographiques de référence ;
- Les outils nécessaires à la réalisation, à l'IGN, d'applications (vectorielles et matricielles) aux échelles conceptuelles comprises entre le 1/10 000^e et le 1/50 000^e à partir des données topo-géographiques de référence.

La méthode de conception choisie par le groupe de travail SGISR est la méthode MECOSIG (Méthode de Conception de SIG) définie par [PANTAZIS D. & DONNAY J.-P. 1996]. Cette méthode se veut ouverte, c'est-à-dire qu'elle est susceptible d'être utilisée dans divers environnements, tout en ayant une structure permettant un enrichissement ultérieur éventuel.

2.1.3 Catalogue de données SGISR

Une étape de la conception de SGISR est l'élaboration d'un catalogue de données. Celui-ci a été réalisé selon la norme ISO 19110. Nous avons eu l'occasion de consulter une version sous la forme d'une base de données Access©. Ce catalogue reprend l'ensemble des données qui apparaîtront dans SGISR. Il nous a donc été nécessaire d'étudier son contenu lors de l'étape d'appariement pour rechercher l'ensemble des données nécessaires à l'élaboration du 1/50 000^e. La figure 7 représente un extrait de celui-ci.

On peut y voir que les données sont rassemblées en domaines différents : réseau routier, réseau de chemin de fer, réseau haute tension,... Nous utiliserons également cette notion de domaine par la suite pour structurer notre travail.

Le catalogue est divisé en trois parties principales. Premièrement, les « Features » (Objets) pour lesquels sont systématiquement repris : une définition française et néerlandaise, un code, une liste d'attributs, un type géométrique, les relations que cet objet entretient avec les autres objets. Dans une seconde partie sont décrits tous les attributs avec également une définition et la liste de leurs valeurs possibles. Enfin, la dernière partie utile pour notre travail reprend chaque valeur d'attribut avec son code ainsi qu'une description textuelle. Sont également reprises dans le catalogue, les associations entre objets.

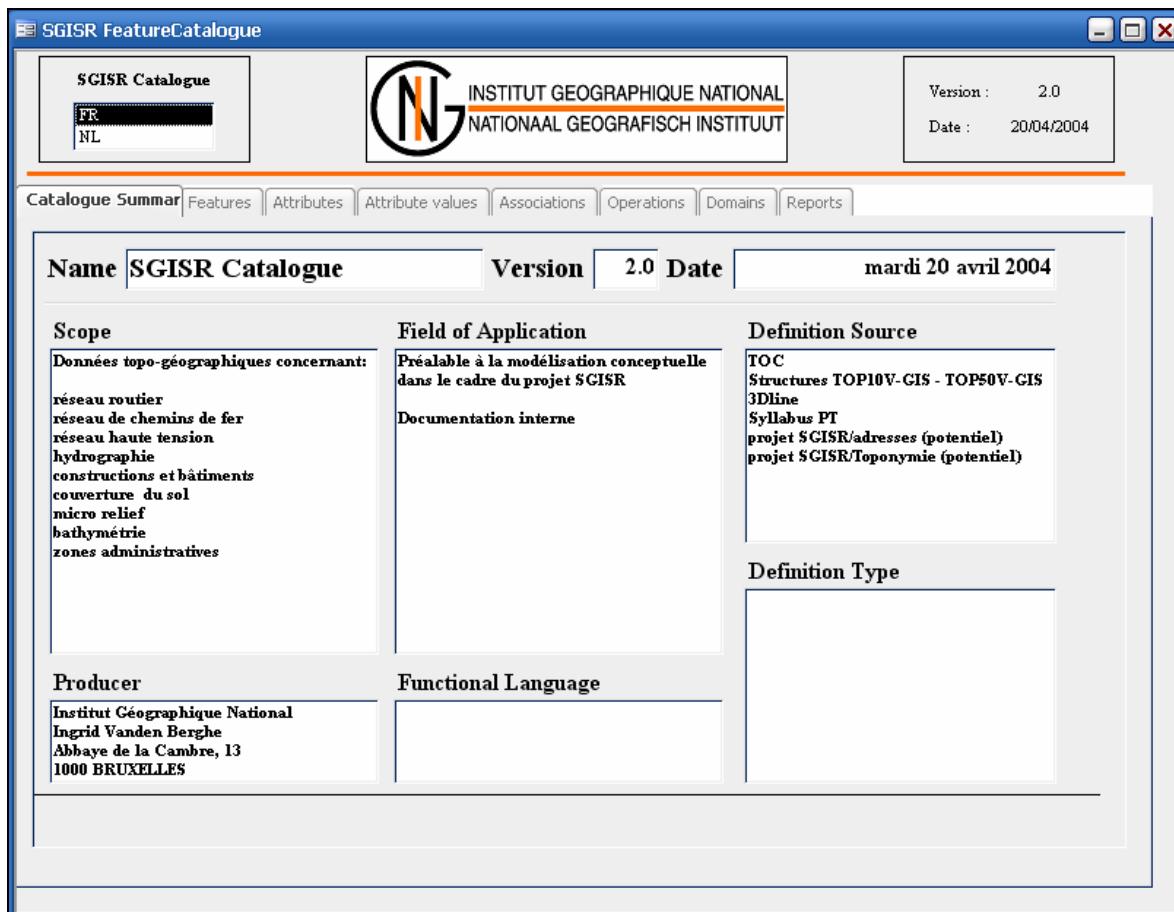


Figure 7 : Extrait du catalogue des données SGISR [IGN 2004, *Feature catalogue Main Version*].

2.1.4 Description du modèle conceptuel

Le modèle conceptuel des données de SGISR est divisé en domaines qu'il est possible de rapprocher de la notion de « package » en UML. Dans le cadre de son stage de fin d'études, [PRUVOST J.-Ch. 2004] a contribué à élaborer le MCD SGISR en se basant sur le catalogue de données préalablement établi.

Lors de l'établissement du MCD, la topologie et plus particulièrement les relations topologiques ont suscité de nombreuses questions. En effet, pour concevoir une base de données géographiques la plus complète possible, il aurait été nécessaire de faire apparaître toutes les relations topologiques possibles entre les objets. Seulement, la plupart des relations topologiques présentes au niveau du schéma, sont de type « intersecte », « traverse », etc. L'IGN a donc décidé de laisser cette partie aux logiciels de gestion du SIG qui intègrent des méthodes permettant les requêtes topologiques ou géographiques à partir des coordonnées des objets présents dans la base. Des matrices de relations topologiques peuvent également être utilisées par la suite.

Une étude approfondie du modèle de SGISR a été nécessaire pour se représenter l'organisation des données dans le but de créer un modèle des données du 1/50 000^e assez proche pour faciliter l'appariement.

Une version réduite du MCD est fournie en annexe 1.

On peut y voir que le MCD est divisé en 6 domaines :

- **Administrative division** qui représente les limites administratives du territoire, ainsi que les zones maritimes et les bornes frontières.
- **Landcover** qui décrit l'utilisation du sol ainsi que les détails caractéristiques du paysage.
- **Structure** qui regroupe les différentes constructions et zones du territoire. Les constructions sont les bâtiments, les éléments construits du réseau,... Les zones quant à elles sont par exemple les domaines militaires, les zones de connexion du réseau, les terrains de sport, les stations d'essence,...
- **High-voltage network** qui englobe tous les objets du réseau à haute tension.
- **Microrelief and bathymetry** représente les données du microrelief ainsi que la bathymétrie. Notons que l'altimétrie est gérée dans le projet DEM.
- **Network Hydrography, Rail transport, Road** qui est le domaine le plus important dans le MCD puisqu'il reprend l'ensemble des réseaux hydrographique, routier et de chemin de fer.

En plus de ces 6 domaines sont définis les « data types » qui énumèrent l'ensemble des valeurs possibles que peuvent prendre les attributs ainsi que leur code. Ce sont les mêmes que dans le catalogue des données SGISR.

2.2 Données géographiques du 1/50 000^e 2^e édition.

2.2.1 TOC (Table Of Contents)

Le projet TOC qui a débuté en 2001 au sein de l'IGN a été un préalable au projet SGISR. Comme le français et le néerlandais sont utilisés couramment au sein de l'IGN, plusieurs problèmes de traduction existaient. Ainsi, certaines structures qui n'existent que dans une partie linguistique du pays portent des noms dans la langue locale. [IGN 2001, *La charte du projet TOC*]

L'objectif de ce projet a été d'établir une base de données commune des éléments topographiques et cartographiques utilisés dans les filières de production des données au 1/10 000^e et au 1/50 000^e. Le projet consiste à leur donner un nom et une définition unique et ce, en néerlandais et en français.

Il a été nécessaire d'être exhaustif pour mener à bien le projet. Ainsi, chaque élément repris dans les légendes des cartes devait être repris dans le contenu du TOC. De plus, pour éviter toutes ambiguïtés, une relation univoque est nécessaire entre chaque objet et sa définition. Enfin, il est évident que les définitions en français et en néerlandais doivent avoir la même signification.

Environ 530 éléments cartographiques et topographiques pour lesquels une définition était souhaitable ont été sélectionnés. Une fois ces éléments identifiés avec une appellation correcte, les synonymes utilisés à l'IGN ont été également repérés. Ces synonymes sont mentionnés avec la définition de l'élément topographique ou cartographique. [IGN 2001, *TOC définitions*]

Pour que les définitions restent compréhensibles et parce que certaines définitions s'appliquent seulement dans un certain cadre, il a souvent été nécessaire de faire plusieurs définitions spécifiques au lieu d'une générale. Pour ces définitions spécifiques, il est indiqué le domaine d'application dans lequel chaque définition est valable. Par exemple : la définition de *plage* stipule qu'on se trouve sur terrain sablonneux. C'est de fait le cas en Belgique, mais pas dans d'autres pays où l'on rencontre aussi des plages de galets. C'est pourquoi, avec la définition, il est indiqué que le domaine d'application concerné est la Belgique.

2.2.2 Top50V-Gis – Liste des objets à modéliser

La représentation actuelle de la carte à l'échelle du 1/50 000^e est effectuée à partir d'un SIG géo-relationnel implémenté sous ArcInfo©. L'information spatiale de cette base de données est représentée sous forme de points, de lignes ou de polygones stockés dans différentes couches d'informations. A chaque élément sont liés des attributs qui permettent de le caractériser.

L'ensemble des données est stocké dans 18 couches d'informations. Des types géométriques différents ne pouvant pas être mélangés dans une même couche, on distingue souvent des couches –PNT (contenant des objets ponctuels), -LINE (contenant des objets linéaires) et –POLY (contenant des objets surfaciques). Pour le codage des données, dix thèmes sont différenciés. Il s'agit des données administratives, de l'altimétrie, de l'électricité, des surfaces du sol, de l'hydrographie, du paysage, du chemin de fer, des routes, des structures et de la végétation. La plupart des thèmes correspondent à des couches d'informations. Quelques exceptions sont à remarquer, par exemple la couche « occupation du sol » n'a pas de thème propre. Elle reprend des objets polygonaux des thèmes végétation, surface du sol, hydrographie et structure. Il arrive que des objets soient repris dans plusieurs couches à la fois. On constate donc qu'il existe une certaine redondance au niveau de ces données. Cela s'explique par le fait que les données peuvent être distribuées par couches. Il faut donc, pour

que l'information fournie soit complète, rassembler certains éléments de plusieurs couches. L'ensemble des couches est présenté à la figure 8.

Chaque type de données possède un identifiant particulier. Celui-ci est commun aux données du 1/10 000^e et 1/50 000^e. Il possède une structure qui permet de retrouver pour chaque objet à quelle couche d'information il appartient ainsi que sa catégorie, sous-catégorie et numéro d'ordre. Par exemple, l'hôpital codé comme ST210 est un élément du thème structure (ST***) qui est un service (**2**), de type hôpital (**1*). Ces codes étant un repère important pour le personnel de l'IGN, nous tenterons dans la mesure du possible de les conserver lors de la modélisation de ces données (cfr. chapitre 3). Les données contiennent également des codes réservés aux valeurs inconnues (9) et non applicables (8).

La liste de l'ensemble des objets présentés dans le Top50V-Gis actuel (2^e édition) est reprise dans un document interne à l'IGN. Ce document ne se présente pas sous la forme d'un catalogue de données comme pour la BDG SGISR. Il s'agit de la liste des objets par domaines (Dataset) et couches de chaque type géométrique (FeatureClasses). Les attributs relatifs à chaque couche ainsi que les valeurs possibles y sont également repris. [IGN 2005, *Structure et codage des données de la 2^{ème} édition 1 : 50 000*]

Un deuxième document, tiré d'une base de données Access 2000® reprend la définition des objets ainsi que les critères de sélection utilisés lors de la seconde édition. Les figures 9 et 10 représentent un exemple de chacun des deux documents décrits.

Figure 8 : Tableau des couches géométriques du TOP50V-Gis par domaine.

Domaine	Dataset	Type géométrique	Feature Class
DONNÉES ADMINISTRATIVES	ADMIN	Lignes	ADMILINE
		Polygones	ADMIPOLY
ALTIMÉTRIE	ALTI	Lignes	ALTILINE
		Points	ALTIPNT
		Lignes	BATHYLINE
		Points	BATHYPNT
OBJETS COMPLEXES	CPL	Polygones	CPLPOLY
ELECTRICITÉ	ELEC	Lignes	ECLINE
		Polygones	ECPOLY
		Points	ECPNT
HABITAT	HABITAT	Polygones	HABPOLY1
		Polygones	HABPOLY2
		Polygones	HABPOLY3
		Polygones	HABPOLY4
HYDROGRAPHIE	HYDRO	Lignes	RIVERLINE
		Lignes	HYDROSTRUCTURE
		Polygones	HYDROPOLY
		Points	HYDROPN
UTILISATION DU SOL	LANDUSE	Polygones	LANDUSE
RÉSEAU FERROVIAIRE	RAILTRANS	Lignes	RAIL
		Lignes	CABLETRANS
		Points	RAILPNT
RÉSEAU ROUTIER	ROADTRANS	Lignes	ROAD
		Lignes	OTHERWAY
		Points	ROADPNT
PAYSAGE	LANDSCAPE	Lignes	SCAPELINE
		Polygones	SCAPEPOLY
		Points	SCAPEPNT
STRUCTURES	STRUCTURE	Lignes	STRUCLINE
		Polygones	STRUCPOLY
		Points	STRUCPNT

ELEC	
Elecline	
Type :	
Domaine : TypeElecnetLignes	
EL110 : Ligne à haute tension EL111 : Ligne à haute tension souterraine	
KVA (Kilovolt Capacity Attribute) :	
Domaine : KvaElecnetLignes (0,70,150,220,380)	
Signification : Voltage de la ligne à haute tension.	
Contenu : Valeurs réelles	

Figure 9 : Extrait du document de spécification du TOP50V-Gis [IGN 2005, Structure et codage des données de la 2^{ème} édition 1 : 50 000].

TYPE	EL110	Ligne à haute tension
DESCRIPTION		
l'ensemble des installations techniques, telles que les pylônes à haute tension, ainsi que les câbles souterrains ou aériens des lignes à haute tension, qui transportent d'un endroit à un autre un courant électrique sous haute tension.		
COMPLEMENTS		
CRITERES DE SELECTION ED2: Toutes celles qui font au moins 70KV et dont la longueur est supérieure à 500 m. Généralisation possible en suivant la carte CPTE (au 1/300000) en présence de plusieurs lignes parallèles très proches.		

Figure 10 : Extrait de la description des types du TOP50V-Gis [IGN 2004, Critères de sélection du 1 :50 000 2^{ème} édition].

Aux attributs qui décrivent les caractéristiques des objets s'ajoutent des attributs propres à la représentation cartographique du 1/50 000^e. Il s'agit de « SYM » et « GRA ». L'attribut « SYM » est utilisé pour donner une information quant à la symbolisation de l'objet. Par exemple, les routes sont normalement symbolisées en fonction de la largeur. Mais certaines d'entre elles sont symbolisées en fonction du nombre de bandes de circulation qu'elles possèdent, sans tenir compte de leur largeur. L'attribut « SYM » permet alors de faire cette distinction. Il permet également de masquer certains éléments sur la carte, comme par exemple, un segment de route qui traverse un bâtiment, ou encore aide à la représentation des symboles des ponts. L'attribut « GRA » quant à lui permet de « déclasser » une route lorsque plusieurs segments sont fort proches et qu'il n'est pas possible de les représenter dans leur classe d'origine faute de place. La route est alors représentée par un symbole de catégorie inférieure.

3 Élaboration d'un modèle conceptuel pour les données du 1/50 000^e

Pour pouvoir effectuer la généralisation conceptuelle, nous avons vu qu'il était nécessaire d'élaborer un modèle de données correspondant à la gamme d'échelle pour laquelle nous souhaitons travailler. Le présent chapitre va donc consister, à partir de la liste des classes du TOP50V-Gis, à créer le modèle des données du 1/50 000^e.

Pour cette partie, une grande collaboration a été nécessaire avec l'IGN afin de répondre parfaitement à leurs exigences. Une étape de validation par les responsables du projet GenMap a été nécessaire.

La démarche utilisée est la suivante. Nous avons modélisé séparément chaque domaine pour rechercher toutes les relations que les objets entretiennent entre eux. Ils ont été modélisés sur base des classes du TOP50V-Gis que nous avons préalablement représentées selon le formalisme UML.

Comme on peut le voir sur la figure 11, nous avons également ajouté en surimposition les « liens » qui existent entre le modèle que nous créons et les classes actuelles. De cette manière, notre démarche permet d'être exhaustive.

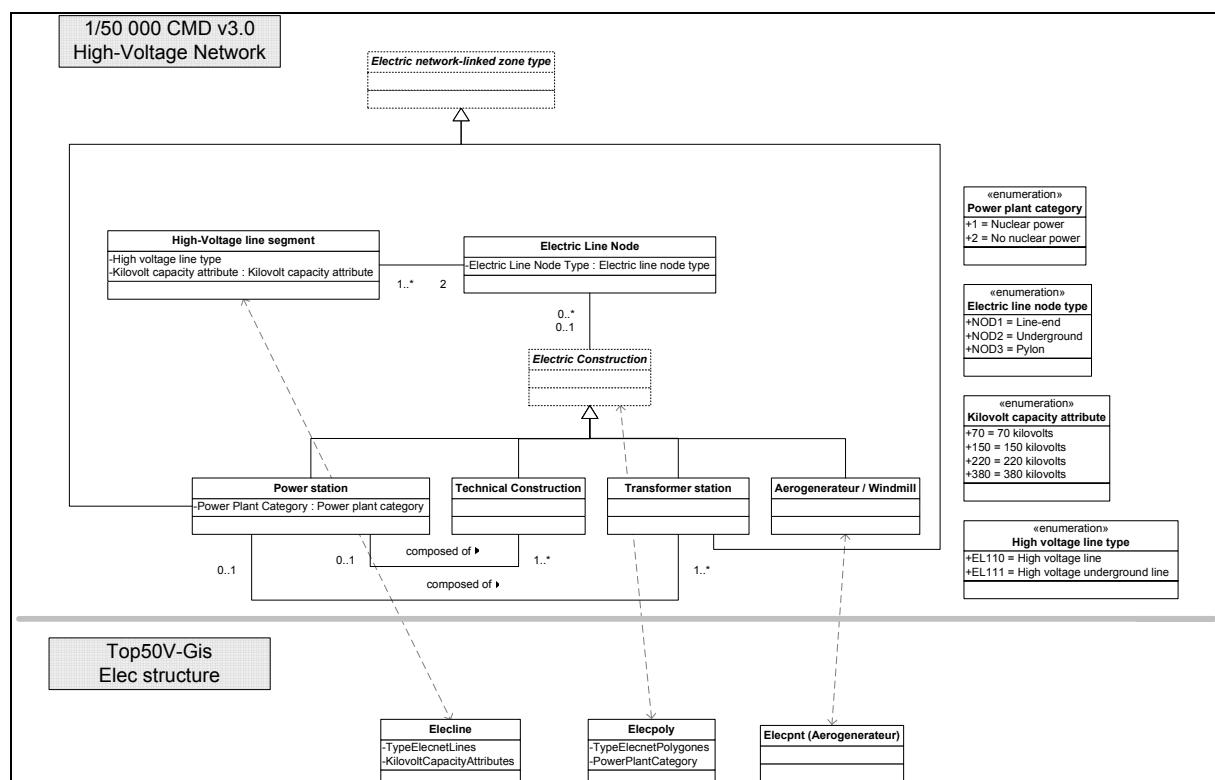


Figure 11 : Modélisation conceptuelle du domaine « High Voltage Network » à partir du Top50V-Gis.

Par souci de lisibilité du modèle, nous ne représenterons pas toutes les relations topologiques, mais uniquement les plus importantes. Comme dans le cas de l'élaboration du modèle SGISR, nous partons sur le postulat que ces relations peuvent facilement être retrouvées grâce à des requêtes géographiques et topologiques des SIG-logiciels. Il n'est, dès lors, pas nécessaire de toutes les représenter. Seules seront modélisées les relations qui permettent d'obtenir une topologie de base (organisation en arcs et noeuds des segments de réseau) ainsi que les relations entre les objets dont l'existence sur la carte est définie par un autre, comme c'est le cas pour les bornes kilométriques qui n'ont de raison d'apparaître que si elles sont à coté d'un segment de voirie.

3.1 Prescriptions à l'élaboration du modèle

3.1.1 Choix de conserver une même sémantique

L'IGN nous a imposé plusieurs contraintes pour la création du modèle. Premièrement, il nous est nécessaire de conserver un maximum la sémantique actuelle du 1/50 000^e. Il faut continuer à pouvoir produire des cartes similaires à celles produites actuellement. Notons que certains changements ont quand même été effectués en accord avec les responsables du projet GenMap, notamment lorsque par exemple, des attributs n'étaient plus utilisés ou lorsque des sources n'existaient plus. L'étape de généralisation conceptuelle doit normalement redéfinir une nouvelle sémantique en correspondance avec la gamme d'échelle à laquelle on souhaite arriver. Comme dans notre cas, la représentation cartographique existe déjà, tout l'enjeu sera de pouvoir modéliser les données actuelles sans modifier (de trop) leur sémantique.

3.1.2 « Ressemblance » à SGISR

Comme nous l'avons montré dans l'état de l'art, pour éviter les conflits d'appariement, il est nécessaire que les deux modèles soient le plus proche possible du point de vue du formalisme, de la sémantique et de l'abstraction. En effet, il serait trivial d'apparier une BDG avec elle-même. Pour faciliter la suite du travail, nous avons décidé de créer un modèle le plus ressemblant possible à SGISR. Comme il sera possible de le constater, certaines parties seront communes aux deux modèles car elles décrivent des phénomènes identiques. Le formalisme utilisé est le même (UML) ainsi que la méthode pour concevoir les relations. En effet, il ne serait pas intéressant d'inclure plus de relations topologiques dans les données du 1/50 000^e qu'il n'en existe dans SGISR car il serait impossible des les apparier par la suite.

3.2 Modélisation des domaines du 1/50 000^e

Dans la suite de ce chapitre, nous allons passer chaque domaine en revue et expliquer brièvement la manière dont ils sont modélisés. D'une manière générale, les valeurs d'attributs de type « unknown » ne seront pas reprises ainsi que les doublons pour respecter les règles en matière de modélisation. Notons enfin que le modèle est rédigé en anglais de la même manière que SGISR. Cela permet d'avoir un langage commun entre tous les intervenants au sein de l'IGN.

3.2.1 High-Voltage Network

Le réseau électrique à haute tension est construit sur base des objets « High-voltage line segment » (Segment de ligne à haute tension) et « Electric line node » (Nœud de ligne électrique) (fig. 13). Ceux-ci construisent une topologie dans le réseau par une association qui les relie. Comme dans le cas d'un réseau électrique, il n'existe pas de polyligne fermée, un segment sera toujours connecté à deux nœuds. Ce qui se traduit par la cardinalité [2] sur l'association. Cette cardinalité correspond en fait à une double dépendance fonctionnelle (comme s'il y avait deux relations avec une cardinalité de un), ce qui signifie qu'il ne sera pas nécessaire de créer une table pour représenter l'association dans le modèle relationnel.

Les « Power station » (Centrale électrique), « Technical construction » (Construction technique), « Transformation Station » (Station de transformation) et « Wind Turbine » (Aérogénérateur) ont été généralisés en une classe « Electric construction » qui sera implémentée pour retrouver plus facilement une classe de type « Electropoly » comme dans le *Top50V-Gis* qui est une couche reprenant toutes les surfaces au sol des bâtiments ou des zones qui ont un rapport avec le réseau électrique. Cette généralisation est exclusive et totale. Cela signifie respectivement qu'une instance ne peut appartenir en même temps à deux classes filles et que toutes les instances de la classe mère sont réparties dans les classes filles.

On remarque que les différenciations des « Power Station » de SGISR n'existent pas dans le 1/50 000^e. Seul apparaît un attribut supplémentaire : « Power plant category » (Catégorie de centrale), utilisé pour savoir si la centrale est de type nucléaire ou pas.

Deux associations « composed of » apparaissent au niveau du schéma car elles déterminent une condition obligatoire pour la représentation des « Power station ». En effet, seules les centrales électriques qui possèdent un « Technical construction » et une « Transformation station » sont représentées.

La figure 12 donne un aperçu de la structure des données telles qu'elles sont modélisées. Un nœud est obligatoire à l'entrée d'une ligne dans chaque zone ou bâtiment. Cependant un

bâiment ou une zone n'est pas nécessairement traversé par une ligne à haute tension. L'analyse du jeu de données TOP50V-Gis dont nous disposions laissait apparaître cette structure, nous l'avons conservée pour ne pas devoir reconstruire une nouvelle topologie.

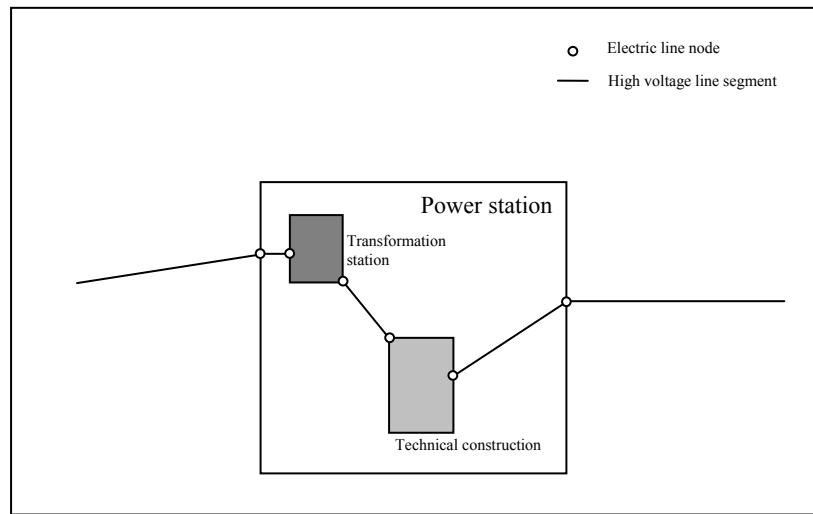


Figure 12 : Schéma des données de « High-voltage network » telles que représentées par le modèle.

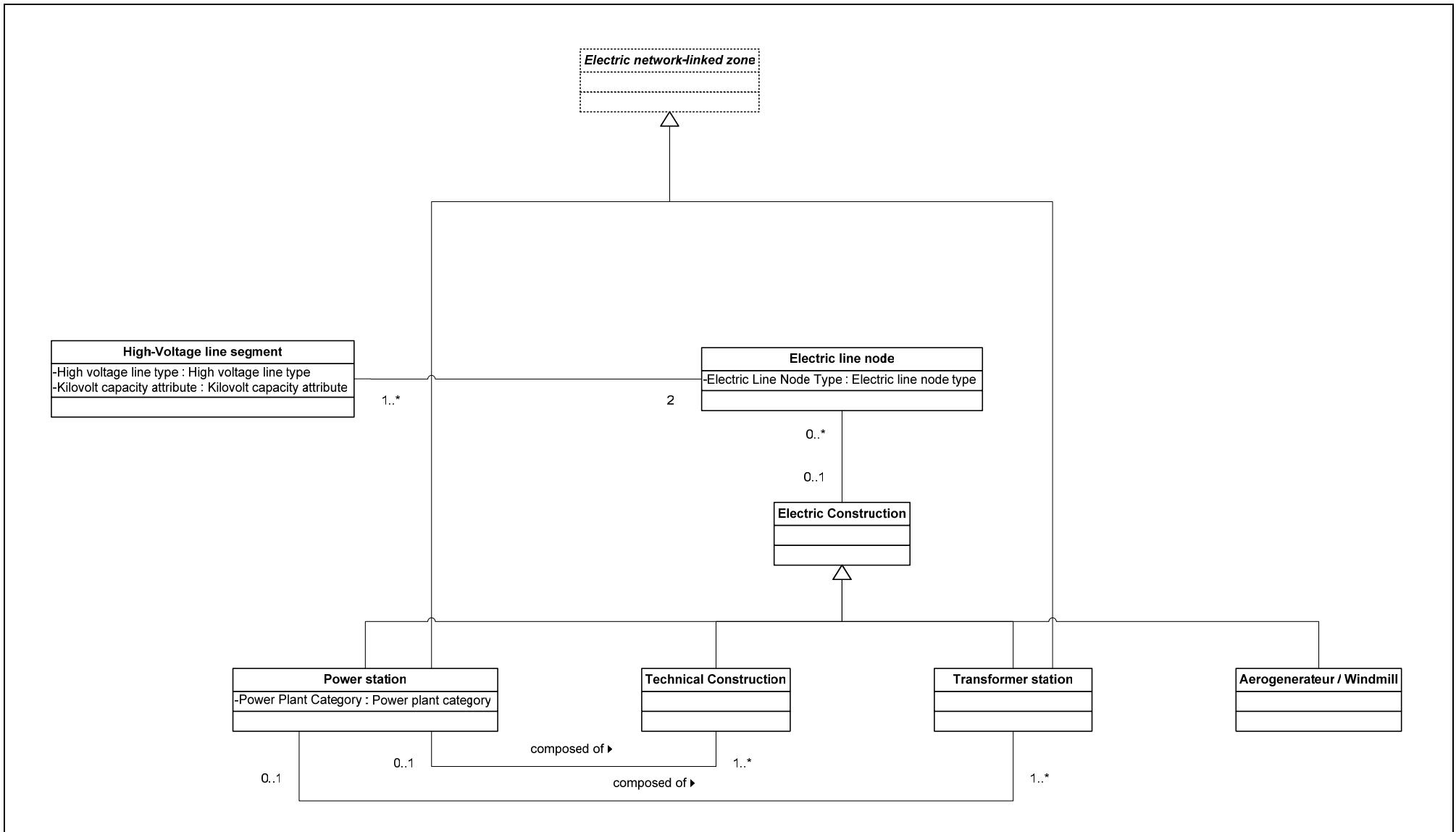


Figure 13 : Modèle conceptuel du domaine « High Voltage Network ».

3.2.2 Administrative division

Le domaine « Admin » (fig. 14) reprend les données administratives de la carte. Il est quasi identique à son correspondant dans SGISR. L’unité principale est la « Municipality » (Commune) qui se compose respectivement en « District » (District), « Province » (Province), « Region » (Région) et enfin en « State » (État). En plus de chaque relation de composition sont ajoutées des associations directionnelles qui précisent quelle « municipality » sera le chef lieu du niveau supérieur.

