BASES DE DONNÉES GÉOGRAPHIQUES ET CARTOGRAPHIQUES À DIFFÉRENTS NIVEAUX DE DÉTAIL

Par Anne RUAS – Sébastien MUSTIÈRE

Laboratoire COGIT Institut Géographique National

On parle de représentation multiple lorsque l'on représente de différentes façons une même entité du monde réel et que ces représentations sont en relation. Par exemple on peut parler de représentation multiple lorsqu'on représente une même entité 'hôpital' par 1/ un objet ponctuel de la classe 'Hôpital' et de nom 'Lapeyronie' et par 2/ un objet complexe de nom 'Lapeyronie' et composé d'un ensemble de bâtiments à l'intérieur d'une enceinte et que ces deux représentations sont mises en relation. La représentation multiple facilite la gestion conjointe de différentes bases de données géographiques ayant différents niveaux de détail. La notion de niveau de détail est ici à rapprocher de la notion d'échelle en cartographie.

En cartographie, le mot 'échelle' est porteur de deux sens. Au sens mathématique, il s'agit du rapport entre la taille «terrain» et la taille de représentation. Généralement, les objets cartographiques ont une position conforme à l'échelle de représentation et une taille conforme aux contraintes d'interprétation des signes. Par ailleurs le mot 'échelle' correspond aussi à l'échelle de raisonnement. Lors d'un changement d'échelle en cartographie, on opère à la fois un changement mathématique et un changement de contenu pour permettre un autre raisonnement sur l'espace géographique représenté. Ceci se reflète sur la légende, en particulier sur la définition des données transcrites sur la carte.

Qu'en est-il au niveau des Bases de données ? Comment décrire le niveau de détail d'une base de données ? Dans ce contexte, à quoi correspond la généralisation ? A quoi sert le concept de Représentation Multiple ?

Ce papier est extrait de (Ruas 2004) et est enrichi dans sa partie 4 par les deux auteurs.

1 Le niveau de détail d'une base de données

Pour constituer une base de données il faut répondre aux deux questions : Quoi ? (Que représenter ?) et Comment ? (Comment représenter numériquement ce Quoi ?). Ce qui est retenu (le quoi) est organisé en un schéma conceptuel composé – suivant l'approche Objet - de classes, d'attributs et de relations. Ces classes décrivent les types d'information qu'il peut y avoir dans la base et les attributs décrivent l'information associée aux objets des classes. On associe aux attributs et aux relations des informations permettant de mieux les définir (cardinalité, type, domaine de valeurs possibles). On ajoute dans le schéma des contraintes entre valeurs d'attributs appelées contraintes d'intégrité (par exemple : si 'Nombre de voies' = 1 alors 'Type' ∉ {Nationale, Autoroute}). A ce niveau, on spécifie aussi la ou les géométries acceptées pour les objets. Par exemple on impose que la géométrie d'une route soit linéaire et on peut autoriser que la géométrie d'un bâtiment soit surfacique ou ponctuelle. On peut relier le type de géométrie à certaines valeurs d'attributs (par exemple si 'Type de Bâtiment' = 'Eglise' alors 'type de géométrie' = 'point', sinon 'type de géométrie' = 'surface').

Le schéma conceptuel traduit la catégorisation que l'on fait du monde réel. Son analogue sur une carte est le répertoire des informations de la légende.

Une fois que le schéma conceptuel et ses contraintes associées sont définis conformément aux spécifications de sélection et de représentation, on les traduit en schéma logique qui est la traduction de ce schéma sur le SIG (Système d'information géographique) ou le SGBD (Système de gestion de base de données) dont on dispose pour représenter physiquement les données. Selon l'outil utilisé, tous les éléments du schéma et toutes les contraintes ne pourront pas être traduites facilement.

Les niveaux de détail

On pourrait penser que les contraintes graphiques liées à la perception sont éliminées lors de la numérisation des données sous forme de bases de données. Il n'y a en ef fet aucun problème de 'place' pour représenter l'information. On a généralement tendance à dire qu'une base de données n'a pas d'échelle (en tant que rapport de réduction au sens mathématique), pourtant elle possède un niveau de détail sémantique et géométrique.

Le niveau de détail sémantique

Le niveau de détail sémantique est traduit par le schéma de données et les règles de sélection associées au schéma. Là encore, on retrouve les deux aspects r elatifs au contenu : le quoi (ce que l'on retient) et le comment (avec quel niveau de détail on décrit ce quoi). Le quoi permet de préciser par exemple si les trottoirs sont représentés dans une base, s'ils sont assimilés aux routes. Pour le comment, il faut définir le niveau de détail avec lequel on décrit les attributs des objets. Par exemple, pour la largeur d'une route, on peut la décrire au centimètre près, au mètre ou par intervalles. De même lorsqu'on décrit le nombre d'habitants d'une commune, on peut le décrire à l'habitant près, à la dizaine, centaine ou au millier. Cette précision de description est la résolution sémantique de la base.

Le niveau de détail sémantique permet de dire : ce que l'on retient du monde réel et avec quelle résolution sémantique on décrit ce que l'on retient.

Niveau de détail géométrique

Pour décrire le niveau de détail géométrique d'une base de données, on parle souvent de résolution. La résolution correspond en principe à la plus petite unité mesurable ou représentable.

Pour le mode maillé, la résolution correspond à la taille du pixel (figure 1). On parle de la résolution d'une image, d'un scanneur, d'une imprimante, d'un écran. Ce qui est plus petit que cette taille ne peut pas être correctement décrit. Le pixel prend une valeur unique qu'elle 'étend' à la surface qu'il couvre même si cet espace contient plusieurs informations. Cette valeur est soit la moyenne des valeurs radiométriques du pixel (cas des images), soit un choix d'une valeur dominante. Dans ce dernier cas des objets distants vont devenir connexes et certaines superpositions imposent le choix d'une valeur dominante. Dans le mode maillé, la résolution limite la précision géométrique et on choisit en principe une résolution qui s'adapte au niveau de détail recherché pour décrire la géométrie des objets et leurs relations spatiales.

