



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de La Recherche Scientifique
Université Larbi Ben M'hidi Oum El Bouaghi
Faculté Des Sciences Exactes et des Sciences de La Nature et de la Vie
Département des Sciences de La Nature et de la Vie



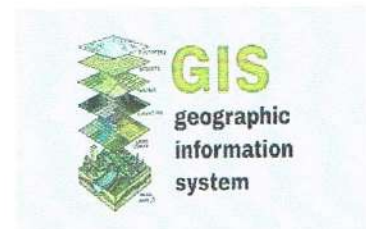
Polycopié de cours

Systèmes d'information géographique (S.I.G)

Destinée aux étudiants en master II :
Filière : Ecologie et environnement
Option : Ecologie des milieux naturels

Présenté par :

Docteur : KHAMMAR Hichem



Année universitaire : 2021/2022

AVANT PROPOS

Parce que les systèmes d'information géographique (SIG) intègrent progressivement l'héritage de la production cartographique classique, ils sont souvent assimilés ou réduits à des outils de cartographie assistée par ordinateur. Parce qu'ils reposent nécessairement sur la technologie informatique, ils sont aussi parfois assimilés à des ordinateurs spécialisés, à leurs périphériques et à leurs données.

En réalité, la dimension technique, économique et sociale des SIG va bien au-delà de ces aspects concrets et visibles de leurs activités, et il est important de recadrer celles-ci en rappelant les concepts généraux qui sont à la base des SIG.

Voici donc un polycopié à l'évidence utile. Il est consacré à l'étude des différents concepts autour de l'information géographique. Il prend en compte le fait que les étudiants seront confrontés lors de leurs futures activités à des problématiques de gestion d'information géographique.

Le polycopié présente les différentes clés qui leur permettront de remplir soit des fonctions techniques autour du SIG soit de gérer des projets traitant de données géographiques sans avoir à les manipuler (gestion de prestataires, maîtrise d'ouvrage, management d'équipe SIG, composante SIG d'un projet plus général, etc.).

Ainsi à l'issue de ce cours, les étudiants seront capables de:

- Comprendre la notion de l'information géographique numérique
- Comprendre les concepts de bases des SIG
- Découvrir les fonctionnalités des SIG
- Utiliser efficacement les outils SIG de traitements de données urbaines mis à leur disposition dans le cadre de leur vie professionnelle
- Découvrir la variété de domaines d'application
- Pratiquer sur le logiciel SIG Map Info

Pour illustrer les différents concepts abordés en cours, les TPs réalisés sous MapInfo permettent de se confronter concrètement aux concepts et aux respects de certains principes communs à tous les logiciels de gestion de l'information géographique.

Ce document peut aider les étudiants dans le domaine SNV pour réaliser leurs projets et surtout les écologistes et les étudiants d'aménagement de territoire

Introduction générale

L'accroissement continu de la démographie et l'amélioration recherchée des conditions de vie poussent l'homme à améliorer l'exploitation du sol et augmenter sa production. Toutefois, le sol est vulnérable et occupe une surface limitée. Par conséquent, il doit être géré de la manière la plus rationnelle possible. Pour gérer, il faut connaître. Autrement dit, l'utilisation rationnelle de la surface disponible suppose l'existence, d'une part, de données sous forme d'*inventaires* complets et fiables et, d'autre part, d'*outils* permettant de manipuler, de gérer et d'analyser de telles données.

L'évolution rapide et continue que connaît, depuis une trentaine d'années, la technologie relative à la saisie et au traitement des données géographiques a favorisé le développement de chaînes de production et d'exploitation entièrement numériques. Des volumes considérables de données numériques sont produits chaque jour, constituant ainsi des inventaires sous forme de bases de données géographiques nationales, régionales et locales.

Récemment, et grâce à des outils d'analyse spatiale puissants tels que les Systèmes d'Information Géographique (ou SIG), les données géographiques numériques trouvent des applications multiples en matière de planification, d'aménagement, d'urbanisme, de transport, de sécurité et de sauvegarde de l'environnement. En effet, en offrant la possibilité de superposer des thèmes géographiques et de réaliser des opérations d'analyse spatiale sur base de données multisource et multimédia, les SIG constituent un outil d'aide à la décision.

Le présent document constitue le support destiné aux élèves ingénieurs en Topographie dans le cadre du cours Systèmes d'Information Géographique. Il présente de manière explicite les concepts et les principes de base sur lesquels sont fondés ces nouveaux outils. Il traite les différents volets à savoir : les définitions, le passage des Systèmes d'Information aux SIG, les Bases données et leurs modèles, les différentes opérations de saisie de données nécessaires à la construction des bases, l'exploitation de ces dernières en termes de modélisation et d'analyse spatiale.

Introduction aux Systèmes d'Information Géographique : définitions et concepts de base

I.1. De la carte aux systèmes d'information géographique

La connaissance du territoire a été, depuis toujours, une préoccupation des sociétés. Qu'il s'agisse, comme aux époques anciennes, de repérer des itinéraires de migration, de délimiter les droits d'usage des sols, de localiser des ressources, ou plus récemment d'analyser les conséquences des phénomènes climatiques, de prévoir des récoltes ou de guider la trajectoire de missiles, les hommes ont toujours eu besoin, de représenter graphiquement des événements ou des informations. C'est la raison pour laquelle la carte est un outil très ancien, dont les évolutions techniques, sociales et politiques correspondent à celles des civilisations: cadastres égyptiens et romains, cartes marines de l'époque des grandes découvertes, cadastre napoléonien, etc.

Avant que les ordinateurs ne soient utilisés dans les processus cartographiques, les bases de données spatiales étaient disponibles sous forme de dessin ou de tracé sur une portion de papier ou de film. Ceci avait des conséquences importantes sur la collection, le traitement et l'utilisation de l'information contenue dans ces bases de données: perte d'informations (classification, généralisation, ...), manipulation difficile (changement d'échelle, de projection, ...), etc.

Les procédés manuels d'établissement de ces cartes sont lents et très coûteux, mais ceci ne constitue pas un inconvénient tant qu'il s'agit de cartes dont la période d'utilisation s'étend sur plusieurs dizaines d'années. Mais récemment, il y a eu un tel besoin de cartes et d'informations sur l'évolution de notre environnement que les techniques cartographiques classiques deviennent totalement inadéquates: une carte pourrait être périmée avant même d'être publiée. Cette demande accrue de cartes a en partie provoqué l'automatisation de certaines tâches liées au processus cartographique et la création de bases de données cartographiques dans lesquelles résident, ensemble, données cartographiques et non cartographiques. En effet, depuis les années soixante, la cartographie a connu une profonde mutation liée à l'avènement de l'informatique, dont le premier grand attrait était de pouvoir gérer de très grandes quantités de données ainsi que d'automatiser des tâches répétitives.

Plus récemment, la photographie aérienne et plus précisément l'imagerie satellitaire permettent de suivre l'évolution du paysage, l'avancement de la désertification ou de l'érosion, les incendies forestiers, etc. Cependant, les produits livrés par les satellites et les détecteurs aéroportés nécessitent un certain nombre de traitements et de corrections pour les remettre sous

forme d'images et par la suite sous forme de cartes. Ceci fait l'objet de la télédétection.

Le besoin de marier la télédétection, les procédés géodésiques et topographiques, la cartographie et d'autres disciplines est rendue actuellement possible grâce au développement des Systèmes d'Information Géographique.

I.2. Définitions et concepts

I.2.1. Système

Systeme: "[...] ensemble organisé, d'éléments intellectuels. Ensemble coordonné de pratiques tendant à obtenir un résultat [...] ou présentant simplement une certaine unité. Ensembles de pratique, de méthode et d'institutions formant à la fois une construction théorique et une méthode pratique [...] ensemble possédant une structure [...] ou constituant un tout organique. Système dynamique, où intervient la notion du temps (opposé à statique)", d'après le Petit Robert (dictionnaire de la langue française).

Selon Gualtierotti (1991), "un système est fait de choses et de relations entre ces choses". Le Moigne (1990) définit la notion de système général comme "quelque chose (n'importe quoi, présumé identifiable) qui dans quelque chose (environnement) pour quelque chose (finalité du projet) fait quelque chose (activité = fonctionnement) par quelque chose (structure = forme stable) qui se transforme dans le temps (évolution)".

I.2.2. Donnée et information

Information: "Renseignements sur quelqu'un, sur quelque chose = enquête, examen, investigation, étude - Élément ou système pouvant être transmis par un signal ou une combinaison de signaux."

Donnée: "Ce qui est donné, connu, déterminé dans l'énoncé d'un problème, et qui sert à découvrir ce qui est inconnu. - Ce qui est admis, connu ou reconnu, et qui sert de base à un raisonnement, de point de départ pour une recherche (point de départ, élément de base). - Élément, renseignement.". (D'après le Petit Robert)

Les deux termes *données* et *informations* sont habituellement utilisés pour décrire le même concept. En effet, chez les francophones, la distinction entre ces deux termes est rarement faite: ils sont employés l'un pour l'autre (et réciproquement). Cependant, chez les Anglo-Saxons, les termes *données* (*data*) et *informations* font souvent l'objet d'une distinction: les premières sont généralement

"brutes de fonderie", les secondes représentatives de savoir et d'expertise (intelligentes et renseignées) [Salgé dans Rouet, 1991].

Pour Chevallier (1983), l'*information* est un élément de savoir, ayant une signification propre, et susceptible d'être représentée par des données pour pouvoir être stockée, traitée ou transmise. Tandis que la *donnée* est un phénomène physique porteur d'information (représentation d'information sous forme codée).

En distinguant les deux termes, Rouet (1991) précise qu'une *donnée* devient une *information* au moment même où elle est directement utile pour prendre une décision, pour réfléchir à une stratégie, ou simplement pour agir. Pantazis et Donnay (1996) ajoutent: "[...] une information peut être une connaissance dont la création / naissance nécessite un ou plusieurs traitements ou opérations sur d'autres données ou informations déjà stockées sur un support quelconque. Le fait d'être stockées sur un support, fait de ces informations, des données."

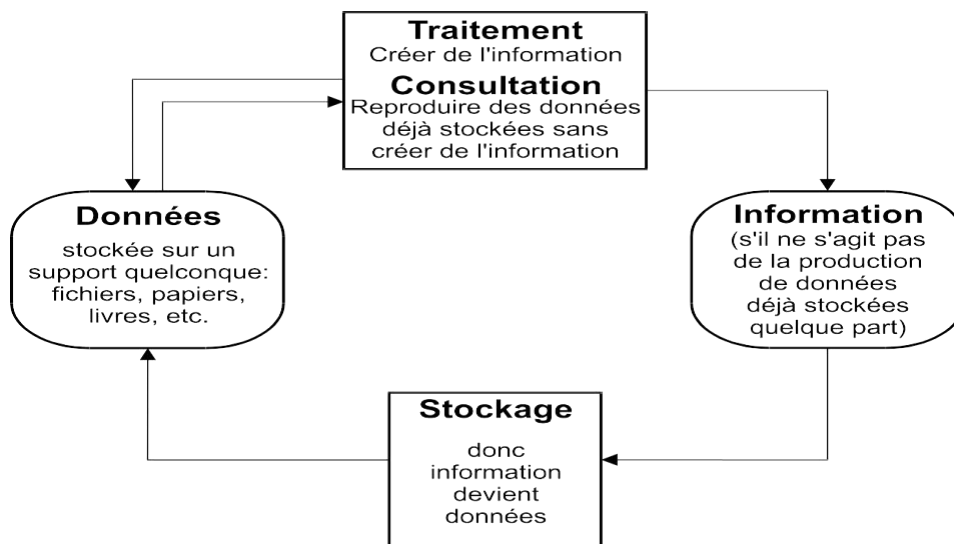


Figure I.1. *Information et données*

I.2.3. *Système d'information*

Un Système d'Information (ou S;I.) est un ensemble de processus s'appliquant à des données brutes afin de produire une information pertinente en vue d'une prise de décision.