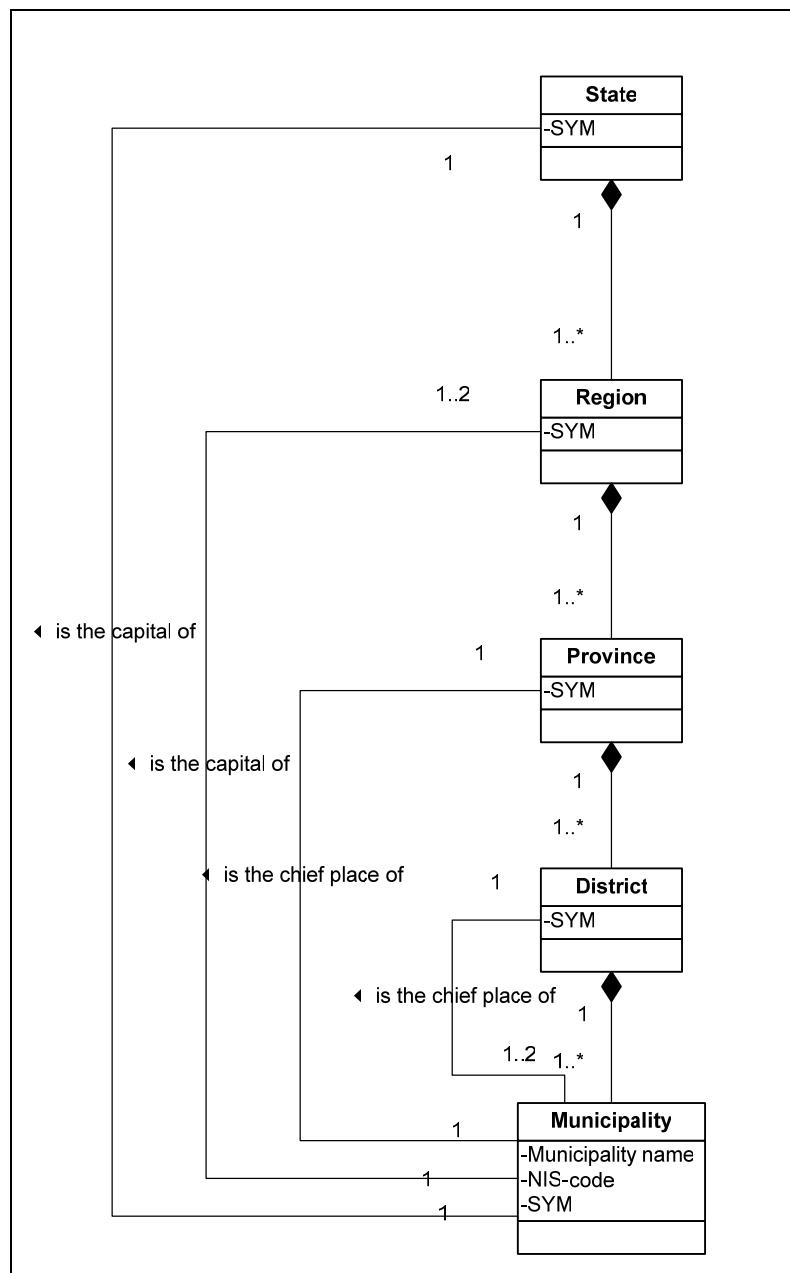


Figure 14 : Modèle conceptuel du domaine « Admin ».

A chaque objet géographique est ajouté un attribut « SYM » pour indiquer une information supplémentaire quant à la symbolisation. Cette valeur fictive ne devrait pas se situer sur un modèle de données géographiques. Cependant, nous avons voulu reprendre l'ensemble des données nécessaires à l'élaboration de la carte pour pouvoir remplir pleinement l'objectif du mémoire. Nous plaçons donc, comme le souhaite l'IGN, ces données relatives à la symbolisation dans le modèle.

La représentation actuelle des limites dans le TOP50V-Gis est toujours effectuée en conservant la limite la plus générale. Une limite communale qui est aussi une frontière n'est actuellement pas conservée. Dans le modèle tel qu'il est défini, toutes les limites sont conservées, il s'agira d'effectuer une opération de sélection pour la représentation cartographique. Cette opération supplémentaire ne pose pas de problèmes au service en charge de la représentation cartographique. De plus, les données sont structurées de manière identique dans SGISR, une représentation par limite serait beaucoup plus contraignante pour l'appariement.

3.2.3 Landuse et landscape

Les deux domaines « Landuse » et « Landscape » du *Top50V-Gis* ont été regroupés dans un nouveau domaine appelé « Landcover » (fig. 15). Une partie des éléments de « Lanscape » ont été rejettés dans le domaine « Alti ».

La représentation de « Landuse » (Occupation du sol) est effectuée sous la forme de polygones qui ne se superposent pas (un seul type d'occupation de sol par endroit). Ceci nous a permis de regrouper toutes les occupations en un seul objet géographique et d'énumérer tous les types possibles. Par rapport à SGISR, le nombre de types est réduit, nous en discuterons lors de l'appariement des deux modèles.

Deux objets supplémentaires composent le domaine. Il s'agit de « Isolated tree » (Arbre isolé) et de « Row of trees » (Rangée d'arbres) qui se généralisent en une classe abstraite « Landscape vegetation » (Végétation de paysage). Cette généralisation est exclusive et totale.

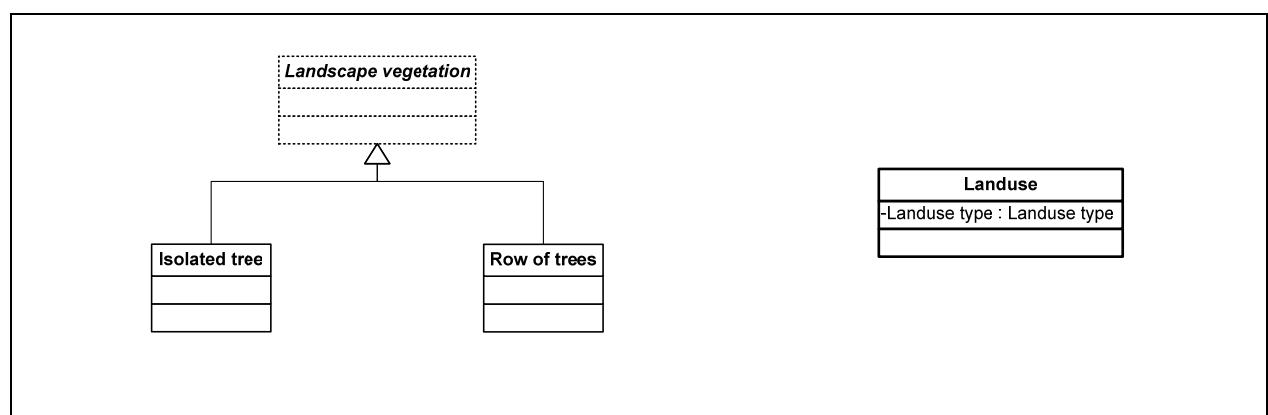


Figure 15 : Modèle conceptuel du domaine "Landcover".

3.2.4 Rail transport network

Le réseau de chemin de fer (fig. 16), comme l'ensemble des réseaux, est un élément essentiel de la représentation du 1/50 000^e. Le modèle est construit autour de celui-ci. La partie du chemin de fer est assez simple, elle est constituée de segments et de nœuds. Une association relie les arrêts pour voyageurs aux segments. Un ensemble de segments s'agrègent en une classe « Numbered Railway » (Ligne de chemin de fer numérotée).

Chaque segment est également qualifié par les attributs « Railway existence category » qui définit l'état de la voie, et par « Lane track number » qui précise si la voie est simple ou multiple (dans le cas de voies multiples, un seul segment est conservé).

Dans le domaine Railway du TOP50V-Gis est également stocké un objet « Cable transport » (Transport par cable) qui reprend soit les téléphériques ou les bandes transporteuses industrielles. Le modèle inclut donc un objet géographique y faisant référence. D'autres relations arrivent à ce domaine, nous en parlerons lors du regroupement des domaines sur le modèle général. La classe « Railway traveler stop » (Arrêt pour voyageur du chemin de fer) possède une association de topologie avec les segments de réseau. Comme précisé précédemment, l'existence de cette classe dans la représentation n'a de raison d'être que si elle est associée au réseau, cette relation ne nous paraît donc pas superflue.

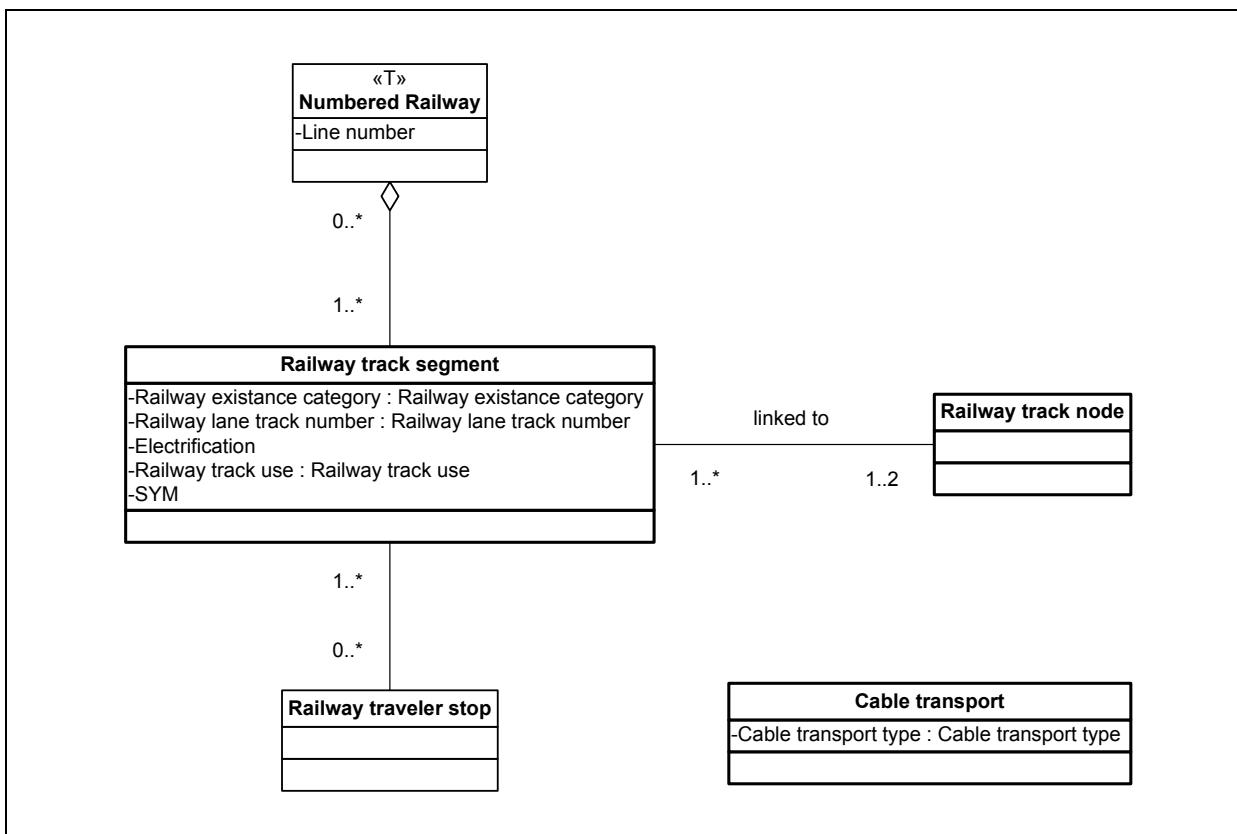


Figure 16 : Modèle conceptuel du domaine "Rail transport network".

3.2.5 Road transport

Comme pour le chemin de fer, les routes s'axent autour des segments et des nœuds. Deux types de segments sont distingués (fig. 18). Il s'agit du « Road segment » (Segment de route) et du « Other way segment » (Autre segment de voirie). Ces deux classes représentent respectivement les axes de routes principales et les autres chemins (chemins de terre, pistes cyclables, ...), la distinction est effectuée sur base de la facilité à emprunter la voie et sur son revêtement.

Les nœuds se distinguent en deux catégories. Ils représentent soit un petit rond point dont le diamètre est inférieur à un seuil de représentation, soit un simple nœud dans le réseau (fig. 17). Certains de ces nœuds « ordinaires » se regroupent avec quelques segments par une relation de composition pour créer une entité de « Large Roundabout » (Grand rond-point). En effet, il semblait étrange de conserver l'information des ronds-points non représentables à cause de leur petite taille et de ne rien savoir quant aux ronds-points plus grands. Nous avons alors décidé de créer une classe reprenant l'ensemble des éléments composant ceux-ci. La classe « Large Roundabout » est précédée du stéréotype « T » qui précise qu'il s'agit d'une classe textuelle. Celle-ci ne reprendra que les identifiants des entités participant aux grands ronds-points.

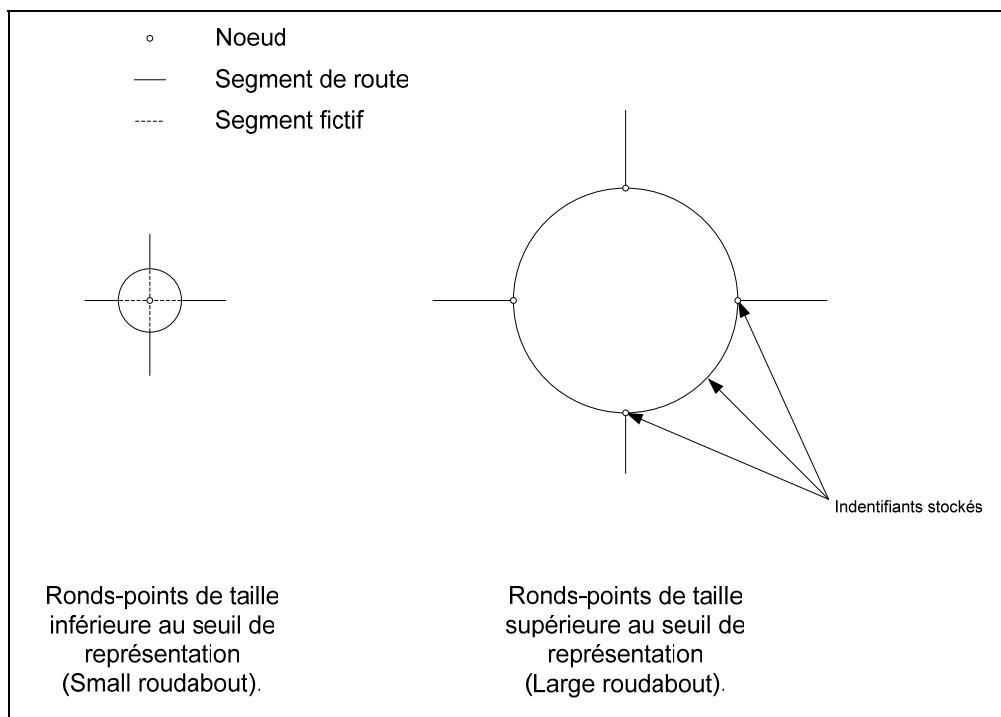


Figure 17 : Représentation des ronds-points dans le modèle conceptuel 1/50 000.

Certains des segments de routes se composent également en une classe «Numbered Road » (Route numérotée) dans laquelle on retient leurs noms : « Road name » et numéro national : « Road number ».

Les attributs qualifiant les segments de routes sont : la classe de largeur « Road width », les états « Road existence category », leur type de revêtement selon les codes OTAN : « Road surface type », leur utilisation : « Road use », la présence ou non de berme centrale : « Road median » et enfin deux attributs utilisés pour la symbolisation : « SYM » et « GRA » comme expliqué précédemment. Notons que le 1/50 000^e est une échelle militaire, elle utilise par conséquent des catégories particulières comme les codes OTAN, il s'agit en fait des valeurs X, Y et Z qui définissent le type de revêtement en fonction de la praticabilité par divers types de temps.

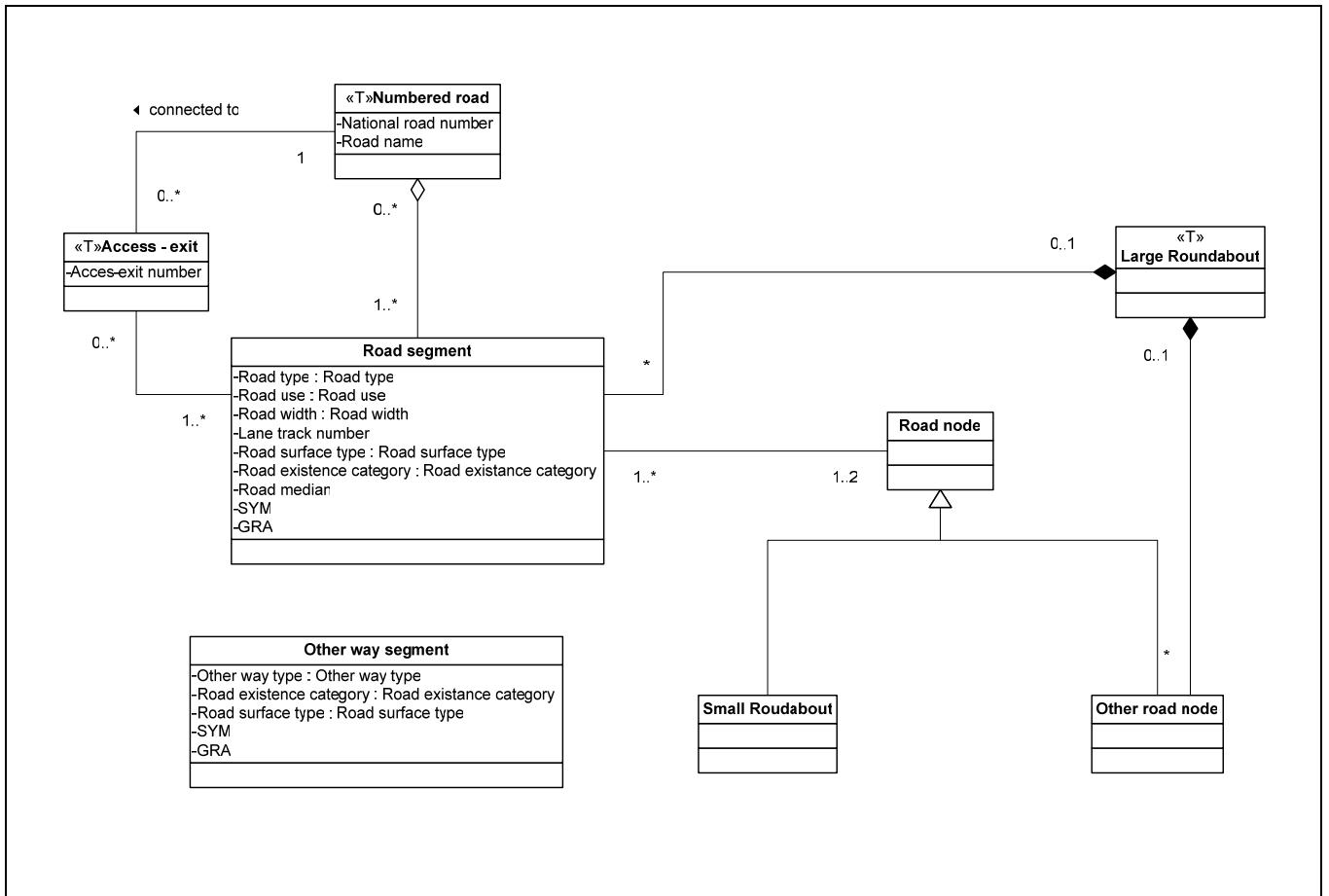


Figure 18 : Modèle conceptuel du domaine "Road network".

3.2.6 Watercourse

Le réseau hydrographique est composé de segments de rivières connectés aux nœuds par une association définissant une dépendance fonctionnelle (fig. 19). Les nœuds entretiennent des relations avec les « Water surface » (Surface d'eau). Celles-ci n'ont pas toujours lieu, ce qui se traduit par les cardinalités [0...*]. Les « Water surface » sont par exemple des étangs, lacs, l'estuaire de l'Escaut,...

Aux segments sont également associés les surfaces de rivières : « Watercourse surface ». En effet, au-delà d'une certaine largeur, les rivières sont représentées par leurs berges et non plus

classifiées en fonction de leur largeur. De plus, y sont stockés les élargissements de rivières, ces objets propres à l'échelle du 1/50 000^e représentent une excroissance locale du réseau qui le fait passer à plus de 50 mètres de largeur. Ces objets tendent à disparaître de la représentation, ils ne sont pas repris dans les données de référence.

La classe « Hydraulic structure » (Structure hydraulique) représente les constructions intervenant sur le réseau, comme les écluses, les barrages, ascenseurs à bateaux. Tous ces objets présents le long du réseau hydrographique sont à opposer aux « Coastal structure » (Structure côtière) qui représentent ceux situés le long de la côte (brise-lames, ...).

L'objet géographique « Water point » (Point d'eau) reprend l'ensemble des objets hydrographiques de types ponctuels comme les sources, balises, fanaux,... Il pourrait sembler étrange que les sources ne soient pas en relation avec le réseau hydrographique. Il en est ainsi car les sources qui alimentent directement le réseau hydrographique ne sont pas représentées au 1/50 000^e.

Enfin, « Wetland » (Zone humide) qui vient en surimposition sur la représentation fait l'objet d'une classe à part entière. Il n'aurait pas été logique au niveau conceptuel les regrouper dans « Landuse » car ils peuvent se superposer avec d'autres affectations de sol.

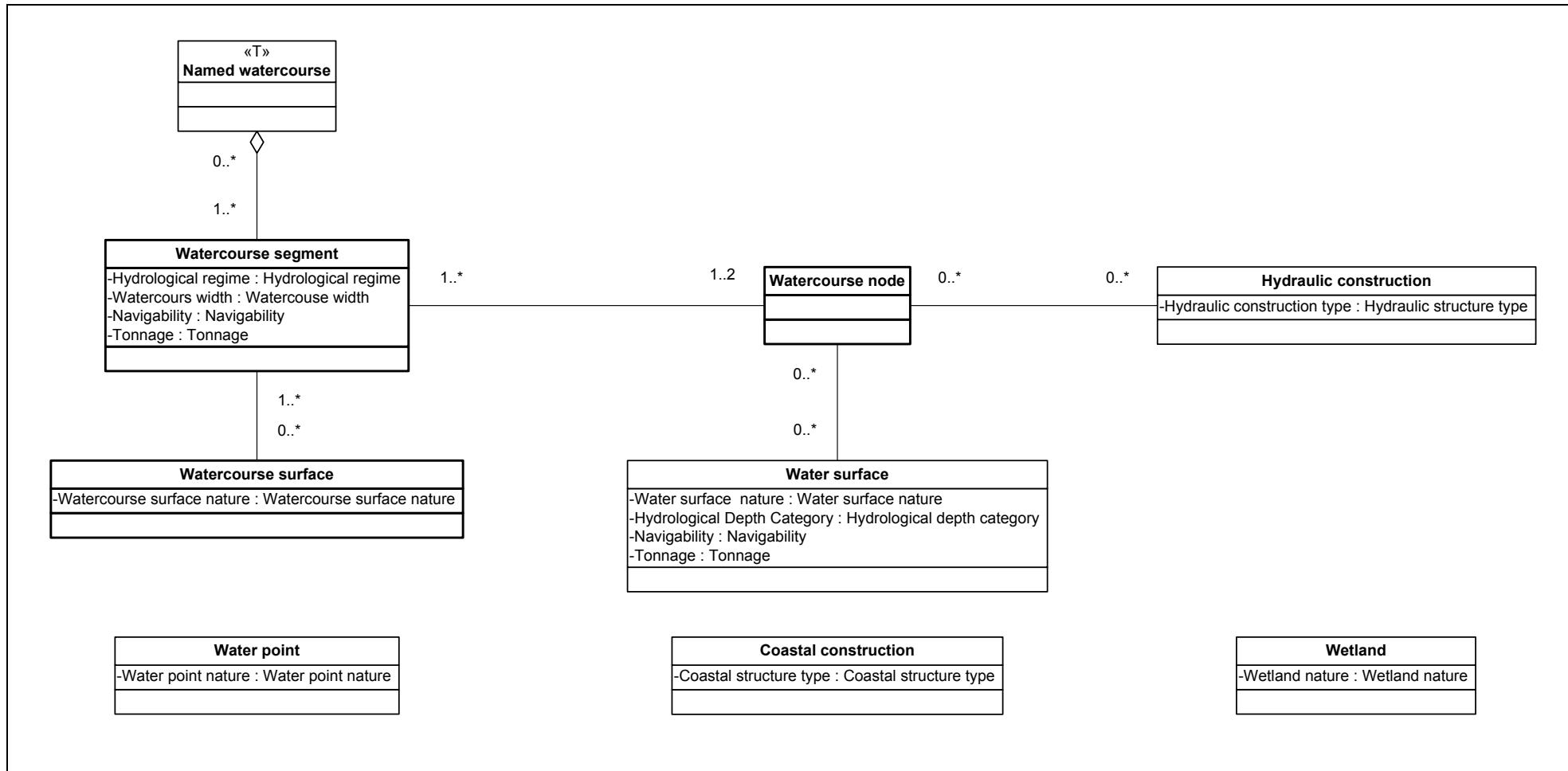


Figure 19 : Modèle conceptuel du domaine "Hydro network"

3.2.7 Network

Un certain nombre d’éléments étant communs aux réseaux, ils apparaissent à un niveau supérieur. Les segments, nœuds et éléments principaux se généralisent hors des trois « sous-réseaux ». Les « Principal network element » (Élement principal de réseau) sont associés aux « Kilometric marker » (Borne kilométrique). Les segments quant à eux sont doublement associés aux « Bridge » (Pont). Une première association qui s’appelle « bridged by » signifie que le segment passe en dessous du pont et la seconde appelée « carried by » signifie que le segment se situe sur le pont. La classe « Bridge » est précédée du stéréotype « T » pour préciser qu’il s’agit uniquement d’éléments textuels. En effet, dans la représentation du 1/50 000^e, les ponts sont représentés comme des segments de réseau, ils possèdent juste un attribut supplémentaire. Cette classe sert uniquement à reprendre le niveau du pont et l’identifiant du segment qui fait office de pont. Le modèle est présenté figure 20.

3.2.8 Structure

Le domaine structure est divisé en deux grands groupes d’objets, les « General construction » (Construction générale) et les « Zone » (Zone). Ces objets se spécialisent chacun en plusieurs classes (fig. 21).

L’objet « General construction » se spécialise en « Construction » (Construction) et « Building » (Building). La distinction est effectuée sur base de la fonction du bâtiment. Les « Building » sont considérés comme des bâtiments où l’homme peut entrer pour y effectuer une activité. Les « Construction » regroupent les bâtiments dans lesquels l’homme n’effectue pas d’activité.

Des classes abstraites sont ajoutées pour que les objets géographiques soient au même niveau de filiation dans les relations de généralisation. Toutes les relations sont exclusives et totales.

Une association conditionnelle est effectuée entre les « Building » et les « Railway track segment » uniquement lorsque l’attribut « Building type » (Type de building) représente une gare. Une association conditionnelle signifie qu’elle n’est effectuée que lorsqu’elle remplit la condition énoncée, qui dans ce cas est le fait d’être une gare. Il en est de même pour les zones, lorsqu’elles-ci est un « Motorway carpark » (Parking autoroutier).

La classe « Electric Network-linked zone» (Zone de connexion du réseau électrique) ne reprenant que des objets en relation avec le domaine « High-voltage network », nous avons décidé de l’y déplacer (voir modèle complet).

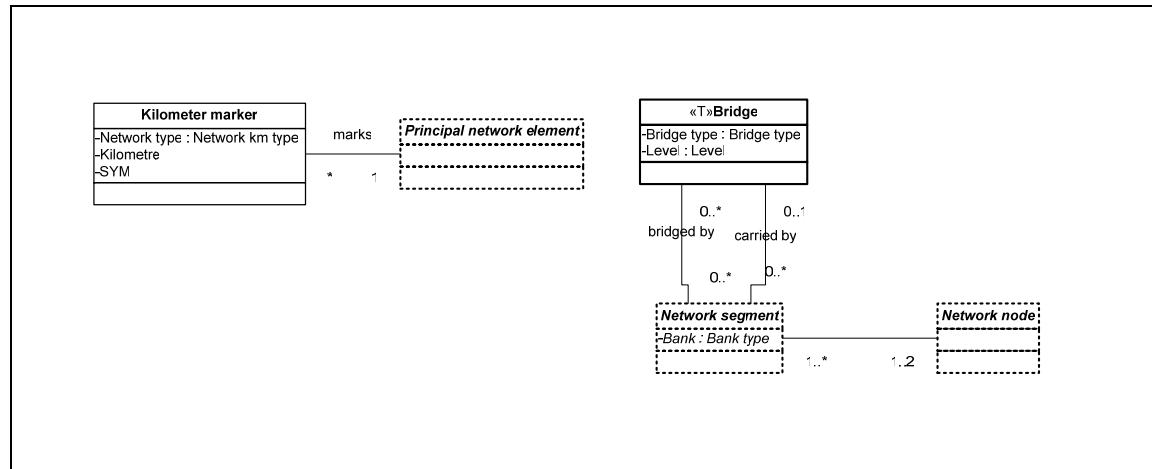


Figure 20 : Modèle conceptuel du domaine « Network »

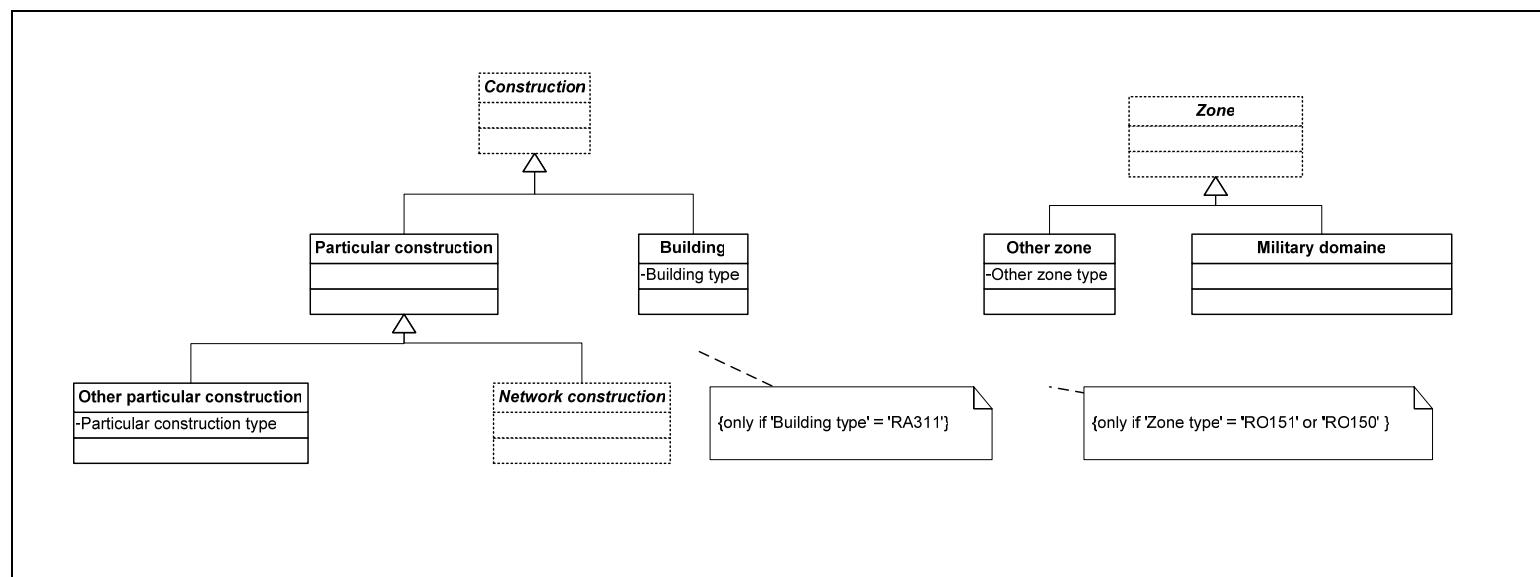


Figure 21 : Modèle conceptuel du domaine "Structure".