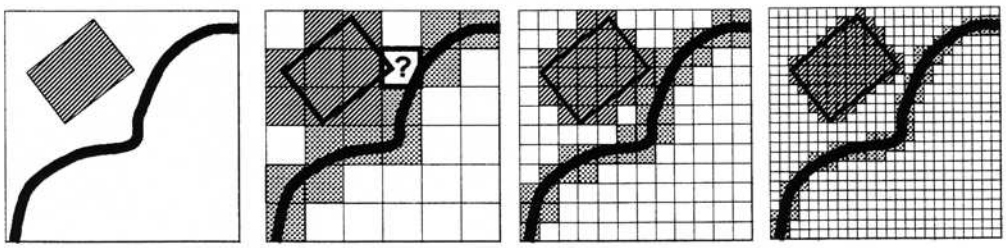

Figure 1 : *La résolution des données maillées : représentation d'une route et d'un bâtiment.*

Pour des données vectorielles, par analogie avec le mode maillé, le terme résolution pourrait sous-entendre *la distance minimale entre deux points consécutifs d'une ligne,* puisqu'on passe de la maille au point. Le terme de résolution, pris dans ce sens, ne nous renseigne que peu sur le niveau de détail géométrique de la base. Prenons l'exemple d'une ligne décrite par une succession de points. La figure 2 illustre une même courbe décrite par des résolutions dif férentes. Même si, dans les deux cas, les points sont toujours situés sur la courbe, on voit que les écarts entre la polyligne et la courbe sont beaucoup plus importants sur la figure de droite que sur celle de gauche. C'est en ef fet avant tout l'écart entre les lignes qui compte et non la distance entre les points. Il est donc essentiel de connaître la fonction d'interpolation utilisée pour savoir si le pas est ou non approprié, compte tenu de l'écart planimétrique attendu.

Figure 2 : *Une courbe décrite par une succession régulière de points*

La distance idéale entre les points qui décrivent une courbe (dans le cas d'une interpolation linéaire) n'est pas régulière mais dépend de la sinuosité de la courbe à décrire et de l'écart planimétrique attendu. Plus la courbure est importante, plus il faut de points pour bien retracer la forme de la ligne. Contrairement au mode maillé - qui décrit une partition de l'espace - il n'y a pas de lien clairement défini entre l'interdistance minimale entre les points, la précision de position et la capacité à retracer les détails d'une forme. Le terme résolution est donc ambigu. L 'expression 'la résolution de cette base est de 1 mètre' peut être compris de diverses façons : 1/ les points représentés sont au minimum distants de 1 mètre; ou 2/ l'écart entre la polyligne tracée par interpolation entre les points et la géométrie de l'entité réelle est inférieure à 1 mètre; ou encore 3/ les détails géométriques ayant une taille inférieure à 1 mètre de côté ne sont pas représentés. Pour nous, la résolution géométrique d'une base devrait être comprise comme un ordre de grandeur qui renvoie aux ordres de grandeurs décrits par Y. Lacoste (Lacoste 1980 p.23). Pour décrire le niveau de détail géométrique d'une base de données, il faudrait ajouter à la résolution - vue comme un ordre de grandeur - les concepts de précision géométrique et de granularité. La précision géométrique qualifie l'écart entre la position réelle et la position représentée. Cet écart varie dans l'espace, et on en donne souvent une valeur moyenne. La granularité décrit la taille des plus petites formes géométriques décrites dans la base. Il s'agira pour une ligne de la taille minimale d'un virage ou pour un bâtiment de la taille minimale d'un décrochement.

Les niveaux de détail

Une base de données géographique est décrite par une composante sémantique et une composante géométrique. Au niveau sémantique, le niveau de détail de la base est défini par son schéma conceptuel qui inclut les règles de sélection (les conditions sémantiques et géométriques que doivent remplir les objets pour être représentés dans la base)

et une résolution sémantique sur les attributs. Au niveau géométrique, le niveau de détail est défini par la résolution qui donne l'ordre de grandeur géométrique des phénomènes présents dans la base, la précision géométrique qui décrit l'écart entre la position terrain et la position représentée, et la granularité qui décrit la taille des plus petites formes représentées. Dans une base de données, il y a en principe une cohérence entre les niveaux de détails sémantique et géométrique.

2 Les bases de données géographiques et cartographiques

Définitions

D'un côté, l'un des objectifs majeurs des géomètres-cartographes est de définir la position 'exacte' des objets géographiques dans un système de référence, d'un autre côté, pour permettre aux humains de se constituer une image mentale d'une réalité géographique, il faut représenter cette information sur une carte (écran ou papier), à une échelle donnée. Les contraintes de lisibilité et d'interprétation imposent alors une modification très significative de la taille et de la position relative des objets.

Lorsqu'on parle de base de données géographiques, à quoi fait-on référence ?

Une solution simple est de distinguer les bases de données géographiques des bases de données cartographiques selon la définition suivante :

- une base de données géographiques regroupe des données géographiques 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 décrits dans la base. Ces concepts sont décrits par le schéma de 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 est conforme 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. Ces concepts sont décrits par le schéma de données et par les spécifications.

Cette terminologie est employée par (Brassel et W eibel 1988) qui utilisent le terme de *modèle de paysage numérique* (Digital Landscape Model - DLM) pour décrire les bases de données géographiques, et le terme de *modèle cartographique numérique* (Digital Cartographic Model - DCM) pour décrire les bases de données cartographiques. Cette distinction DLM / DCM est principalement utilisée en Allemagne. Cette définition a le désavantage de dissocier la géographie de la cartographie alors que la cartographie est partie intégrante de la géographie. Cela sous-entendrait aussi qu'une base de données cartographiques n'est pas une base de données géographiques. De plus, dans le langage courant les deux types de bases sont souvent appelés bases de données géographiques sans distinction.

