

3 L'INFORMATION GEOGRAPHIQUE

Bengt Rystedt, Suède

Traduction : Henri Desbois et Emmanuel Devys

Harmonisation : Felix de Montety

3.1 Introduction

On appelle « information géographique » une information liée à une position géographique. Cette position, ou localisation, s'exprime sous une forme numérique qui permet son traitement informatique. La longitude et la latitude sont les informations les plus utiles. La localisation est traitée plus en détail dans le chapitre suivant. Une façon simple de décrire comment les ordinateurs traitent l'information géographique est de se la représenter sous forme de couches superposées qui décomposent le paysage (voir figure 3.1).

Certaines couches correspondent à la topographie, d'autres aux divisions administratives, aux routes, aux lacs, aux rivières, et ainsi de suite. On peut ajouter d'autres couches pour représenter des données thématiques sur la géologie, l'utilisation du sol, et la végétation. La figure 3.1 montre le principe d'un modèle numérique de paysage structuré en couches. Ce mode d'organisation de l'information géographique en couches a été employé pour la première fois dans les années 1960 pour l'inventaire des terres du Canada, et a été développée et utilisée comme base pour l'aménagement du territoire et la gestion des ressources naturelles du pays.

Les couches contiennent l'information géographique mais elles doivent aussi être liées à divers attributs stockés dans des tables de bases de données relationnelles. Un objet surfacique de la base est lié à ses attributs par un numéro unique que l'on appelle identifiant. Le traitement informatique de l'information géographique a fait un progrès décisif quand Jack Dangermond a découvert qu'on pouvait manipuler simultanément la géométrie dans une base de données

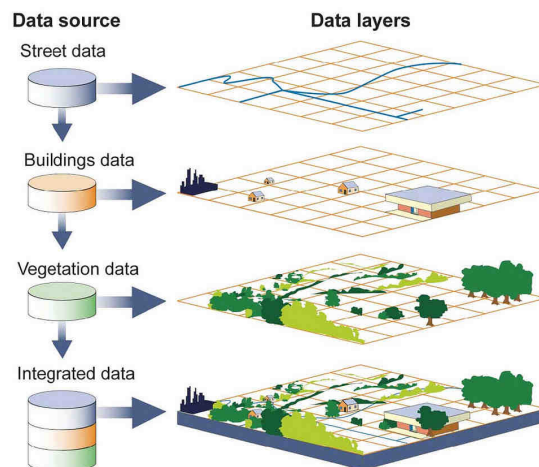


Figure 3.1. Illustration des principes d'une modélisation par ordinateur d'un paysage. Chaque couche contient à la fois des données de localisation et un ensemble d'attributs. Source :

<http://education.nationalgeographic.com/education/photo/new-qls/>

et les attributs dans une autre. Il appela son système ARC/INFO, ARC pour la géométrie et INFO pour la base de données relationnelle stockant les attributs. De nombreux autres systèmes ont été mis au point par la suite.

3.2 Le modèle de données

Avant que l'information géographique soit utilisable pour l'analyse et la cartographie, elle doit être structurée dans un modèle de données approprié. La figure 3.1 montre une ébauche d'un tel modèle structuré en couches. Il faut également définir tous les objets à inclure dans la base. Ces objets sont constitués d'éléments qui peuvent être des points, des lignes, ou des surfaces.

L'élément le plus important d'un modèle de données géographiques est sa topologie, c'est à dire la façon dont les éléments de base s'agencent pour former des

réseaux et des surfaces. Dans un réseau tel qu'un réseau routier, les points aux extrémités des segments sont appelés des nœuds, et la topologie décrit quelles routes sont connectées à quels nœuds (figure 3.2).

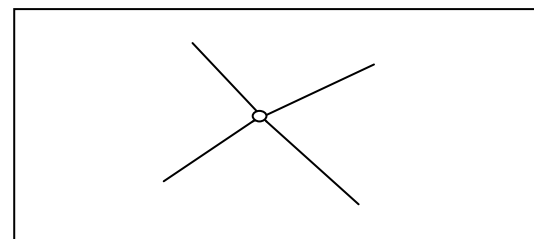


Figure 3.2. Principe de représentation d'un réseau routier, avec quatre routes connectées à un nœud. Les nœuds et les routes doivent avoir un identifiant, et peuvent aussi avoir des attributs.