3.2.9 Complexe

Le domaine « Complexe » (Complexe) du Top50V-Gis illustre bien la redondance présente dans les données actuelles. En effet, il n'est pas du tout repris dans le modèle conceptuel que nous avons conçu car il ne contient que des objets déjà représentés dans d'autres domaines. Il sera néanmoins toujours possible de recomposer par après une couche de données reprenant ces données pour les utilisateurs intéressés.

3.2.10 Altimetry

Bien qu'il s'agisse d'un modèle de données géographiques, nous avons quand même inclus un domaine reprenant l'altimétrie. Le modèle SGISR en possède également un mais uniquement pour la bathymétrie et le micro relief. Dans notre application, nous préférons reprendre l'ensemble des données utiles à la création de la carte. Nous y reprenons donc en plus les « Altimetry contour » (Courbe de niveau) (fig.22).

La bathymétrie est identique à celle du modèle SGISR. Reste les objets du « Micro relief » (Micro relief) et les « Dune zone » (Zone de dune) qui n'intègrent pas la classe « Landuse » pour les mêmes raisons que la classe « Wetland ».

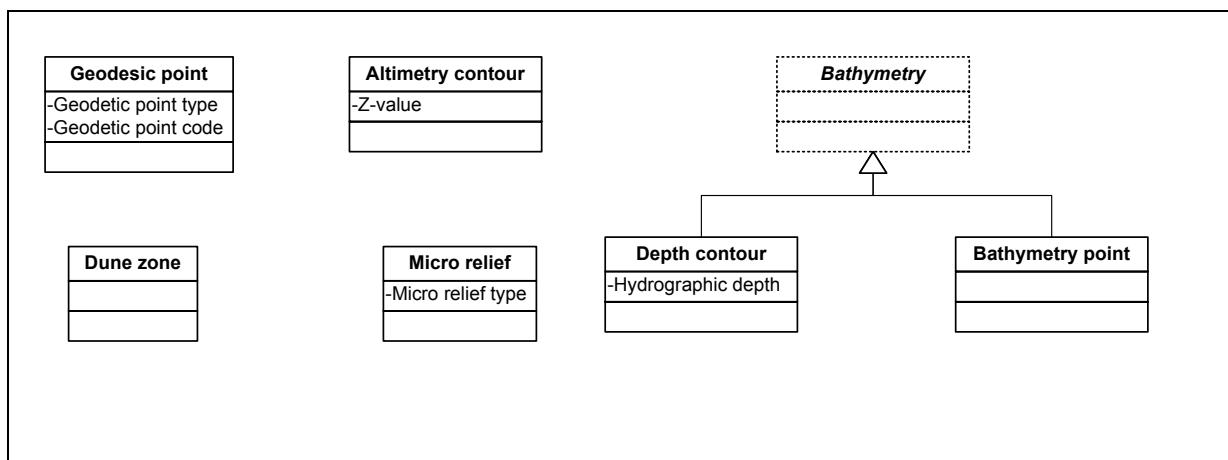


Figure 22 : Modèle conceptuel du domaine "Altimetry".

3.2.11 Autre

Une classe supplémentaire est insérée dans le modèle de données (fig. 23). Cette classe n'existe pas actuellement dans les données du 1/50 000^e, mais elle permettra dans le futur de faire le lien avec la base de données de la toponymie. Il s'agit de « Fictive toponymic point » (Point fictif de toponymie).

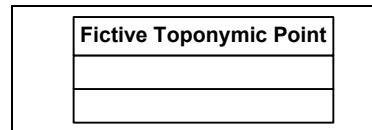


Figure 23 : Modèle conceptuel de la classe "Fictive toponymic point".

3.3 Modélisation des relations entre domaines

L'étape suivante à la réalisation du modèle conceptuel des données du 1/50 000^e consiste à rechercher les relations que les différents objets des différents domaines entretiennent entre eux. Il s'agit principalement de relations de généralisation et d'associations.

Nous avons déjà eu un aperçu lors de la modélisation du domaine « Network » des diverses généralisations qui interviennent pour les éléments du réseau. Des relations de généralisation sont également effectuées vers la classe « Network construction » du domaine « Structure ». Sont repris dans cette généralisation : « Cable transport », « Kilometer marker » et « Hydraulic structure » (fig. 24).

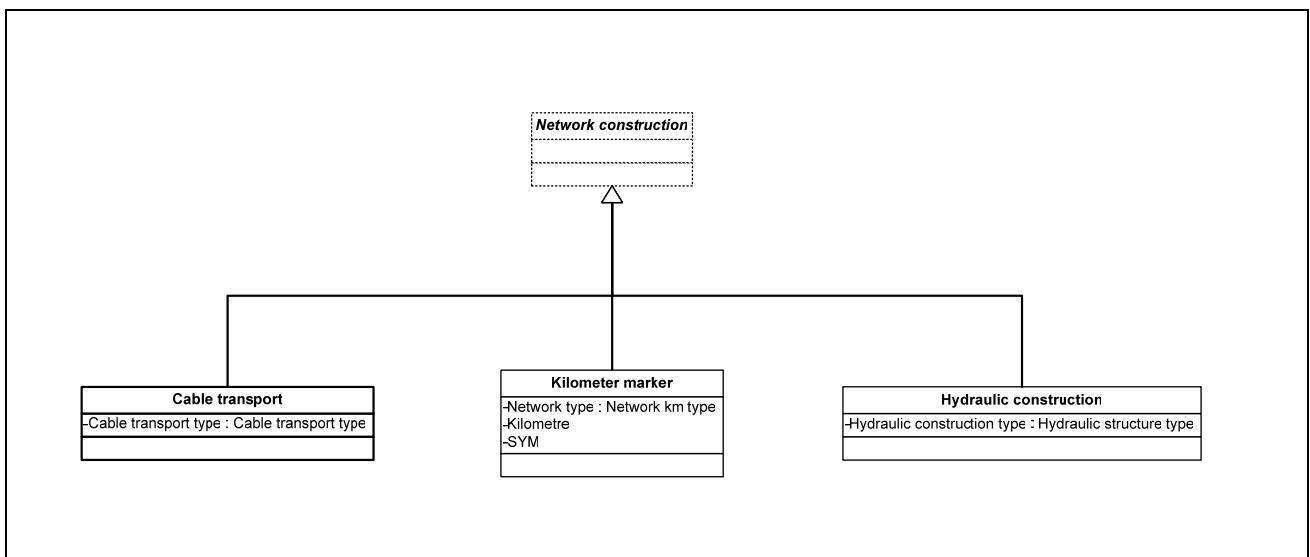


Figure 24 : Modélisation des relations entre les domaines : "Network construction"

Une autre généralisation est effectuée sur les « Zone » (fig.25), ainsi « Electric network-linked zone» est une classe fille de « Zone » au même titre que « Other zone » ou « Military domain ».

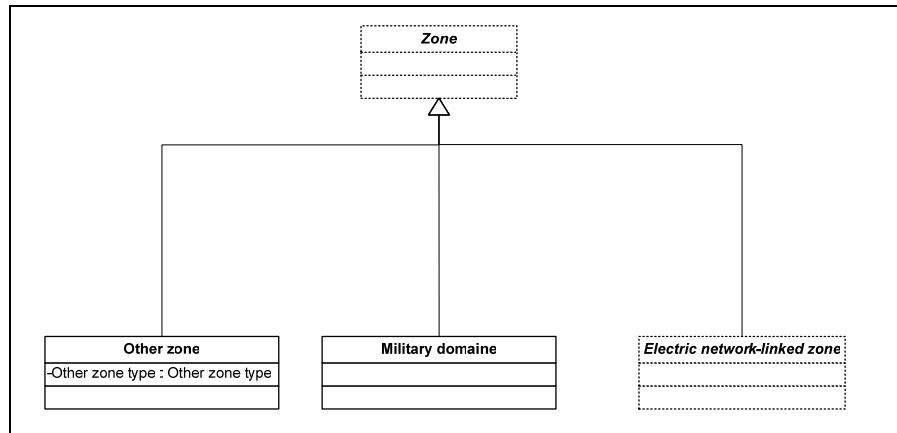


Figure 25 : Modèle conceptuel des relations de généralisation de "Zone".

Des associations conditionnelles sont également effectuées. Nous avons déjà discuté le cas des gares et des parkings autoroutiers. Les « Geodesic point » sont également associés aux « General construction » uniquement lorsque le point géodésique est situé sur une construction remarquable.

Les conditions des associations ne sont pas définies dans un langage formalisé en UML. Nous avons essayé d'être le plus clair possible en les exprimant. On peut en voir la forme sur la figure 26.

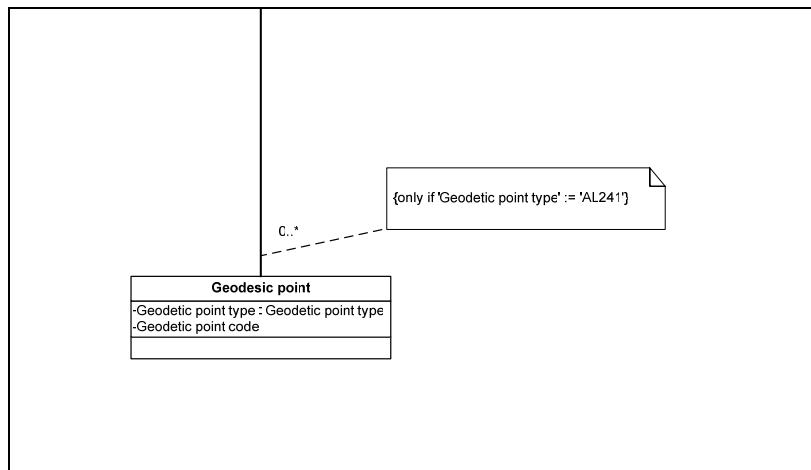


Figure 26 : Exemple d'une association conditionnelle.

3.4 Énumérations

Chaque attribut peut prendre un certain nombre de valeurs possible en fonction de son type. Dans la majeure partie des cas du modèle conceptuel, il s'agit d'un certain nombre d'énumérations pour lesquelles un code est défini. L'ensemble des énumérations est présenté sur le modèle conceptuel complet (fig. 27).

<table border="1"> <thead> <tr> <th style="text-align: center;">«enumeration»</th></tr> <tr> <th style="text-align: center;">Watercourse width</th></tr> </thead> <tbody> <tr> <td>+0 = Represented by bank</td></tr> <tr> <td>+10 = 0 - 3 meters</td></tr> <tr> <td>+20 = 3 - 15 meters</td></tr> <tr> <td>+30 = 15 - 25 meters</td></tr> <tr> <td>+40 = 25 - 50 meters</td></tr> </tbody> </table>	«enumeration»	Watercourse width	+0 = Represented by bank	+10 = 0 - 3 meters	+20 = 3 - 15 meters	+30 = 15 - 25 meters	+40 = 25 - 50 meters	<table border="1"> <thead> <tr> <th style="text-align: center;">«enumeration»</th></tr> <tr> <th style="text-align: center;">Tonnage</th></tr> </thead> <tbody> <tr> <td>+300</td></tr> <tr> <td>+600</td></tr> <tr> <td>+1350</td></tr> <tr> <td>+2000</td></tr> <tr> <td>+3000</td></tr> </tbody> </table>	«enumeration»	Tonnage	+300	+600	+1350	+2000	+3000	<table border="1"> <thead> <tr> <th style="text-align: center;">«enumeration»</th></tr> <tr> <th style="text-align: center;">Water point nature</th></tr> </thead> <tbody> <tr> <td>+HY150 = Water points (source, fountain, well)</td></tr> <tr> <td>+HY240 = Small lock</td></tr> <tr> <td>+HY261 = Siphon culvert entrance</td></tr> <tr> <td>+HY271 = Signal light</td></tr> <tr> <td>+HY272 = Beacon</td></tr> <tr> <td>+HY274 = Lighthouse</td></tr> </tbody> </table>	«enumeration»	Water point nature	+HY150 = Water points (source, fountain, well)	+HY240 = Small lock	+HY261 = Siphon culvert entrance	+HY271 = Signal light	+HY272 = Beacon	+HY274 = Lighthouse
«enumeration»																								
Watercourse width																								
+0 = Represented by bank																								
+10 = 0 - 3 meters																								
+20 = 3 - 15 meters																								
+30 = 15 - 25 meters																								
+40 = 25 - 50 meters																								
«enumeration»																								
Tonnage																								
+300																								
+600																								
+1350																								
+2000																								
+3000																								
«enumeration»																								
Water point nature																								
+HY150 = Water points (source, fountain, well)																								
+HY240 = Small lock																								
+HY261 = Siphon culvert entrance																								
+HY271 = Signal light																								
+HY272 = Beacon																								
+HY274 = Lighthouse																								

Figure 27 : Exemple d'énumération.

Elles se représentent à l'aide du stéréotype « enumeration ». Nous avons choisi de conserver les codes de valeurs du Top50V-Gis, la correspondance avec les codes de SGISR sera effectuée lors de l'appariement.

3.5 Modèle conceptuel des données du 1/50 000^e

Le modèle conceptuel complet est fourni en annexe 2.

3.6 Remarques suite à la validation de l'IGN

Comme nous l'avons déjà précisé au début du chapitre, l'élaboration du modèle conceptuel a été le fruit d'une collaboration intense avec l'IGN. La méthode de développement a été itérative pour chaque fois combler les manques et corriger les erreurs. Le modèle présenté dans le chapitre est la dernière version (v3.1), les remarques ont été incluses dans le travail. Celles-ci portaient essentiellement sur des souhaits de représentations plus précises comme pour le cas des ronds-points de grande tailles ou des problèmes sémantiques de définition de classe comme dans le cas de la différenciation des « Road segment » et « Otherway segment ». Au final, toutes les demandes ont été prises en compte et intégrées.

4 Appariement des MCD SGISR et 1/50 000

4.1 Introduction

Maintenant que nous disposons d'un modèle conceptuel des données à l'échelle conceptuelle du 1/50 000^e, nous allons pouvoir réaliser les étapes suivantes de la généralisation conceptuelle ; il s'agit de :

- l'appariement au niveau des schémas ;
- l'énonciation des critères de sélection sémantique ;
- l'énonciation des critères de sélection géométrique.

Beaucoup de travaux portent sur l'intégration de bases de données géographiques. L'intégration consiste à « fusionner » deux bases pour n'en obtenir qu'une seule en sortie.

Comme il est possible de le constater sur la figure 28, l'intégration requiert un travail d'une part au niveau des schémas et d'autre part au niveau des données. La démarche qu'il nous est nécessaire d'exécuter dans ce travail peut sembler similaire. Nous n'avons cependant pas d'intégration à effectuer, le but n'étant pas d'obtenir une seule base. Seules quelques étapes sont semblables. Si nous reproduisons le même schéma appliqué à notre étude, il est possible de situer où se trouve le présent chapitre dans le processus de généralisation conceptuelle.

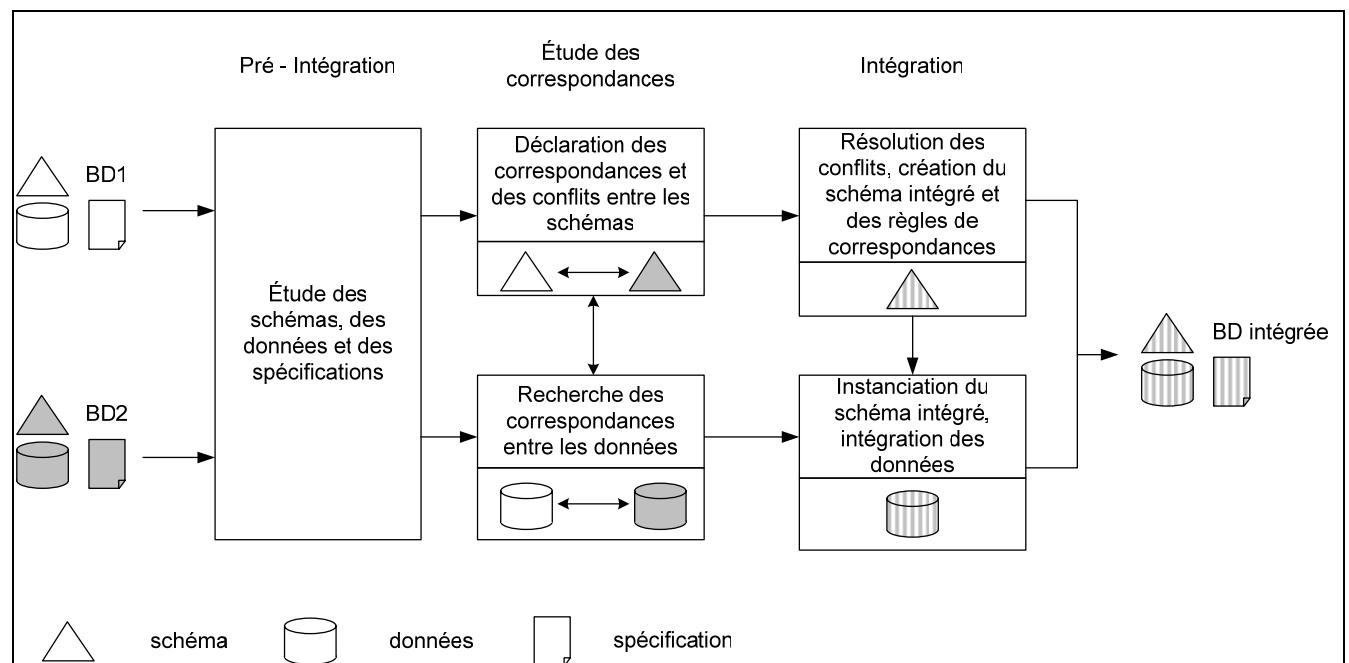


Figure 28 : Processus d'intégration de bases de données. [Sheeren D. 2005, fig.18]

Nous allons déclarer les correspondances entre les schémas ainsi que les règles de sélections sémantiques et géométriques, ce qui correspond à l'encadré pointillé sur la figure 29.

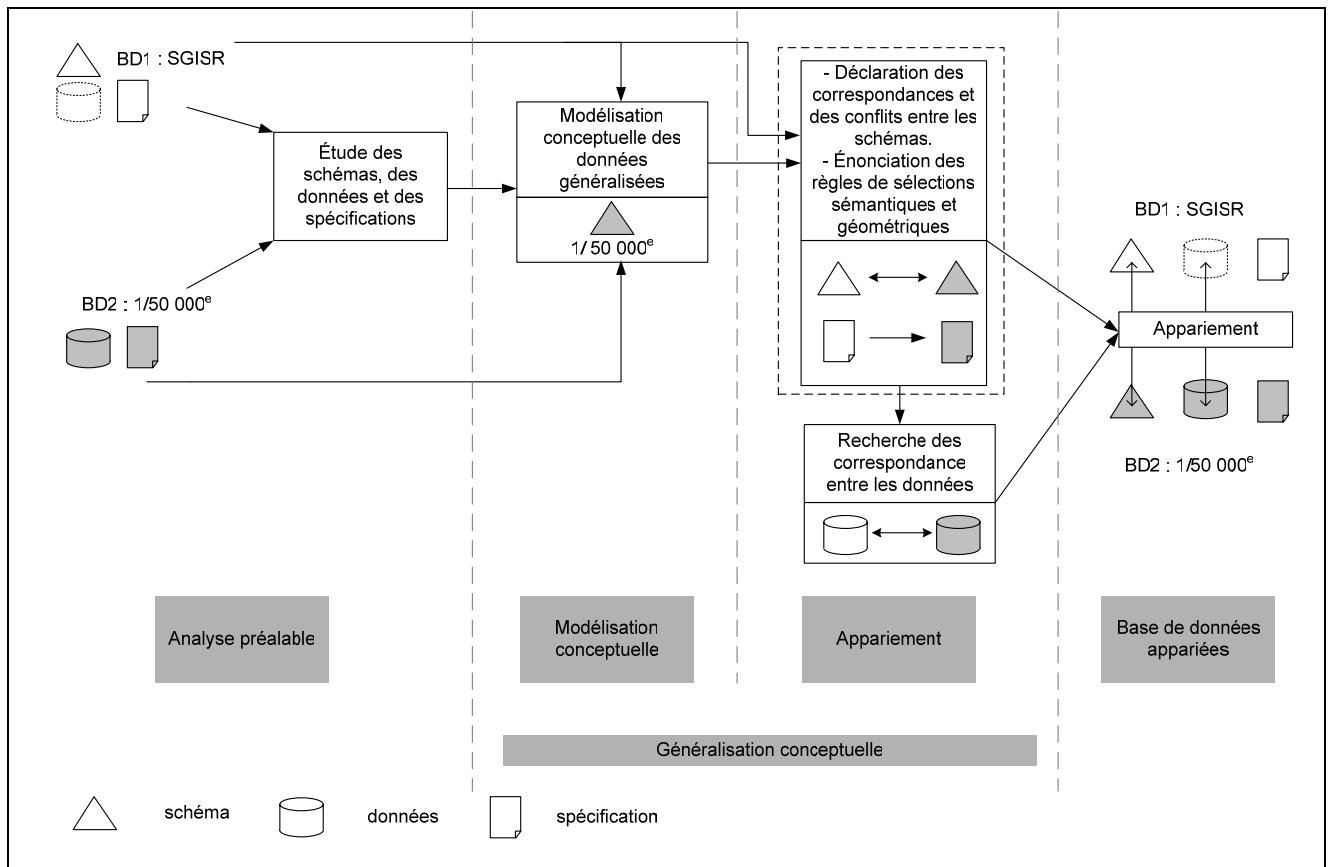


Figure 29 : Processus de généralisation conceptuelle.

Pour déclarer les correspondances entre deux schémas en vue d'une intégration, il existe un langage défini par [SPACCAPIETRA et al. 1992]. Celui-ci s'appuie sur la notion d'*Assertions de Correspondance Interschémas* (ACI). Nous allons tenter de les utiliser dans notre cas et également d'y inclure les contraintes de sélections sémantiques et géométriques.

4.2 Assertion de correspondances Inter-schémas (ACI)

4.2.1 Définition générale

Les ACI servent à identifier les correspondances entre deux schémas et également à détecter les conflits qui pourraient exister entre ceux-ci. « Pour apparier les deux modèles, il convient de préciser quels sont les éléments qui expriment les mêmes phénomènes du monde réel et comment ils se correspondent dans les bases. » [SHEEREN D. 2005, p. 32]. Par exemple, la classe « Route » d'une base de données pourrait correspondre à la classe « Voie » de la seconde. Il est alors nécessaire de préciser dans quelle mesure ces deux classes représentent le même phénomène.

Il est important de remarquer que les ACI définissent les correspondances en intention, cela signifie qu’elles travaillent au niveau des types. Par opposition, on pourrait travailler en extension c'est-à-dire au niveau des instances de la base de données. Les correspondances ne sont pas définies comme, par exemple : *l’objet² <intégration de bases de données> du schéma 2 correspond à l’objet <intégration de base de données> du schéma 1*, mais comme : *chaque objet du schéma 2 correspond à l’objet du schéma 1 qui a la même valeur d’attribut A*. [PARENT C. & SPACCAPIETRA S. 1996, pp. 5-6] Cet exemple permet de faire la différence entre la déclaration en intension ou en extension. Le premier cas s’appuie directement sur une instance particulière de la première base pour la comparer à une instance de la seconde, tandis que le second cas utilise le type « objet » de la première base et le compare au type « objet » de la seconde en précisant quel attribut les met en correspondance.

La forme générale d’une ACI est la suivante :

ACI BD1.Élément ₁ <Rel> BD2.Élément ₂
AIC <i>Identifiants correspondants</i>
AAC <i>Attributs correspondants</i>

Les spécifications suivantes doivent être fournies pour définir complètement une ACI :

- ***quels sont les éléments en correspondance :***

Les éléments en correspondance peuvent être soit des occurrences ou des valeurs (classe, attribut, valeur), soit des chemins reliant les occurrences et/ou les valeurs (relation). Dans le premier cas, les éléments liés sont simplement désignés par leur nom qualifié (comme dans l'exemple général). Dans le second cas, les chemins sont notés en énumérant les éléments qu'ils traversent. Par exemple, une association de la première base reliant les routes et les ponts est en correspondance avec une association de la seconde reliant les chemins et ponts. Il arrive dans certains cas qu'il faille préciser un sous-ensemble d'éléments concernés par l'assertion. Dans l'exemple précédent, il faudrait préciser que l'assertion est valable pour les routes de la première base dont le type est « chemin ».

² Dans le texte initial, l'auteur parle d' « article » parce qu'il fait référence à un exemple de gestion de bibliothèque.

- ***comment leurs extensions potentielles sont liées :***

Chaque assertion décrit avec quelle relation ensembliste les deux ensembles d'objets du monde réel sont en relation. Il peut s'agir d'une relation d'équivalence (\equiv), d'inclusion (\supseteq), d'intersection (\cap) ou de disjonction (\neq) des deux ensembles.

- ***comment les instances correspondantes sont identifiées :***

Dans la mesure du possible, les éléments des deux bases doivent être liés par un identifiant correspondant. Une clause de spécification de la correspondance entre les instances est ajoutée. Le système doit être capable de trouver dans une base l'objet correspondant à un objet donné dans l'autre base. Dans le cas de données géographiques, ce genre d'identifiant n'existe quasi jamais. Un autre moyen a été développé pour mettre les instances en correspondance. Il s'agit de la clause « *avec géométrie correspondante* » (AGC) développée par [DEVOGELE T. 1997, p. 84]. Dans la forme générale, les instances en correspondance sont décrites par la clause AIC signifiant « *avec identifiant correspondant* ».

- ***comment leurs représentations sont liées :***

La représentation des éléments en correspondance possède généralement des propriétés communes, au-delà de celles utilisées pour leur identification. Il s'agit de la clause « *avec attributs correspondants* » (AAC) dans la forme générale.

Dans l'énonciation des ACI, nous utilisons certains opérateurs logiques comme le « **ET** » : « **A** », le « **OU** » : « **V** » et enfin le « **NOT** » : « **¬** ». La base de données géographiques du 1/50 000^e est représentée par « **50K** » tandis que celle des données de référence est appelée par son nom « **SGISR** ».

4.2.2 Conflits détectés et représentés grâce aux ACI

Lors de la déclaration des ACI, un certain nombre de conflits peuvent être identifiés entre les deux schémas. [DEVOGELE T. 1997] propose une taxonomie des conflits dans le cas d'une intégration de données géographiques. Nous en reprendrons uniquement les conflits qui interviennent dans notre application.

1. Conflits de définitions de classes :

Les conflits de définitions des classes sont nombreux. Ils ont été regroupés en trois groupes :

a. Conflits de classifications :

« Les conflits de classifications apparaissent lorsque deux classes sémantiquement liées peuvent décrire des phénomènes du monde réel différents, c'est-à-dire lorsqu'un même ensemble de phénomènes du monde réel est classé différemment d'un schéma à l'autre » [DEVOGELE T. 97, p.58]. Un exemple de ce type de conflit est par exemple la distinction au niveau du nombre de voies de chemin de fer dans SGISR et au 1/50 000^e (fig. 30). Le premier modèle fait la distinction entre les voies simples et doubles tandis que le second identifie des voies simples ou multiples. Dans ce cas, on parlera de **conflit de groupement**.

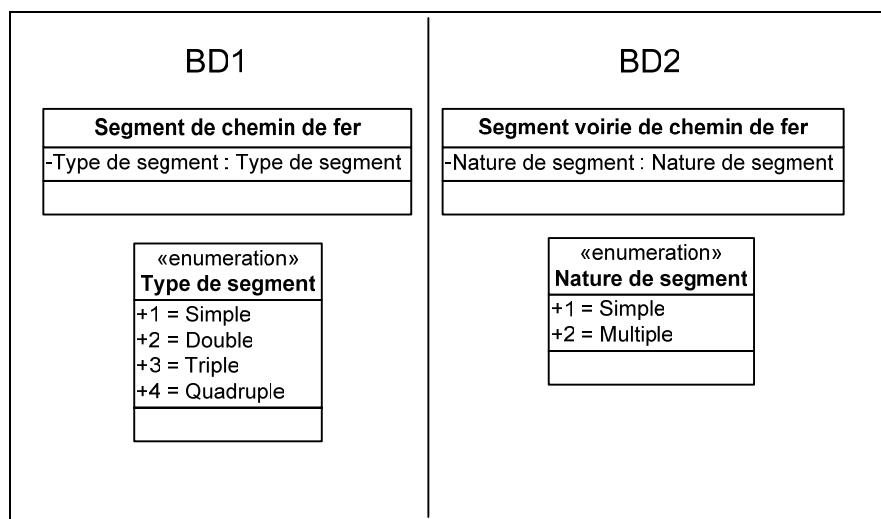


Figure 30 : Exemple de conflit de classification.

Un **conflit de résolution** apparaît au niveau des points d'eau du 1/50 000^e qui regroupe trois classes dans les données de références (sources, fontaines et puits) en une. Enfin, un **conflit de données / métadonnées** intervient lorsqu'une information de classification correspond à une donnée dans une base et à une métadonnée dans l'autre. Par exemple, les talus sont repris comme des objets dans les données sources tandis qu'ils sont des attributs dans le 1/50 000^e.

L'expression de ces conflits aux moyens des ACI s'effectue lorsque les deux ensembles en correspondance ne sont pas liés par la relation ensembliste d'équivalence ($RE \neq \equiv$). Il est cependant possible d'exprimer plus directement ces différences de schémas en ajoutant l'ensemble des éléments (classes, relations) correspondant de chaque base entre parenthèses. L'ACI obtenue est de type :

ACI BD ₁ .Élément ₁ <RE> BD ₂ .(Élément ₂₁ , Élément ₂₂ ,...)
ou
ACI BD ₁ .(Élément ₁₁ , Élément ₂₁)<RE> BD ₂ .(Élément ₂₁ , Élément ₂₂ ,...)

b. Conflits de critères de spécifications :

Pour pouvoir représenter les phénomènes du monde réel par une instance ou une classe, il est indispensable de se fixer des critères de sélection. Ceux-ci déterminent dans quelles conditions le phénomène est représenté ou pas dans la base. Il existe un conflit de critère de spécifications (plus précisément, conflit de critères de sélection) lorsque dans les deux bases, les critères pour la représentation ne correspondent pas. Par exemple, dans les données de référence, toutes les rivières sont représentées tandis que dans les données généralisées ne sont pas représentées celles qui sont inférieures à 100 mètres et connectées à une seule extrémité du réseau. Il en est de même pour les ronds-points représentés dans les données généralisées par un simple nœud en deçà d'un certain seuil (fig. 31).

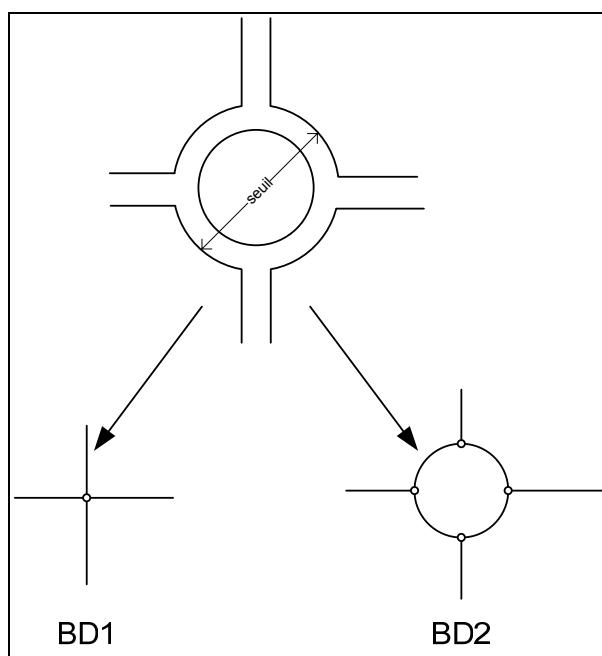


Figure 31 : Exemple de conflit de critères de spécification dans le cas de grands ronds-points.