Une autre solution serait de dire qu' une base de données géographiques possède un espace de signification qui est constitué des concepts géographiques décrits dans la base. Ces concepts sont décrits par le schéma de données et par les spécifications.

Il y a deux types de bases de données géographiques :

- les bases de données géographiques *conformes* (BDG) qui regroupent des données géographiques dont la position de chaque objet est conforme à la précision géométrique et à la granularité de saisie,

- les bases de données cartographiques (BDC) qui regroupent 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.

Niveau de détail et Gamme d'échelles

Lorsque les bases de données géographiques ont été pensées puis constituées dans les années 1980, l'aspect cartographique semblait ne pas poser de problèmes majeurs. Il suffirait d'associer à chaque objet une symbolisation en fonction de sa nature. Force est de constater que cela n'a pas été facile. En reprenant l'exemple de l'IGN, la constitution de 'lignes de production' de cartes à partir de bases de données est un processus long qui passe généralement par un projet qui analyse puis développe un ensemble d'outils informatiques visant à minimiser le temps passé en corrections interactives. La carte au 1: 25 000 réalisée à partir de la BDTopo (base de données métrique) est sémantiquement plus riche que la carte traditionnelle mais, d'une part, de nombreuses heures de corrections interactives sont nécessaires pour respecter les contraintes de lisibilité et, d'autre part, la lisibilité de certains objets - dont les bâtiments- est moins bonne que sur les cartes traditionnelles. De même, l'IGN étudie la réalisation des cartes au 1: 100 000 à partir de la BDCarto (base de données décamétrique) et ne sait toujours pas réaliser de cartes au 1: 50 000 à partir de la BD Topo. On est donc encore loin de la 'carte à la carte', celle qui répondrait à nos besoins. Ce que l'on pourrait faire facilement, c'est une carte au 1: 15 000 à partir de la BDTopo (parce que cela correspond à l'ordre de grandeur des données à cartographier) et un 1: 50 000 à partir de la BDCarto avec une légende ressemblant au 1: 100 000 (parce que la BDCarto a été saisie en grande partie à partir des cartes au 1 : 50 000 filtrées). Les processus de fabrication de cartes à partir de bases de données s'avèrent être tellement coûteux, que les instituts de cartographie (en France, en Espagne, en Allemagne) créent des 'bases de données cartographiques' dérivées des 'bases de données géographiques' pour minimiser les opérations de mise en position et de symbolisation aux seuls objets qui seront mis à jour. Ces bases de données cartographiques sont des bases de données toujours vectorielles mais dont les objets ont subi des dilatations, des simplifications et des déplacements pour se conformer aux contraintes graphiques liées à l'échelle de représentation finale.

Si tout le monde s'accorde pour dire qu'une base de données n'a pas d'échelle, on pourrait ajouter qu'elle est faite pour une gamme d'échelles, qu'elle peut être facilement cartographiée lorsque l'échelle de la carte est proche de l'échelle de saisie et que, sinon, des processus complexes de généralisation sont à mettre en œuvre pour pouvoir représenter correctement l'information dérivée de la base. On se rapproche des concepts d'échelle caractéristique et d'échelle limite liées à un phénomène. L'échelle caractéris-

tique est l'échelle la plus appropriée pour rendre sensible un phénomène en fonction du niveau de son étude (Comité Français de Cartographie 1970). L'échelle limite est l'échelle inférieure ou supérieure à l'échelle caractéristique en deçà ou au-delà de laquelle la représentation d'un phénomène perd sa signification (Comité Français de Cartographie 1970). Les informations géographiques comprises dans une base de données décrivent des phénomènes qui ont chacun une échelle caractéristique en fonction d'un besoin ; et collectivement la base de données peut être utilisée dans une gamme d'échelles en deçà et au-delà de laquelle la représentation perd sa signification.

Le processus de généralisation introduit en figure 3 a deux objectifs distincts :

o Permettre la fabrication d'une base de données cartographiques à partir d'une base de données géographiques conforme. Les deux bases géographiques ont le même espace de signification ou niveau de détail.

o Permettre la fabrication d'une base de données géographiques conforme ayant un certain niveau de détail à partir d'une autre base de données géographiques conforme dont le niveau de détail est supérieur.

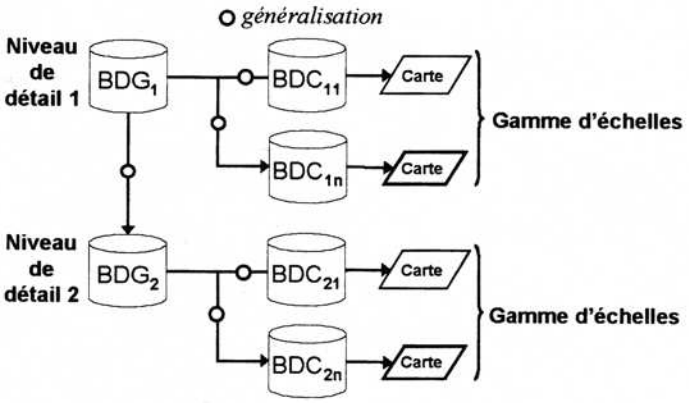

Figure 3 : *Bases de données géographiques conformes et cartographiques.*

En d'autres termes, une base de données correspond à une échelle de raisonnement ou d'analyse (une échelle est toujours 'médiatrice d'une intention' (Racine 1981 p.155)) et peut être représentée cartographiquement dans une gamme d'échelles mathématiques compatible avec l'ordre de grandeur de l'ensemble des objets correspondant à cette analyse. Au-delà de cette gamme d'échelles, on ne peut plus représenter les données sans introduire de nouveaux concepts géographiques qui définissent (parfois implicitement) l'existence d'une nouvelle base de données géographiques.