Dans le cas des surfaces, chaque surface en jointe plusieurs autres. Lorsqu'on suit une limite dans un sens, il existe toujours une surface à gauche et une autre à droite. Lorsqu'on décrit la topologie d'un élément surfacique, chaque segment de sa limite est en double, une fois pour chaque sens, avec une surface à gauche et une autre à droite. Cela peut sembler superflu, mais c'est nécessaire pour que le système soit adapté à l'analyse géographique dans un Système d'Information Géographique (SIG). La figure 3.3 montre une commune divisée en deux paroisses représentées par des polygones.

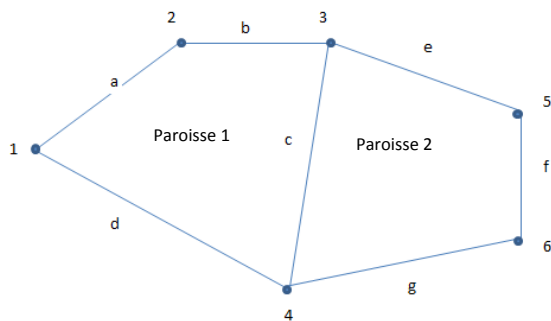


Figure 3.3. Une commune divisée en deux paroisses. En suivant la frontière de chaque paroisse dans le sens horaire, on peut voir que le segment « c » est parcouru dans deux sens, alors que les autres ne le sont que dans un seul.

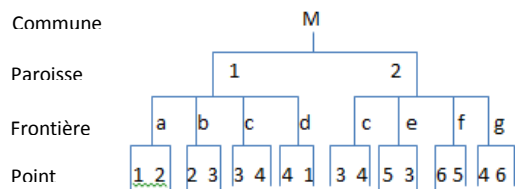


Figure 3.4. La structure hiérarchique des deux paroisses de la figure 3.3. On voit que le segment « c » figure deux fois, et que tous les points figurent deux fois, sauf les points 3 et 4, qui figurent 4 fois.

L'ensemble des limites administratives peut comprendre les nations, les comtés, les communes, les paroisses et les parcelles foncières. Cela signifie également que tous ces échelons peuvent être représentés par des surfaces dont les limites sont stockées dans la base (par exemple sous forme d'une structure hiérarchisée comme dans la figure 3.4). La figure montre aussi que les segments et les points sont enregistrés plusieurs fois, si bien que l'augmentation de la taille de la base est plus rapide qu'un accroissement linéaire.

Nous avons jusqu'à présent vu une structure de données, et suggéré que ces données devaient être stockées dans une base. L'usage d'une base de données relationnelle est une façon très répandue d'organiser une base de données. Cela signifie que les données sont stockées dans des tables, et que des relations établissent des liens entre les différentes tables. Une base de données relationnelle pour l'exemple précédent est présentée dans les tables ci-dessous. Le nombre de colonnes dépend ici du nombre de paroisses. Les coordonnées sont seulement des approximations.

Table 3.1. Tables d'une base de données relationnelle. Les coordonnées x et y sont données à titre d'exemple.

Commune	Paroisse 1	Paroisse 2
Nom de la commune	Nom de la paroisse	Nom de la paroisse

Surface	Segment	Segment	Segment	Segment
Paroisse 1	a	b	c	d
Paroisse 2	c	e	f	g

Frontière	Point	Point	Paroisse 1	Paroisse 2
a	1	2	1	
b	2	3	1	
c	3	4	1	2
d	4	1	1	
e	3	5	2	
f	5	6	2	
g	6	4	2	

Point	Coord X	Coord Y	Segment	Segment	Segment
1	80	229	a	d	
2	221	121	a	b	
3	375	119	b	c	e
4	372	295	c	d	g
5	517	127	e	f	
6	544	228	f	g	