La description de ce type de conflit au niveau des ACI, s'effectue en déclarant le critère de sélection dans une clause sélection. Notons que cette sélection peut être d'ordre sémantique ou géométrique. L'ACI est alors du type :

ACI $BD_1.\text{Élément}_1 \equiv \text{SELECTION}(critère\ de\ sélection)BD_2.\text{Élément}_2$

Par exemple dans le cas du conflit décrit ci-dessus, l'ACI serait du type :

ACI $50K.\text{Segment de rivière} \equiv \text{SELECTION}(\neg(\text{Impasse}(\text{SGISR}.\text{Segment rivière})$
 $\wedge L(\text{SGISR}.\text{Segment rivière}) < 100\text{ m})) \text{ SGISR}.\text{Segment rivière}$

« *Impasse* » est une fonction qui permet de vérifier si les nœuds du segment sont d'ordre 1. « *L* » est une fonction qui renvoie la longueur du segment. Nous utiliserons beaucoup ce type de déclaration d'ACI pour exprimer les critères de sélections sémantiques et géométriques.

c. Conflits de fragmentation :

Les abstractions des phénomènes du monde réel en objets peuvent fragmenter un phénomène en plusieurs objets suivant les valeurs et les attributs. La segmentation des objets n'est pas nécessairement identique d'une base à l'autre. Il existe un conflit de fragmentation lorsqu'un objet d'une base correspond à plusieurs objets dans l'autre base, ou quand un ensemble d'objets dans une base correspond à un ensemble d'objets dans l'autre base sans qu'il soit possible d'établir une relation bijective entre les objets. Par exemple, une route est segmentée en fonction d'un changement d'attribut sur une longueur suffisante, dans les données de références, cette longueur est de 50 mètres tandis qu'elle est de 200 dans les données dérivées (fig. 32).

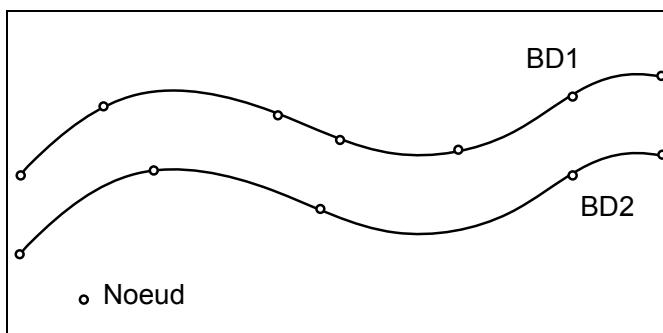


Figure 32 : Conflit de fragmentation entre deux tronçons de routes correspondants.

L'expression de ces conflits au niveau des ACI s'effectue en déclarant des éléments virtuels qui sont un regroupement d'un certain nombre d'éléments existant d'une base en correspondance avec un élément d'une autre. L'élément virtuel se définit en utilisant l'expression « *SET* » et un couple d'entiers représentant les cardinalités minimales et maximales de l'ensemble formé. L'ACI est de type :

ACI BD₁.Élément₁ <RE> BD₂.SET([N : M] Élément₂)

2. Conflits de structure :

Sont regroupés dans les conflits de structure les conflits de structure classique et ceux de stockage de l'information.

a. Conflits de structure classique :

Un conflit de structure survient lorsque les éléments en correspondance sont décrits par des concepts différents. Il peut s'agir de classes, d'attributs et de relations. Un exemple peut être facilement trouvé dans nos données. Il s'agit des talus qui sont représentés par une classe dans les données SGISR et par un attribut du segment de route dans les données dérivées (fig. 33).

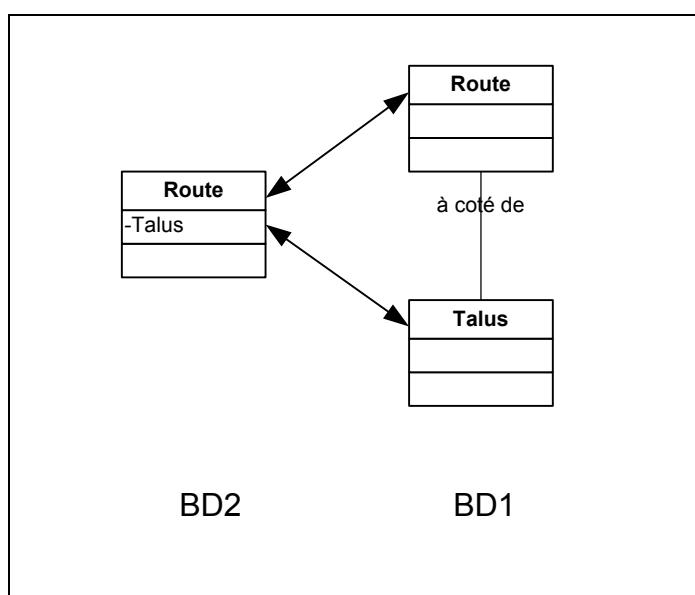


Figure 33 : Exemple de conflit de structure.

b. Conflit de stockage de l'information :

Dans les BDG, une certaine quantité d'information reste implicite, comme des relations de topologies ou de proximité. Dans certains cas, une base peut faire référence à une information implicite d'une autre. C'est le cas notamment des bermes centrales (fig. 34). Celles-ci sont représentées de manière explicite dans les données dérivées alors qu'elles sont implicites dans les données sources (espace stérile ou couvert de végétation entre deux voies).

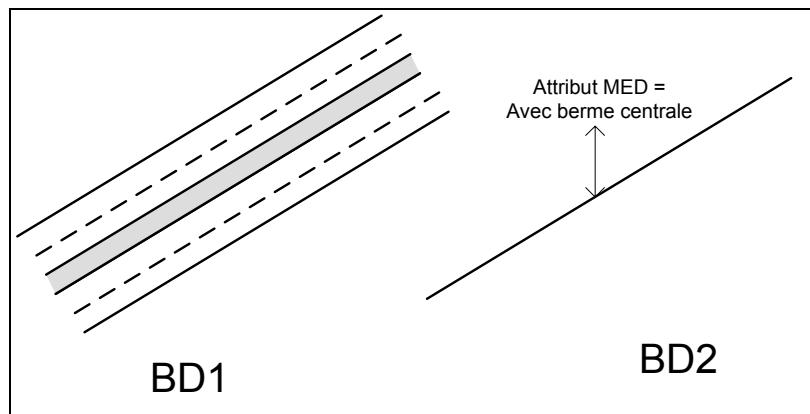


Figure 34 : Exemple de conflit de stockage d'information.

Ce genre de conflits est difficilement exprimable au niveau des ACI. Il est alors nécessaire d'inclure des fonctions qui travaillent sur la géométrie des objets et qui testent par exemple si une chaussée du même nom est proche et quelle est la nature de ce qui sépare les deux chaussées. Ce genre de processus est donc développé au niveau de l'appariement des instances des bases de données géographiques et non au niveau de l'appariement des schémas. Néanmoins, dans notre application, nous disposons dans ce cas précis d'une association entre les segments séparés par une berme centrale, ce qui permet d'éviter ce processus et de déclarer la correspondance dans l'ACI.

3. Conflit de description sémantique et géométrique :

Ils résultent d'une différence de propriété dans les classes ou attributs en correspondance. Ils peuvent porter sur les types de classes ou sur leurs attributs. Un exemple de conflit de description sémantique serait le « type de revêtement des segments de routes » décrits comme étant « Surface solide » ou « Gravier » dans les données de références tandis que les données généralisées utilisent les catégories OTAN (code militaire prenant les valeurs X,Y ou Z. Ils dépendent de la praticabilité de la voie en fonction des intempéries). Un conflit de description géométrique est repéré dans la représentation des chapelles qui sont des polygones dans les données de références et des points dans les données généralisées (fig. 35).

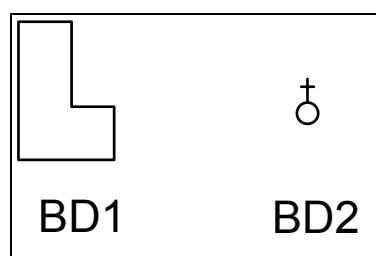


Figure 35 : Exemple d'un conflit de description géométrique.

4.2.3 Intégration et fonctions

La plupart du temps, dans le cas d'une intégration, la résolution de ces conflits ne pose pas trop de problèmes. Dans notre cas, nous ne devons pas résoudre tous les conflits. En effet, nous recherchons les différences entre les deux schémas qui surviennent suite au changement de niveau d'abstraction. Des conflits de critères de spécification par exemple devront être gérés mais les bases ne peuvent pas être modifiées.

Il arrive qu'il ne soit pas possible d'établir directement les correspondances entre les classes ou les attributs. Pour les attributs, par exemple, il est possible que l'information contenue dans une base corresponde à une combinaison d'informations contenues dans plusieurs attributs de la seconde. Dans ce cas, aux déclarations peuvent s'ajouter des fonctions préalablement définies. La clause « avec attributs correspondants » (AAC) devient alors :

AAC $f(\text{BD}_1.\text{Classe}_1.\text{Attribut}_1)$ <Rel> $g(\text{BD}_2.\text{Classe}_1.\text{Attribut}_1)$
--

Les fonctions peuvent soit être standard comme la somme, la moyenne, le minimum, le maximum ou des fonctions plus évoluées.

4.3 Exploitations des ACI pour la déclaration des contraintes de sélection

Dans le cas d'une généralisation conceptuelle par modèle, il est nécessaire de définir les critères de sélections sémantiques et géométriques des éléments qui constitueront la nouvelle base. Nous allons tenter d'utiliser les principes définis précédemment pour intégrer ces critères dans la déclaration des correspondances.

4.3.1 Critères de sélections sémantiques

Les critères de sélection sémantique semblent les plus simples à mettre en œuvre. Il nous est cependant nécessaire de différencier un critère intervenant au niveau des classes d'un critère intervenant au niveau des attributs.

La déclaration d'un critère de sélection au niveau des classes peut s'effectuer en reprenant la structure de l'ACI utilisée pour représenter un conflit de critère de spécification avec la clause *SELECTION*. La sélection s'effectue sur une (ou une gamme de) valeur d'un attribut de la classe. Celle-ci est précisée dans la clause *SELECTION*. L'ACI devient alors :

ACI BD₁.Elément₁ ≡ **SELECTION(BD₂.Elément₂, Attribut₂₁ = "valeur")BD₂.Elément₂**

Ceci permet de déclarer que, par exemple, la classe reprenant les lignes de végétation dans les données dérivées ne contiendrait que les lignes de type « rangées d’arbres ». Nous avons également à faire au phénomène inverse où une classe des données généralisées regroupe plusieurs classes des données de référence. Ce cas peut également se déclarer au niveau des ACI de la même manière que les conflits de classifications. Dans ce cas, on précise pour une classe des données généralisées à quelles classes elle correspond dans les données de référence.

ACI BD₁.Élément₁ ≡ BD₂.(Élément₂₁, Élément₂₂)

Dans le cas des « segments de sentiers » des données généralisées, par exemple, ils regroupent les « segments de sentier » et les « pistes cyclables » des données sources.

Dans certains cas, ces opérations ne s'effectuent pas au niveau des classes, mais sont précisées au niveau des attributs ou même au niveau des énumérations des valeurs possibles prises par les attributs. Dans ce cas, on précise pour chaque attribut ou chaque énumération à quoi il ou elle correspond dans l'autre base. Les ACI obtenues sont de type :

ACI BD₁.Élément₁ ≡ **BD₂.Élément₂**

AAC BD₁.Élement₁.Attribut₁ =

SELECTION(BD₂.Élément₂.Attribut₂₁ = "valeur")BD₂.Élément₂.Attribut₂₁

ACI BD₁.Élement₁ ≡ BD₂.Élement₂

AAC $BD_1.\text{Élément}_1.\text{Attribut}_{11} = BD_2.\text{Élément}_2.\text{Attribut}_{21}$

énumérations

Attribut₁₁

Attribut₂₁

"valeur₁" = SELECTION (BD₂.Élément₂.Attribut₂₁ = (" valeur₁ " \wedge " valeur₂ "))

Dans la clause énumération, nous reprenons chaque valeur d'attribut et nous précisons avec quoi elle est en correspondance. Il est également possible à ce niveau de détail de préciser des critères de sélection, ou, comme dans l'exemple précédent de préciser qu'une modalité de la première base correspond au regroupement de deux autres de la seconde.

4.3.2 Critères de sélections géométriques

Les critères de sélection géométrique s'effectuent de la même manière que les sélections sémantiques. Des fonctions sont utilisées pour calculer soit la longueur, la superficie, la plus grande diagonale, ... d'un objet. Ces fonctions nécessitent d'être préalablement définies. Dans

notre cas, nous nous limitons aux fonctions de base qui sont appelées « L » et « S » respectivement pour le calcul de la longueur d'un segment ou de la superficie d'un polygone. Elles appellent des valeurs en mètres et en mètres carrés, l'unité n'est pas inscrite dans l'ACI. La clause utilisée est *SELECTION* suivie du nom de la fonction utilisée et du critère :

SELECTION L((BD₁.Élément₁)> seuil m)BD₁.Élément₁

Elle peut également être utilisée au niveau des énumérations lorsqu'un objet n'est pas repris en dessous d'une certaine valeur seuil pour un de ses attributs :

ACI BD₁.Élément₁ ≡ BD₂.Élément₂

AAC BD₁.Élément₁.Attribut₁₁ = BD₂.Élément₂.Attribut₂₁

énumérations

Attribut₁₁

Attribut₂₁

"valeur₁"= SELECTION S("valeur₁")> seuil m²

4.3.3 Changements de dimensions

Dans le cas d'un changement de dimension lors du passage d'un objet d'une base de données à une autre, nous utilisons une fonction de réduction de dimension. Celle-ci permet soit de passer d'une entité linéaire à une entité ponctuelle, soit d'une entité zonale à une entité linéaire, soit d'une entité zonale à une entité ponctuelle (fig. 36). Le cas du passage d'une entité ponctuelle à une autre sorte d'entité géométrique est rare dans le cas d'une généralisation. Nous nommons cette fonction REDUC_DIM_X_Y, où l'on remplace respectivement le X et le Y par le type de géométrie duquel on vient et celui vers lequel on va (avec Z pour zonal, L pour linéaire et P pour ponctuel).

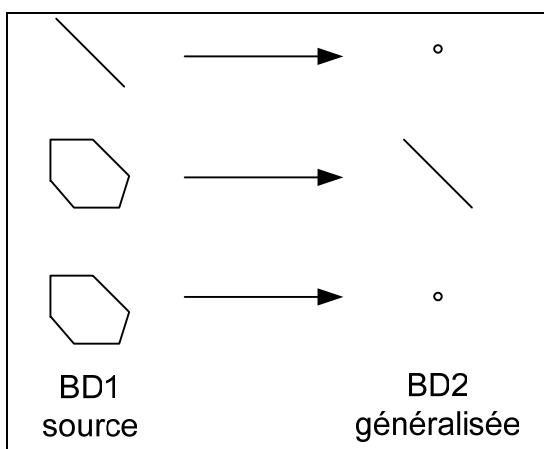


Figure 36 : Réduction de dimension lors du passage d'une BDG source à une BDG généralisée.

4.4 Définition des ACI pour l'ensemble des objets du modèle SGISR et 1/50 000

Dans la suite du chapitre, nous allons rechercher pour chaque objet du modèle du 1/50 000^e quelles sont les correspondances dans le modèle SGISR. Nous effectuons la comparaison dans le sens 1/50 000^e – SGISR par souci d'exhaustivité dans notre démarche. De cette manière nous sommes certain de reconstruire tous les objets du 1/50 000^e. Dans la présentation des ACI, nous reprenons les classes en correspondance et nous y ajoutons un lien pointillé, qui ne répond à aucun formalisme, pour indiquer si la correspondance est de type [1 : 1], [1 : n] ou [n : m]. La situation idéale serait de toujours avoir plus de données dans SGISR que dans le 1/50 000^e et de n'effectuer que des sélections. Nous verrons que ce n'est pas toujours le cas et que certains éléments du 1/50 000^e doivent être recomposés.

4.4.1 High-Voltage Network

Cette ACI met en correspondance la classe « High-Voltage line segment » du 1/50 000^e et la classe « High-voltage line segment » de SGISR (fig.37). Intéressons nous à la manière dont les données sont représentées dans les deux bases pour pouvoir les mettre en correspondance. Dans le cas de SGISR, comme on peut le voir sur la figure 37, les lignes haute tension sont composées de trois segments qui se regroupent en une seule ligne triphasée, tandis que dans le 1/50 000^e, les lignes à haute tension ne sont représentées que par un seul segment. Nous avons ici le cas typique d'un conflit de fragmentation.

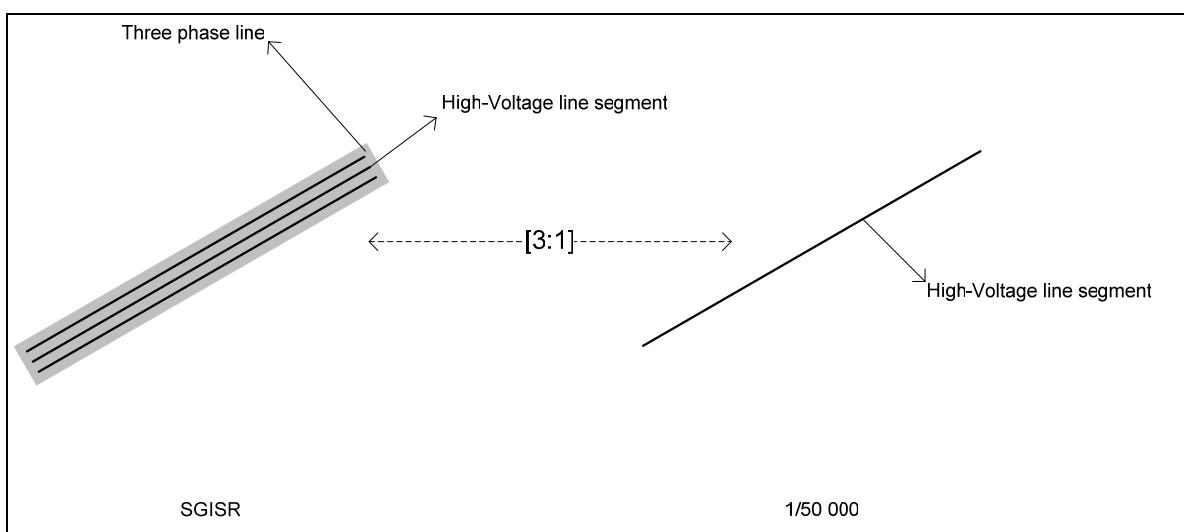


Figure 37 : Structure des lignes haute tension dans SGISR et le 1/50 000^e

Celui-ci est exprimé au niveau de l'ACI par la clause « SET » comme nous l'avons vu précédemment. De plus, une contrainte de sélection géométrique doit également apparaître, en effet, les segments inférieurs à 500 mètres ne sont pas représentés, nous incluons donc en plus un critère de sélection basé sur une fonction de calcul de longueur dans l'ACI.

Les codes des valeurs d'attributs sont mis en correspondance au niveau de l'énumération des types. D'une manière générale, le 1/50 000^e sera représenté dans une couleur turquoise, tandis que SGISR sera représenté dans une couleur verte, les critères de sélection et opérateurs particuliers seront en rouge.

Notons qu'une contrainte n'est pas formalisée au niveau de l'ACI, il s'agit d'une généralisation possible si des lignes sont fort proches, selon une carte de réseau à une échelle plus petite. Cette contrainte est plus d'ordre de généralisation structurelle que conceptuelle. En effet, elle est là uniquement pour permettre une meilleure lisibilité de la carte. Il n'est donc pas nécessaire de la reprendre dans l'ACI (fig. 38 et 39).

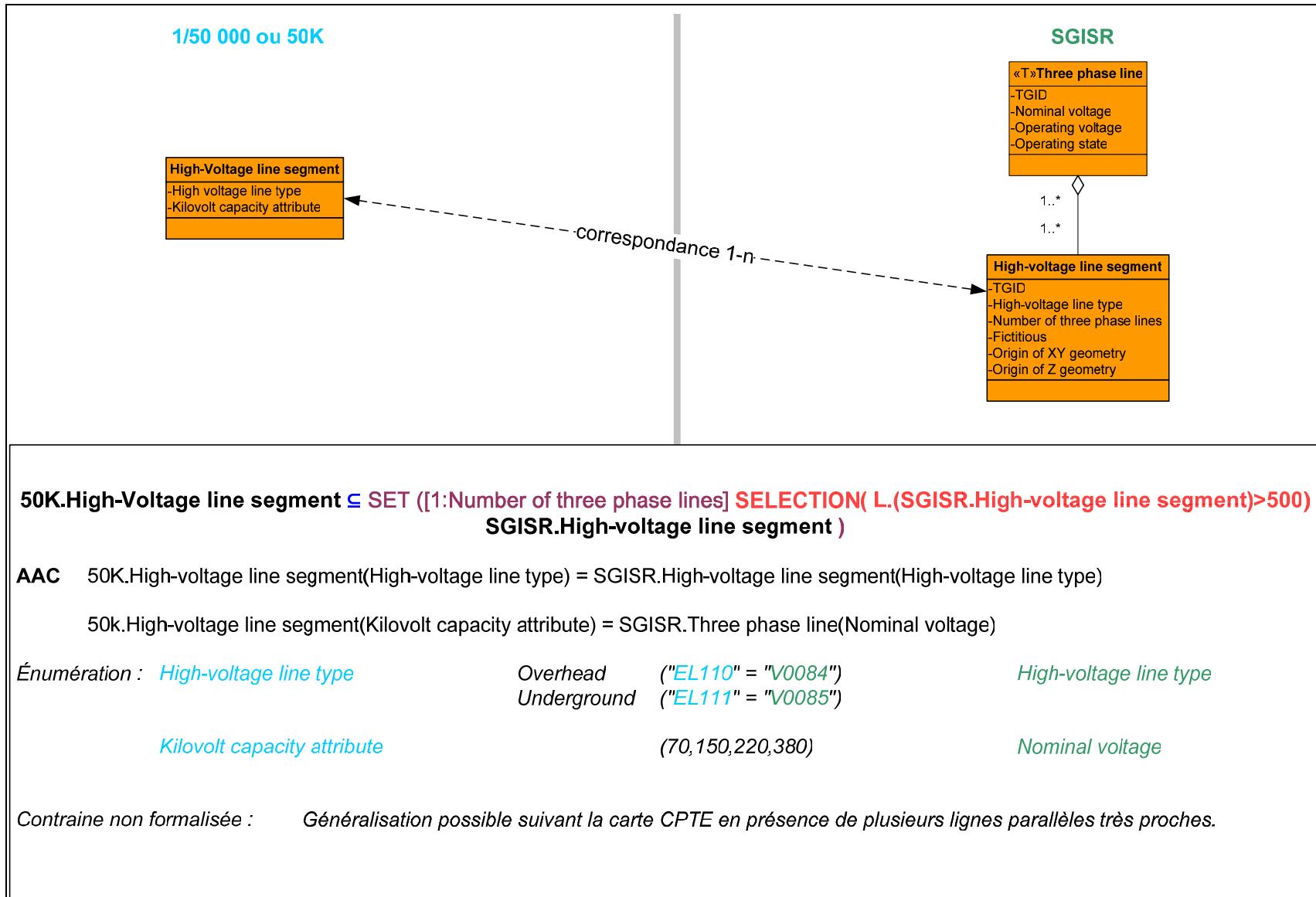


Figure 38 : ACI High-voltage network 1/3

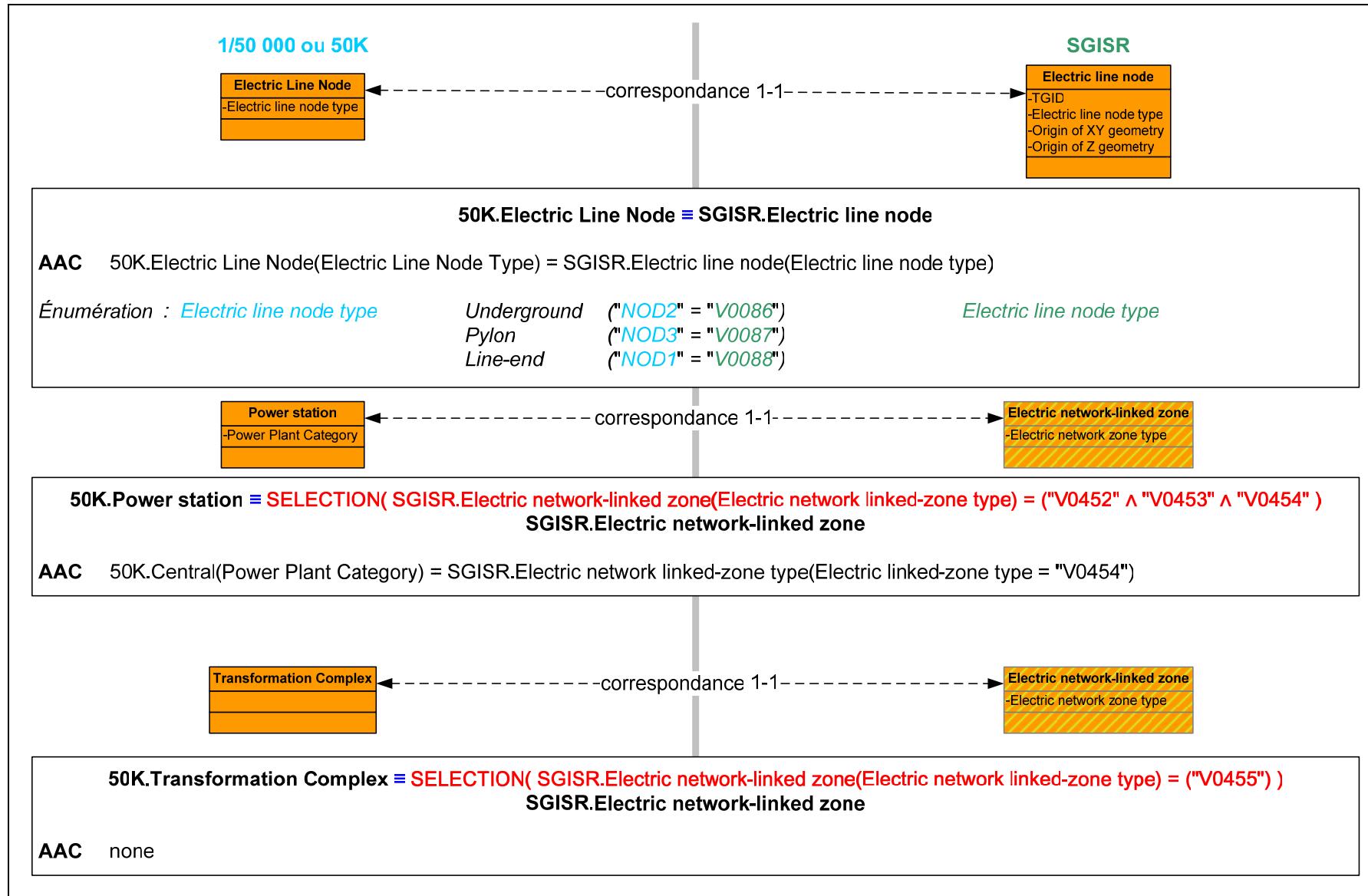


Figure 39 : ACI High-voltage network 2/3.

Les nœuds du réseau électrique sont tout à fait en correspondance dans les deux modèles, il n'y a donc rien de spécial au niveau de l'ACI. Les « Power station » (Centrale électrique) et les « Transformation complex » (Complexe de transformation) sont des regroupements de diverses modalités possible de la classe « Electric network-linked zone type » (Zone de connexion du réseau électrique). Elles sont identifiées avec une clause *SELECTION*.

Il en est de même pour les « Wind turbine » (Aérogénérateur) qui correspondent à une valeur d'attribut des « Other particular construction » (Autre construction particulière) (fig.40). Il s'agit d'un conflit de structure, les deux objets ne sont pas représentés par les mêmes concepts dans les deux bases. Cependant, celui-ci ne prête pas à conséquence dans notre application, il suffit d'effectuer une sélection sur les objets nécessaires dans les données de références.

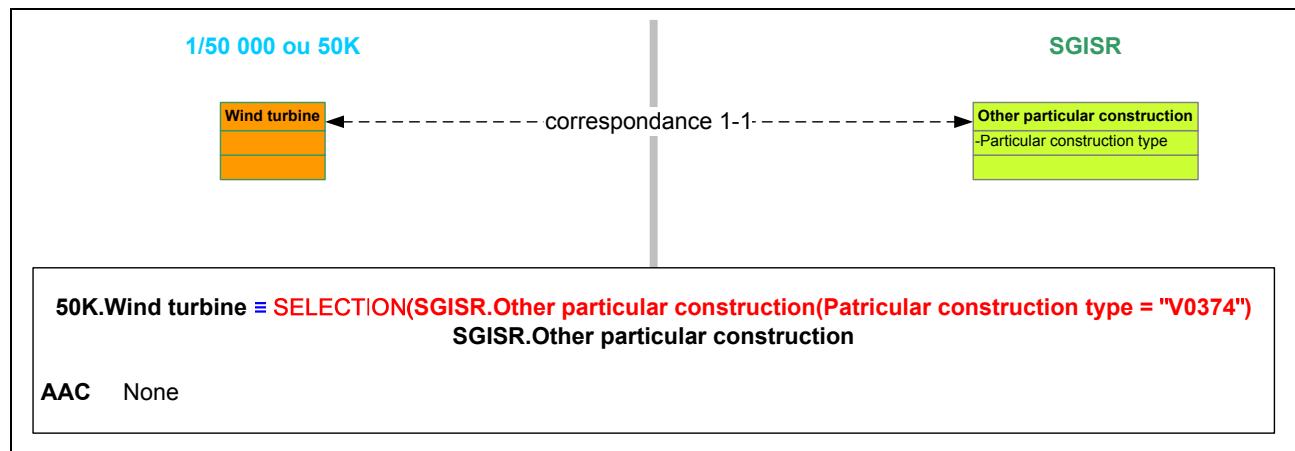


Figure 40 : ACI High-voltage network 3/3.

4.4.2 Administrative division

Le cas des divisions administratives peut sembler simple de prime abord, cependant comme l'unité de base n'est pas la même dans les deux modèles, les correspondances entre ceux-ci ne seront pas directes. En effet, SGISR utilise le secteur statistique qui est composé en unités de plus en plus grandes pour arriver finalement à représenter l'état. Dans le cas des données généralisées, l'unité de base est la commune. Il est donc nécessaire de chaque fois effectuer le lien en se basant sur une sélection de tous les secteurs statistiques composant soit la commune, le district, la province, la région ou l'état. La clé utilisée pour ce regroupement est le code NIS qui permet de faire la distinction. Il ne serait pas possible d'effectuer directement les liens entre les mêmes entités car dans SGISR les classes qui se composent à partir des secteurs statistiques sont de type textuel tandis que dans les données dérivées, elles sont de types géographiques (Feature Class). Les ACI obtenues sont représentées à la figure 41 et 42. On peut y voir que l'appariement se fait sur le « NIS-Code » qui dans sa structure propre permet de retrouver la commune, la région et la province. Le lien pourra peut-être être effectué directement sur les classes correspondantes lors de l'implémentation de la base de données SGISR, il serait alors nécessaire d'étudier ce qui sera exactement représenté dans les classes textuelles (« Municipality », « District », « Province », « Region » et « State »).