4 Les Généralisations

Dans le chapitre sur la généralisation cartographique (Bertin 1998(1967) p.300) de données thématiques, Jacques Bertin distingue la généralisation conceptuelle de la généralisation structurale:

- la généralisation conceptuelle est traduite par un changement d'implémentation et implique une nouvelle conceptualisation (« 'puits de mines' se transforme en 'bassin houiller'»),
- la généralisation structurale conserve la conceptualisation mais opère une 'simplification de la répartition'.

Par ailleurs, Bertin précise que la rédaction des fonds de cartes (qui est notre centre d'intérêt) implique à la fois les deux types de généralisation. Selon la définition de Bertin, *la généralisation conceptuelle traduit l'apparition et la disparition de concepts géographiques* et la généralisation structurale modifie les formes et les positions sans modifier l'existence de ces concepts ainsi que les signes utilisés pour les représenter.

De nombreux cartographes décomposent la généralisation en trois types d'opération (Beguin et Pumain 1994 p.17-18):

- **la sélection** élimine l'information non pertinente,
- **la schématisation** simplifie les formes (schématisation structurale) et change le mode d'implémentation (schématisation conceptuelle),
- **l'harmonisation** 'recherche un équilibre entre les besoins de simplification, le souci de maintenir une localisation précise des phénomènes représentés et les choix graphiques opérés'.

Dans les articles traitant de généralisations faites à partir de bases de données géographiques, de nouveaux termes apparaissent tels que la généralisation de modèle (*model generalisation / geographical information abstraction*), la généralisation de données ou la généralisation cartographique (*map, cartographic, graphic generalisation*). «*Geographical information abstraction mainly concerns managing geographical meaning in databases and map generalization mainly concerns structuring map representations* «(Nyerges 1991 p.62). Ces termes - *managing geographical meaning* et *structuring map representation* - ont donné lieu à des interprétations différentes selon les auteurs. Nous nous concentrons ici sur la généralisation de 'base de données' puis sur la généralisation liée à la symbolisation cartographique avant de faire le point sur ces notions.

La généralisation de modèle (ou généralisation de BD conforme)

Nyerges (Nyerges 1991) décompose l'opération d'abstraction en quatre opérations classiques : la réification [1] (nommé *classification* dans le texte initial), l'association, la généralisation (au sens base de données) et l'agrégation.

- La réification est le fait de définir les concepts présents dans la base sous forme de classe et d'être capable de distinguer le sens de chaque classe (*'how one distinguishes the meaning of one class from another* » (Nyerges 1991 p.77)). Par exemple, on définira les classes *Routes* et *Rivières* en définissant 1/ le sens des concepts de Route et de Rivière dans cette base et 2/ ce qui sera décrit sur les routes et sur les rivières.

- L'association est le fait de mettre en relation des concepts en décrivant le sens de ces relations. Par exemple on spécifiera la relation possible entre des entités ponts, routes et rivières.

- La généralisation est le fait de se définir un concept plus global qui englobe différents concepts sémantiquement proches. L'opération contraire est la spécialisation. Lorsqu'on dispose de différentes classes en relation de généralisation/ spécialisation on parle de hiérarchie entre les classes. Une classe générale décrit globalement dif férents concepts qui sont décrits plus précisément dans les 'sous' classes (ou classes spécialisées). Par exemple des classes *Rivières* et *Routes* peuvent être généralisées en une classe *Voie de communication.* On obtiendra alors une hiérarchie entre des concepts géographiques.

- L'agrégation est le fait de combiner des entités pour constituer une nouvelle entité qui décrit un autre concept. Par exemple, on peut créer le concept de *Réseau hydrographique* constitué à partir de toutes les entités connexes de type *Rivières.* Une agrégation peut être définie comme une association forte entre **1** entité vers **n** entités qui la composent. On parle de relation de composition.

On notera bien que la généralisation/spécialisation est une relation entre classes alors que l'agrégation est une relation entre entités. On notera également que cette vision est cohérente avec l'approche Objet en représentation des connaissances et en base de données.

Dans le papier de Nyerges, la généralisation - vue comme une abstraction de l'information géographique- consiste en la définition d'un nouveau schéma conceptuel ainsi que son instanciation (i.e. le fait de mettre des objets dans la base définie par ce nouveau schéma). Si on se place dans la configuration où l'on dispose d'une base de données initiale (décrivant un certain niveau de détail géographique), cela va se traduire 1/ par la définition d'un nouveau schéma conceptuel qui retrace un autre niveau de détail puis 2/ par le 'transfert' de données de la base initiale vers la nouvelle base en cours de constitution. Les opérations pour réaliser le nouveau schéma à partir du premier schéma et pour transférer les données sont les opérations de réification, d'association (de mise en association), de généralisation et d'agrégation. Effectivement, on va redéfinir les classes (par exemples on va supprimer des classes, supprimer des attributs) et des relations (certaines relations seront abandonnées). Des classes créées pourront être des généralisations de classes (la classe *Végétation* est une généralisation des classes *Bois* et *Vergers*) (figure 4). Lorsque l'on redéfinit les classes, on redéfinit les règles de sélection des objets de la classe, par exemple en disant que *' les bâtiments dont la surface réelle est inférieure à 80m2 ne seront pas représentés dans la base'*. Ce faisant, il y a un *filtrage* de données qui est une généralisation. De même de nouveaux concepts vont apparaître (donc de nouvelles classes) qui seront construits à partir d'une agrégation d'objets simples d'un autre type. Par exemple, on peut ajouter la classe *Zone Urbaine* dont les instances seront calculées (construites) par *agrégats* de bâtiments proches et de taille importante.