Selon Rolland (1986), un S.I. peut être défini comme un artefact, un objet artificiel greffé sur un objet naturel qui peut être une organisation. Il est conçu pour mémoriser un ensemble d'images de l'objet réel à différents moment de sa vie; ces images doivent être accessibles par les partenaires de l'organisation qui s'en servent pour décider des actions à entreprendre dans les meilleurs

conditions. Un S.I. est, en quelque sorte, une extension de la mémoire humaine qui amplifie le pouvoir de mémorisation des acteurs de l'organisation et facilite leur prise de décision[Rolland, 1986].

I.3. Systèmes d'Information Géographique

I.3.1. Définitions et concepts de base

Qu'est ce qu'un système d'information géographique ?

Une large diversité de définitions, dont certaines sont citées ci-dessous, se proposent de répondre à cette question.

Clark (1986): Système de développement et d'élargissement du domaine d'utilisation des cartes.

Aronoff (1989): Ensemble de procédures utilisées pour conserver et traiter de l'information à référence géographique.

Burrough (1986): Ensemble d'outils puissants permettant l'acquisition, le stockage, l'édition aussi bien que la transformation et l'analyse des données spatiale relatives au monde réel.

Ducker (1979): Un type particulier de systèmes d'information dont la base de données contient des données reliées à des entités physiques, des activités ou des événements localisés et assimilables aux formes géométriques de points, de lignes et de zones.

Rhind (1989): Système composé de matériel et de logiciels ainsi que des procédures pour supporter l'acquisition, la gestion, la manipulation, la modélisation et la représentation des données à référence spatiale, pour résoudre des problèmes complexes de planification et d'aménagement.

Enfin, l'Union Internationale des géographes précise que le SIG n'est pas en lui-même une discipline, mais plutôt une assise commune pour le traitement de l'information et les domaines qui utilisent les techniques de l'analyse spatiale.

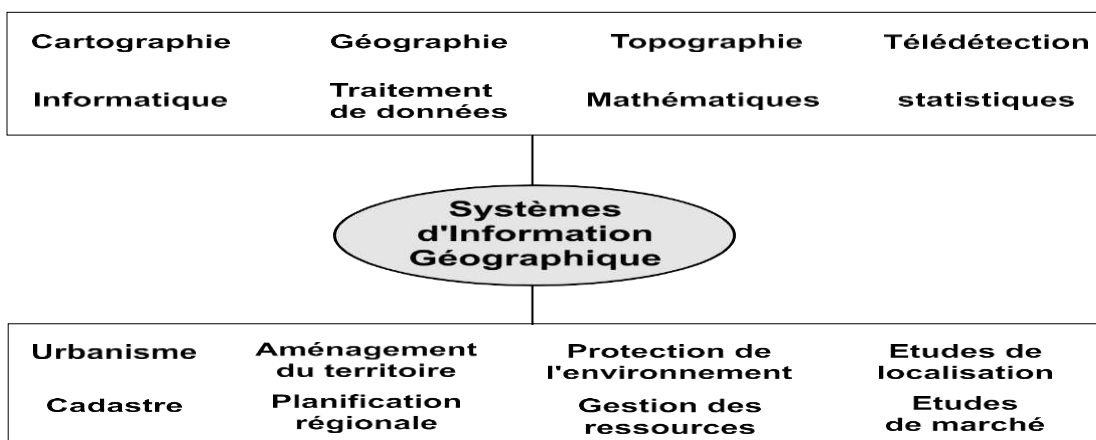


Figure I.2. Multidisciplinarité des SIG

Un SIG peut être conçu comme un outil d'analyse de l'information dans un contexte spatial. En effet, la raison d'être d'un SIG est de permettre l'analyse simultanée de plusieurs variables spatialisées sur base de données multisource et multimédia. Quand un certain nombre de cartes monodisciplinaires sont disponibles pour une zone d'étude déterminée, une combinaison de l'information multidisciplinaire est nécessaire pour une supervision globale. Cette opération peut être réalisée de deux manières:

La première est manuelle, elle représente un procédé classique mis en valeur par Mc Harg (1969). Elle consiste simplement à superposer les copies transparentes des cartes sur une table lumineuse (Fig. I.3). Les limites de cette méthode sont principalement: le nombre limité de cartes qui peuvent être superposées simultanément pour une analyse visuelle et, les difficultés d'obtenir à une même échelle les différentes copies transparentes.

La deuxième méthode est une approche assistée par ordinateur. Les données sont alors enregistrées sous forme numérique. Des programmes conçus pour des opérations d'analyse permettent de réaliser ce que Mc Harg faisait avec la superposition des transparents, mais d'une façon plus rapide et reproductible.

Cartes sur support
transparent

Cartes en format
numérique

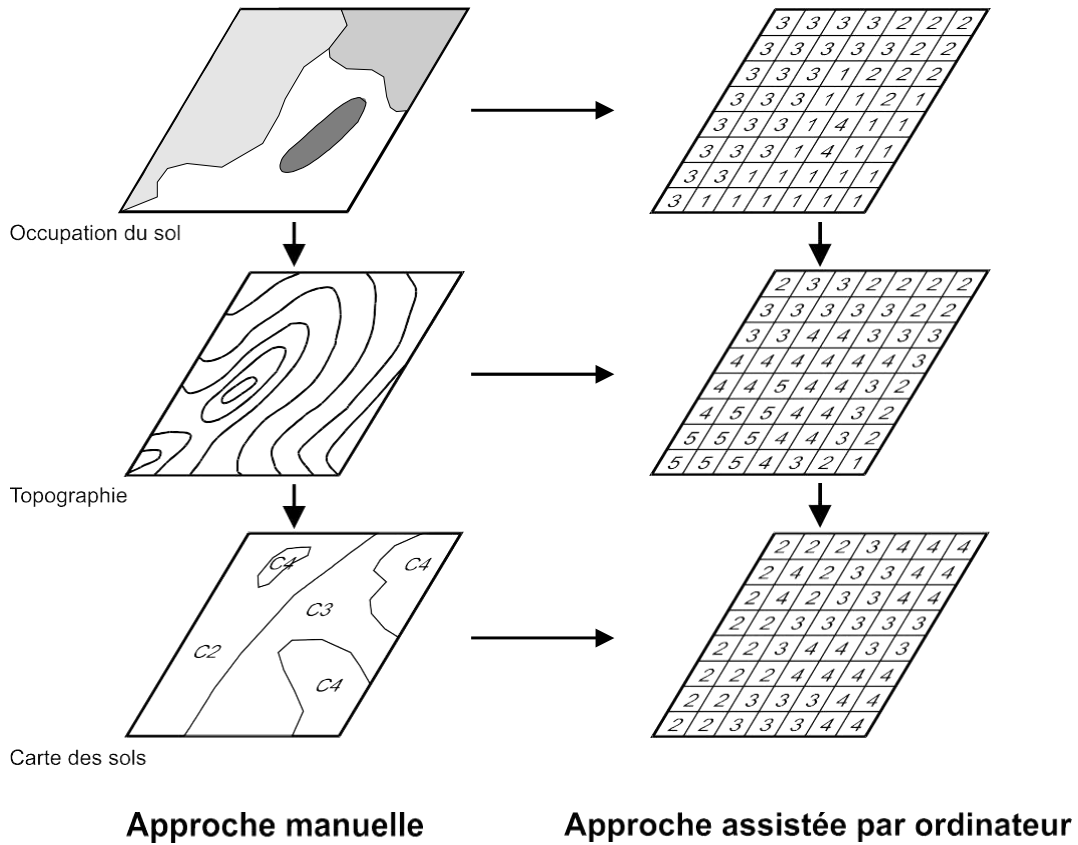


Figure I.3. Combinaison de l'information multidisciplinaire:
Approches manuelle et assistée par ordinateur

I.3.2. Composantes d'un SIG

Selon Burrough (1986), un SIG consiste en trois composantes principales à savoir: une composante "*Hardware*" (matériel), une composante "*Software*" (logiciel) qui se présente sous forme d'une collection de programmes ou modules spécifiques pour chaque type d'applications, et un *aspect organisationnel* convenable.

- **Hardware:** la partie "*Hardware*" se compose d'un ordinateur ou unité centrale de traitement (Central Processing Unit ou CPU) qui est liée à une unité disque permettant la sauvegarde des données et des programmes, d'un digitaliseur ou tout autre dispositif permettant la conversion analogique / numérique, ainsi que des moyens de sortie (écran graphique, traceur, imprimante, etc.). Aussi, une unité de bandes magnétiques peut se révéler nécessaires pour des opérations de stockage ou de communication avec d'autres systèmes.

Software: la partie "*Software*" d'un SIG comprend cinq modules de base qui assurent les opérations suivantes:

- Saisie et vérification des données;
- Gestion et sauvegarde de la base de données;

- Sortie ou représentation des données;
- Transformation des données;
- Interaction avec l'utilisateur (interface).

Figure I.4. Représentation schématique d'un SIG

Aspect organisationnel: les deux composantes *Hardware* et *Software*, même ultra puissantes, ne garantissent pas l'efficacité du SIG qu'elles composent. Pour assurer l'efficacité d'un tel système, les différentes composantes, y compris les différents acteurs (opérateurs, utilisateurs, gestionnaires, etc.) doivent être placées dans un contexte organisationnel approprié. Pour ce

faire, une étape d'étude et de conception est indispensable. Cette étape, préliminaire à toute acquisition de matériel ou de logiciels, permet de cerner le problème concerné, de définir les objectifs et de déterminer les moyens qu'il faudrait mettre en œuvre pour assurer l'efficacité du système. La conception et la mise en place des SIG fera l'objet du dernier chapitre.

I.3.3. SI / SIG

SIG = SI (G) ?

Traitement SI - SIG

la génération (production de l'information),

- Tables / cartes

l'acquisition et la mémorisation,

- Codes barres, clavier / numérisation
- Bases de données alphanumériques / images satellitaires, BD

géographiques la communication et la diffusion,

- Statistiques, tableaux / cartes, plans

...la mise à jour / maintenance,

- Clavier, codes barres, ... / numérisation, à l'aide des photo satellitaires

...la gestion,

- Données alphanumériques / données mutisource et

multimédia l'intégration d'informations et données,

- Données alphanumériques / données multimédias

Traitements de base d'un SIG

Traitements de base d'un SI classique:

- Accès, saisie, stockage, édition et tenue à jour, diffusion et échange

+ *Spécificités géographiques:*

- Saisie: interfaces avec les périphériques spécialisés (table à numériser, scanner, station totale, GPS, etc.)
- Mémorisation, Stockage: structures de données propres à l'aspect géométriques

des données géographiques (topologie, TIN, quadtrees) et compactage des données images

- Edition et tenue à jour : fonctions d'édition cartographique (mosaïque, généralisation)

- Diffusion, échange: standards d'échange (ex. Edigéo) et formats des données

vectérielles et images (ex. DXF, BNA, TIFF)

1.3.4. SIG / CAO

Il est clair qu'un SIG diffère d'un système graphique. En effet, ce dernier ne tient pas compte des attributs et des relations topologiques qui peuvent caractériser les entités graphiques considérées. En plus, la différence majeure entre un SIG et les technologies modernes qui s'y apparentent, particulièrement les systèmes de Conception Assistée par Ordinateur (CAO, CAD pour Computer-Aided-Design), concerne le grand volume et la diversité de données ainsi que la nature spécialisée des méthodes d'analyse spatiale utilisées dans un SIG. Cependant, vu l'importance du marché des SIG, les grands constructeurs des produits CAO ont franchi la limite du domaine de la CAO, et commencent à proposer des modules SIG qui se greffent sur les systèmes CAO correspondant

Exemples:

AutoCAD Map de Autodesk Inc sur AutoCAD;

Arc-CAD de ESRI Inc. sur Auto CAD;

Geo-Graphics de Bentley Systems Inc. sur MicroStation.

1.3.5. Intérêt et domaines d'application des SIG

Il ressort de ce qui précède qu'un SIG est beaucoup plus qu'un moyen de codage, de sauvegarde et d'extraction de données. Dans un sens plus large, les données dans un SIG représente un modèle du monde réel. Le traitement de ces données permet de fournir des produits pouvant servir de base pour la prise de décision dans divers projets concernant l'aménagement du territoire, l'analyse des processus environnementaux, l'étude de l'impact des modifications dues à l'activité humaine, etc.