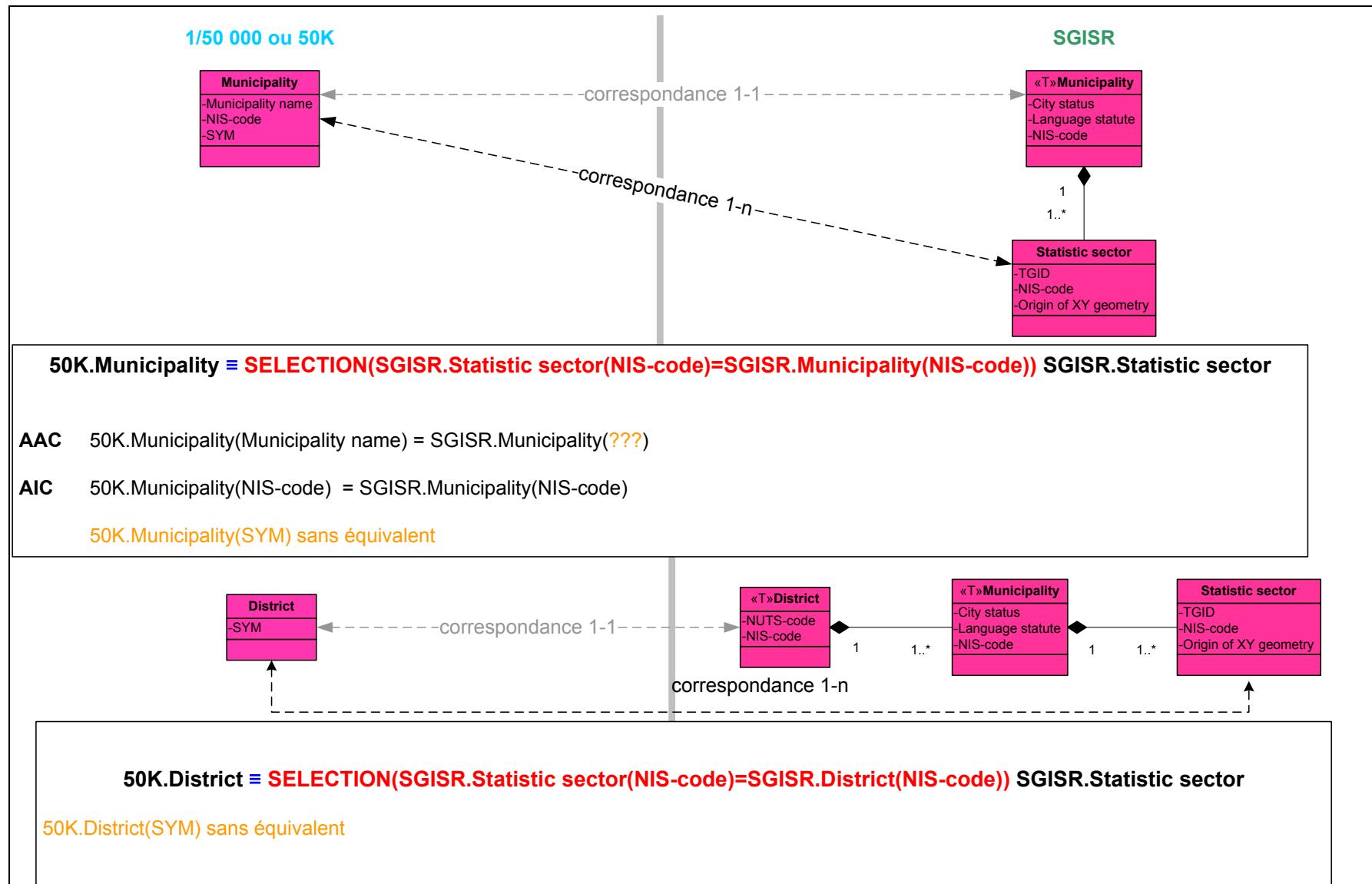


Figure 41 : ACI Administrative division 1/2.

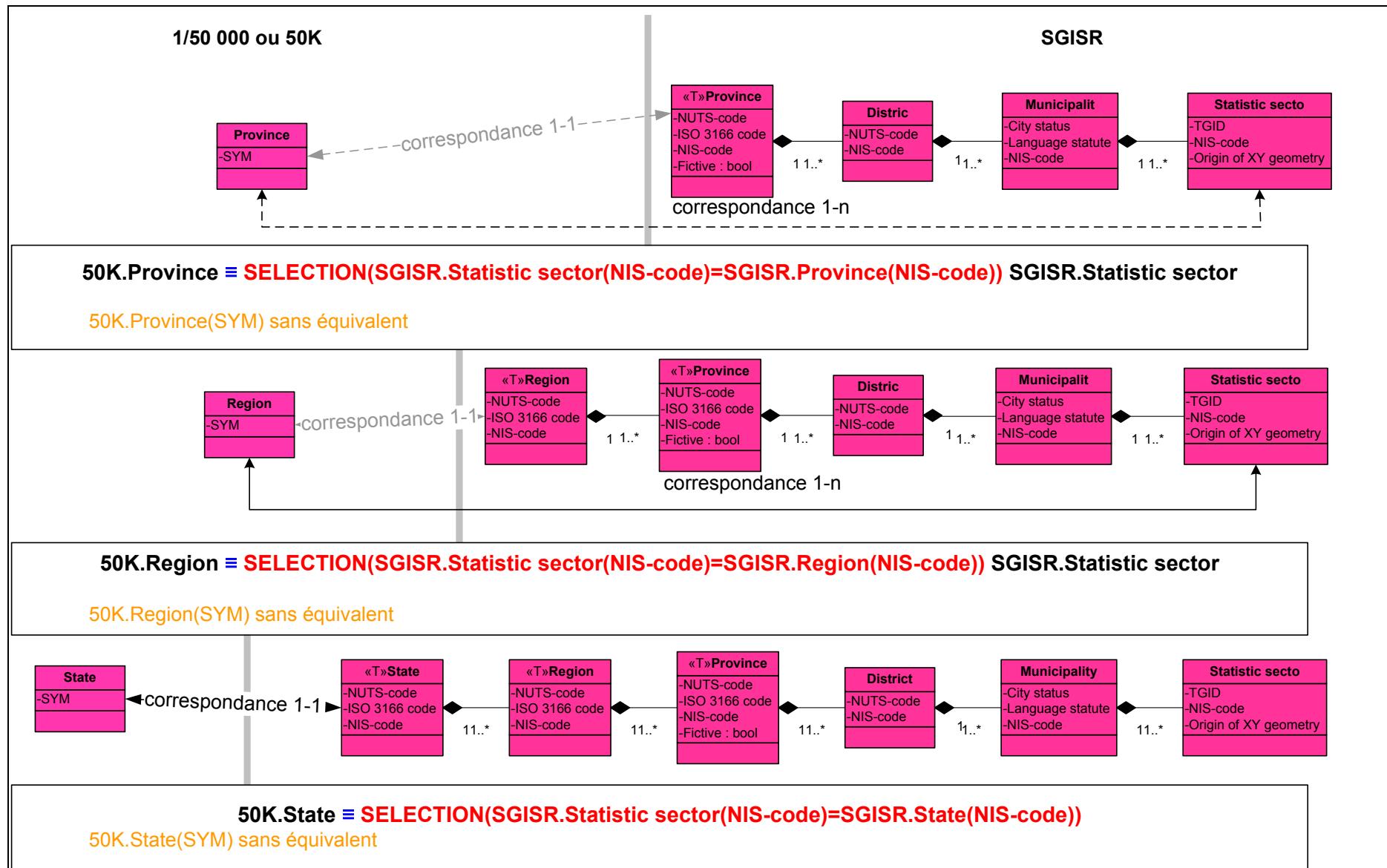


Figure 42 : ACI Administrative division 2/2.

Dans les ACI, on peut constater que l'attribut « SYM » est sans correspondance dans SGISR. Comme nous l'avons vu précédemment, il s'agit d'un attribut réservé à la généralisation cartographique du 1/50 000^e que les responsables du projet GenMap souhaitaient conserver dans la base de données. Les valeurs de cet attribut ne pourront donc pas être reprises de SGISR. Les noms des communes ne sont également pas stockés dans la base de données de référence alors qu'ils le sont dans les données généralisées. Il sera nécessaire d'effectuer un lien avec la base de toponymie s'il est nécessaire de retrouver les noms (les noms étant déjà tous présents dans les données du 1/50 000^e et comme ils ne changent que très rarement, il n'est peut-être pas nécessaire d'effectuer ce lien supplémentaire).

4.4.3 Landcover

La classe « Landuse » (Occupation de sol) est typique des conflits de classification (fig. 43). Les éléments qui la composent doivent être repris de deux classes de SGISR (« Parcel » et « Wetland »). Les valeurs précises à reprendre sont donc énumérées au niveau de l'ACI. De plus, pour cette classe, il y a beaucoup de critères de sélection géométrique. Ils sont énoncés au niveau des énumérations car les valeurs de sélection sont différentes en fonction du type d'occupation du sol. Les sélections sont essentiellement effectuées avec la fonction « S » qui renvoie la surface du polygone en question, les valeurs reprises dans les ACI sont en mètres carrés.

Un cas est cependant plus complexe qu'une « simple » sélection. Il s'agit des « Lawn » (Pelouse) qui n'ont pas le même critère de sélection si elles sont situées dans le parc d'un château, un complexe sportif ou dans des ruines. Dans ce cas, il est nécessaire d'inclure une procédure qui teste si le polygone est dans ou hors cette zone particulière et alors seulement d'effectuer la sélection.

Pour ce qui est des « Isolated tree » (Arbre remarquable) et des « Row of trees » (Rangée d'arbres), les correspondances sont directes (fig. 44). Seule un critère de sélection géométrique est nécessaire pour les rangées d'arbres afin que seules celles supérieures à 300 mètres soient reprises.

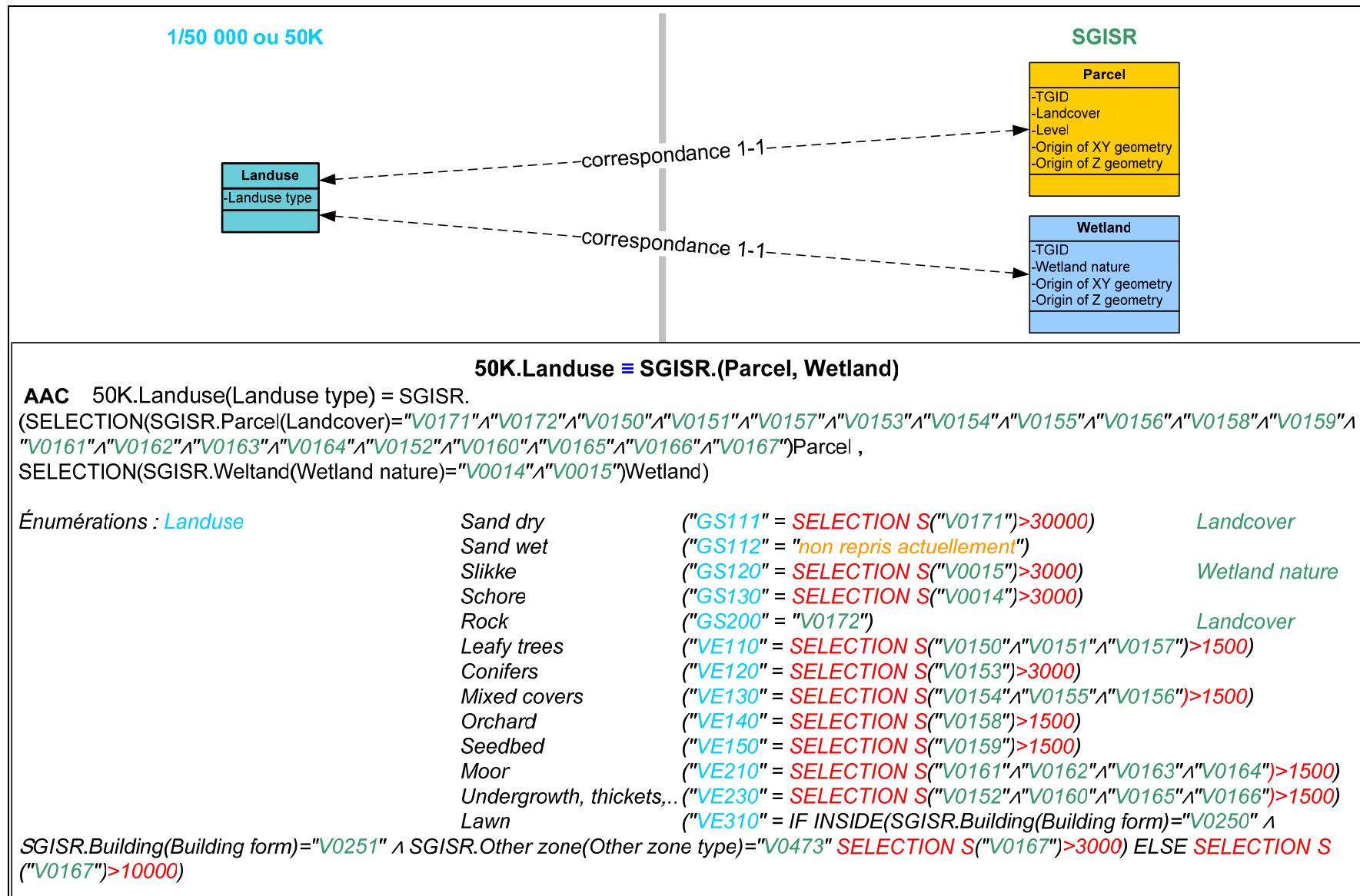


Figure 43 : ACI Landcover 1/2.

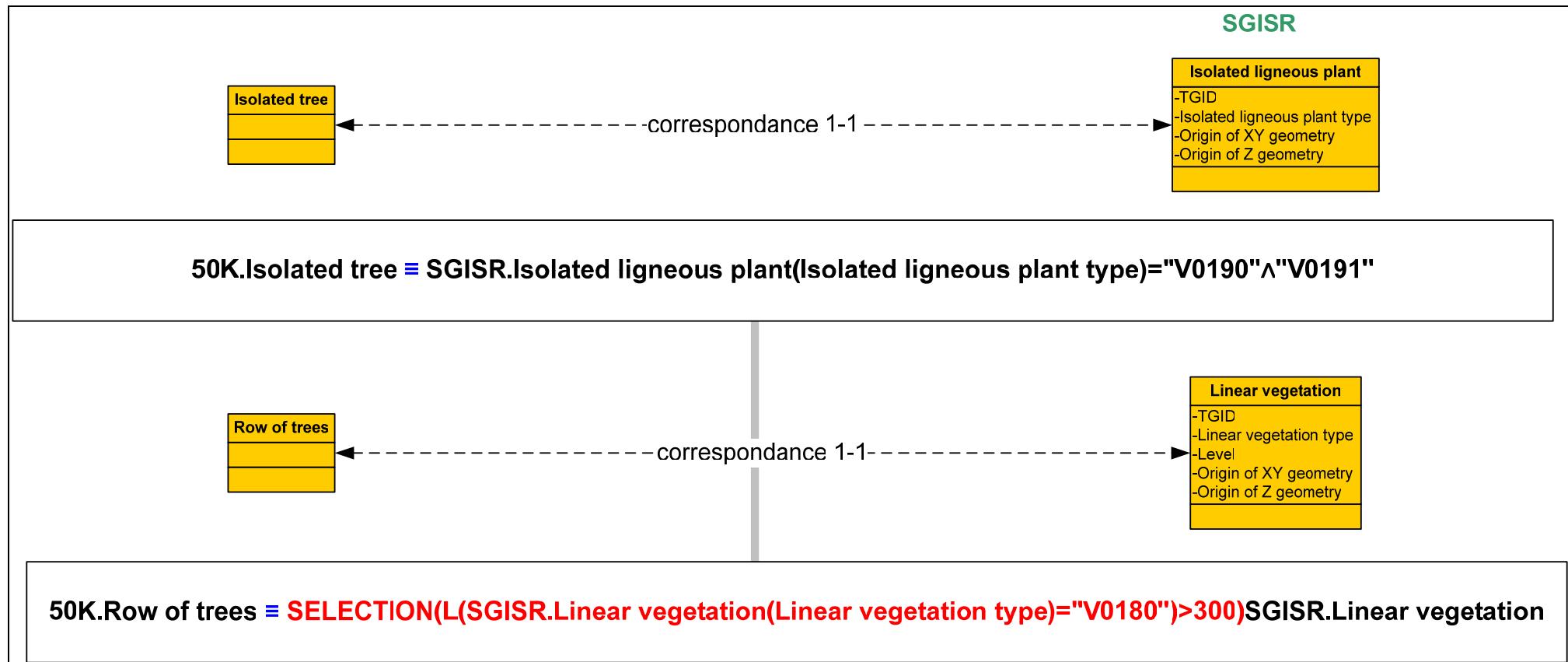


Figure 44 : ACI Landcover 2/2.

4.4.4 Rail transport network

Comme dans le cas des lignes à haute tension, le réseau de chemin de fer n'est pas représenté de la même manière dans les deux bases de données. Comme on peut le voir sur la figure 45, les lignes sont regroupées dans les données généralisées. Le cas est typique d'un conflit de fragmentation au niveau des classes et également d'un conflit de groupement au niveau des attributs.

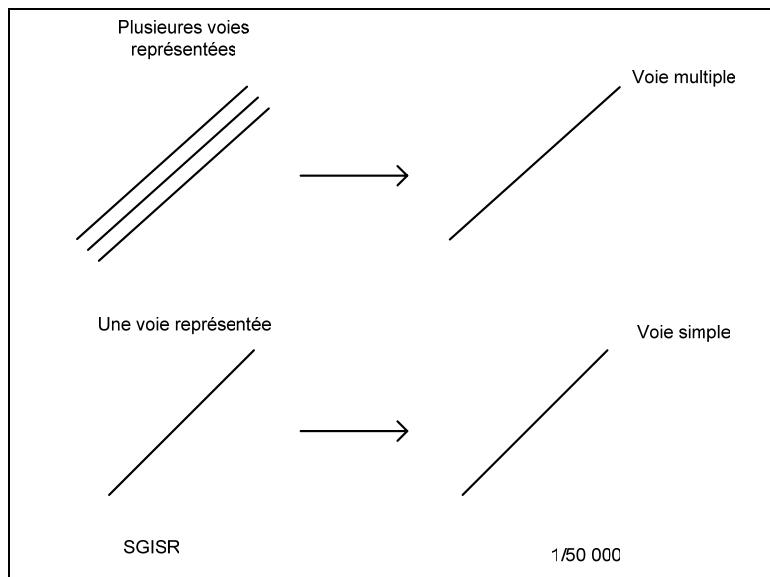


Figure 45 : Représentation des voies de chemin de fer.

Les critères de sélection géométriques ne s'effectuent pas de la même manière en fonction du type de voie de chemin de fer, les sélections seront donc différenciées. Comme on peut le constater sur la figure 46, l'attribut « Railway lane track number » (Nombre de ligne de voie ferrée) est défini en fonction de la valeur de la cardinalité de la clause *SET*, en effet, si celle-ci est [1 :1], la valeur de l'attribut 1/50 000^e sera « Simple way » (Voie simple), sinon elle sera « Multiple way » (Voie multiple). Notons également qu'une valeur n'est pas disponible dans les données de référence, il s'agit de « Secondary way railway track » (Segment de réseau de chemin de fer secondaire). Les responsables du projet GenMap nous ont assuré qu'il doit apparaître dans les données de SGISR. Comme dans les cas précédents, l'attribut « SYM » est sans correspondant.

Les autres correspondances du domaine sont de types [1:1] (fig. 47).

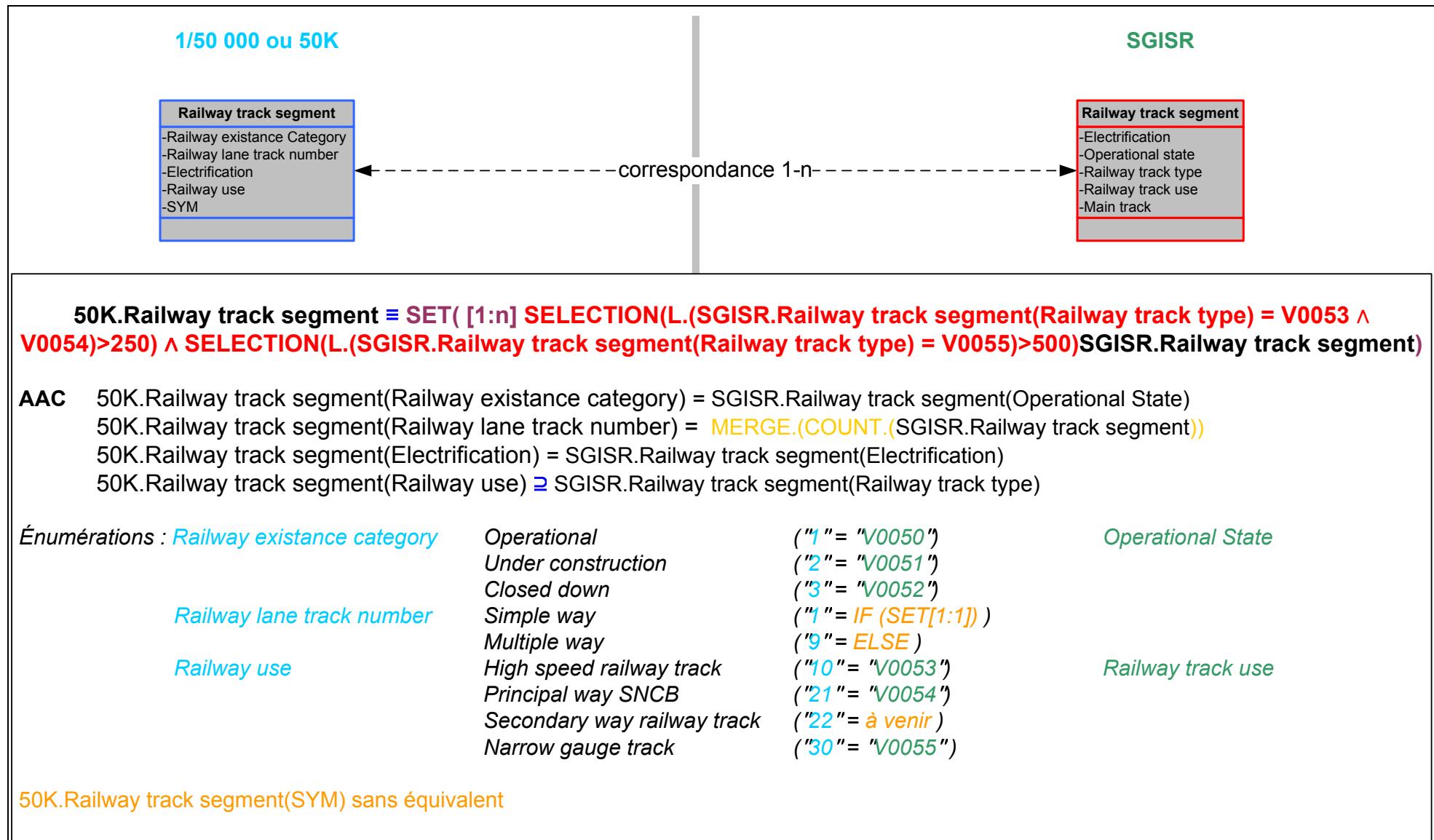


Figure 46 : ACI Rail transport network 1/2.

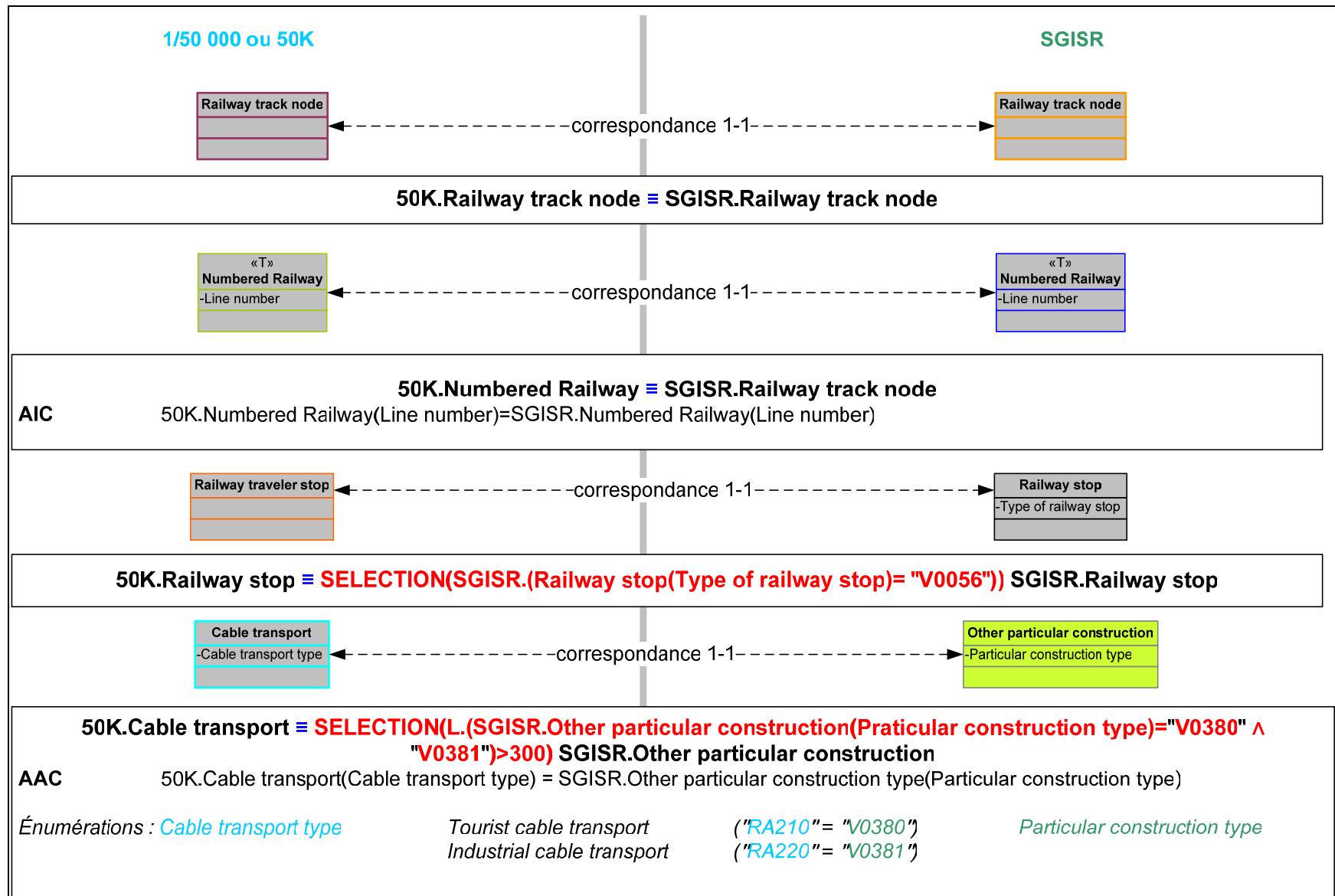


Figure 47: ACI Rail transport network 2/2.

Dans le cas des « Railway traveler stop » (Arrêt du chemin de fer pour voyageurs), nous effectuons une sélection sémantique sur la classe « Railway stop » (Arrêt du chemin de fer) pour éliminer les arrêts de tram et de métro qui ne sont pas repris dans les données dérivées. Le « Cable transport » (Transport par câble) est mis en correspondance avec certaines modalités de « Other particular construction type » (Autre construction particulière). « Cable transport » représente les bandes transporteuses industrielles ainsi que les téléphériques. Dans tous les cas, ils ne sont repris que si leur longueur est supérieure à 300 mètres, ce qui est exprimé par une sélection géométrique et la fonction « L ».

4.4.5 Road transport

Les segments de routes sont moins différenciés dans le 1/50 000^e que dans SGISR, un certain regroupement aura donc lieu entre les classes. Comme SGISR est une base de données topogéographiques, si une route est composée de deux voies de circulation séparées (cas des autoroutes), elles seront toutes deux représentées. Le 1/50 000^e ne représente qu'un axe dans ce cas avec un attribut sur cet axe qui précise si il y a ou pas une berme centrale entre les deux voies de circulation (fig. 48). Nous sommes dans le cas d'un conflit de fragmentation qui n'apparaît que, lorsque dans SGISR, il y a l'association « road segment to road segment » (segment de route connecté à un segment de route). La représentation dérivée fusionne alors les deux axes en un seul.

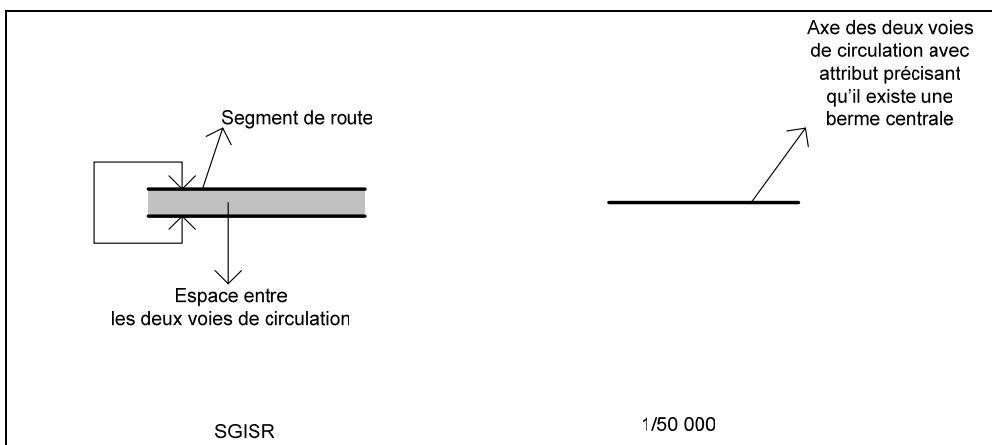


Figure 48 : Représentation des segments de routes dans SGISR et au 1/50 000^e.

L'attribut « Road median » (Berme centrale) est mis en relation avec l'association « road segment to road segment » de SGISR auquel s'ajoute une contrainte de sélection géométrique qui oblige que la berme soit d'au moins 300 mètres de longueur pour être représentée.

SGISR stocke la largeur réelle des routes, il est alors nécessaire d'effectuer une classification sur cette largeur pour la représentation dérivée (fig. 50). Il sera sans doute nécessaire de revoir les classes de largeur lors de l'établissement des critères des sélections des éléments de SGISR en fonction de la précision avec laquelle sera donnée la largeur de route.

Les « Other way segment » (Autre segment de voie) sont un regroupement de deux classes de SGISR, il s'agit de « Dirt road segment » (Segment de chemin de terre) et de « Foot/Cycleway segment » (Segment de voie non carrossable) (fig. 51). Une contrainte de sélection géométrique est d'éliminer tous les segments inférieurs à 1 kilomètre. Cependant, il est également nécessaire de conserver les segments de taille inférieure qui mènent à un détail topographique remarquable. Cette contrainte ne peut être formalisée au niveau de l'ACI, elle sera traitée manuellement.

Les sorties, noms de routes et les nœuds ordinaires s'apparent facilement, il s'agit de correspondances directes (fig.52). Cela se complique pour la gestion des ronds-points (fig. 53). En effet, dans SGISR, le croisement entre les voies est géré par des surfaces de connexion, celles-ci contiennent l'ensemble des segments et des nœuds qui font partie du croisement. Dans le cas des données du 1/50 000^e, il n'y a ni surface de route, ni surface de connexion. Il est donc nécessaire lors de l'appariement de repérer quelles surfaces de connexions correspondent à un rond-point. Ensuite, il faut identifier quels segments et nœuds sont compris dans cette surface. Enfin, il reste à appairer cet ensemble soit à un seul nœud si le rond-point est en dessous du seuil de représentation, soit à un ensemble de nœuds et de segments si la taille est supérieure. La procédure se résume à la figure 49.