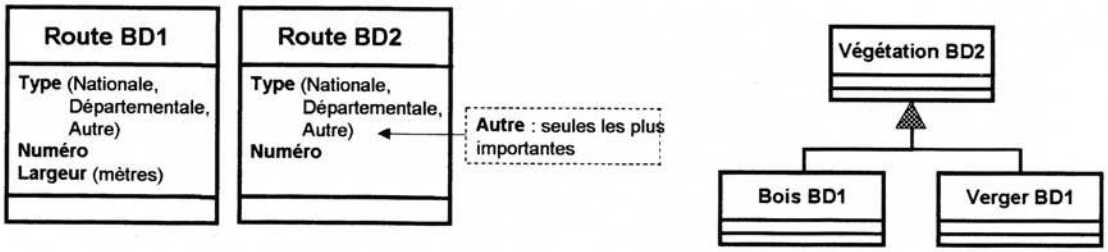

Figure 4 : Nouvelle réification: perte d'attribut (largeur) et redéfinition de contenu (autre) (à gauche)
Création d'une nouvelle classe (Végétation) par généralisation (à droite).

Même si il est *juste* de dire que les opérations de réification, d'association, de généralisation et d'agrégation sont au cœur de l'opération d'abstraction qui consiste à passer d'une représentation numérique du monde à un certain niveau de détail vers une nouvelle représentation de niveau de détail moindre, cela ne suffit pas pour définir ce qui se passe au niveau des données. En particulier,

- les règles de réification changent même si le nom de la classe est identique : une instance de la classe route ou bâtiment n'a donc pas la même signification dans les deux bases (figures 5),

- les agrégats qui représentent l'émergence de nouveaux concepts (de nouvelles classes) doivent être également définis.

[1] La *réification* est le fait de matérialiser un concept (par un objet, une classe) pour pouvoir le manipuler . La réification permet de représenter un concept.
La *classification* est le fait de rassembler des objets similaires en classes.

Comme on l'a vu précédemment, lors des changements de niveau de détail, de nouveaux concepts apparaissent qui sont traduits par l'apparition de nouvelles classes dans le schéma. Cela peut être le cas de la classe *Îlot* qui remplace les bâtiments dans les zones très denses. Dans ce cas, une nouvelle classe apparaît qui est un agrégat conditionnel de bâtiments (figure 5). Un objet de la classe *Îlot BD2* est un Îlot dense en bâtiments (dans BD1) qui serait cartographiquement représenté par une zone grisée sur une carte par exemple. Dans ce cas, les bâtiments représentés dans BD2 sont ceux qui ne sont pas inclus dans un îlot de *Îlot BD2*.

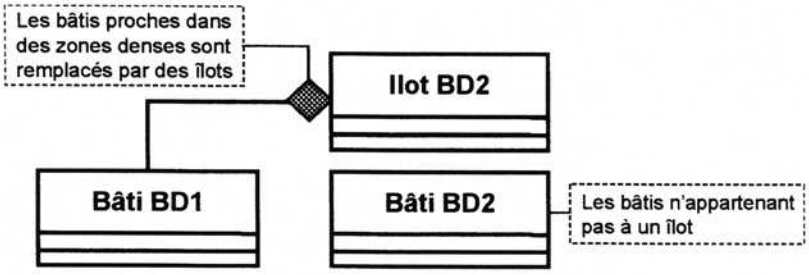

Figure 5 : Apparition des îlots dans la nouvelle base

Si on veut décrire ce qui se passe lors des changements de niveaux de détail entre une base de données et une autre, il est nécessaire d'ajouter des concepts sous forme de classes qui permettent d'expliquer les modifications. Pour plus de concision, on peut adopter une modélisation en représentation multiple (chapitre 4).

Si on reprend les définitions d'une base de données géographiques conforme, lors d'une généralisation dite de modèle, il faut définir :

- **Un nouveau schéma conceptuel et ses contraintes** :

  - L'émergence de nouvelles classes obtenues par généralisation de classes et par agrégation d'objets. Les conditions d'agrégation doivent être décrites au mieux puisque généralement ce sont des considérations contextuelles qui régissent les mécanismes d'agrégation (par exemple tous les bâtiments ne s'agrègent pas en zone urbaine).

  - De nouvelles règles de sélection qui définissent les conditions que doivent remplir les objets pour être représentés dans la nouvelle base de données. Ce point est particulièrement délicat à décrire lorsque l'information n'est plus exhaustive mais que l'on sélectionne les objets les plus importants et que cette importance se définit de façon contextuelle. Ce sera le cas pour de nombreux types d'objets tels que les routes ou les rivières.

  - La résolution sémantique au niveau des attributs, c'est-à-dire le niveau de détail de description des attributs (i.e. les contraintes de domaine).

- **Une nouvelle description géométrique** qui se décompose en quatre volets :

  - Les règles de représentation de la géométrie (par exemple le fait de dire que les rivières sont représentées par leur axe et non par leurs bords si leur largeur terrain est inférieure à 50m).

  - La résolution qui donne l'ordre de grandeur géométrique des phénomènes présents dans la nouvelle base et qui est généralement traduite dans les règles de classification (conditions géométriques que doivent remplir les objets pour être représentés dans la base).

  - La précision géométrique qui décrit l'écart entre la position terrain et la position représentée.

  - La granularité qui décrit la taille des plus petites formes représentées.

Le tout - le schéma conceptuel contraint et la description géométrique - doit donner une description de l'espace géographique cohérente avec l'espace réel, c'est-à-dire que les données de cette nouvelle base doivent collectivement décrire un espace de signification géographique. En guise d'exemple, nous insisterons surtout sur le besoin de maintien des relations de positionnements relatifs (distance et topologie) qui sont autant importantes sur une carte que dans une base de données : sur une carte ou dans une base de données, si deux routes ne se croisent pas dans la réalité, elles ne doivent pas se croiser dans leur représentation.