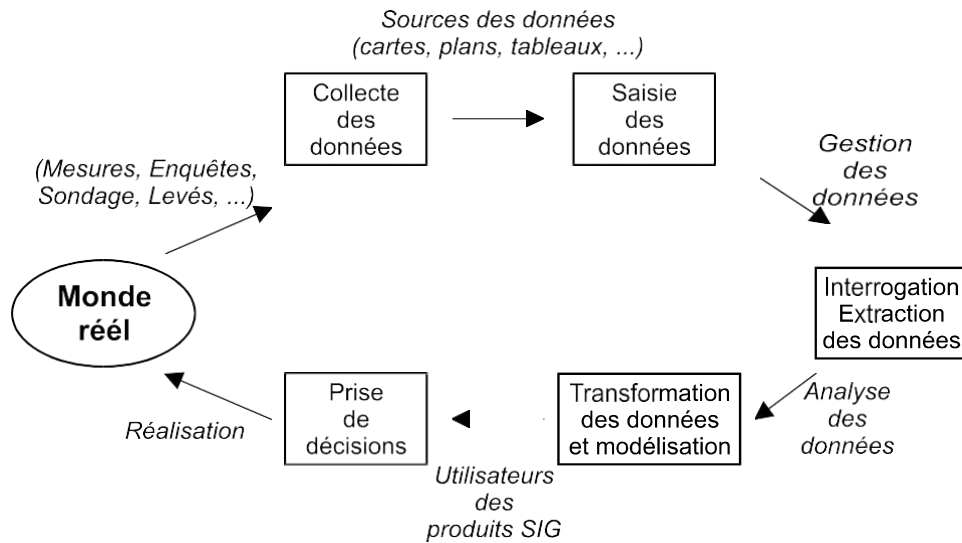


Figure I.5. SIG comme un modèle du monde réel

Domaines d'application des SIG

Les SIG peuvent répondre aux attentes d'utilisateurs relevant de divers domaines allant de la gestion des ressources naturelles à la prise de décisions socio-économiques ou politiques.

- **Génération et gestion de données à référence spatiale**

- Agence cartographique, cadastre, fournisseur de données géographiques publics ou privés
- Données en mode vectoriel ou image, avec attributs limités
- Traitements généralement simples
- Importance de la visualisation cartographique

- **Gestion de réseaux de distribution**

- Distribution d'eau, d'énergie, communication, égouts
- Secteur public, semi-public (régies, intercommunales), grands concessionnaires privés
- Donnée à grande échelle et de précision, en mode vectoriel
- Traitements spécialisés
- Visualisation graphique simple

- **Gestion de réseaux de transport**

- Transports publics routiers ou ferroviaires, gestion de flotte de grands transporteurs routiers privés, schémas de circulation urbains, etc.

- Données vectorielles à moyenne ou petite échelle, de précision limitée
- Traitements spécialisés (trafic engineering, accessibilité, etc.), parfois en mode image
- Visualisation graphique simple (graphes, réseaux)

- **Gestion du territoire (aménagement, environnement, etc.)**
 - Ministères, associations diverses (parcs naturels, sociétés d'industrialisation, etc.), bureaux d'étude privés (études d'impact, ...)
 - Très grande diversité de données, de grande à petite échelle, en mode vecteur et éventuellement en mode image
 - Traitement spécialisés, voire exploratoires
 - Visualisation cartographique complexe

- **Gestion de ressources naturelles**
 - Foresterie, exploration minière et pétrolière, etc.
 - Données à moyenne et petite échelles, en mode image et souvent en mode vecteur, de précision limitée mais d'une grande diversité
 - Traitements spécialisés et exploratoires (simulations, etc.)
 - Visualisation complexe (géo-référenciation délicate, spatio-cartes, ...)

- **Géo-marketing**
 - Grandes sociétés de distribution, banques (réseau d'agences), gestion de parcs immobiliers et fonciers, etc.
 - Données à grande et moyennes échelles de précision limitée, en mode vectoriel, avec de nombreux attributs socio-économiques
 - Géométriéférenciation fréquente par adresse postale
 - Traitements spécialisés et surtout exploratoires
 - Visualisation cartographique simple (cartes statistiques, etc.)

- **Enseignement, recherche**
 - SIG n'est pas centré sur l'organisation, il ne doit pas informatiser les données et les tâches quotidiennes de l'organisation (école, laboratoire)

- SIG-logiciel utilisé comme une boîte à outils (tout comme un logiciel de traitement statistique ou de traitement de texte)
- Critères de sélection basés sur:
 - . Les possibilités de traitements, sur tous les types de données
 - . Les fonctionnalités de visualisation cartographique
 - . La facilité d'apprentissage (didacticiels)
 - . Le prix (shareware, versions dégradées pour l'enseignement, etc.)
 - . Les possibilités de développement (recherche)
 - . Les standards du marché (initiation des étudiants)

Fonction	Type de problèmes à résoudre
Localisation	- Qu'il y a-t-il à tel endroit ? - Où est localisé tel phénomène ?
Conditions	- Où est localisé tel phénomènesoutelles contraintes ?
Tendance	- Comment tel phénomène a-t-il évolué ? - Comment a évolué tel endroit ?
Accessibilité	- Quel est le chemin le plus intéressant pour joindretel ou tel endroit ?
Comparaison	- Tel phénomène est-il corrélé à l'autre ? - Tel phénomène explique-t-il la répartition de tel autre ?
Modélisation	- Que se passe-t-il en tel endroit si tel ou tel phénomène est modifié de telle ou telle façon ?

II.1. INTRODUCTION

La représentation des données géographiques peut être réalisée suivant l'un des deux modes fondamentaux: un mode **vectorel** et un mode **raster** (maillé, matriciel ou cellulaire).

Pour le mode vectorel, l'unité logique de base dans un contexte géographique correspond à une ligne (courbe de niveau, rivière, limite d'une zone, etc.) ou à un segment de ligne (ou vecteur). Une série de localisations ponctuelles (x, y) sont enregistrées le long de la ligne. Les points peuvent être considérés, dans une telle organisation de données, comme des lignes de longueur nulle (une seule paire x,y).

Pour le mode raster, l'unité logique de base est une cellule (ou "pixel", picture element) qui est une unité de surface. La zone étudiée se trouve ainsi subdivisée en pixels de même et de même dimension.

II.2. MODE VECTORIEL

Pour décrire l'information géographique localisée, on utilise généralement 3 niveaux d'informations:

- le niveau *géométrique*, qui décrit la géométrie des objets géographiques
- le niveau *sémantique*, qui décrit la nature des objets géographiques
- le niveau *topologique*, qui décrit les relations topologiques (connexité ...) entre les objets géométriques.

II.2.1 .Les données géométriques

Les données géométriques sont composées de points, de lignes et de polygones.

Une ligne peut être décrite par

- un ensemble de segments (un segment est une droite composée de 2 points)
- un ensemble de courbes plus complexes (cubiques ...)
- un ensemble de segments et de courbes

Un polygone est constitué par un ensemble de lignes fermées. Tout polygone peut être "troué" par d'autres polygones (une île sur un lac, une cour à l'intérieure d'une maison, ...).

II.2.2. Les données sémantiques

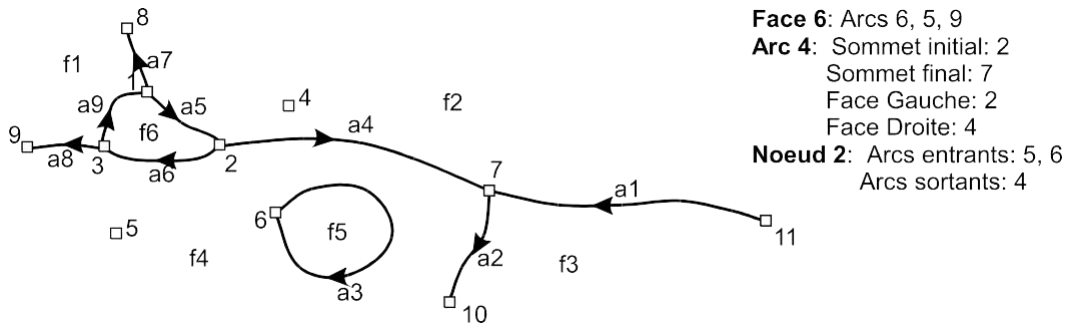
Les données sémantiques sont la description non géométrique des données géographiques. Elles peuvent représenter des entités simples (bâtiment, portion de route, carrefour) ou des entités plus complexes (ensemble de bâtiments, route, échangeur ...). On pourra associer à chaque objet des informations permettant de les décrire: largeur d'une route, nom, numéro, ...

II.2.3. Les données topologiques

On peut représenter des données géographiques sous la forme d'un graphe. Un graphe est composé d'arcs et de noeuds:

- Un **arc** est compris entre un sommet initial et un sommet final.
- Un **noeud** possède des arcs entrants et des arcs sortants. Les arcs entrant sont orientés vers le noeud, donc ce noeud est leur sommet final. Les arcs sortants partent de ce noeud, qui est leur sommet initial
- Il y a un noeud à chaque intersection d'arcs.

Les relations de connexité dans un graphe sont appelées **relations topologiques**. Lorsqu'on présente des données planaires on parle de graphe planaire auquel on peut associer la notion de **face**. Ainsi, un arc possède une face gauche et une face droite.



Les structures permettant d'organiser les données vectorielles peuvent être regroupées, selon qu'elles tiennent compte ou non des relations topologiques qui régissent les objets géographiques, en deux catégories: les structures topologiques et les structures non topologiques

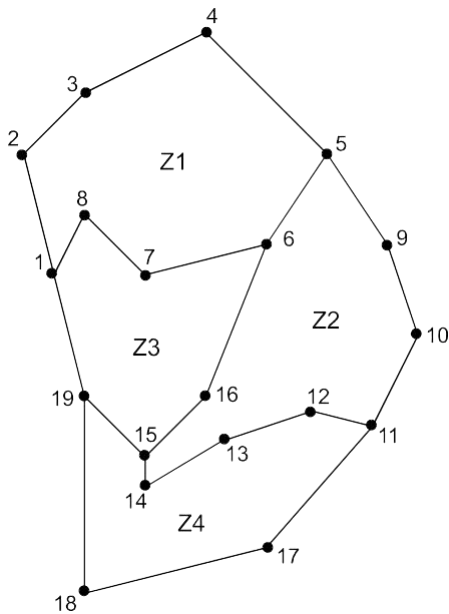
II.3. STRUCTURES NON TOPOLOGIQUES

II.3.1. Structure séquentielle simple

Description

Dans le fichier constituant la base de données spatiales, chaque entité (polygone) est décrite, indépendamment des autres, par une série de paires de coordonnées. La fermeture des zones est explicite. Le séparateur de zones correspond à un couple de coordonnées de valeur particulière (-1, par exemple). La fermeture peut être

- Explicite: le point initial est répété à la fin de la zone dans le fichier
- Ou réalisée par voie logicielle.



X	Y	X	Y	X	Y
x1	y1	x12	y12	x19	y19
x2	y2	x13	y13	x1	y1
x3	y3	x14	y14	-1	-1
x4	y4	x15	y15	x19	y19
x5	y5	x16	y16	x15	y15
x6	y6	x6	y6	x14	y14
x7	y7	x5	y5	x13	y13
x8	y8	-1	-1	x12	y12
x1	y1	x1	y1	x11	y11
-1	-1	x8	y8	x17	y17
x5	y5	x7	y7	x18	y18
x9	y9	x6	y6	x19	y19
x10	y10	x16	y16	-1	-1
x11	y11	x15	y15		

Fichier des zones:

Les Avantages

- Simplicité (préparation du fond réduite au minimum).
- Les points constituant une zone sont accessibles en bloc et d'une manière séquentielle. Ceci facilite l'écriture des algorithmes tel que la fusion de deux polygones ou de traitements cartographiques par plages tel que l'hachurage.

Les inconvénients

- Redondance de l'information (presque la totalité des points sont enregistrés plus d'une fois)
- © Exigences importantes au niveau de la capacité de stockage.
- double saisie du même point

- ⊙ Risque d'une superposition inexacte lors du tracé du fond ou de l'hachurage des zones.
- Temps De Saisie Et De Trace Important
- Mise à jour difficile
- Toute modification ultérieure ⊙ une réorganisation généralement complexe du fichier.