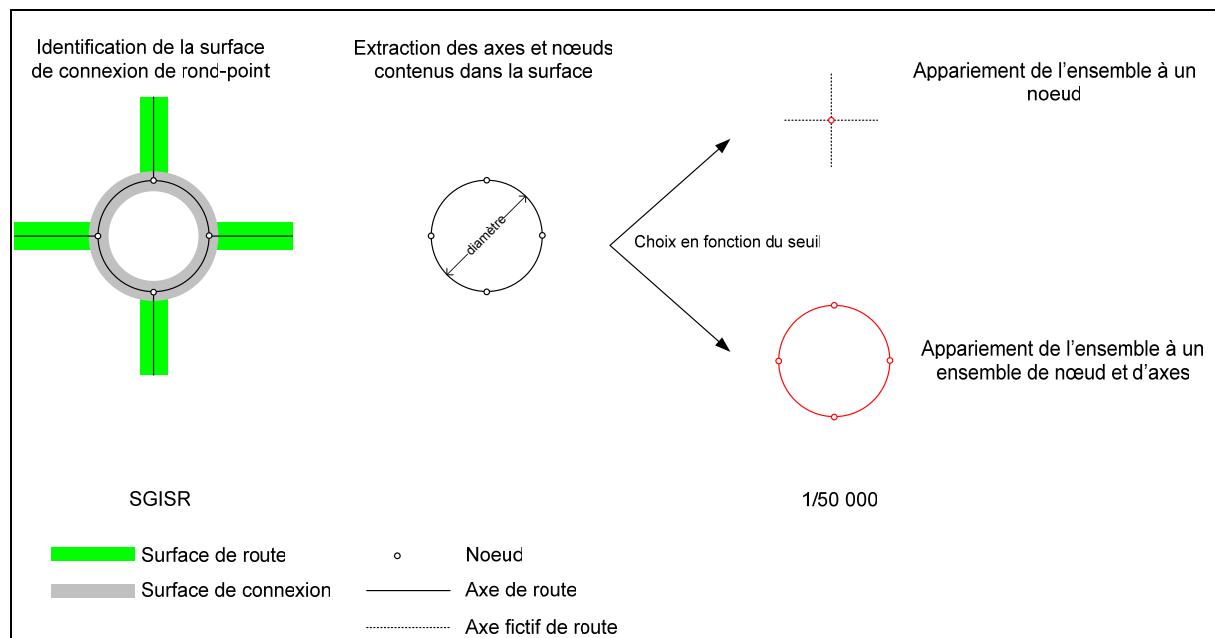


Figure 49 : Procédure d'appariement des ronds-points.

Cela se traduit dans l'ACI par la fonction « IN » qui teste si un objet est contenu dans un autre. Le seuil n'est actuellement pas défini car il n'est pas identifié dans les données de référence, les critères de sélections étant toujours en cours d'élaboration. Le cas des petits ronds-points (inférieurs au seuil) pourrait être traités de la même manière, nous choisissons de le simplifier au regroupement des nœuds qui composent le rond-point source.

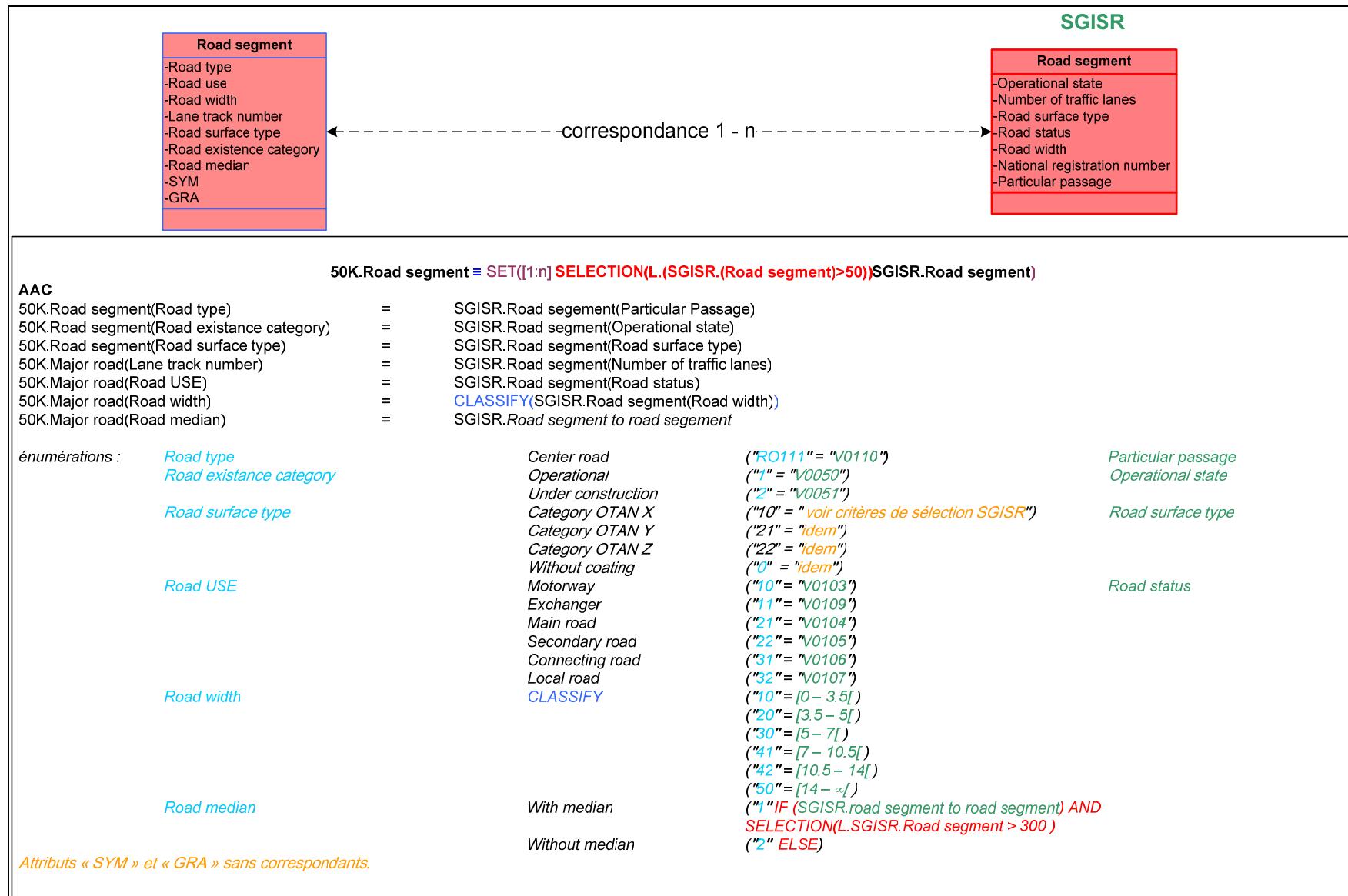


Figure 50 : ACI Road network 1/4.

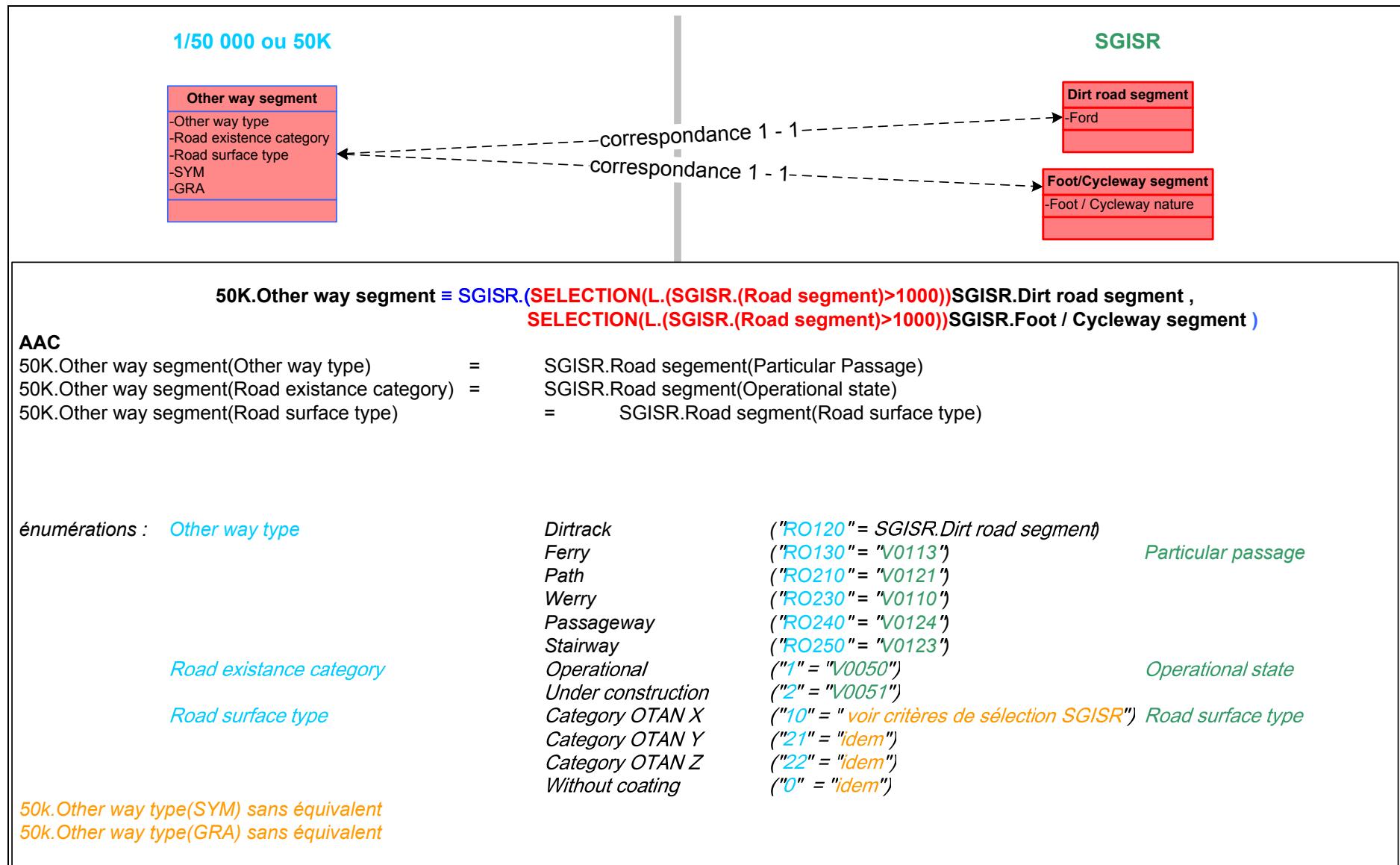


Figure 51 : ACI Road network 2/4.

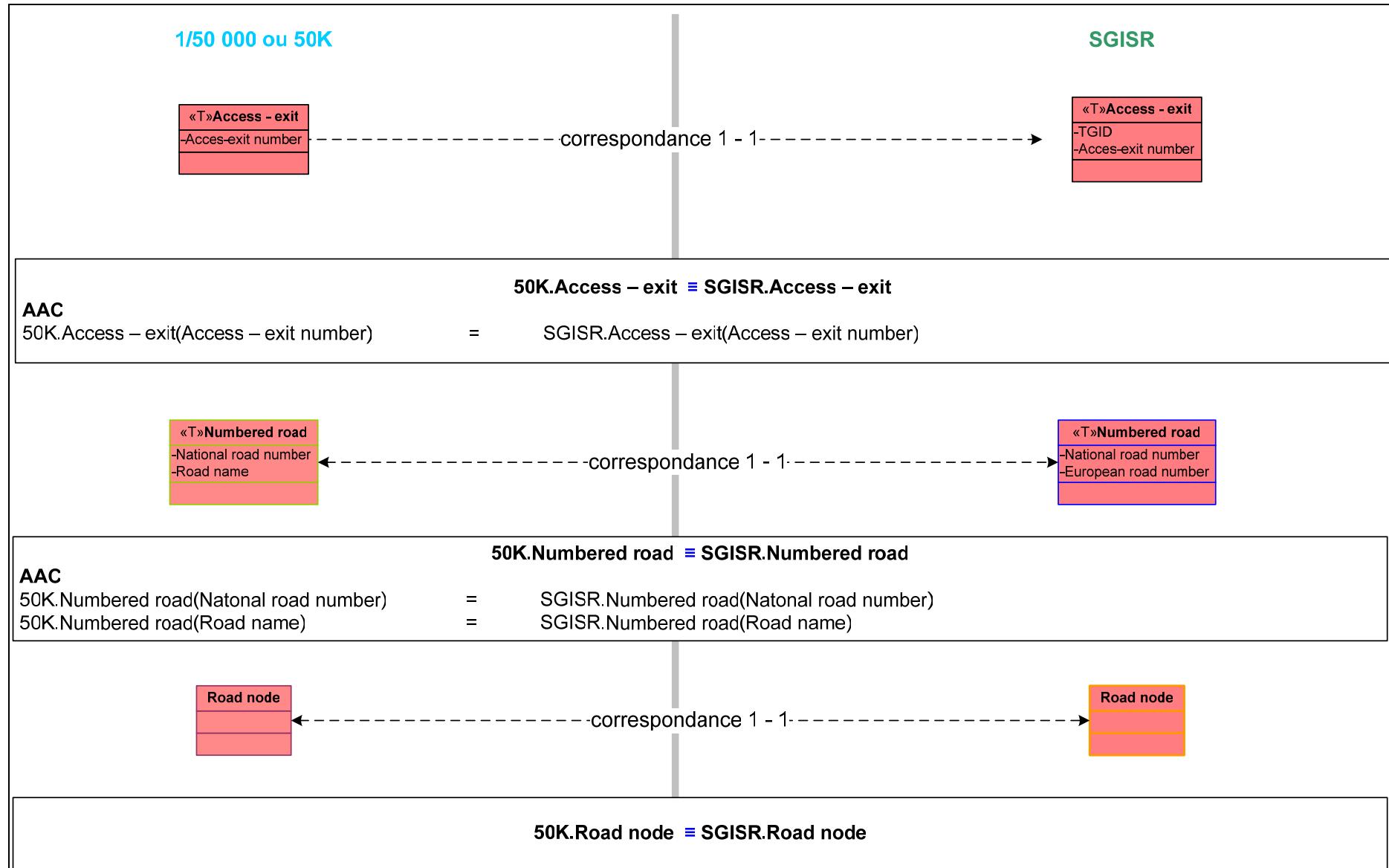


Figure 52 : ACI Road network 3/4.

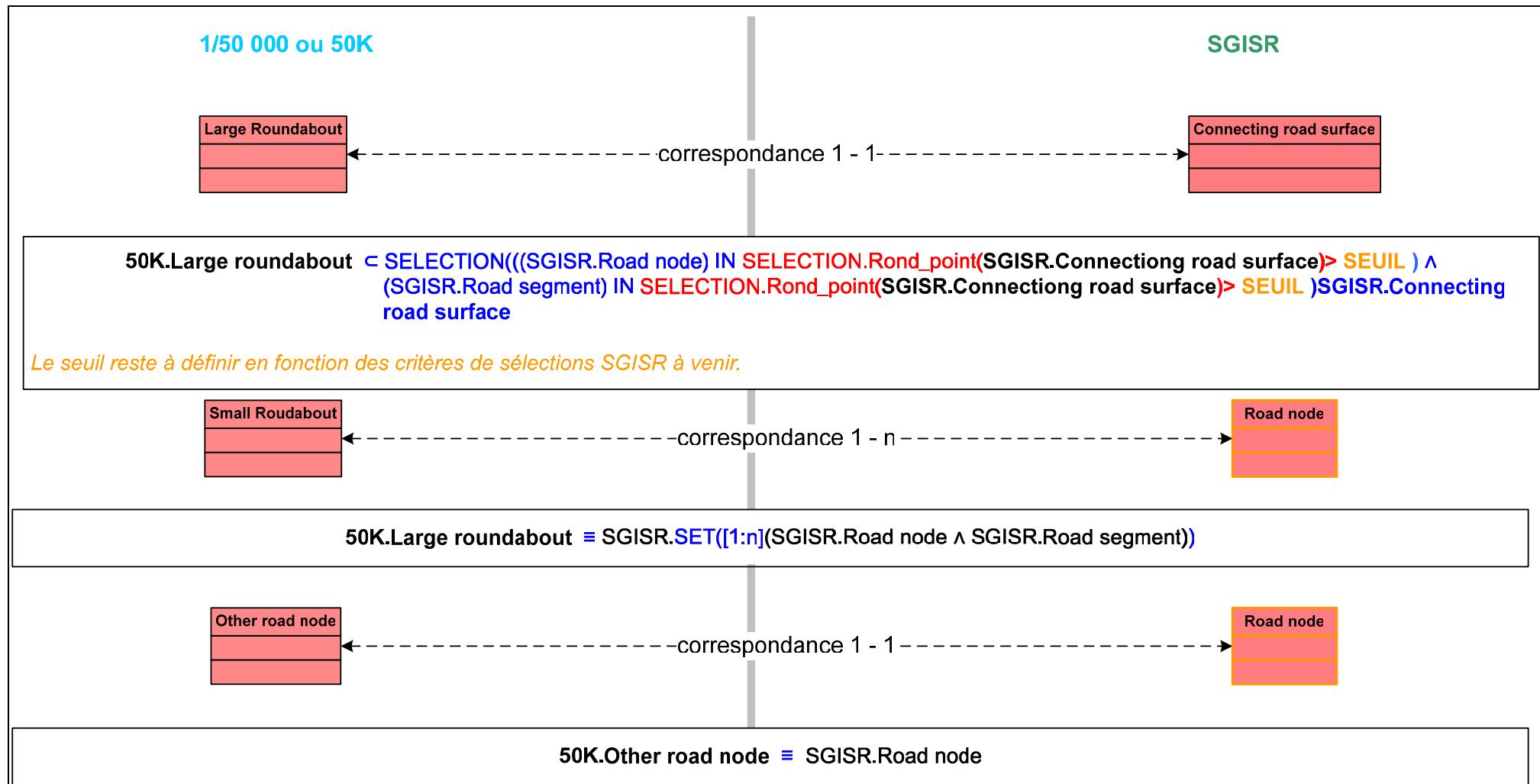


Figure 53 : ACI Road network 4/4.

4.4.6 Hydrography

Le réseau hydrographique est un exemple type des critères de sélections géométriques. En effet, seuls les segments qui sont supérieurs à 100 mètres et connectés au réseau par les deux extrémités sont conservés. Il est donc nécessaire d'utiliser deux fonctions de sélections, la première : « Impasse » qui teste si un des nœuds est de degré 1, la seconde : « L » qui calcule la longueur du segment. L'ACI obtenue est reprise dans la figure 54.

La classe « Watercourse surface » (Surface de cours d'eau) dans le 1/50 000^e n'est représentée que si le cours d'eau a plus de 50 mètres de large. Dans ce cas, il n'est plus représenté par un symbole mais par ses berges. Il est donc nécessaire de rechercher dans SGISR seulement les surfaces qui contiennent des segments dont la valeur de leur attribut « largeur » est supérieure à 50 mètres. Nous utilisons la fonction « IN » pour sélectionner les bonnes surfaces (fig. 55). La classe « Watercourse surface » possède également un attribut qui définit le type de surface : « Wartercouse surface » (Surface de cours d'eau) ou « Windening » (Élargissement). L'élargissement intervient sur un cours d'eau qui n'a pas une largeur supérieure à 50 mètres, mais qui s'élargit localement à plus de 50 mètres et qui permet ainsi le croisement de bateaux. Cet élément n'est pas apparié au sein de l'ACI car d'une part, il n'est pas représenté par les berges de la rivière et d'autre part il tend à disparaître des représentations. L'appariement s'effectuera au cas par cas au niveau des instances.

Les nœuds du réseau du 1/50 000^e sont en correspondance directe avec ceux de SGISR (fig. 56). Les « Hydraulic construction » (Construction Hydraulique) s'apparentent directement, il est juste effectué une sélection sur la valeur « Construction on pillars » (Construction sur pilotis) comme on peut le remarquer dans l'ACI représentée figure 57.

Comme nous l'avons déjà précisé lors de l'élaboration du modèle conceptuel, les sources ne sont pas représentées dans le 1/50 000^e si elles sont connectées au réseau hydrographique. Cette propriété peut être retrouvée dans les données de référence grâce à l'association « represented by » qui existe entre « Water point » et « Water node ».

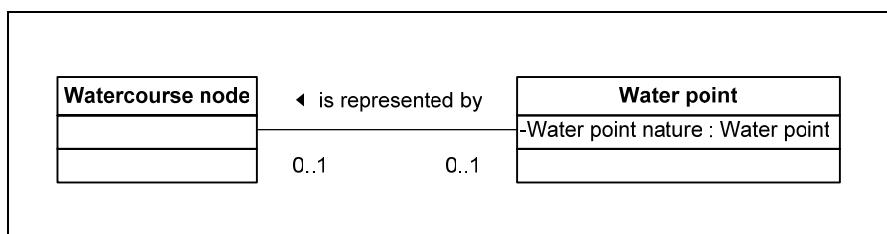


Figure 54 : Extrait du domaine Hydrography du MCD SGISR.

Lorsque cette association n'a pas lieu, cela signifie que le point d'eau n'est pas en contact avec le réseau car il n'en est pas un nœud. Dans l'ACI, nous sélectionnons donc uniquement les sources pour lesquelles il n'y a pas d'association. Notons que la classe « Water point » du 1/50 000^e reprend plus que la classe correspondante, il est nécessaire d'y ajouter des objets comme les siphons, les phares,... ceux-ci sont repris de la classe « Hydraulic construction ».

Les « Water surface » (Surfaces d'eau) s'apparentent directement aux « Water surface » de SGISR (fig. 58).

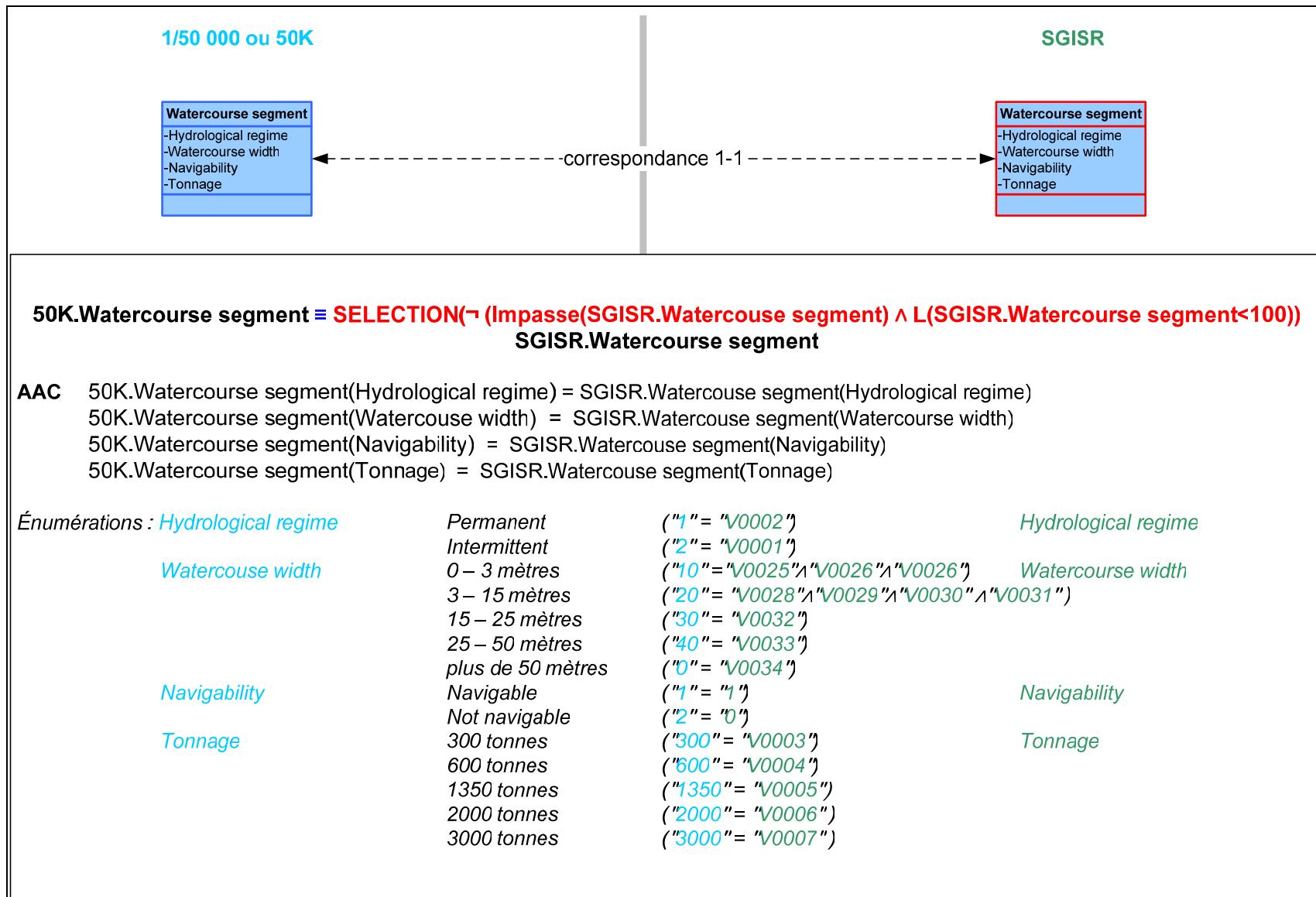


Figure 55 : ACI Hydrography 1/4.

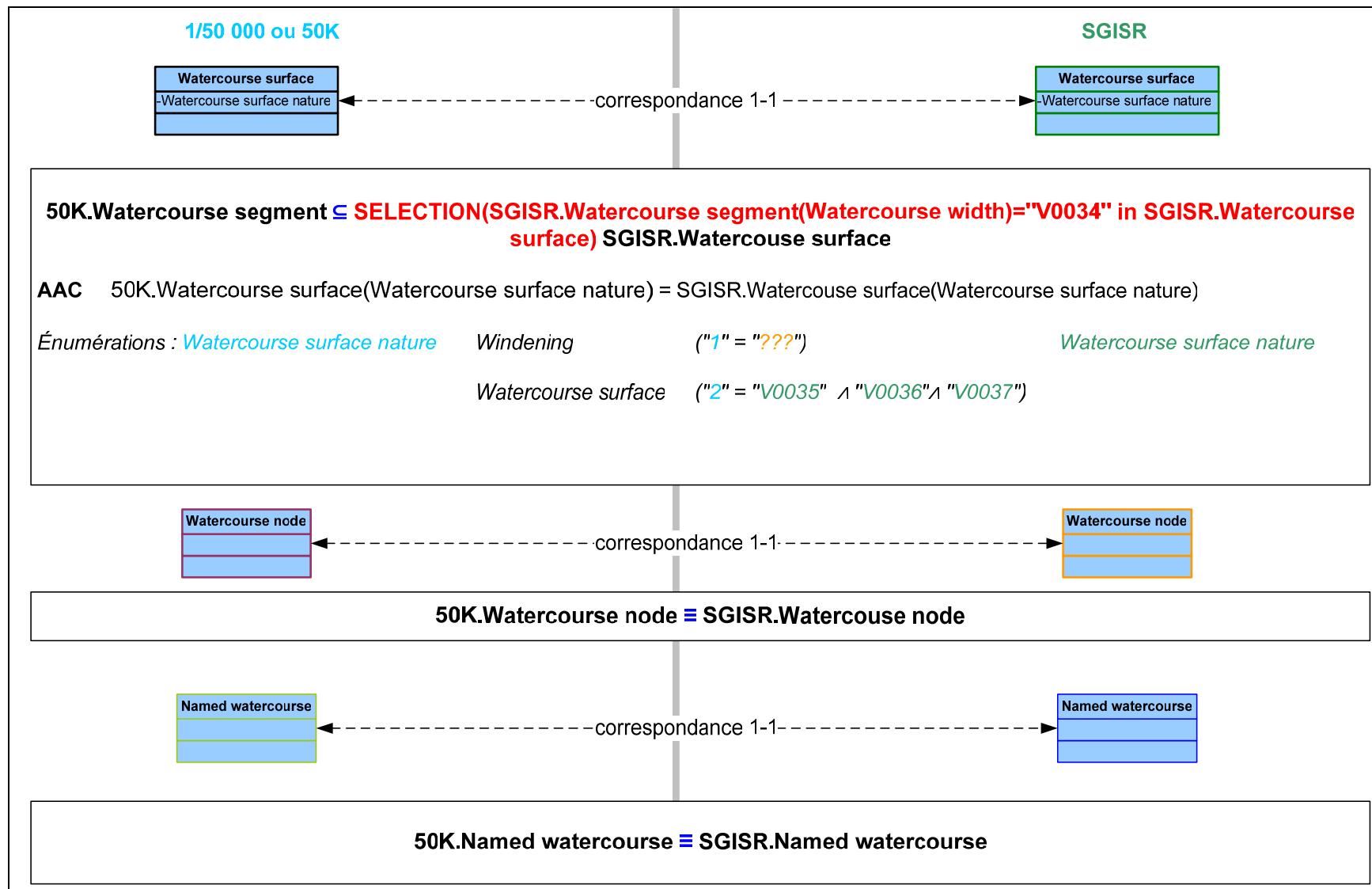


Figure 56 : ACI Hydrography 2/4.

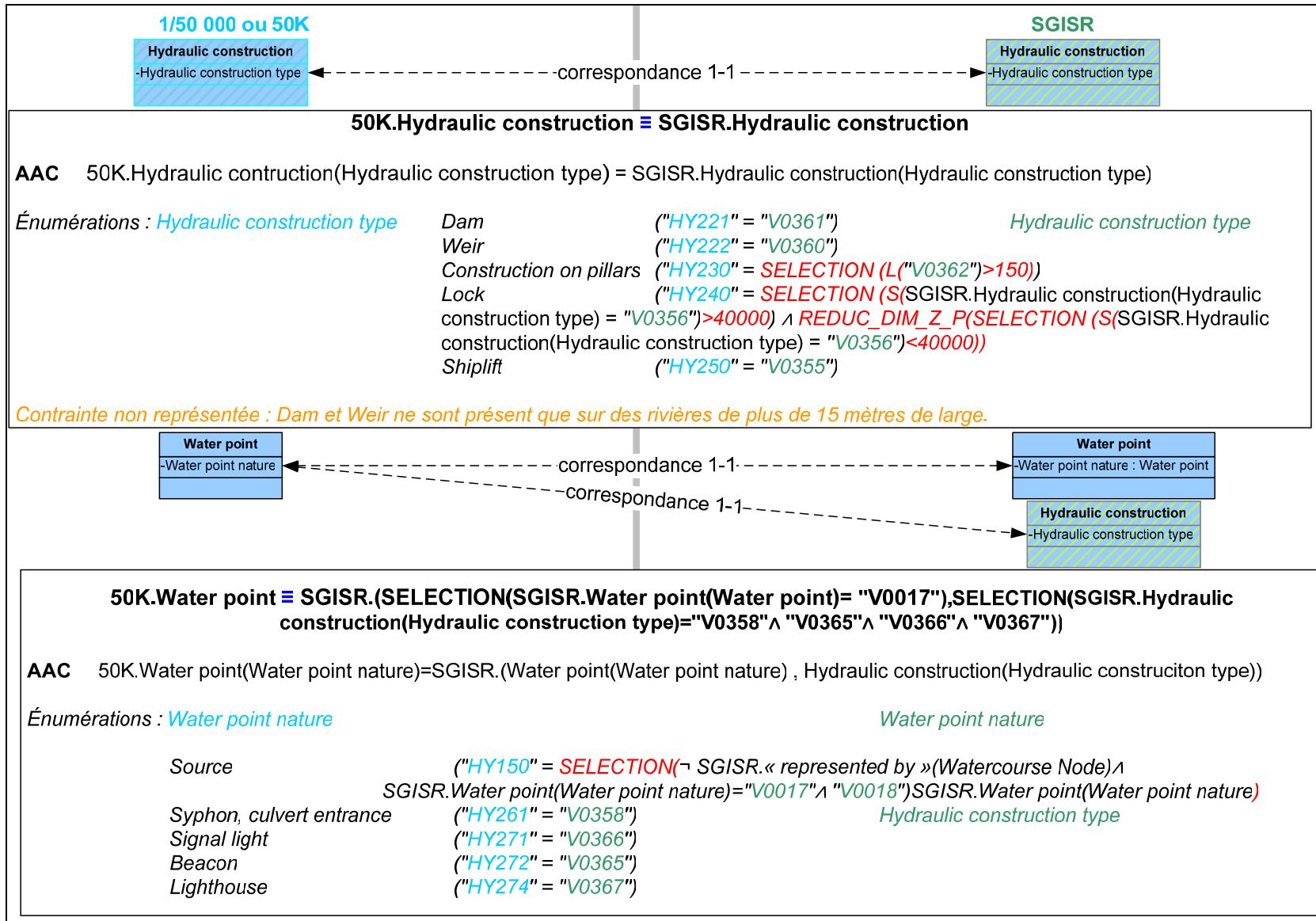


Figure 57 : ACI Hydrography 3/4.