Pour résumer, la généralisation de modèle (ou de base de données conforme) redéfinit les règles de sélection (le *quoi* : la définition de ce qui est dans la base) et les règles de représentation (le *comment* : la représentation dans la base les objets qui sont retenus ou créés par agrégation). Les règles de sélection et les règles de représentation sont explicitées (ou plutôt à expliciter) au niveau du nouveau schéma de données et induisent des modifications au niveau des données. Dans de très nombreux cas (la sélection, l'agrégation, le changement d'implémentation), ces opérations requièrent une analyse contextuelle des données soit pour savoir quels sont les objets à éliminer ou à agréger , soit pour contrôler que les opérations faites sur les objets de base (telles que les changements d'implémentation ou le changement de résolution) ne génèrent pas des incohérences avec l'environnement de ces objets.

Généralisation cartographique et généralisation graphique

'The difference between the model view and the cartographic view of generalization is the possibility for database manipulation in the former case, independent of cartographic representation' (Muller et al 1995 p.6). La généralisation cartographique est définie comme celle qui intègre les contraintes liées à la représentation graphique (' *cartographic generalization (...) deals with graphic representation* ' (Muller et al 1995 p.5). Elle intègre donc les seuils de lisibilité et déclenche les opérations de dilatation et de déplacement

('An operation such as feature displacement is typically cartographic' (Muller et al 1995 p.5).

Par ailleurs, selon la définition traditionnelle, la généralisation cartographique est le processus qui permet de passer d'une échelle à une autre échelle plus petite. Elle intègre donc la généralisation conceptuelle (que l'on retrouve dans la généralisation dite de modèle) et la généralisation structurale (Bertin 1998(1967) p.300). Pour sortir de cette ambiguïté, certains auteurs (dont (Harrie et Sarjakoski 2002 p.233)) parlent de généralisation graphique qui est restreinte à la prise en compte des contraintes de lisibilité et de la légende finale de la carte.

Les contraintes de lisibilité (visibilité et interprétation) et la prise en compte de symboles (telle que la largeur des routes – qui correspond à une dilatation jusqu'à une valeur fixée en fonction de la nature de la route) induisent :

- le grossissement des petits objets surfaciques,
- le déplacement des objets trop proches (produisant des anamorphoses locales),
- certaines éliminations locales d'objets lorsque le déplacement n'est pas suffisant pour résoudre les problèmes d'espace,
- l'amplification ou l'élimination de certaines formes dans la géométrie des objets soit parce que le symbole des objets est empâté en raison de sa largeur (c'est souvent le cas des routes sinueuses) soit parce que l'échelle de la carte ne correspond pas à la granularité de la base (c'est le cas lorsqu'on est resté dans la gamme d'échelles qui n'induit pas l'ajout de nouveaux concepts).

La figure 6 présente une synthèse des trois généralisations en reprenant le schéma de la figure 3.

- La **généralisation de BD** conforme ou de modèle est appliquée entre bases de données géographiques conformes. Il n'y a pas de contraintes liées à la symbolisation et aux tailles minimales (mais les objets de la base peuvent être par ailleurs contraints par des tailles minimales).
- La **généralisation graphique** est provoquée par l'utilisation de symboles et par la mise à l'échelle. Les produits source et généralisé sont dans la même gamme d'échelles. La généralisation graphique est essentiellement limitée à des anamorphoses et des simplifications géométriques.
- La **généralisation cartographique** (ou généralisation) inclut en même temps l'apparition de nouveaux concepts et la mise à l'échelle graphique.

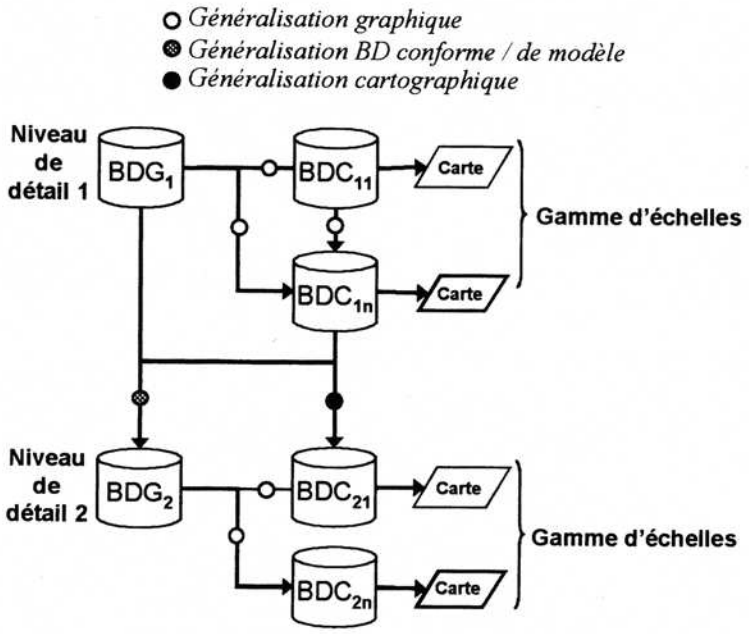

Figure 6 : Les différents types de généralisation

Nous avons fait figurer le processus de généralisation cartographique directement entre des bases de données cartographiques (BDC) parce que dans certains cas, les producteurs partent de bases de données cartographiques pour en dériver directement d'autres bases de données cartographiques, ce qui correspond au processus traditionnel de généralisation cartographique.

4 L'apport de la représentation multiple de données géographiques et cartographiques

Une unique base de données de référence ?

Les chapitres précédents illlustrent la diversité des bases

de données nécessaires à une bonne analyse et une bonne gestion de l'information géographique. Auparavant la carte faisait à la fois office de lieu de recensement de l'information géographique (avec la précision et la complétude nécessaire) et de médiateur à travers le langage cartographique. Désormais, on préfère séparer ces aspects au sein même des bases de données, en différenciant les bases de données géographiques conformes des bases de données cartographiques. De plus, les besoins d'analyse à différents niveaux de détail restent inhérents aux données géographiques et se reflètent au niveau des bases de données par la co-existence de différentes bases à différents niveaux de détail.