Exemples d'utilisation

Logiciel SYMAP (1^{ère} version, années 60, laboratoire Harvard).

La structure séquentielle simple ne constitue pas une structure de travail C'est une structure de transfert de données entre logiciels (DXF, BNA, VEC, etc.).

II.3.2. Dictionnaire de points

Description

l'inconvénient majeur de la structure séquentielle simple provient de la double opération desaisie. La structure utilisant un dictionnaire de points constitue une solution intéressante à ce problème. Il ne s'agit cependant que d'une amélioration de la première structure.

- le 1^{er} fichier (dictionnaire) ne contient aucune redondance.
- le 2^{ème} fichier (pointeurs) comporte une forte redondance d'informations, mais cette fois, elle porte sur des étiquettes.
- le second fichier est indépendant de la numérisation proprement dite
- la fermeture des zones est explicite et le séparateur correspond à un pointeur de valeur négative.

Avantages

- saisie unique de tous les points du fond de carte (⊙ pas de problèmes de superposition)
 - réduction du temps de saisie
 - l'ordre des points du dictionnaire n'a pas d'importance
- ⊙ créer plusieurs bases de données sur base d'un seul dictionnaire (provinces, arrondissements, communes) ⊙ dictionnaire au niveau le plus fin.

Inconvénients

- Préparation du fond est un peu plus lourde.
- Si la saisie est unique, le tracé est toujours doublé: qualité du dessin ⊙, temps du tracé ⊙

- Deux types d'accès (double lecture pour la reconstruction des zones): le 1^{er} séquentiel, le 2^{ème} est direct (incompatibilité avec certains périphériques de mémoire externe).

Exemple d'utilisation

la structure utilisant un dictionnaire de points a été utilisée par plusieurs programmes de cartographie. Le plus célèbre est le programme CALFORM, postérieur à SYMAP distribué par le laboratoire Harvard.

II .3. STRUCTURES TOPOLOGIQUES

Des structures plus performantes sont qualifiées de topologiques car elles assimilent le fond de carte à un graphe et elles retiennent et utilisent un certain nombre de règles topologiques qui s'appliquent aux éléments du graphe.

Les points du fond de carte sont les sommets, ou les noeuds, du graphe. Les segments reliant deux sommets sont les arcs ou les arrêtes. Enfin, les zones sont constituées d'une séquence d'arrêtes et apparaissent, par conséquent, comme des circuits dans le graphe.

La distinction des différents éléments du graphe conduit, dans la pratique, à la création d'autant de fichiers séparés. Les relations topologiques entre les éléments se traduisent quant à elles, par une série de pointeurs entre les fichiers. En fait, seul le fichier stockant les sommets est explicitement lié à l'opération de numérisation, les autres fichiers n'étant constitués que de pointeurs.

II.3.1. La structure DIME

Le concept DIME (Dual Indépendant Map Encoding), développé par le "U.S. office of the census", est à la base d'un des premiers et plus célèbres systèmes opérationnels faisant référence à une structure topologique de données spatiales. Il a été conçu pour permettre l'édition et la gestion des bases de données décrivant les réseaux urbains de voirie et les unités statistiques intra-urbaines utilisés pour les recensements américains dès 1970.

Description

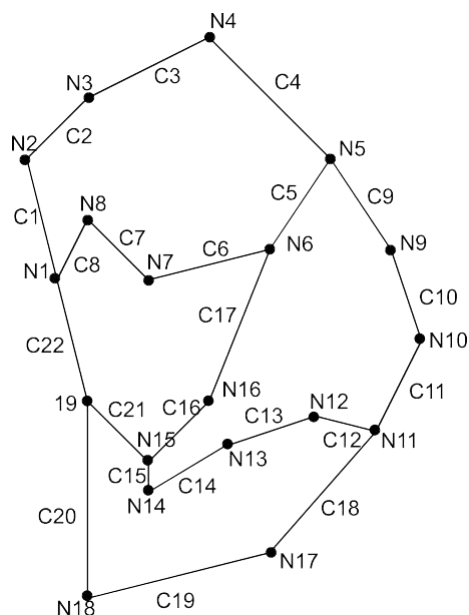
L'élément charnière de cette structure est l'arrête qui correspond le plus souvent à une face d'un

îlot urbain. Les lignes complexes (rues courbes, lignes de chemin de fer, etc) sont représentées par une série de segments successifs. La disposition des arrêtes détermine des blocs, ou îlots, pour lesquels les données du recensement sont agrégées.

Fichier des noeuds

Etiquette	X	Y
N1	X _{N1}	Y _{N1}
N2	X _{N2}	Y _{N2}
N19	X _{N19}	Y _{N19}

N3	X _{N3}	Y _{N3}
...
N17	X _{N17}	Y _{N17}
N18	X _{N18}	Y _{N18}



Fichier des chaînes

Etiquette	Varianles Urbanist.	Noeud de Départ	Noeud de Destinat..	Zone gauche	Zone droite
C1	...	N1	N2	0	Z1
...
C5	...	N5	N6	Z2	Z1
...
C20	...	N18	N19	0	Z4

C21	...	N19	N15	Z3	Z4
C22	...	N19	N1	0	Z3

La base de données correspondante se résume à un fichier de coordonnées, mémorisant les sommets du graphe, et à un fichier d'arrêtes. Chaque enregistrement de ce dernier fichier correspond à une arrête et comprend, tout d'abord, plusieurs variables urbanistiques telle que le nom de la rue, pour le tronçon correspondant à l'arrête. Ensuite, des variables de type pointeur: deux pointeurs vers les deux sommets fixant les extrémités de l'arrête et deux codes identifiant les deux blocs de part et d'autre de l'arrête.

Avantages

Cette structure permet une édition aisée de la base de données et elle est bien adaptée au traitement et à la mise à jour des données de recensement.

Inconvénients

Pour des traitements cartographiques diversifiés, la structure DIME est trop rigide. La reconstitution de la frontière des blocs, la reconnaissance des éléments linéaires simples, etc. demandent de nombreuses manipulations et un temps assez considérable. En outre, certains types de traitements, tels que la généralisation des lignes par exemple, sont pratiquement à exclure.

II.3.2. La structure POLYVRT

Le système intitulé POLYVRT (POLYgon conVeRTer) a été développé par le laboratoire Harvard mais, en fait, il synthétise plusieurs travaux originaux préalables. Il utilise une distinction poussée des éléments spatiaux constituant le fond de la carte et a servi de référence à plusieurs autres systèmes développés par la suite.

Description

L'élément de base de cette structure reste linéaire. Il ne s'agit pas d'un simple segment mais d'une "**chaîne**". Les chaînes peuvent être constituées de plusieurs segments successifs. Une chaîne peut donc contenir plusieurs points intermédiaires qui doivent être distingués des **noeuds** fixant ses

extrémités. Un point intermédiaire n'appartient qu'à une seule chaîne, tandis qu'un noeud est généralement partagé par plusieurs chaînes. Enfin, une chaîne ne peut appartenir, au plus, qu'à deux zones adjacentes. Les zones sont définies par une séquence de chaînes. A la limite, une seule chaîne peut constituer la frontière, dans le cas d'une île par exemple (ici, en outre, les deux noeuds fixant les extrémités de la chaîne coïncident)

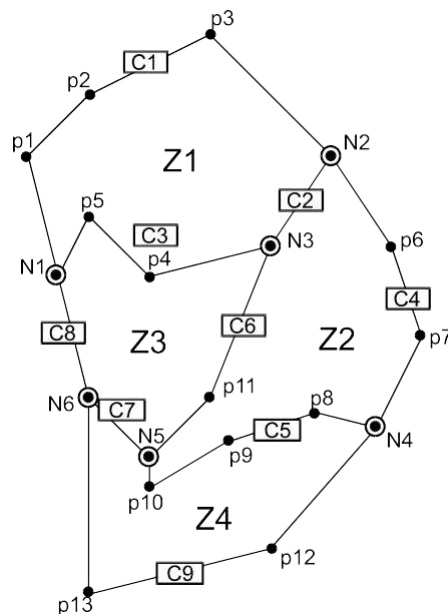
Le fichier de chaînes contient des enregistrements constitués uniquement de pointeurs:

- Vers les noeuds situés aux extrémités,
- Vers les éventuels points intermédiaires,
- Vers les deux zones situés de part et d'autre de la chaîne.

Le fichier de zones comprend, pour chaque zone, les pointeurs vers les chaînes nécessaires à la reconstitution des frontières.

Fichier des zones

Etiquette	Liste de pointeurs de chaînes
Z1	C1 C2 C3
Z2	C4 C5 C6 C2
Z3	C3 C6 C7 C8
Z4	C9 C7 C5



Fichier des chaînes

Etiquette	Nombre de points	Noeud de départ	Noeud de destination.	Zone gauche	Zone droite
C1	3	N1	N2	0	Z1
C2	0	N2	N3	Z2	Z1
C3	2	N3	N1	Z3	Z1
C4	2	N2	N4	0	Z2
C5	3	N4	N5	Z4	Z2

C6	1	N5	N3	Z3	Z2
C7	0	N5	N6	Z4	Z3
C8	0	N6	N1	0	Z3
C9	2	N4	N6	0	Z4

Avantages

Ce système est beaucoup plus riche que tous les précédents et il permet d'envisager tous les traitements cartographiques classiques. Toutes les corrections et mises à jour de la banque de données spatiales sont réalisées très aisément. On peut éviter le doublement du tracé, inévitable avec les deux premières structures, en utilisant le seul fichier de chaînes comme "input" de la procédure de tracé.

la structure supporte simultanément tous les types d'éléments spatiaux puisqu'à côté des zones, utilisant toute la structure, il est possible de manipuler des éléments ponctuels et linéaires simples, n'utilisant qu'une partie de la structure POLYVRT.

Inconvénients

Ce système possède plusieurs inconvénients mais la plupart sont inhérents à toute structure topologique. C'est ainsi que la préparation du fond de carte, précédant la numérisation, est particulièrement lourde. Il faut distinguer et étiqueter les noeuds, les chaînes et les zones et il faut marquer les points intermédiaires. D'autres part, la structure implique un accès aléatoire à tous les fichiers constituant la base de données. L'accès aux enregistrements étant effectué par l'intermédiaire de plusieurs pointeurs, le temps nécessaire à la lecture d'information, pour la reconstitution d'une frontière par ex., peut devenir non négligeable.

La structure POLYVRT de base a été améliorée, par la suite, pour supporter des entités géographiques particulières tels que les îles et les lacs. Dans ces deux derniers cas, la solution constitue d'introduire un nouveau type de chaîne: chaîne invisible.

II.4.1. Structures particulières spécifiques à l'information altimétrique

II. 4.1.1. Structure en Réseau Triangulaire Irrégulier

La structure en Réseau Triangulaire Irrégulier (ou RTI) (en anglais: TIN pour Triangulated Irregular

Network) a été développée au début des années 70. A l'origine, il s'agissait simplement de construire une surface à partir d'un ensemble de points répartis de façon irrégulière. Les systèmes commerciaux utilisant les RTI ont commencé à faire leur apparition dans les années 80, sous la forme de fonctions de construction d'isophyses intégrées dans les logiciels de SIG ou de cartographie numérique.

Description

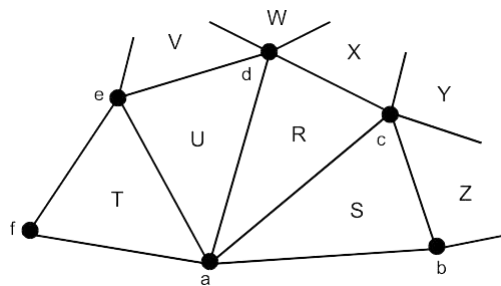
Deux méthodes peuvent être considérées pour encoder des réseaux triangulaires: 1) triangle par triangle 2) en considérant les sommets par rapport à leurs voisins.

Description par triangles

Dans ce cas, un enregistrement contient :

- Un numéro de référence pour le triangle ;
- Les coordonnées x, y, z des trois sommets;
- Les numéros de référence des triangles adjacents.