3.3 Trouver des coordonnées dans une base de données

Les tables ci-dessus sont organisées selon l'ordre des identifiants des objets. Chaque table correspond à un fichier dans la base, si bien qu'il est assez facile de trouver un objet. C'est plus difficile avec les coordonnées. La coordonnée x est définie comme une distance sur l'axe de l'équateur au pôle (nord ou sud). La coordonnée y correspond à la distance, sur un axe est-ouest, à partir d'un méridien de référence, en fonction de la projection choisie (pour plus de détails, voir chapitre 9). On ne peut évidemment pas organiser les coordonnées en tables. Une solution de ce problème est d'utiliser des quadrees. On commence par diviser la zone en quatre carrés, puis chaque carré à nouveau en quatre de sorte qu'on arrive à 16 carrés, puis on continue jusqu'à avoir une seule paire de coordonnées dans chaque carré. On utilise le système binaire pour attribuer des identifiants aux carrés. Après la première division, nous avons les nombres 00, 01, 10, et 11. Avec les quadrees, il est facile de trouver les coordonnées d'un point simplement en cliquant sur l'écran. On ne donne pas ici d'exemple de quadree. Pour en savoir plus, on recommande de se reporter à Worboys et Dunckham (2004).

3.4 Modèle d'information

Une base de données géographique doit être fondée sur le monde réel et répondre à des besoins d'analyse spécifiques. Examinons par exemple un système de gestion de la fibre optique d'un quartier. Il faut prendre en compte des objets comme les propriétés immobilières, les propriétaires ou les occupants, la localisation des câbles, les contrats de gestion et les coûts. L'analyse des besoins doit être discutée avec les futurs utilisateurs du système. Il est également important de conserver une trace de tout le processus. Les étapes et la documentation associée sont montrées dans la figure 3.5.

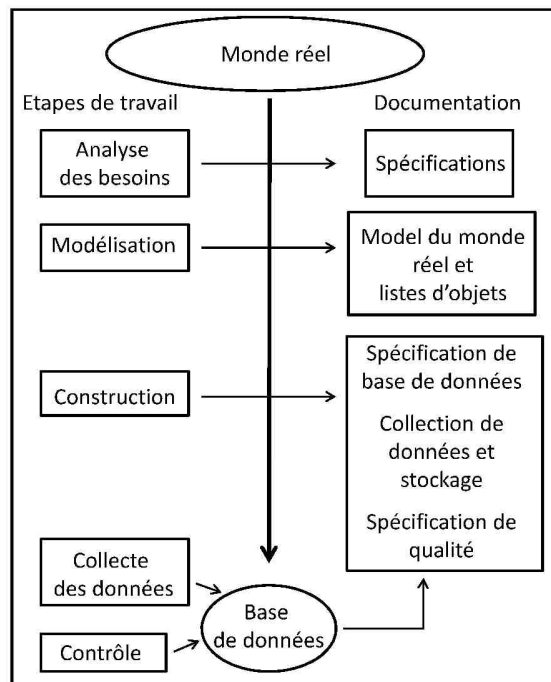


Figure 3.5. A gauche, étapes du processus de la modélisation de l'information ; à droite, documents associés à chaque étape.

La modélisation de l'information a fait l'objet de recherches nombreuses ces dernières années. Un comité technique de l'ISO (ISO TC 211) a élaboré et publié de nombreuses normes qui ont été utilisées par tous les producteurs d'information géographique. Dans notre exemple sur la fibre optique, certains types d'objets ont été mentionnés. Leurs spécifications sont précisées dans le catalogue d'objets.

3.5 Métadonnées et qualité

Nous n'énumérerons pas toutes les normes de l'information géographique mais nous nous contenterons d'une courte description des métadonnées et des informations de qualité. Les métadonnées

décrivent le type de données contenues dans la base et donnent les informations utiles en fonction du type d'application supposée.