Pour assurer au mieux la cohérence entre ces bases, et pour optimiser la gestion des mises à jour, l'idéal du producteur de données centralisé semblerait être de ne maintenir qu'une seule base de données géographique très détaillée, et d'en dériver automatiquement et à la demande toutes les autres bases géographiques ou cartographiques, à différents niveaux de détail. Cependant, cela n'est en pratique ni possible ni souhaitable :

- La première difficulté est technique. Par définition, les généralisations permettent de passer d'une base de données à une autre. Mais ces généralisations ne sont pas toutes entièrement automatisables : la complexité de ces processus requiert toujours l'expertise des cartographes.
- Un deuxième frein provient des données elles-mêmes : différentes bases peuvent refléter différents points de vue qui ne se dérivent pas les uns des autres (on ne peut entièrement dériver le point de vue foncier du cadastre à partir du point de vue topographique de l'IGN, ni inversement).
- Un troisième problème vient des sources et des rythmes de mise à jour. Les données à faible niveau de détail sont plus faciles à mettre à jour que celles plus détaillées, et sont donc plus souvent actualisées. Si on veut conserver cet avantage des données peu détaillées, on ne peut pas toujours se permettre d'attendre que les données détaillées soient mises à jour pour propager ces évolutions aux autres données.
- Enfin, un dernier aspect du problème provient plus directement de l'utilisation des données. Une fois que des données ont été dérivées (automatiquement ou non) à partir d'une base de données de référence, il peut être important de conserver ces données dérivées et de les mettre à jour (plutôt que de mettre à jour les données détaillées et de réitérer la dérivation à la demande) : cela garantit à l'utilisateur de ces données dérivées une meilleure cohérence entre les différentes versions.

Plus généralement, si il existe divers producteurs de données, et que chacun souhaite conserver la maîtrise de ses bases et de leur mise à jour pour assurer le respect de ses propres contraintes, alors le scénario d'existence d'une unique base de données de référence de laquelle se dériverait toutes les autres bases devient encore moins pertinent.

La représentation multiple pour une meilleure gestion de la diversité des données géographiques

Un scénario plus réaliste pour gérer la diversité des bases de données est la co-existence de deux grandes approches : la généralisation à la demande quand cela est possible, et l'intégration des différentes bases dans un système à représentation multiple.

La représentation multiple est le principe de modélisation de bases de données qui consiste à représenter au sein d'une base plusieurs représentations informatiques d'une même entité du monde réel, explicitement mis en relation (cf. figure 7). Dans les bases de données classiques, une entité du monde réel est représentée par un seul objet. Dans un modèle à base de représentations multiples, une même entité peut être représentée par plusieurs objets, qualifiés alors d'homologues et explicitement reliés. Ces différents objets sont gérés au sein d'une même base. En pratique, ces objets peuvent aussi être stockés dans différentes bases physiquement séparées. Mais ceci peut être rendu transparent pour l'utilisateur qui accède à ces bases comme à une base unique à travers des systèmes de bases de données dits fédérés.

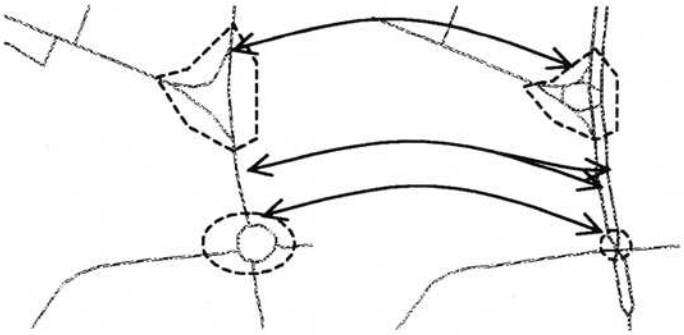

Figure 7 : Deux représentations d'un même réseau routier , et quelques relations entre objets homologues.

La représentation multiple permet d'expliciter les liens entre les différentes représentations des objets du monde réel. Cela est utile pour relier les objets 'géographiques' à leurs équivalents 'cartographiques', mais aussi pour représenter les liens entre les objets homologues reflétant différents niveaux de détail et différents points de vue.

La représentation multiple dans les bases de données apparaît particulièrement utile dans le domaine des bases de données géographiques, comme le reflète divers travaux à ce sujet (Kilpeläinen 2000, Vangenot et al. 2002, Bédard et Bernier 2002).

Utilités de la représentation multiple

La représentation multiple, comme nous l'avons évoqué, est une approche très intéressante pour le producteur de données, et en particulier pour lui permettre de gérer les mises à jour. Elle doit permettre d'introduire les mises à jour sur une base, et de les propager plus ef ficacement aux autres bases : au minimum, il devient possible d'indiquer dans les autres bases quels objets doivent être mis à jour à leur tour , et au mieux il est possible d'automatiser cette propagation. Ceci concerne en particulier la propagation des mises à jour d'une base de données géographique conforme vers une base de données cartographique qui en est dérivée (et donc au final une mise à jour accrue des cartes), mais aussi la propagation d'une base vers une autre moins détaillée (Kilpeläinen 2000, Badard et Lemarié 2000).

Un deuxième intérêt de la représentation multiple est de permettre la mise en œuvre d'opérations de contrôle de la qualité des bases. En effet, une fois les données de différentes bases mises en correspondance, on peut vérifier la qualité d'une base par rapport à une autre base plus précise servant de référence. Plus généralement, si aucune base ne peut être vue comme une référence par rapport à une autre, cela permet au moins d'étudier la cohérence des données (Egenhofer et al. 1994, Sheeren et al. 2004). La représentation multiple est donc un moyen d'assurer la cohérence entre les différentes bases de données, ou du moins de détecter certaines incohérences.