Un sommet peut appartenir, en moyenne, à six triangles. Ces coordonnées tridimensionnelles seront donc répétées six fois. Cette redondance peut être évitée en créant un fichier de sommets distinct auquel on fait référence dans le fichier des triangles.



Triangle	Sommets	Triangles adjacents
R	$a(x,y,z), c(x,y,z), d(x,y,z)$	S, X, U
S	$a(x,y,z), c(x,y,z), b(x,y,z)$	Z, R
etc.		

Description par sommets

L'alternative consiste à enregistrer pour chaque sommet les informations suivantes:

- un numéro d'identification;

- Les coordonnées x, y, z ;
- Les références (pointeurs) aux sommets voisins.

Cette structure constituait la structure originelle des RTI développée par Peucker et ses collaborateurs en 1978

Sommet	Coordonnées	Sommets connectés
a	x, y, z	f, e, d, c, b, ...
b	x, y, z	a, c, ...
c	x, y, z	a, d, b, ...
etc.		

Le choix de la méthode à utiliser dépend principalement de l'usage recherché, par exemple:

- les analyses de pente utilisent la première méthode;
- les opérations d'interpolation des courbes de niveau utilisent plutôt la seconde structure.

II.4.1.2. Structure topologique pour les courbes de niveau

Structure développée afin de répondre aux besoins de traitements complexes de l'information altimétrique, telle que l'extraction des lignes caractéristiques de la morphologie du terrain, le traitement des courbes de niveau extrêmes dépourvues de points cotés, etc. Une telle structure assure, d'une part, une organisation efficace des données de base et, d'autre part, la disponibilité d'informations concernant le voisinage dans le sens altimétrique permettant, par exemple, l'accès direct à la courbe de niveau immédiatement inférieure ou immédiatement supérieure.

La structure présentée ici a été développée au sein de la section de topographie de l'I.A.V. Hassan II. La description complète de cette structure est réalisée à l'aide de quatre fichiers: un fichier des entités (courbes de niveau et points cotés), un fichier de coordonnées, un fichier relatif au voisinage inférieur et un fichier relatif au voisinage supérieur. La liaison entre ces fichiers permet de décrire explicitement la topologie régissant les différentes entités altimétriques.

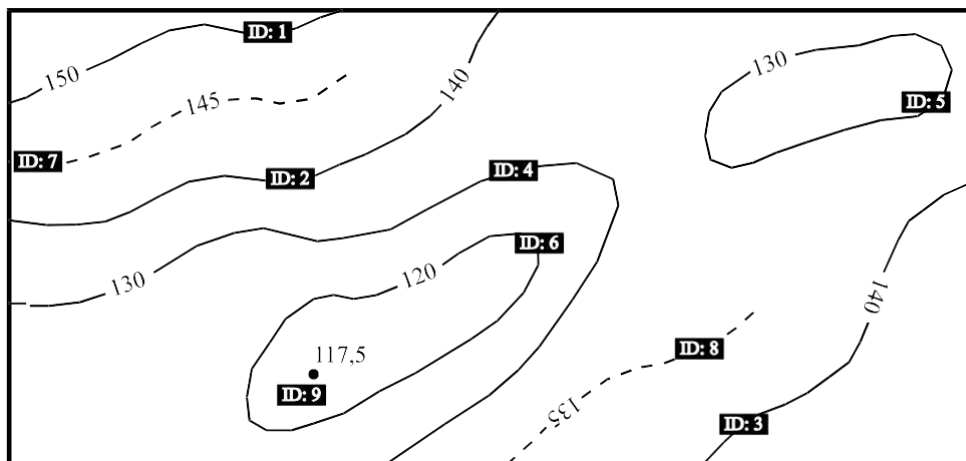
Le fichier des entités est le fichier principal de la structure. Il fournit, d'une part, les caractéristiques de chacune des entités altimétriques et, d'autre part, les éléments de liaison, ou pointeurs, permettant de connecter le fichier des entités aux autres fichiers de la structure. Ce fichier est

constitué d'une succession d'enregistrements dont chacun se rapporte à une entité altimétrique. Le contenu explicite d'un enregistrement du fichier des entités se présente comme suit:

- **Identifiant (ID)**: c'est un numéro unique permettant d'identifier chacune des entités altimétriques et, par la suite, d'accéder aux différents renseignements qui la caractérisent.
- **Altitude (ALT)**: c'est l'altitude lue sur la carte concernant l'entité altimétrique en question.
- **Nombre de points (NP)**: c'est le nombre de points qui constituent l'entité considérée. Dans le cas d'un point coté, ce nombre est égal à 1.
- **Adresse du point de départ (APD)**: c'est un pointeur vers le fichier des coordonnées. Il permet d'accéder aux coordonnées du premier point de l'entité altimétrique considérée et, grâce à l'information contenue dans *NP*, à celles de tous les autres points constituant cette entité.
- **Forme (FOR)**: ce paramètre renseigne la forme ou l'allure d'une courbe de niveau et a pour valeur "F" ou "O" selon que la courbe de niveau considérée est fermée (les coordonnées des premier et dernier points de la courbe coïncident) ou ouverte. Dans le cas d'un point coté, ce paramètre a pour valeur "P".
- **Continuité (CON)**: ce paramètre permet de savoir si l'entité considérée est continue ("C") ou interrompue ("I"). Evidemment, ce paramètre ne concerne que les courbes de niveau.
- **Longueur (LON)**: c'est la longueur en mètres des entités courbes de niveau. Ce paramètre est utilisé, d'une part, dans la détermination du voisinage et, d'autre part, dans le calcul d'autres paramètres tels que la pente générale du terrain (cf. chapitre 8). Il a pour valeur zéro dans le cas d'un point coté.
- **Adresse du voisinage supérieur (AVS)**: il s'agit d'un pointeur vers le fichier du voisinage supérieur. Il permet d'accéder directement aux entités constituant le voisinage supérieur d'une courbe de niveau ou d'un point coté.
- **Nombre d'entités voisines supérieures (NVS)**: c'est le nombre d'entités constituant le voisinage supérieur d'une courbe de niveau ou d'un point coté.
- **Adresse du voisinage inférieur (AVI)**: il s'agit d'un pointeur vers le fichier du voisinage

inférieur. Il permet d'accéder directement aux entités constituant le voisinage inférieur d'une courbe de niveau ou d'un point coté.

- *Nombre d'entités voisines inférieures (NVI)*: c'est le nombre d'entités constituant le voisinage inférieur d'une courbe de niveau ou d'un point coté



Extrait d'une carte topographique 1/50.000 (équidistance: 10 m)

Contenu du fichier des entités de la structure topologique proposée

ID	ALT	NP	APD	LON	FOR	CON	NVI	AVI	NVS	AVS
1	150,0	8	1	2749,3	O	C	2	1	0	0
2	140,0	15	9	4754,0	O	C	2	2	2	1
3	140,0	9	24	2862,7	O	C	3	3	0	0
4	130,0	21	33	6893,6	O	C	1	4	3	2
5	130,0	12	54	3505,2	F	C	0	0	3	3
6	120,0	13	66	4852,2	F	C	1	5	1	4
7	145,0	7	79	2563,4	O	I	1	6	1	5
8	135,0	7	86	2571,6	O	I	2	7	2	6
9	117,5	1	93	0,0	P	P	0	0	1	7

Fichier du voisinage inférieur

Adresse	ID des voisins inférieurs
1	7, 2
2	4, 5
3	4, 5, 8
4	6

Fichier du voisinage supérieur

Adresse	ID des voisins supérieurs
1	1, 7
2	2, 3, 8
3	2, 3, 8
4	4

5	9		5	1
6	2		6	2, 3
7	4, 5		7	6

Fichier des coordonnées des points constituant chacun des éléments du fichier des entités

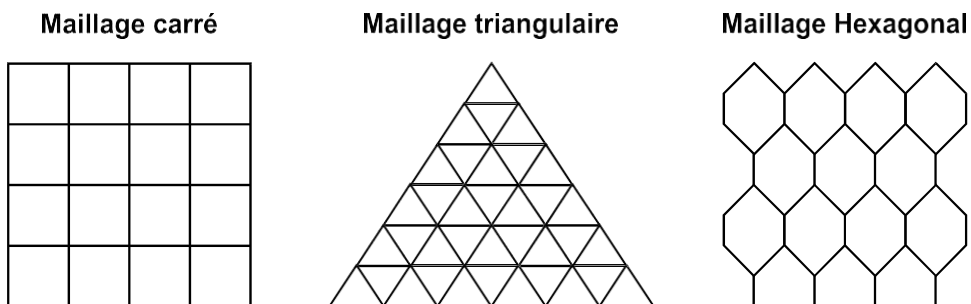
Adresse	X	Y	Adresse	X	Y
1	x1 ₁	y1 ₁	-	-	-
2	x1 ₂	y1 ₂	-	-	-
3	x1 ₃	y1 ₃	88	x8 ₃	y8 ₃
4	x1 ₄	y1 ₄	89	x8 ₄	y8 ₄
5	x1 ₅	y1 ₅	90	x8 ₅	y8 ₅
6	x1 ₆	y1 ₆	91	x8 ₆	y8 ₆
-	-	-	92	x8 ₇	y8 ₇
-	-	-	93	x9	y9

II.2. LES STRUCTURES DITES "RASTER"

II.2.1. Concepts de base

Différents types de maillage:

Les structures des données raster consistent en une subdivision de la zone étudiée selon un maillage régulier, dont l'élément de base est une unité de surface. Trois types de maillage sont généralement considérés: le maillage carré ou rectangulaire, le maillage triangulaire et le maillage hexagonal.



A. Maillage hexagonal

L'avantage principal du maillage hexagonal régulier vient du fait que toutes les cellules voisines d'une cellule donnée sont équidistantes à partir du centre de cette cellule. Ceci offre

une facilité pour les fonctions de recherche et d'extraction ce qui n'est pas le cas pour le maillage carré où, les voisins selon la diagonale ne sont pas à la même distance, par rapport au centre de la cellule en question, que les voisins selon les quatre directions cardinales.

B. Maillage triangulaire

La caractéristique commune pour tous les maillages triangulaires, réguliers ou irréguliers, est que les triangles n'ont pas tous la même orientation. Ceci fait que plusieurs procédures, y compris la comparaison unicellulaire, qui sont simples à réaliser avec les deux autres types de maillage sont plus complexes. Toutefois, cette même caractéristique procure au maillage triangulaire l'unique avantage de la représentation du relief.

C. Maillage carrée

C'est le type de maillage le plus utilisé pour deux raisons:

- manipulation facile par les ordinateurs
- compatibilité avec les dispositifs d'acquisition et de représentation de données spatiales. Chaque pixel de la grille est référencé par un numéro d'ordre de la ligne et de la colonne auxquelles il appartient au sein de la matrice. En plus, il contient un nombre qui représente la valeur de l'attribut ou un pointeur vers une table d'attributs.

Les limitations de la grille régulière sont principalement dues aux larges volumes de données et par conséquent à la capacité mémoire exigée. Cependant, quand les données sont constituées principalement d'entités zonales, les exigences de stockage peuvent être considérablement réduites à l'aide des techniques de compression.

III.5.1.1. Représentation des éléments géographiques

Dans les structures raster, un point est représenté par un seul pixel ou une seule cellule de la grille, une ligne par un nombre de pixels adjacents suivant une direction, et une zone par une agglomération de pixels voisins

Notion de résolution

La représentation raster assume que l'espace géographique peut être considéré comme une surface

cartésienne plane. Ainsi, chaque pixel se trouve associé avec une parcelle de terrain de forme carrée. La résolution, ou l'échelle des données, est de ce fait, la relation entre la taille du pixel dans la base de données et ses dimensions à la surface de la terre.

Du fait que chaque pixel dans la matrice bidimensionnelle ne peut supporter qu'une valeur, les différents attributs ou variables géographiques doivent être représentés séparément dans des couches différentes. Ceci résulte en une structure tridimensionnelle où chaque point de la carte est décrit par une rangée verticale dans laquelle chaque position supporte la valeur de l'attribut associé à une couche donnée.