Les métadonnées, ou données sur les données, décrivent la base de données et peuvent comprendre :

- Le nom de la base de données
- L'organisation responsable
- La couverture géographique
- La liste des objets du catalogue
- Le système de coordonnées
- Les règles de téléchargement et relatives aux schémas d'application
- Les coûts

Les informations de qualité sont aussi un type de métadonnées et peuvent comprendre :

- *L'origine*, qui indique les sources de base pour les données, le mode d'acquisition, et l'organisation responsable de la donnée.
- *La précision géométrique*, qui indique le degré de précision des coordonnées horizontales et verticales.
- *L'actualité*, qui précise la date des données et informe sur leur mise à jour prévue.
- *L'exhaustivité*, qui dit si tous les objets sont inclus ou non, qui informe sur la justesse de la classification, et qui dit si la topologie est complète (par exemple, si les routes forment un réseau complet).

Le même genre de mesure de qualité peut être fourni pour les attributs attachés aux objets. La qualité peut aussi comprendre des informations sur un système de contrôle qualité. L'information de qualité dans son ensemble permet à un utilisateur de savoir si une base convient à une application donnée.

3.6 Acquisition des données

Une base de données géographique comprend à la fois une géométrie et des attributs. La géométrie peut être en format vecteur ou raster. Le format vectoriel est plus naturel et plus proche de la géométrie qu'on enseigne à l'école. Les données raster sont sous forme de petits carrés appelés *pixels* et sont une représentation moins précise de la géométrie.

Les données géographiques peuvent être acquises de nombreuses manières. Les plus précises sont obtenues par des mesures directes sur le terrain. La numérisation de photographies aériennes ou de cartes est plus commune. Les géomètres créent des systèmes d'administration de la terre, pour lesquels ils mesurent les parcelles foncières et les éléments issus de l'aménagement du territoire comme la localisation des bâtiments, des routes et des ponts. Les réseaux enterrés d'électricité, de téléphone et d'égouts doivent également être relevés. La municipalité archive la localisation de ces réseaux, et un entrepreneur qui veut creuser le sol doit obtenir un permis et se faire remettre une carte avec l'emplacement des réseaux enfouis. Cependant, il n'existe pas de carte publique des tous les réseaux, car des criminels pourraient l'utiliser pour atteindre les intérêts vitaux de la commune. Dans les grandes villes, il existe des tunnels où passent toutes sortes de câbles.

D'autres sources de données géographiques sont les images aériennes et spatiales. Elles sont utilisées pour l'agriculture et la sylviculture, pour évaluer l'état de la végétation et l'utilisation du sol. Google Earth donne une bonne idée des possibilités. Cependant, l'utilisation des images en haute résolution peut être limitée par les

forces armées ou pour diverses raisons. Avec l'augmentation des résolutions, nous risquons d'en voir trop. On est autorisé à voir les images, mais il est interdit de recenser les installations de défense ou de retraiter les données sans une autorisation écrite des autorités qui doivent en être avisées, conformément à la loi nationale.

Pour traiter des données géographiques, il faut avoir un Système d'Information Géographique (SIG) capable de manipuler efficacement ces données (voir chapitre 15). Le résultat du traitement de l'information géographique peut être présenté sous forme de cartes complétées par des tableaux, comme dans un atlas numérique (voir chapitre 7). Lorsqu'on calcule le trajet le plus court entre deux endroits, on obtient une carte montrant le plus court chemin et un tableau qui indique les longueurs de chaque segment du trajet.

Les données géographiques peuvent aussi s'acquérir avec un GPS et un terminal mobile pour enregistrer les données. De retour à la maison, on peut transférer les données dans l'ordinateur, et si le résultat est satisfaisant, les introduire dans le système désiré ou les publier dans open street map (www.openstreetmap.org) pour les rendre accessibles à tous. Pour plus de détails Pour plus de détails sur open street map, voir le chapitre 16.

Bibliographie

Worboys, M.F. et Duckham, M., 2004: *GIS: A Computing Perspective*, Second edition. London: CRC Press. ISBN 0-415-28375-2.

Le texte de ce chapitre est issu d'un guide des bases de données publié en 1994 par Lantmäteriet, l'institut géographique national de Suède