Améliorer la mise à jour et la cohérence des bases a bien sûr des conséquences directes sur leur qualité, et donc sur leurs utilisations. Mais cela n'est pas le seul ef fet bénéfique escompté pour les utilisateurs. La représentation multiple peut et doit permettre d'améliorer les utilisations qui sont faites des données. Elle doit être utilisée pour réaliser de véritables analyses combinant plusieurs points de vue et plusieurs niveaux d'analyse. La mise en relation explicite des données dans les bases doit permettre une navigation entre les objets homologues, navigation qui doit être beaucoup plus riche qu'une simple superposition visuelle des données. Depuis que les cartes existent, l'espace géographique a toujours été analysé selon dif férents points de vue. Nous espérons que la représentation multiple est un moyen efficace pour que l'analyse des données numériques suive au plus près cette approche.

5 Conclusion

Les cartes comme les bases de données géographiques ont différents niveaux de détail. La notion d'échelle mathématique, rapport de longueurs entre la carte et le terrain, n'est pas à proprement parlé une notion définie pour les bases de données. Celles-ci sont mieux définies par les notions de précision géométrique, et de résolution géométrique et sémantique. Ces notions décrivent le « quoi » (ce qui est représenté), et le « comment » (comment cela est représenté). Néanmoins, les bases de données possèdent aussi intrinsèquement une échelle caractéristique à travers le niveau de détail des phénomènes qu'elle met en valeur, ou plus précisément une gamme d'échelles associée. Il est important de bien distinguer les notions de bases de données géographiques conformes à l'espace modélisé, et de bases de données cartographiques qui intègrent des contraintes de lisibilité de la carte. On navigue entre ces bases à différents niveaux de détail par l'intermédiaire des différents types de généralisation : les généralisations de bases de données conformes, les généralisation cartographiques, et les généralisations graphiques. Dans ce contexte, et du point de vue du producteur de données, l'intérêt de la représentation multiple est bien identifié pour la gestion et la mise à jour des bases : la représentation multiple est nécessaire à l'automatisation des processus de généralisation et la généralisation génère des représentations multiples.

Nous pensons que le concept de représentation multiple ouvre la voie à une utilisation plus riche des données géographiques numériques par les utilisateurs géographes. Il devrait permettre une prise en compte plus facile de phénomènes que l'on étudie usuellement indépendamment les uns des autres, faute de pouvoir les étudier conjointement.

Références

Badard, T., and Lemarié, C., 2000, Propagating updates between geographic databases with dif ferent scales. In *Innovations in GIS VII: GeoComputation*, Atkinson, P. and Martin, D.,Eds. (Taylor and Francis, London).

Bédard, Y., and Bernier, E., 2002, Supporting Multiple Representations with Spatial V iew Management and the Concept of «VUEL». Proceedings of Joint W orkshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, ICA Com. on Map Generalisation. Ottawa, CANADA, July 7th-8th.

Beguin M et D. Pumain, 1994, *La représentation des données géographiques*. Paris, Armand Colin

Bertin J., 1998, *Sémiologie graphique*, EHESS ISBN 2-7132-1277-4 (réedition de *Sémiologie Graphique* 1967, Mouton, Gauthier-Villars).

Brassel K., Weibel R., 1988, «A review and conceptual framework of automated map generalization» *IJGIS* Vol. 2, N° 3, p229-244.

Comité Français de Cartographie (CFC), 1970, «Glossaire français de cartographie» Dans le *Bulletin du Comité Français de Cartographie* Fasicule N° 46.

Egenhofer, M.J., Clementini, E., and Di Felice, P., 1994, Evaluating inconsistencies among multiple representations. Proceedings of the Sixth International Symposium on Spatial Data Handling (SDH'94), Edinburgh, Scotland, pp. 901-920.

Harrie L., Sarjakoski Tapani., 2002, «Simultaneous graphic generalization of vector data sets» dans *GeoInformatica* Vol. 6, N° 3, p 233-262.

Kilpeläinen, T., 2000, Maintenance of Multiple Representation Databases of Topographic Data. The Cartographic Journal, vol.37, n.2, December 2000, pp.101-107.

Lacoste Y., 1980, «Les objets géographiques» dans *Cartes et figures de la terre*, Centre Georges Pompidou, Paris, p16-23.

Müller J.-C., Weibel R., Lagrange J.-P., Salgé S., 1995, «Generalization : state of the art and issues» dans Müller , Lagrange and Weibel (ed.) *GIS and Generalization,* London, Taylor and Francis, p 3-17.

Nyerges T., 1991, «Representing geographical meaning» dans B Buttenfield & R McMaster (ed.) *Map Generalization* , Harlow, Longman, p59-85.

Racine JB, 1981, «Problématiques et méthodologie : de l'implicite à l'explicite» dans H. Isuard, JB Racine et H Raynard (ed.) *Problématiques de la géographie* Paris, PUF le géographe.

Ruas, A., 2004, *Le changement de niveau de détail dans la représentation de l'information géographique* . 3 Novembre 2004. Habilitation à Diriger des recherches. Université de Marne La V allée. ftp://ftp.ign.fr/ign/COGIT/HDR

Smaalen J., 1996, «A hierarchical rule model for geographic information abstraction» *7 th international symposium on Spatial Data Handling* (SDH'96, Pays-Bas/Delft), 4B.31-41.

Vangenot, C., Parent, C., Spaccapietra, S., 2002, Modelling and manipulating multiple representations of spatial data. Proceedings of the 10th International Symposium on Spatial Data Handling, pp.81-93.

Sheeren, D., Mustière, S., Zucker, J.-D. 2004, Consistency Assessment Between Multiple Representations of Geographical Databases: a Specification-Based Approach. Proceedings of the 11th International Symposium on Spatial Data Handling, Leicester, UK.