III.5.1.2. Fichier image

- Liste de nombre (valeurs)
 - 1 nombre / pixel
 - Structure séquentielle (en général, voir plus loin)
 - depuis le pixel supérieur gauche
 - ligne par ligne dans l'image
- Enregistrement des nombres:
 - En binaire
 - entiers positifs [0 à 225]: 1 byte / pixel
 - entiers positifs et négatifs [-32768 à 32767]: 2 bytes / pixel
 - réels 7-8 décimales [$1 \cdot 10^{-45}$ à $1.7 \cdot 10^{38}$]: 4 bytes / pixel
 - réels en double précision, 15-16 décimales [$5.0 \cdot 10^{-324}$ à $1.7 \cdot 10^{308}$]: 8 bytes / pixel
 - En ASCII:
 - 1 byte / chiffre + 1 byte séparateur entre 2 nombres (non utilisé en pratique)
- Inexploitable sans une documentation:
 - incrustée dans le fichier image
 - enregistrée dans un fichier séparé ("transparent" pour l'utilisateur)

III.5.1.3. Fichier de documentation

- **Fichier en tête ou header:**
- **Paramètres généraux:**
 - Instructions de lecture de la table:

- Nombre de lignes et de colonnes
- Type d'enregistrement des nombres (byte, entier, réel,)
- Statistiques de l'image (option):
 - ex. fréquence des valeurs présentes dans l'image
 - Table des couleurs (RGB) associée à l'image (option)
- **Paramètres propres aux images géographiques:**
 - relatifs au géocodage:
 - surface de référence (plan, sphère, ellipsoïde)
 - coordonnées extrêmes de l'image et unité des coordonnées
 - résolution des pixels en unités de coordonnées
 - Relatifs aux attributs:
 - unité des valeurs de pixels
 - légende éventuelle associée aux valeurs présentes dans l'image

III.5.2. Structuration des données raster

Il y a plusieurs façons de structurer des données raster. Certaines de ces méthodes sont employées pour un pur compactage de l'information alors que d'autres créent une structuration complète des données. Parmi les plus utilisées:

III.5.2.1. Bitmap

- Structure séquentielle simple:
 - ligne par ligne, du pixel supérieur gauche au pixel inférieur droit
- Taille de l'image sauvegardée
 - $L \times C \times T$

L= nombre de lignes (ou rangées) C= nombre de colonnes

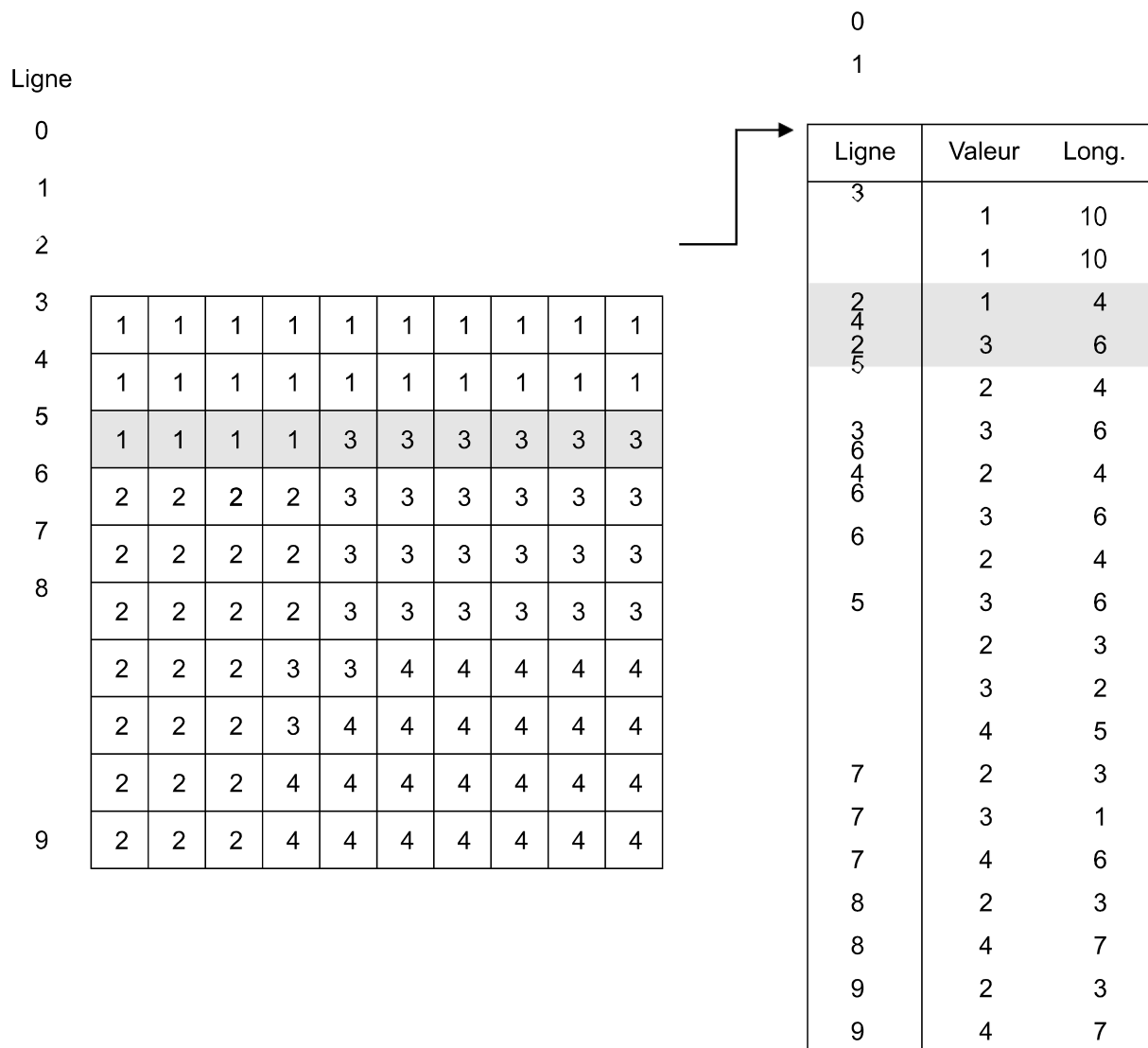
T= nombre de bytes selon le type des valeurs des pixels (byte, entier, réel, ...)

- Inconvénients:
 - Nombreuses répétitions de valeurs successives:
 - cas des masques binaires (0, 1), des images codées en bytes et en entiers
 - Accès lent aux différentes parties de l'image:

- exploitation non optimale de la mémoire tampon à la lecture et à l'affichage de grandes images

III.5.2.2. Run Length Coding (RLC)

- *Run Length Encoding ou coding*
- caractéristiques :
 - Format compressé lors de la sauvegarde du fichier image
 - Exploitation des séquences de valeurs identiques dans l'image
 - Codage de chaque ligne de l'image par une séquence de doublets:
 - Valeur du pixel
 - Nombre d'occurrences successives de cette valeur dans la ligne (fréquence)
 - Applicable aux seules images codées en byte ou en entiers
- Taille de l'image sauvegardée:
 - Coefficient de compression variable:
 - Efficace si fréquence moyenne ≈ 2 pixels
 - Au pire: 200% de la taille bitmap (toutes valeurs successives différentes)
- Inconvénients:
 - Opérations de compression / décompression :
 - Temps de lecture / écriture plus longs
 - Plusieurs variantes, peu documentées



III.5.2.3. Blocs carrés

- Généralisation du principe valeur-fréquence du RLC mais à deux dimensions
- Caractéristiques:
 - Repérage dans l'image des blocs carrés de pixels présentant la même valeur
 - plus petit bloc = 1 pixel
 - Codage de chaque bloc:
 - coordonnées image (ligne et colonne) du premier pixel du bloc
 - longueur (en pixels) du bloc
 - valeur uniforme des pixels du bloc

-
- Applicable aux seules images codées en bytes ou en entiers
- Efficacité de compactage: si la longueur moyenne des blocs ≈ 2 pixels
- Inconvénients:
 - Opérations de compression / décompression
 - temps de compression généralement lent car nombreuses itérations pour identifier les plus grands blocs possibles
 - Structure rarement supportée par les logiciels

III.5.2.4. Tétrarbres

- Structure hiérarchique en arbres quaternaires ou quadrees
- Caractéristiques:
 - Subdivisions successives de l'image en 2^n parties, jusqu'à ce que celles-ci soient homogènes (tous pixels de même valeur)
 - Codage de chaque partie de l'image par un doublet:
 - Ordre de la branche de l'arbre au niveau duquel se situe la partie d'image
 - Valeur unique présentée par les pixels de la partie d'image
 - Applicable, sans perte d'information, aux images en bytes ou en entiers
- **Avantages:**
 - Structure de travail et non seulement une technique de compression
 - Taux de compression variable mais souvent très efficace
 - Possibilité de "généraliser" une image en acceptant comme "homogène" un intervalle de valeurs (aussi applicable à des images codées en réels)
 - Les "quadrees" sont des matrices à résolution variable: le détail est considéré seulement quand il existe dispensant ainsi des exigences excessives de stockage pour les parties de l'image ou le détail est absent.
- **Inconvénients:**
 - Temps de compression / décompression (lecture / écriture)

III.5.2.5. Clés de Peano

- Courbe de remplissage

- Projection bijective d'un espace à n dimensions sur un espace à 1 dim
 - Application à une image: espace à 2 dimensions, où les points = les pixels
 - la courbe passe une fois seulement dans chaque point (pixel)
 - 2 points voisins dans l'espace (image) sont voisins sur la courbe et inversement
 - Il doit être facile de retrouver les voisins d'un point quelconque
 - la courbe doit être stable, même quand l'espace devient grand ou infini
 - la courbe devrait être utilisable dans un espace de résolution variable (ex. quadtree)
- Peano (1890): courbe d'ordre N
 - les points voisins de l'espace ne sont pas toujours voisins sur la courbe
 - les points de la courbe sont définis par une coordonnée: la clé
 - Construction de la clé:
 - coordonnées image (entières) en mode binaire
 - entrelacement des bits des coordonnées

III.5.3. Autres types de courbes: Peano-Hilbert (⊗) Comparaison Vecteur / Raster

Les deux types de structures ont des avantages et des limites. Généralement, la structure vectorielle est plus orientée vers des utilisations qui nécessitent une localisation spatiale précise comme la cartographie topographique et le cadastre. La structure raster est préférée lorsqu'on privilège une information plus détaillée sur les attributs et leur distribution spatiale. La structure raster sera également utilisée lorsque des combinaisons complexes d'attributs sont requises (modélisation et analyse spatiale). Burrough (1986) résume les avantages et les Inconvénients relatifs aux deux types de structures comme suit:

III.5.3.1. Les méthodes vecteur

Avantages:

- Bonne représentation des données phénoménologiques
- Structures compactes de données
- La topologie peut être complètement décrite par un réseau de liens
- Graphiques précis
- L'édition, la mise à jour et la généralisation des graphiques et des attributs sont possibles

Inconvénients

- Structures complexes
- La combinaison de plusieurs cartes vecteur ou la combinaison de ces dernières avec des cartes raster crée des difficultés
- La simulation est difficile
- La représentation graphique peut être coûteuse, particulièrement pour des couleurs et hachures de haute qualité
- La technologie est chère surtout le matériel et les logiciels sophistiqués

III.5.3.2. Les méthodes raster

Avantages

- Simplicité des structures de
- Le processus de superposition ("Overlay") et la combinaison des données cartographiques avec les données de télédétection sont faciles. Facilité des différents types d'analyse spatiale
- La simulation est facile puisque chaque unité spatiale a la même forme et les mêmes dimensions
- La technologie est moins chère et connaît un développement rapide

Inconvénients

- Le volume des données
- La réduction de ce volume en utilisant des cellules plus larges peut conduire à une perte d'information considérable
- La représentation raster est moins esthétique que celle réalisée avec des traits fins
- Les transformations relatives aux projections sont coûteuses en termes de temps, à moins que des algorithmes ou du matériel spéciaux soient utilisés.