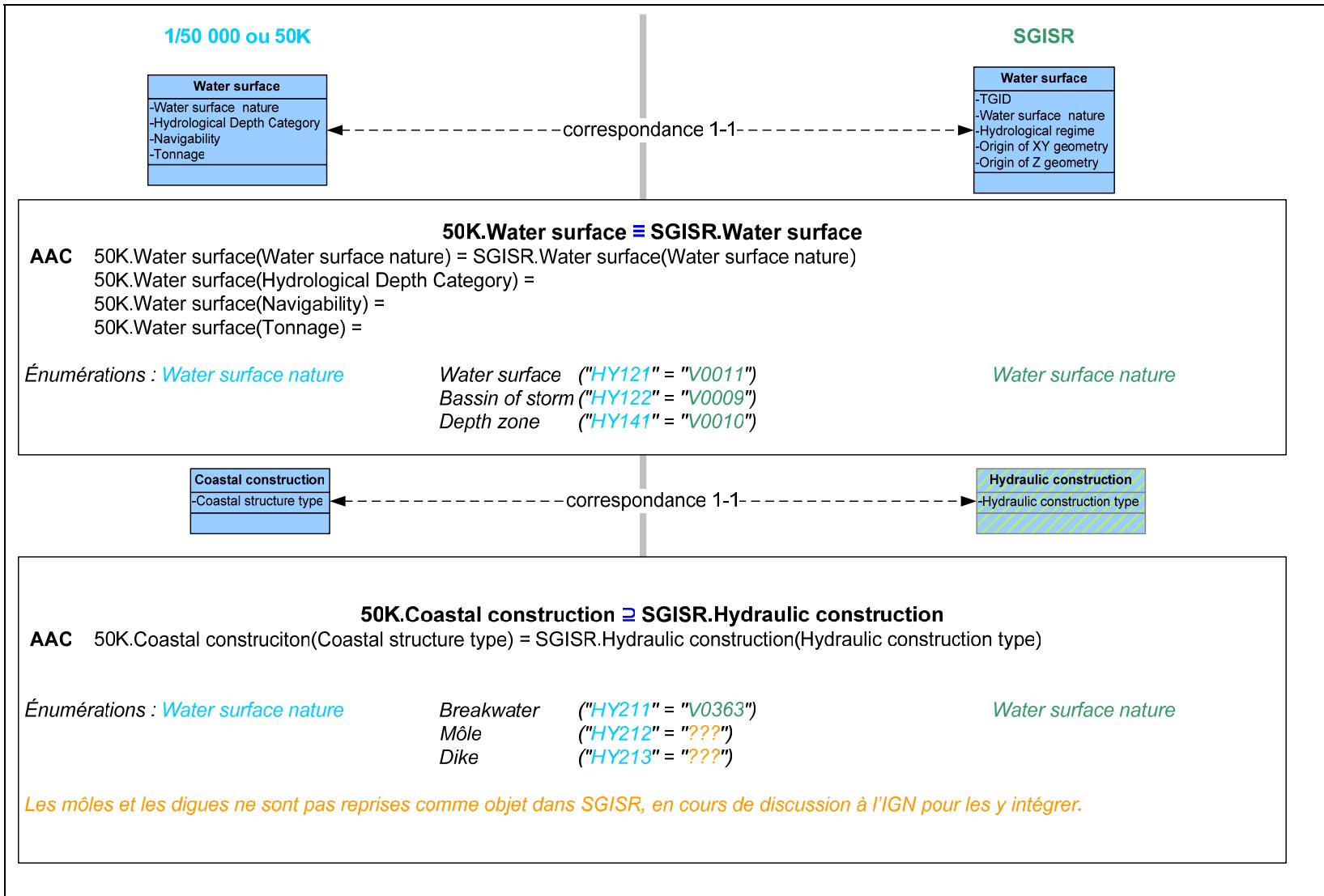


Figure 58: ACI Hydrography 4/4.

4.4.7 Network

Les segments particuliers des réseaux de chemin de fer, hydrographique et routier se généralisent dans les deux modèles en un élément général « Network segment » (Segment de réseau). Cette classe possède un attribut qui précise si le segment est bordé de talus et de quel type. L’attribut utilise la notion de direction de segment, en effet, l’attribut est énuméré comme étant : « remblais à gauche – déblais à droite ». Cette représentation pose plusieurs problèmes pour l’appariement. Premièrement, dans le modèle SGISR, la notion de direction des segments ne devrait pas être stockée de manière logique (dans le sens du kilométrage croissant, de l’écoulement de l’eau, ...). Deuxièmement, les talus ne sont pas des attributs de segment dans SGISR mais des objets, l’appariement au niveau des modèles est donc difficile. Cependant, le groupe de travail SGISR réfléchit à l’intégration de direction sur les segments de SGISR, cela les aiderait notamment pour la position des adresses, ainsi que pour l’exploitation des algorithmes de déplacements par exemple. Dans ce cas, une clause « *Avec géométrie correspondante* » définie par [DEVOGELE T. 1997, p. 84] pourrait être utilisée pour représenter la manière de retrouver les données. Il s’agirait de créer une zone tampon autour des segments, de regarder si cette zone intersecte un talus et si oui de calculer le sens de pente du talus afin de définir s’il s’agit d’un remblais, d’un déblais ou d’une levée de terre (fig. 59) ou d’utiliser la position de la ligne de crête associée à celui-ci. Comme on peut le constater cette procédure serait assez compliquée à mettre en œuvre, la clause n’est alors pas énoncée et remplacée par le processus d’appariement sur les instances des deux bases (fig. 60).

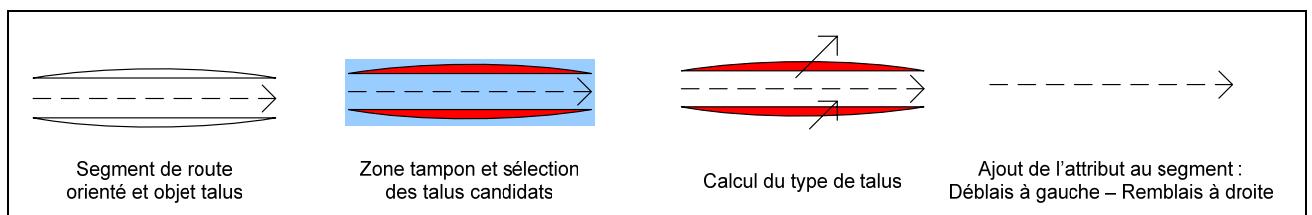


Figure 59 : Exemple de processus permettant le calcul de l’attribut “Bank”.

Les ponts sont également traités de manières différentes d’un modèle à l’autre (fig. 61). En effet, dans le cas de SGISR, ils sont représentés par leurs surfaces au sol, tandis que dans le 1/50 000^e, ils correspondent à une classe textuelle qui précise le type d’accès possible et le niveau. Il est néanmoins possible de retrouver le niveau du pont en recherchant quels segments (1/50 000) sont inclus dans la surface du pont (SGISR). Une analyse est cependant toujours en cours pour enrichir la notion de pont de SGISR car actuellement lorsqu’un segment de chemin pour piétons passe sur un ancien pont carrossable, l’information n’est pas conservée alors qu’elle est très utile pour les militaires.

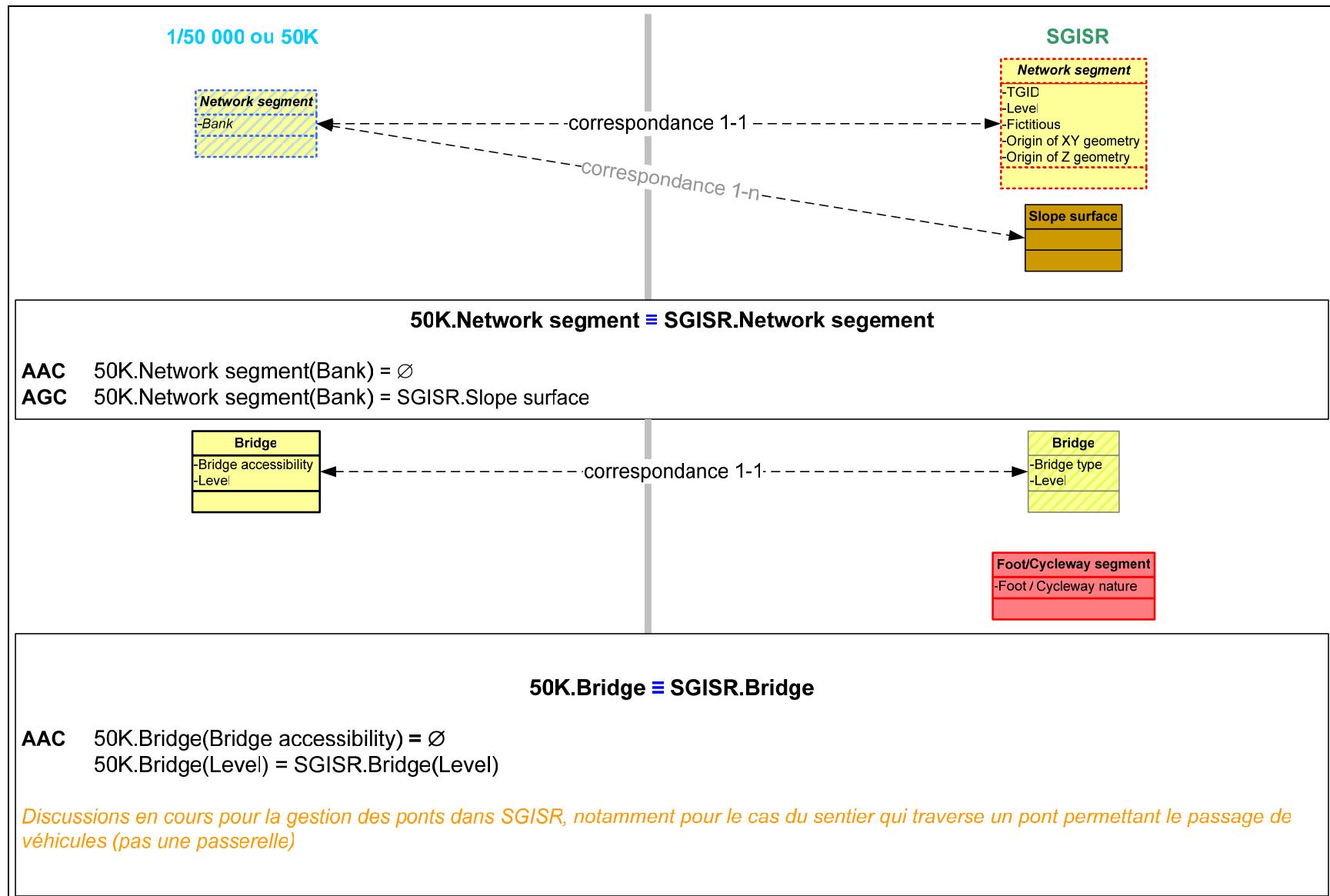


Figure 60 : ACI Network 1/2.

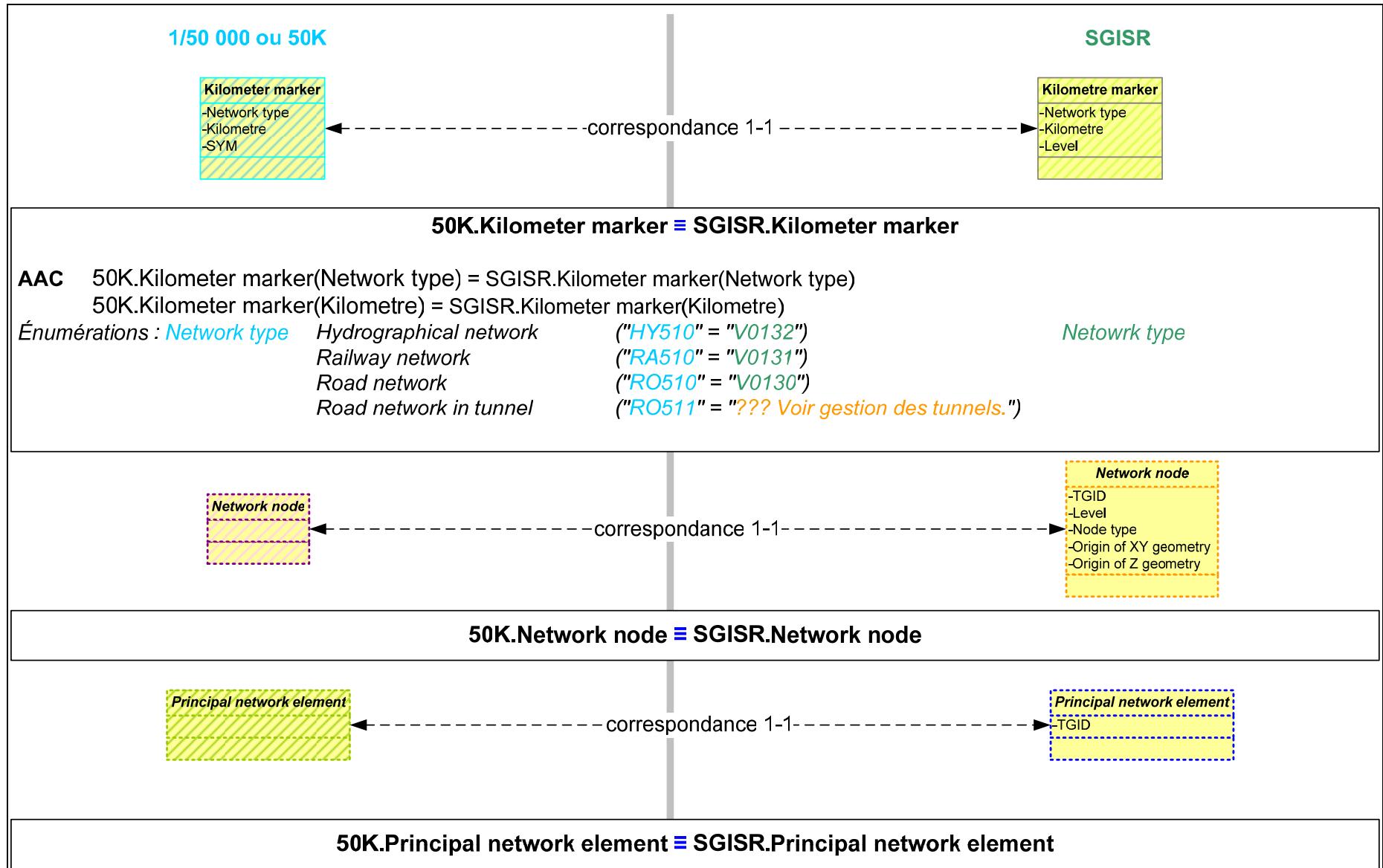


Figure 61 : ACI Netowrk 2/2.

4.4.8 Structure

Jusqu'à présent, nous n'avons pas eu à utiliser la fonction de changement de dimensions que nous avons explicitée précédemment. En effet, l'ensemble des données appariées jusqu'à maintenant possédaient des types géométriques identiques (points, lignes ou polygones). Dans la représentation au 1/50 000^e, le niveau d'abstraction de la réalité étant plus élevé, l'ensemble des structures ne peuvent pas toujours être représentés par des polygones. Une généralisation apparaît dès lors sous la forme d'un passage de polygones à des points. La fonction utilisée sera alors « *REDUC_DIM_Z_P* ». Elle n'intervient pas au niveau de la classe entière mais pour certaines modalités de l'attribut « type ». Lors de l'implémentation du modèle conceptuel en modèle logique, il sera nécessaire de scinder ces classes en fonction de leur type géométrique. Il n'est pas indispensable de l'effectuer actuellement, car les classes sont effectuées en fonction de critères sémantiques.

La classe « Building » (Bâtiment) comporte un certain nombre de constructions pour lesquels les critères de sélections géométriques ou de généralisation ne sont pas les mêmes. Premièrement, les « Train station » (Gare) représentées par leur surface au sol dans SGISR et par un point dans le 1/50 000^e ont une fonction de généralisation « *REDUC_DIM_Z_P* ». Celle-ci permet de représenter le changement qui s'opère avec le changement d'échelle et permet de bien se rendre compte que nous effectuons une généralisation conceptuelle et pas un appariement en vue d'une intégration des deux bases. Les « Industrial, commercial, sport building » (Bâtiment industriel, commercial ou sportif) ne sont pas conservés de la même manière qu'ils soient en « zone d'habitat ouvert » [IGN 2004, *Critères de sélection du 1 :50 000 2^{ème} édition*] ou en agglomération. On conçoit que ces termes sont difficilement explicitables pour la représentation formelle et, qui plus est, dans l'ACI. Nous choisissons donc le critère de sélection le plus précis (bâtiments dont la superficie est supérieure à 600 mètres carrés). Ceux qui ne seront pas nécessaire dans la représentation ne seront pas appariés au niveau des instances. L'ACI de la classe « Building » est présentée à la figure 63.

La classe « Particular construction » (Construction particulière) possède un certain nombre d'objets qui subissent des réductions de dimensions (fig. 64). Il s'agit en particulier des murs qui sont représentés par des polygones dans SGISR et par des polylignes dans le 1/50 000^e. S'y ajoute également un critère de sélection géométrique qui ne conserve que les murs de plus de 300 mètres de long.

Pour ce qui est de l'objet « Water purification plant » (Bassin d'épuration), il est représenté d'une manière différente en fonction de sa géométrie. Il passe d'un objet zonal à soit un objet ponctuel s'il est inférieur à 1500 mètres carrés, soit il reste représenté par un objet zonal. On peut schématiser le passage comme suit à la figure 62.

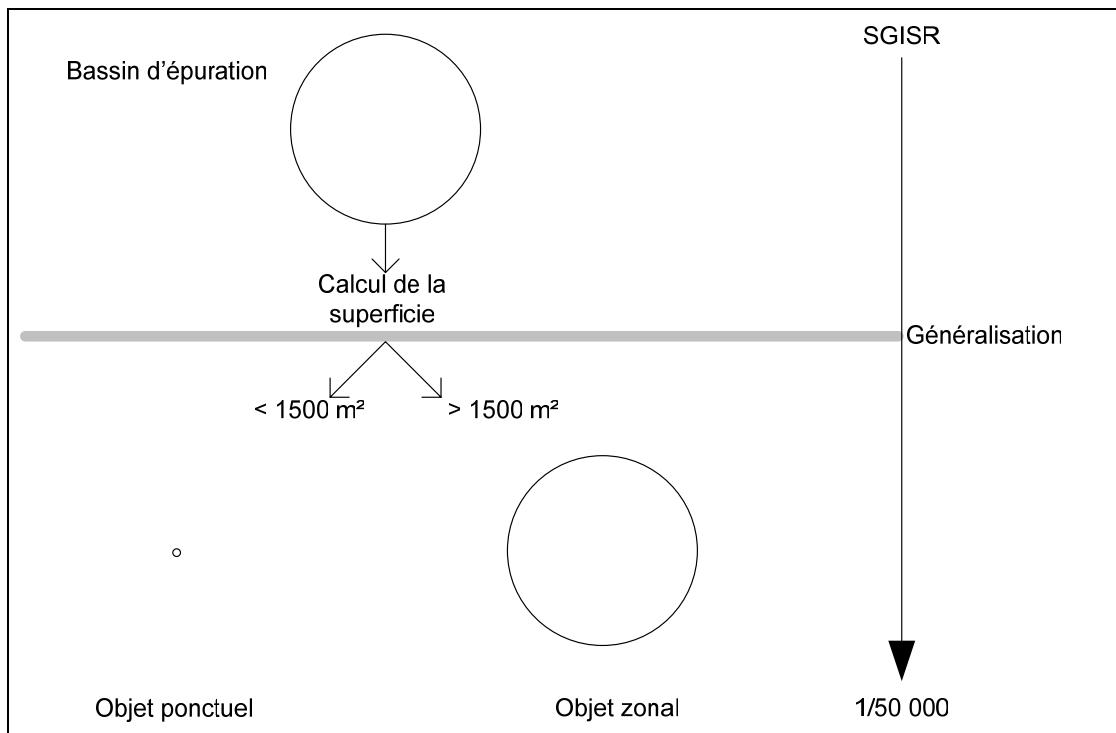


Figure 62 : Généralisation de l'objet "Water purification plant".

La classe « Zone » (Zone) est apparée de la même manière que la précédente. Certaines relations entre les deux modèles sont effectuées moyennant la généralisation de certains objets SGISR en objets ponctuels. Remarquons le lien [2:n] qui peut être fait entre les deux modèles au niveau des « Build districts » (Quartier bâti). Dans ce cas, au moins deux bâtiments se regroupent pour former une zone urbaine. L'objet 1/50 000 sera donc une composition de plusieurs objets de SGISR, ce qui est typique d'un conflit de fragmentation. Il est déclaré comme tel dans l'ACI (fig. 65).

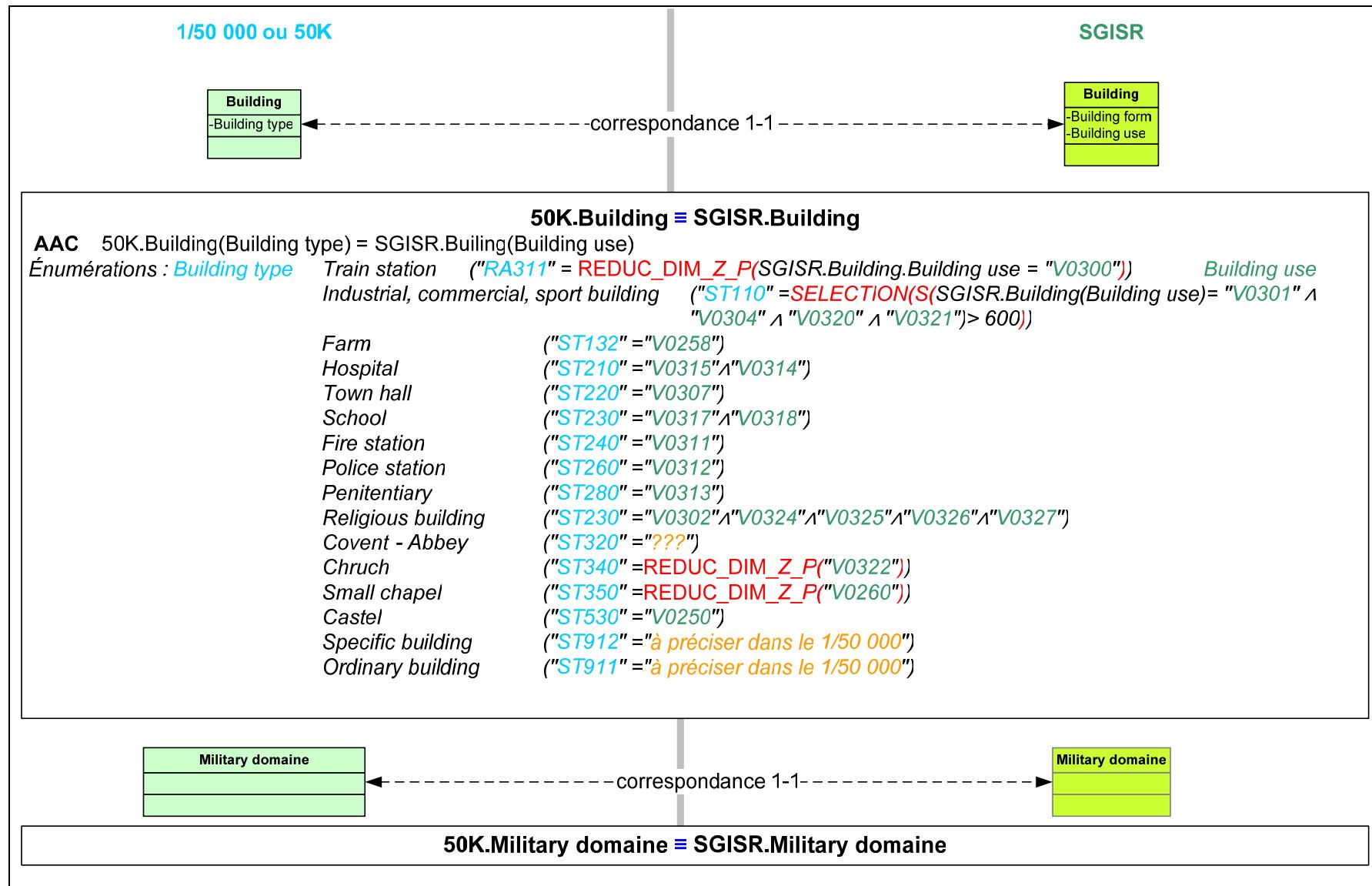


Figure 63 : ACI Structure 1/3.

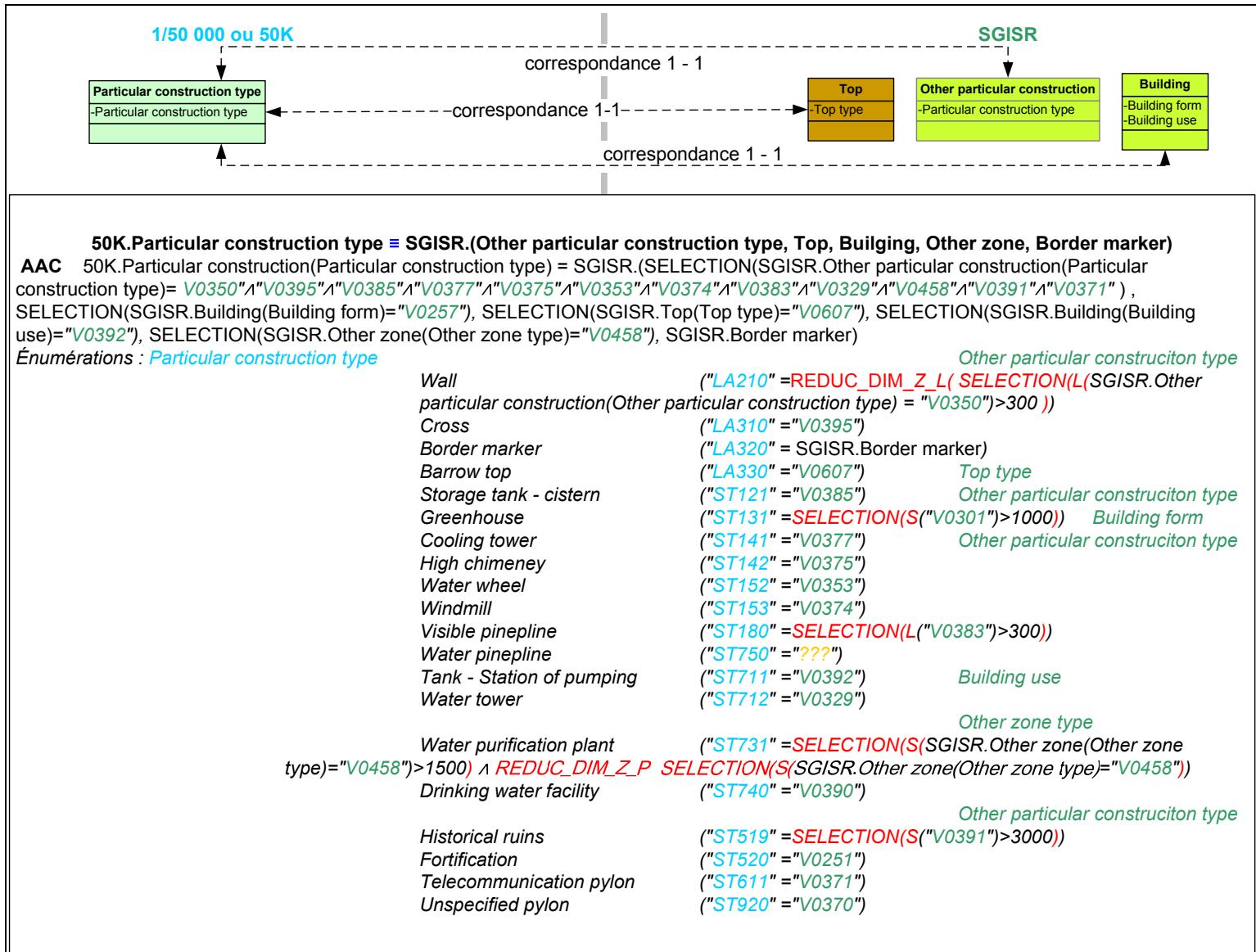


Figure 64 : ACI Structure 2/3.