III.1. Les bases de données numériques

Les bases de données numériques relèvent d'une toute autre logique :

- Conservation des données originales : enregistrement des données indépendant de la représentation graphique ;
- Gestion dynamique : possibilité de mises à jour rapides et continues dans le temps ;
- En plus des autres avantages qu'offre le mode numérique : facilité de traitement et de manipulation (changement d'échelle ou de projection, combinaison de thèmes, etc.), richesse en possibilités graphiques, réduction du coût des réalisations cartographiques, etc.

III. Les SIG et les bases de données

Les tous premiers essais de construction de SIG sont apparus à un moment où la technologie des bases de données n'était pas très répandue et sont partis de zéro en utilisant des outils très limités Comme des systèmes d'exploitation et des compilateurs. Plus récemment, les SIG ont été construits à partir de Systèmes de Gestion de Bases de Données (SGBD) commerciaux.

Chaque système part d'hypothèses ou de modèles concernant les données qu'il manipule. Un SGBD est efficace pour une application particulière dans la mesure où les hypothèses qu'il pose sont compatibles avec la nature des données et des traitements à réaliser. Certains systèmes sont donc plus efficaces que d'autres pour répondre aux besoins des SIG en raison de leur facilité d'ajustement aux modèles de données à référence spatiale.

Trois façons d'utiliser les SGBD dans un SIG

A. Extension du SGBD:

Toutes les données sont accessibles à travers le système de gestion, et doivent correspondre aux modèles retenus par le concepteur du SGBD. On adapte donc les structures de données existantes pour gérer plus ou moins bien le domaine spatial.

B. Solution mixte

La solution mixte consiste en une coexistence de deux modèles pour les données spatiales et thématiques. Certaines données sont accessibles par le biais du système (souvent les tableaux d'attributs et de relations) parce qu'elles correspondent bien au modèle du SGBD. Les autres (le plus souvent des données de géométrie et de localisation) sont accessibles directement

parce qu'elles ne correspondent pas bien au modèle du SGBD.

C. Intégration du domaine spatial dans le SGBD

Mise au point de SGBD comportant les fonctions nécessaires pour gérer les entités spatiales. C'est le type d'avantage que devrait procurer les SGBD orientés - objets qui sont actuellement en cours d'élaboration.

III.1 LES CONCEPTS DE BASES DE DONNEES

Définition

Une base de données est une collection de représentations de la réalité sous forme de données inter-reliées, aussi cohérentes que possible, mémorisées et structurées de manière à faciliter leur exploitation pour satisfaire une grande variété d'applications.

III.1.1. Avantages d'une approche avec des bases de données

- Réduire la redondance (duplication) des données;
- Améliorer la qualité et assurer l'intégrité des données;
- Conserver les définitions sémantiques avec les données (métadonnées);
- Eviter les incohérences (conformité à des modèles, des règles et des normes);
- Réduire le coût de développement des logiciels (beaucoup d'opérations fondamentales sont prises en charge par le SGBD) et éviter de devoir modifier les logiciels d'application à chaque fois qu'on modifie la structure de la base de données;
- Restreindre l'accès aux données et préserver la sécurité des informations.

III.1.2. Procédure de construction d'une base de données

L'objectif principal d'une base de données est de disposer d'un outil de gestion, qu'il s'agisse d'une entreprise, d'une institution ou de phénomènes affectant notre environnement. Pour gérer, il est nécessaire de comprendre et de maîtriser la manière selon laquelle interagissent les différents éléments du système.

- Pour l'entreprise, le concepteur de la BD prend en compte, notamment, l'organisation du personnel, les procédures relevant de l'activité de l'entreprise, le systèmes de comptabilité, etc.
- Pour la gestion d'un bassin versant, il s'agit d'identifier les variables significatives (occupation du sol, relief, etc.) et leur comportement hydrologique.

La procédure qui mène à la construction d'une base de données consiste en plusieurs étapes. En partant du monde réel, on distingue quatre niveaux : le niveau externe, le niveau conceptuel, le niveau logique et le niveau interne.

- *Niveau externe*

Le vrai point de départ est le niveau externe qui présente pour chaque utilisateur, l'ensemble du monde réel qui l'intéresse. Dans certains cas, pour chaque application, on définit un modèle externe qui décrit l'ensemble des données utiles pour celle-ci. On obtient ainsi un ensemble de modèles indépendants.

- *Niveau conceptuel*

Ce niveau est le plus important car il aboutit à l'élaboration d'un modèle conceptuel qui se présente comme la synthèse des modèles externes. On l'appelle conceptuel pour deux raisons: d'une part il est basé sur des concepts bien établis et d'autre part il sert de base à la conception de la base de données.

Le concepteur doit construire un schéma conceptuel de la réalité étudiée permettant une représentation claire, explicite, simple sans biais et communément admise par l'ensemble des utilisateurs.

- *Niveau logique*

L'étape suivante dans la conception d'une BD est la définition du niveau logique, qui est de fait le premier niveau informatique. Il s'agit d'adapter le modèle conceptuel aux exigences des SGBD. Parmi les modèles logiques les plus connus, signalons les modèles hiérarchiques, les modèles en réseaux et les modèles relationnels.

- *Niveau interne*

Le niveau interne correspond au niveau des octets, c'est-à-dire des structures physiques de données en tenant compte des structures des périphériques de stockage, des méthodes d'accès. Généralement, ce niveau n'est pas accessible aux utilisateurs.

III.2.LES SYSTEMES DE GESTION DE BASES DE DONNEES

III.2.1.Les composantes

Types de données

- différents types de valeurs scalaires incluant:
 - les nombres entiers;
 - les nombre réels;
 - les valeurs textuelles (caractères alphanumériques);
 - les dates (année, mois, jour, heure, durée, etc.).
- Certains systèmes plus sophistiqués peuvent inclure des objets, des images, des listes et des photographies parmi les types de données autorisés par le système. Par exemple, une base de données construite pour les pompiers afin de décrire les bâtiments peut comprendre des photographies, des plans des étages, ainsi que l'adresse, le nom du propriétaire, le nombre d'occupants, etc.

Opérations habituelles

Un SGBD permet plusieurs types d'opérations, notamment le tri, l'édition et l'effacement des données, la sélection d'enregistrements en utilisant des critères, la production de rapports et de fichiers d'extraits, etc.

Langage de définition des données (LDD)

C'est le langage utilisé pour décrire le contenu de la BD. Par exemple, les noms d'attributs, les types de données et les métadonnées.

Manipulation des données et langage de recherche

Il s'agit du langage utilisé pour les commandes d'entrée, d'édition, d'analyse, de sortie, de restructuration, etc. Un certain degré de normalisation a été atteint avec le langage SQL (Structured Query Language).

Outils de programmation

Mises à part les commandes et les recherches, on doit pouvoir accéder directement à la base de données à partir des programmes d'application, par le biais, entre autres, de fonctions spécialisées pour gérer les interfaces.

Structures de fichiers

Ce sont les structures internes utilisées pour organiser les données.

III.2.2. Les types de bases de données

Il existe différents modèles de bases de données:

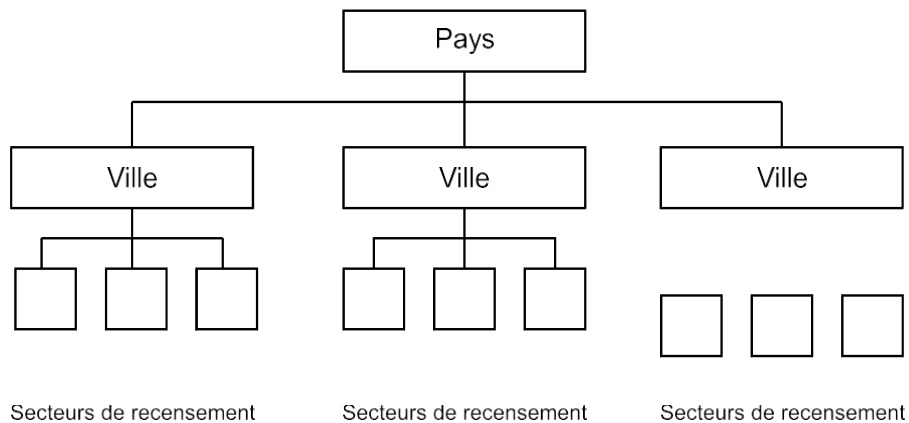
- Tabulaire ("fichier simple") - les données sont sous forme d'un simple tableau de lignes et de colonnes, comme dans un logiciel de tableur.
- Hierarchique: les données sont organisées selon les niveaux d'une hiérarchie.

- En Réseau: les données sont structurées pour former des réseaux entre les entités représentées.
- Relationnel: les données sont reliées entre elles par le biais de relations qui permettent d'établir la correspondance entre des informations distribuées dans plusieurs tableaux.
- Orienté-objet: les données sont structurées autour du concept d'objets qui peuvent être assimilés aux entités du monde réel. Bien que peu répandu aujourd'hui, ce dernier type est appelé à jouer un rôle important dans le domaine des SIG.

III.2.1 LE MODELE HIERARCHIQUE

Les données sont organisées suivant une hiérarchie ou une structure d'arbre. Un élément n'a qu'une seule liaison vers le niveau supérieur. L'élément supérieur s'appelle "parent". Un élément peut être lié à un ou plusieurs éléments inférieurs, appelés "enfants".

Le chemin d'accès à une information suit la branche de l'organigramme. Pour accéder à un élément appartenant au plus bas niveau, on passe nécessairement par tous les niveaux intermédiaires.



Exemple de hiérarchie spatiale

Avantages

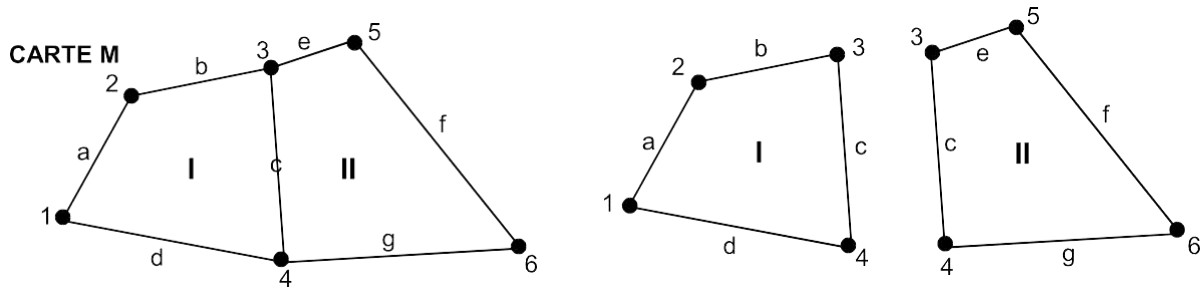
- Structure claire du point de vue présentation;
- Vitesse d'accès élevée.

Inconvénients

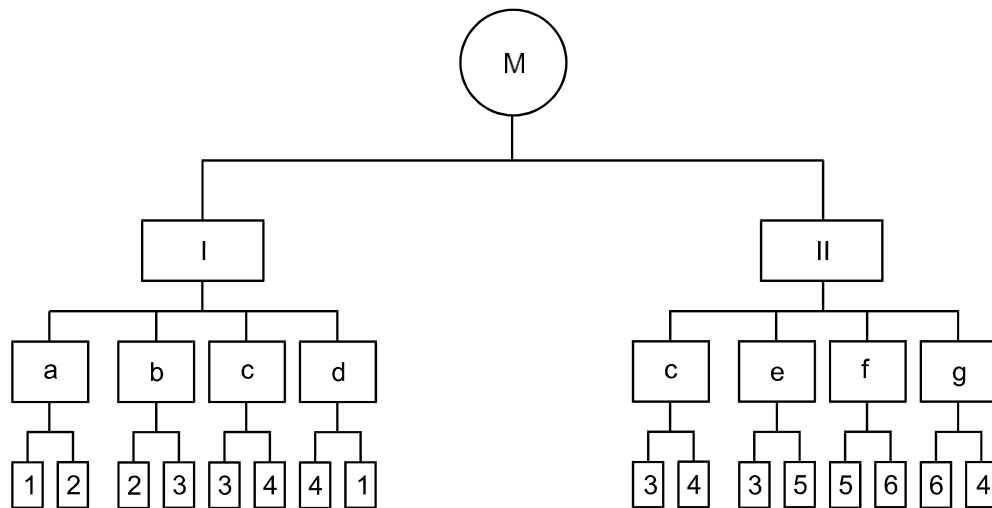
- La structure hiérarchique est inflexible: on ne peut plus définir de nouveaux liens entre les enregistrements une fois que la structure hiérarchique est construite. Dans le cas de données géographiques par exemple, on ne peut plus ajouter de nouvelles relations entre les objets ou définir de nouvelles catégories dans la hiérarchie sans modifier toutes les clés de la base.