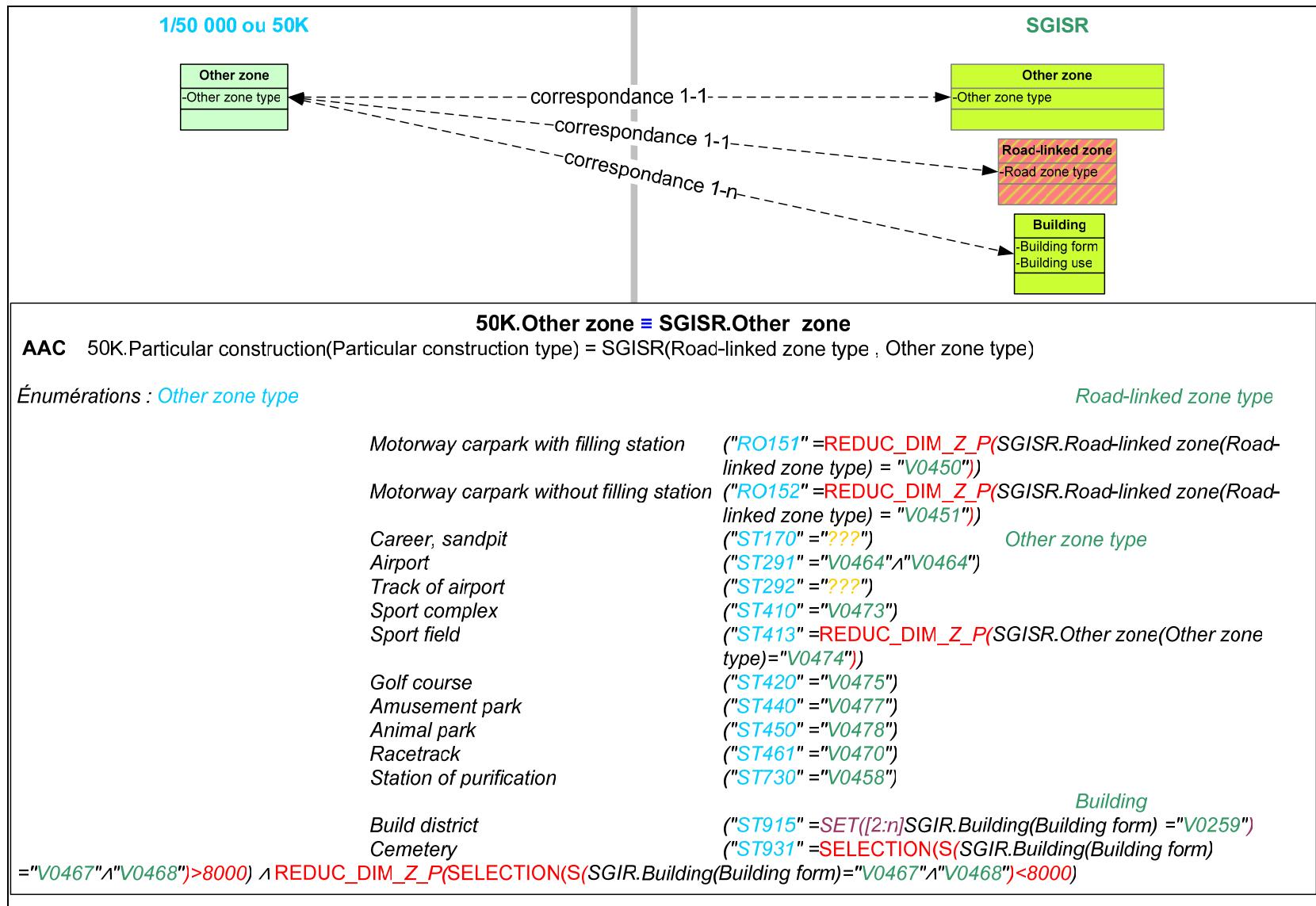


Figure 65 : ACI Structure 3/3.

4.4.9 Altimetry

Le dernier domaine à apparier à l'aide des ACI représente l'altimétrie, la bathymétrie ainsi que les détails caractéristiques du relief. Le premier élément représente les points géodésiques « Geodesic point ». Il est directement en correspondance avec son homonyme dans le modèle SGISR, cependant, les attributs de l'objet 1/50 000 ne peuvent pas être repris de SGISR (fig. 66). Les responsables du projet GenMap prévoient de faire un lien avec la base de données de la géodésie pour ces objets, il sera dès lors possible de retrouver le type et le code des points.

Les autres éléments sont en correspondance directe où un critère de sélection géométrique supplémentaire est nécessaire, comme dans le cas des zones de dunes.

Notons que les courbes de niveau n'ont pas de correspondances dans le modèle (fig. 67). Comme nous l'avons déjà expliqué lors de l'établissement du MCD du 1/50 000^e, les responsables du projet GenMap souhaitaient avoir l'ensemble des données nécessaires à l'établissement de la représentation cartographique. Ces données seront dès lors retrouvées à partir du modèle numérique de terrain de la Belgique (Projet DEM, cfr. 2.1).

4.5 Appariement des associations

Les ACI permettent également de déclarer les correspondances entre les associations des divers schémas (fig. 68). Les relations de généralisation – spécialisation ne sont pas mises en relation parce que l'appariement n'est pas nécessaire pour permettre leur représentation.

Les associations au niveau des réseaux sont appariées directement car elles sont identiques dans les deux modèles. Les deux associations conditionnelles également, même si le cheminement pour retrouver la proximité des gares au chemin de fer semble assez complexe. Il serait dès lors plus intéressant de la recalculer à partir des données avec une fonction de tampon.

Certaines associations n'ont pas de correspondances car les classes ne sont pas organisées de la même manière dans les deux modèles, il s'agit en particulier de la gestion des nœuds routiers et des ronds-points. Il sera nécessaire également de recalculer ces associations lors de l'appariement.

Dans la présentation de l'ACI, les associations ne sont pas nommées car elles ne le sont pas au niveau des schémas, elles sont représentées par les deux classes qu'elles mettent en relation.

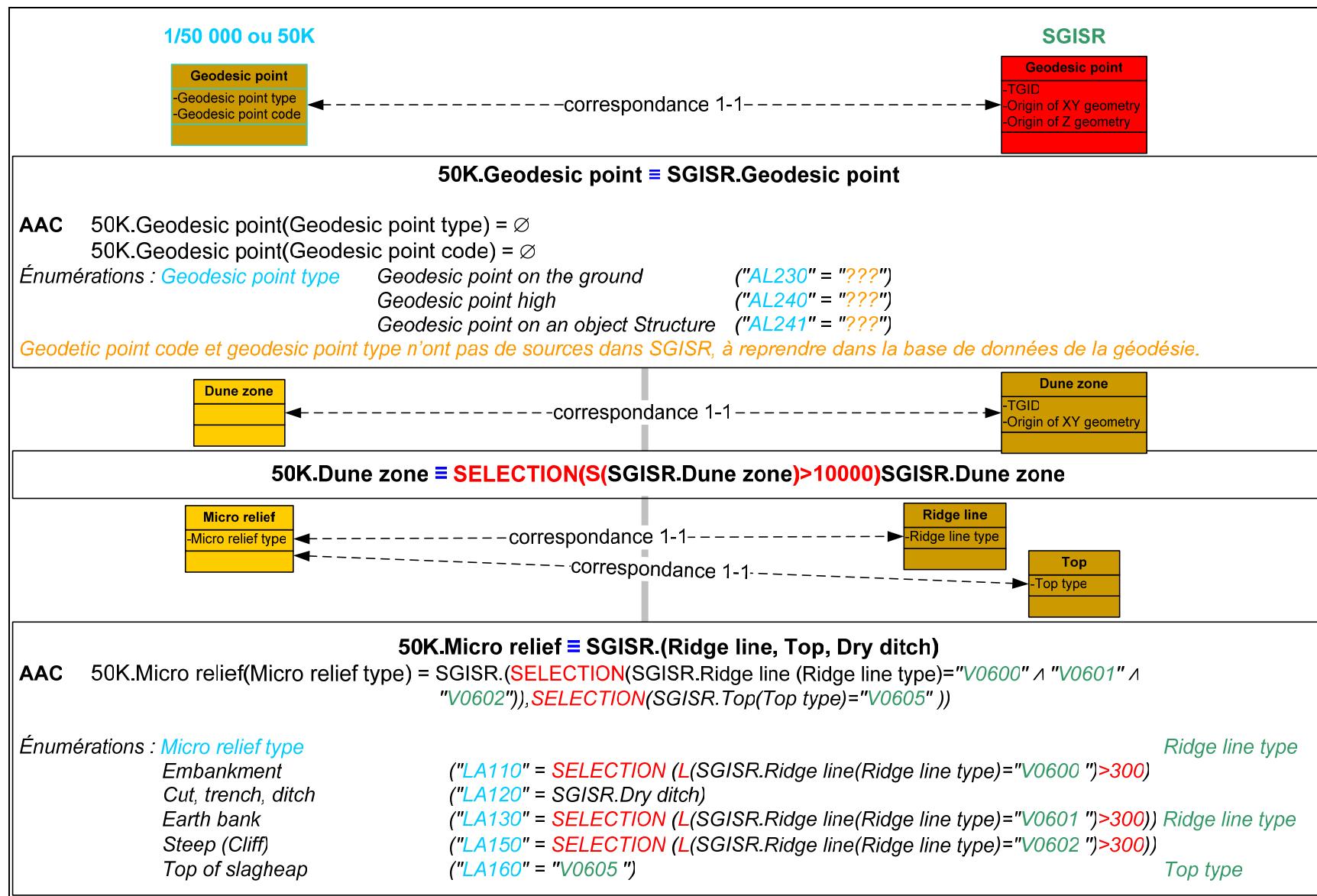


Figure 66 : ACI Altimetry 1/2.

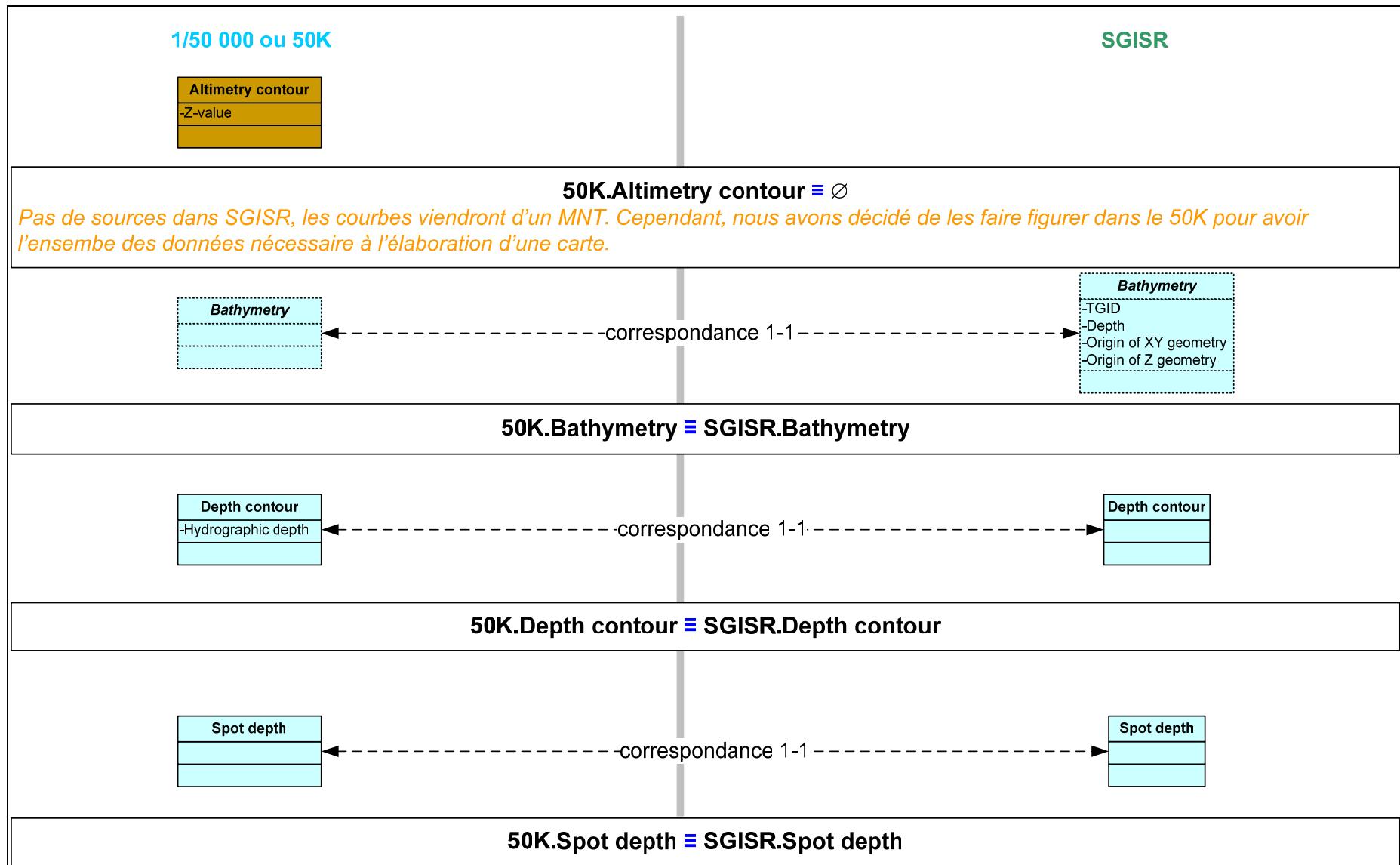


Figure 67 : ACI Altimetry 2/2.

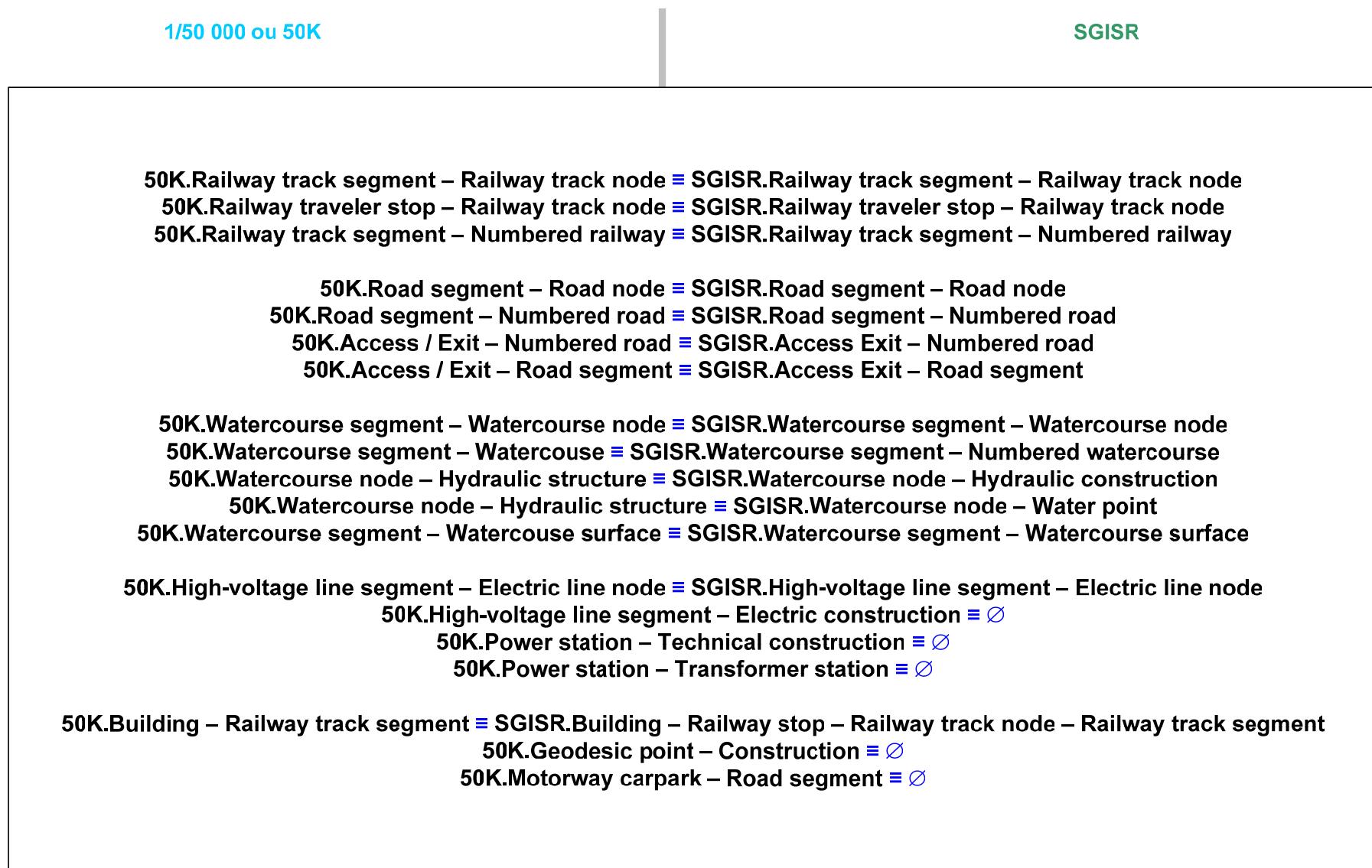


Figure 68 : ACI des associations.

4.6 Conclusions préliminaires sur l'appariement

Nous avons montré qu'il était possible grâce aux ACI, d'une part de formaliser toutes les correspondances entre les schémas, d'autre part, de surcharger celles-ci pour y inclure directement les critères de sélections sémantiques et géométriques qui s'opèrent lors du passage d'un schéma à l'autre. De plus, nous avons énoncé une fonction de simplification géométrique permettant de représenter le changement de dimension de certains éléments lors du passage d'une BDG à une autre.

Comme on peut le constater à la lecture des ACI, les correspondances de la plus grande partie des classes ont été retrouvées. Il en reste néanmoins certaines pour lesquelles il n'est pas possible de trouver directement de relation entre les deux modèles. Ces correspondances nécessitent une méthode d'appariement qui travaille au niveau des instances même des deux bases de données. Nous avons choisi de ne pas plus surcharger les ACI avec des méthodes d'appariement précises car d'une part cela compliquerait considérablement l'interprétation de celles-ci et d'autre part cela rendrait difficile la suite de leur utilisation par des SGBD (système de gestion de base de données). Quant aux surcharges que nous avons apportées aux ACI « traditionnelles », elles ne posent aucun problème (contraintes de sélections sémantiques, géométriques et fonction de réduction de dimension). Il est donc possible d'utiliser les ACI préalablement destinées à l'intégration de bases de données pour représenter certaines étapes de la généralisation conceptuelle.

Ces ACI pourront être utilisées de deux manières. Soit comme une représentation formelle de tous les éléments composant le 1/50 000^e, cette représentation permet de donner un cadre utile aux responsables du projet GenMap pour savoir exactement comment se comporte chaque objet lors du passage d'une BDG à l'autre. Cette représentation sera utilisée lors de l'appariement des instances de la base de données 1/50 000 avec la base de référence. A plus long terme, la formalisation des correspondances sera très utile pour développer des outils de généralisation automatique pour dériver une nouvelle représentation à l'échelle du 1/50 000^e de la base de données de référence. D'une part, comme certains SGBD (par exemple Géo₂¹) peuvent directement interpréter le langage, il serait alors possible, sans trop de difficultés, d'adapter leur interprétation aux nouvelles clauses et fonctions introduites dans le langage. D'autre part, il serait possible de développer un module d'interprétation pour un autre SGBD. L'interprétation du langage par les SGBD permettrait d'effectuer directement des étapes de l'appariement ou de la généralisation automatique.

¹ Géo₂ est un prototype de SGBD orienté objet basé sur le SGBD O₂. Il est notamment utilisé au laboratoire du COGIT.

5 Conclusions

L'avènement de l'informatique dans le monde de la cartographie a entraîné avec lui le stockage des données géographiques dans des bases de données. L'information géographique n'est pas unique, elle peut être envisagée à divers niveaux d'abstraction, dans plusieurs buts, ... bref à diverses échelles. Chaque échelle conceptuelle définit une base de données géographiques qui la représente au mieux.

L'institut géographique national de Belgique, bien conscient de cette réalité, a directement utilisé les outils informatiques à sa disposition. Depuis 2000, il prépare une restructuration interne de ses méthodes de production et de gestion de l'information qui se base sur deux constats. Premièrement, la méthode de gestion des données géographiques la plus efficace serait d'avoir une seule filière de production de données topo-géographiques de référence, une seule collecte d'informations de mise à jour et un modèle conceptuel de données permettant toutes les applications de base de l'Institut. Deuxièmement, il serait indispensable de gérer ces données topo-géographiques de référence de façon continue dans l'espace plutôt que par découpage cartographique. Ces deux observations ont conduit l'IGN à concevoir un nouveau système d'information géographique appelé « Seamless Geographic Information System of Reference » – SGISR.

Le mémoire qui vient d'être réalisé s'insère dans le premier constat qui impose de dériver toutes représentations à petite échelle des données de référence. Deux options s'offrent aux responsables du projet GenMap. La première serait d'effectuer une généralisation automatique complète sur les données de référence pour retrouver l'échelle du 1/50 000^e. Comme nous avons pu le montrer, dans l'état actuel des connaissances, cela n'est pas encore possible. Nous nous sommes alors tournés vers la seconde option qui est la propagation des mises à jour de la base de données de référence vers la base de données à l'échelle conceptuelle du 1/50 000^e. Nous avons plus particulièrement analysé la manière dont pouvaient dériver l'ensemble des données d'une représentation à l'autre, tous les objets de la carte étant susceptibles d'être mis à jour. Ce passage de deux bases de données géographiques à des niveaux d'abstraction de la réalité différents s'appelle la généralisation conceptuelle par modèle. L'objectif de ce mémoire a donc été premièrement de modéliser l'ensemble des données nécessaires à la représentation à l'échelle du 1/50 000^e. Ensuite, il a été nécessaire de vérifier s'il était possible de dériver ces données de celles de référence en appariant les deux modèles. Cet appariement servira dans un premier temps à propager les mises à jour et dans un avenir plus ou moins proche à effectuer une généralisation automatique. Enfin, il restait à énoncer les contraintes de généralisation sémantiques et géométriques nécessaires à la généralisation de chaque objet. Ces étapes furent réalisées avec succès. La démarche d'analyse préalable, de modélisation, et d'appariement correspond aux trois chapitres principaux de ce mémoire.

Nous avons premièrement décrit les données disponibles, qu'il s'agisse du catalogue et du modèle des données de SGISR ou de la liste exhaustive des données présentes à l'échelle du 1/50 000^e. Cette analyse préalable nous a montré l'ampleur de la tâche à réaliser. Cette étape indispensable d'apprentissage des deux échelles est nécessaire pour concevoir les niveaux d'abstractions différents qui existent entre les deux échelles. De plus, elle permet d'assurer la bonne réalisation de l'étape suivante.

Une fois cette analyse préliminaire effectuée, nous avons réalisé en collaboration avec les responsables du projet GenMap le modèle conceptuel des données à l'échelle du 1/50 000^e. La démarche suivie a été de découper l'ensemble des données en domaines reprenant des objets sémantiquement proches. Au total, huit domaines ont été constitués. Chacun d'entre eux a ensuite été modélisé en regard des classes qui lui correspondaient dans le TOP50V-Gis actuel. Les relations qu'entretiennent les classes de domaines différents ont été modélisées par la suite lors de l'intégration des domaines. Cette méthode de travail nous a permis d'effectuer une analyse plus profonde et de ne pas directement travailler sur l'ensemble du modèle où l'ensemble des relations est plus difficile à concevoir. Une méthode itérative a été choisie. Au total 3 versions du modèle ont été nécessaires pour rencontrer les desideratas et exigences de l'IGN.

Une fois le modèle des données généralisées créé, il a été nécessaire de savoir comment chaque objet d'une base pouvait être lié à l'objet ou aux objets correspondants dans l'autre base. Pour ce faire, nous avons utilisé le langage des « Assertions de Correspondance Interschémas » (ACI) développé par [SPACCAPIETRA S. et al. 1992]. Ce langage est utilisé pour déclarer les correspondances en intention entre deux schémas de données. Nous avons montré qu'il était possible d'utiliser et d'adapter ce langage complété par [DEVOGELE T. 1997] pour y inclure les contraintes de généralisation sémantiques et géométriques entre les deux bases. Les ACI de l'ensemble du modèle du 1/50 000^e ont donc été écrites et surchargées pour permettre l'énonciation des règles de sélections sémantiques et géométriques. Des fonctions supplémentaires de généralisation permettant la réduction de dimension d'objets ont été définies. Au final, il apparaît que le modèle de données SGISR contient l'ensemble des données nécessaires pour permettre la représentation à une échelle du 1/50 000^e. Cependant, il faut noter que des adaptations de celles-ci ont du être effectuées en collaboration avec les responsables du projet GenMap. Il s'agit soit de changements sémantiques mineurs (« terrains de football » qui deviennent « terrain de sport »), soit de suppressions d'éléments devenus obsolètes sur la carte (« bornes historiques »), soit de distinctions qui ne seront plus effectuées (« Citerne » et « Silo »). D'autres éléments n'ont pas pu être directement appariés lorsque les conflits entre les deux modèles étaient trop importants, comme par exemple pour les talus. Dans ce cas, il sera nécessaire d'effectuer l'appariement au niveau des instances mêmes des bases de données.

Au final, nous pensons avoir rempli les objectifs fixés pour ce mémoire, mais il est bien évident que le travail ne s'arrête pas là ! Maintenant que les deux jeux de données sont appariés au niveau des types (en intention), il va être nécessaire de mettre en œuvre divers algorithmes des plus simples au plus complexes pour apparier les instances des deux bases de données une fois qu'elles seront instanciées. Il sera également nécessaire de gérer tous les conflits liés soit aux différents modes de production des données, soit aux diverses dates de saisies. Un autre défi majeur à relever va être la propagation des mises à jour. En effet, comme nous l'avons précisé précédemment, une seule filière de mise à jour sera désormais possible, celle de SGISR. Il va donc être nécessaire d'étudier la manière dont vont se propager les mises à jour des données de références par le canevas que nous venons de créer. Ce défi, actuellement commun à l'ensemble des agences cartographiques, permettra de faire entrer la cartographie numérique dans une nouvelle ère, où, seules des données topogéographiques de référence seront acquises et où l'ensemble des représentations dériveront uniquement de celles-ci.

Personnellement, nous pensons que l'IGN est conscient des défis qu'il doit relever pour les années à venir. Le travail que nous avons réalisé au sein de l'équipe de généralisation nous a permis d'intégrer de manière plus concrète ce qu'est le monde du travail et de la recherche. De nombreux contacts et réunions ont été nécessaires pour avoir une vision claire et précise des attentes de l'Institut. Nous n'avons pas la prétention de leur apporter la solution idéale, mais nous pensons que la solution que nous proposons pourrait être exploitée ou réutilisée dans un avenir proche par les responsables du projet GenMap.

6 Bibliographie

Ouvrages de référence

BUTTENFIELD B. & MCMASTER R. (éditeurs) 1991. Map Generalization. Making Rules for Knowledge Representation, Essex : Longman Group, 245 p.

MULLER J.-C., LAGRANGE J.-P. & WEIBEL R. (éditeurs) 1995. Gis and Generalization. Methodology and Practice, London : Taylor & Francis, 257 p.

NAIBURG E. & MAKSIMCHUCK R. 2002. Bases de données avec UML, Collection Référence, Paris : CampusPress, 293 p.

PANTAZIS D. & DONNAY J.-P. 1996. La conception de SIG méthode et formalisme, Collection géomatique, Paris : Hermès, 343 p.

ROQUES P. 2004. UML 2 par la pratique. Études de cas et exercices corrigés. France : Eyrolles, 309 p.

RUAS A. (sous la direction de) 2002. Généralisation et représentation multiple, collection Germes Sciences, Paris : Lavoisier, 391 p.

Articles

BATINI C., LENZERINI M. & NAVATHE S. B., 1996. A Comparative Analysis of Methodologies for Database Schema Integration, *ACM Computing Surveys*, **18**(4), pp. 323-363.

BEAT P. & WEIBEL R. 1999. Using Vector and Raster-Based Techniques in Categorical Map Generalization, *ICA Workshop on Progress in Automated Map Generalization*, Canada: Ottawa, 14 p.

BRASSEL K. & WEIBEL R. 1988. A Review and Conceptual Framework of Automated Map Generalization, *International Journal of Geographic Information Systems*, **2**(3), pp. 229-244.

CHYBICKA I., IWANIAK A. & OSTROWSKI W. 2004. Generalization of the Topographic Database to the Vector Map Level 2 – the components of the Polish National Geographic Information System, *ICA Workshop on Generalisation and Multiple representation*, Leicester, 9 p.

DUCHENE C. 2003, Coordination multi-agents pour la généralisation automatique, *Bulletin d'information de l'IGN*, **74**(3), 10 p.

FECHIR A. & DE WAELE J. 2005. The future production of generalised maps at IGN Belgium, *ICA Workshop on “Generalisation and Multiple representation”*, Espagne : A Coruña, 8 p.

NEUFFER D., BELL M. & WOODSFORD P. 2004. Cartographic Generalisation on the Basis of Model Generalisation, *Symposium Praktische Kartographie Königslutter*, 24 p.

PARENT C. & SPACCAPIETRA S. 1996. Intégration de bases de données : Panorama des problèmes et des approches, *Ingénierie des systèmes d'information*, 4(3), 18 p.

REVELL P. 2004. Building on Past Achievements: Generalising OS MasterMap® Rural Buildings to 1 : 50 000, *ICA Workshop on “Generalisation and Multiple representation”*, Leicester, 14 p.

SPACCAPIETRA S., PARENT C. & DUPONT Y. 1992. Model Independent Assertions for Integration of Heterogeneous Schemas, *The International Journal on Very Large Data Bases*, 1, pp. 81-126.

WARE J. & JONES C. 1998. Conflict Reduction in Map Generalization Using Iterative Improvement, *GeoInformatica*, 2(4), pp. 383-407.

Références non publiées

BADARD T. 2000. Propagation des mises à jour dans une base de données géographiques, thèse de doctorat en Sciences de l'Information Géographique, Université de Marne-la-Vallée, inédit, 114 p.

DEVOGELE T. 1997. Processus d'intégration et d'appariement de Bases de Données Géographiques. Application à une base de données routières multi-échelles, thèse de doctorat, Université de Versailles, Méthodes Informatiques, indédit, 207 p.

IGN 2001. *La charte du projet TOC*, rapport interne, 5 p.

IGN 2002. *Schéma directeur de projet SGISR – version 1.1*, rapport interne, 19 p.

IGN 2002. *TOC définitions*, rapport interne, 106 p.

IGN 2004. *Critères de sélection du 1 :50 000 2^{ème} édition*, document interne.

IGN 2004. *Le projet GenMap*, document interne, 23 p.

IGN 2004. *SGISR Feature catalogue Main Version*, document interne.

IGN 2005. *Structure et codage des données de la 2^{ème} édition 1 : 50 000*, document interne, 20 p.

KMS (Kort & Matrikelstyrelsen), 2004. An Update on Activities at KMS, document interne, 16p.

ORDNANCE SURVEY 2002. *OS MasterMap user guide v2.1*, document interne, 231 p.

PRUVOST J.-C. 2004. Élaboration d'un modèle conceptuel de données à partir d'un catalogue d'objets normalisé et conversion en un modèle logique compatible avec le modèle ArcGis, rapport de stage, Ecole Nationale des Sciences Géographiques, inédit, 58 p.

RUAS A. 2004. Le changement de niveau de détail dans la représentation de l'information géographique, habilitation à diriger des recherches, Université de Marne-la-Vallée. Spécialité Sciences de l'Information Géographique, indédit.

SHEEREN D. 2005. Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales, thèse de doctorat en Informatique – Intelligence Artificielle, Université de Paris 6, 292 p.

Sites Internet

IGN 2004. *Structure et codage des données TOP10V-GIS et TOP50V-GIS*, site de l'IGN (http://www.ign.be/Common/articles/CA_Top10V-GIS_TOP50V-GIS/restruct_doc.htm), consultation le 22 novembre 2004, 63 p.

IGN 2005. *Considérations conceptuelles sur la nouvelle carte de base 1 : 10 000*, site de l'IGN (http://www.ngi.be/Common/articles/ADG_Bsm/considerationsFR.pdf), consultation le 15 juillet 2005, 17 p.

IGN 2005. *Le système d'information géographique et la carte au 1 : 10 000*, site de l'IGN (http://www.ign.be/Common/articles/CA_Bsm/articleFR.pdf), consultation le 15 juillet 2005, 25 p.

IGN France 2001. Descriptif Technique de la BD CARTO®, site de l'IGN France (http://www.ign.fr/telechargement/MPro/produit/BD_CARTO/DT_BDCARTO.pdf), consultation le 15 juillet 2005, 21 p.

IGN France 2001. Descriptif Technique de la BD TOPO PAYS® version 1.2, site de l'IGN France (http://www.ign.fr/telechargement/MPro/produit/BD_TOPO/JT_Agglo/DT_BDTOPO_Pays_1_2.pdf), consultation le 15 juillet 2005, 21 p.

Liste des annexes

Annexe 1. Modèle conceptuel des données SGISR.....	104.
Annexe 2. Modèle conceptuel des données 1/50 000.....	hors mémoire.
Annexe 3. CD-ROM contenant les modèles conceptuels	hors mémoire.

Annexe 1