- Les liens ne sont possibles que verticalement;
- les seules relations géographiques que l'on peut coder facilement sont "est contenu dans" ou "appartient à".
- L'accès aux données est facile par l'utilisation de l'attribut clé, mais difficile pour les autres attributs. Dans les applications géographiques, il est facile de trouver un enregistrement à partir de son niveau géographique (état, région, province, ville, secteur de recensement), mais cela devient difficile si on utilise un autre type d'attribut. Par exemple la liste des municipalités situées au nord de l'Oued Draa et regroupant des populations de 500.000 habitants au moins.
- La redondance des données stockées dans la base.

exemple: considérons une carte M formée de deux polygones I et II.



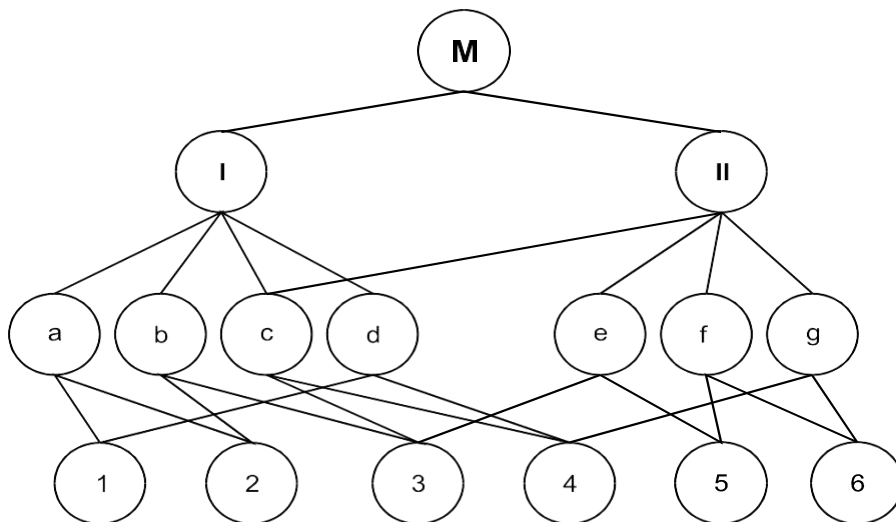
Organisation de la carte M selon le modèle hiérarchique



III.2.2.LE MODELE EN RESEAU

Ce modèle permet aux entités d'avoir des parents multiples et un nombre quelconque d'enfants. Ainsi, le modèle réseau présente une meilleure flexibilité que le modèle hiérarchique pour supporter les relations spatiales complexes.

Exemple: représentation de la carte M selon le modèle réseau.



Avantages

- Le modèle réseau a une plus grande flexibilité que le modèle hiérarchique pour la manipulation des relations spatiales complexes parce qu'il permet de créer des liens entre des entités indépendantes.
- La redondance est moins importante que dans le cas du modèle hiérarchique.

Inconvénients

- Les liaisons entre les entités du même type ne sont pas permises.
- Le modèle en réseau n'est pas très utilisé pour les SIG en raison de la plus grande

flexibilité qu'offre le modèle relationnel.

III.2.3.LE MODELE RELATIONNEL

Le modèle relationnel a été proposé pour la première fois par E. F. Codd, chercheur chez IBM, en 1970. Ce modèle offre une approche flexible des liens entre les enregistrements, ce qui permet une modélisation plus satisfaisante des relations spatiales complexes pouvant exister entre les objets représentés dans la base de données. Il s'agit du plus populaire des modèles de SGBD et il est largement employé pour les SIG:

- Il correspond à la partie INFO du logiciel ARC/INFO;
- A l'empres du logiciel System/9;
- Plusieurs SIG reposent sur ORACLE;
- Plusieurs SIG sous environnement PC utilise la structure Dbase III ou Dbase IV.

Les données sont organisées sous forme d'enregistrements simples contenant un ensemble ordonné de valeurs d'attributs. Ces derniers sont regroupés dans des tables bi-dimensionnelles qui constituent les relations. Chaque table ou relation forme souvent un fichier séparé. Un système de clés est utilisé comme des codes uniques pour identifier les enregistrements dans chaque fichier.

Horizon No.	Profile No.	Profondeur	Texture	structure
1	8609	15	txt_3	str_2
2	8608	58	txt_5	str_4
3	8606	120	txt_3	str_1
4	8603	67	txt_2	str_1

- Deux enregistrements (lignes) ne sont jamais identiques.
- L'ordre des lignes est sans importance.
- L'ordre des colonnes doit être respecté.

Exemple: représentation de la carte M selon le modèle relationnel.

Carte M	M	I	II	
	I	a	b	c
				d

Polygones	II	c	e	f	g
------------------	----	---	---	---	---

Lignes	I	A	1	2
	I	b	2	3
	I	c	3	4
	I	d	4	1
	II	e	3	5
	II	f	5	6
	II	g	6	4
	II	c	4	3

Avantages

- Modèle de base de données relativement flexible.
- Capacité de répondre à toutes à toutes les interrogations basées sur la logique booléenne et les opérations mathématiques.
- L'ajout et l'élimination d'informations sont faciles puisqu'il s'agit juste d'ajouter ou d'effacer un enregistrement.

Inconvénients

- Le désavantage principal de ce modèle réside dans le fait que plusieurs opérations font appel à la recherche séquentielle à travers les fichiers. Avec de larges BD, ceci peut demander un temps énorme même sur des ordinateurs rapides.

III.2.4.LE MODELE ORIENTE VERS LES OBJETS

Les modèles Orientés Objets (OO) sont nés de la convergence de deux domaines:

- les bases de données qui éprouvaient le besoin d'intégrer de nouveaux concepts et outils permettant de dépasser certaines limites du relationnel.
- les langages de programmation objets tels que C++.

Deux points essentiels caractérisent les bases de données OO:

- le modèle de données permet de représenter des structures de données complexes, ce qui répond aux besoins des applications non classiques (imagerie médicale, systèmes d'information géographique, CAO/DAO, etc.).

- les données et les traitements ne sont plus séparés, un objet est décrit par des attributs et par des méthodes qui représentent son comportement.

ANALYSE DES DONNÉES ET MODÉLISATION

V.1. INTRODUCTION

La raison d'être des SIG est de permettre l'analyse des données spatiales et leurs attributs contenus dans la base de données. La différence majeure entre un SIG et un système de CAO ou DAO est la possibilité de manipuler et transformer les données originales pour répondre à des questions (ou requêtes) particulières concernant le monde réel. Le tableau ci-dessous présente une typologie générale des requêtes et traitements effectués dans le cadre des SIG. Dans ce tableau, la complexité des traitements requis s'accroît à chaque ligne et devient maximale pour les fonctions d'aide à la décision.

Exemples de requête	Traitements	Usage
Qu'y a-t-il à tel endroit ?	Localisation	Inventaire localisé
Où trouve-t-on tel genre de phénomène ?	Distribution	Analyse thématique
Qu'est-ce qui a changé depuis...?	Évolution	Analyse temporelle
Que se produirait-il si...?	Modélisation	Simulation des processus
Quelle est la meilleure façon de...?	Optimisation	Aide à la décision

A ne pas oublier:

- La conception de la base de données d'un SIG doit impérativement tenir compte des questions que les utilisateurs peuvent poser !
- Une réponse correcte ne peut être obtenue qu'à travers une formulation appropriée de la question !

V.2. LE LANGAGE DE RECHERCHE STRUCTURÉ (SQL)

La manière de formuler une requête (ou une recherche) diffère selon les systèmes et les possibilités du langage de recherche considéré. Le langage de recherche structuré SQL (Structured Query Language) est utilisé par de nombreux logiciels et tend à devenir une norme dans le processus de traitement des données, bien qu'il soit encore trop limité concernant la recherche spatiale.

Exemples de structure de phrases SQL

SELECT <nom(s) de l'attribut> **FROM** <tableau> **WHERE** <condition de recherche> par exemple
 SELECT nom FROM liste WHERE note = "A"

qui se lit en français: recherche dans liste le nom de tous les étudiants ayant obtenu un "A".

SELECT nom, adresse FROM client WHERE âge < 40

(Sélectionne, à partir de la table client le nom et l'adresse de tous les clients dont l'âge est inférieur à 40 ans).

```
SELECT FROM occupation WHERE classe = "U"
```

(Recherche dans occupation la classe = "U". Les objets sont alors simplement sélectionnés pour représentation; aucun attribut thématique n'a été extrait lors de la recherche).

Les opérateurs du langage SQL

- comparaison: >, <, =, >=, <=, <>

- arithmétique: +, -, *, / (champs numériques uniquement !)

- booléen: and, or, not, xor

Les extensions SQL pour les recherches spatiales

Certains systèmes sont spécialement conçus pour que les recherches spatiales s'effectuent sous SQL. Exemple: l'opérateur WITHIN (à l'intérieur de)

```
SELECT <objets> WITHIN <zone spécifique>
```

Le critère de ces recherches spatiales peut faire appel à un rayon donné à partir d'un point, un rectangle ou encore un polygone irrégulier.

V.3. LES OPÉRATEURS BOOLÉENS

La logique de Boole ou booléenne (George Boole, 1815-1864) est utilisée pour pouvoir combiner plusieurs conditions, par exemple:

WHERE âge > 25 **AND** Ville = "Rabat" Les principaux opérateurs de la logique booléenne sont:

AND, OR, XOR, NOT. Ils sont utilisés pour vérifier si une condition est vraie ou fausse. Le tableau ci-dessous résume l'utilisation de ces opérateurs (1 = vrai, 0 = faux):

A	B	NOT A	A AND B	A OR B	A XOR B	A NOT B
1	1	0	1	1	0	0

10	0	0	1	1	1
01	1	0	1	1	0
00	1	0	0	0	0

Considérant l'exemple d'une base de données contenant des unités thématiques de sols. Chaque unité peut avoir des attributs décrivant la texture et le pH du sol. Si A représente les sols à texture argileuse et B les sols dont le pH > 7, on peut dériver les expressions logiques suivantes:

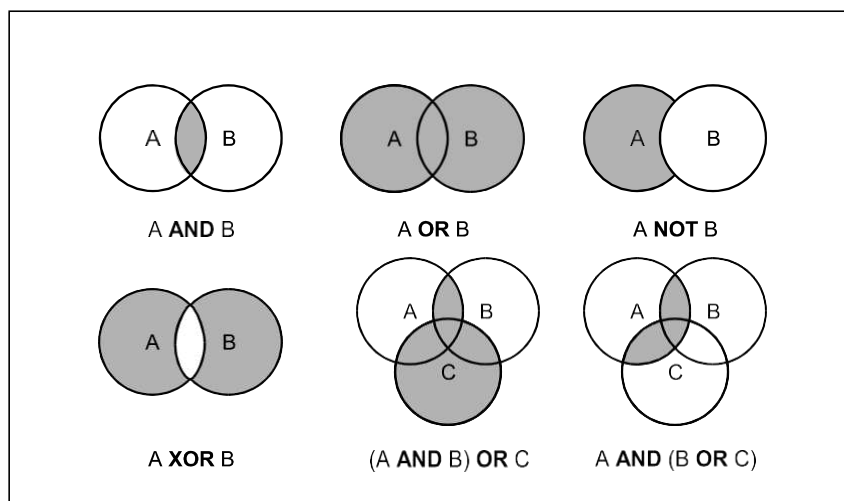
A **AND** B : sélectionne toutes les unités présentant un sol argileux et un pH > 7

A **OR** B : sélectionne toutes les unités caractérisées soit par un sol argileux soit un pH basique, séparément ou en combinaison

A **XOR** B : sélectionne toutes les unités présentant soit un sol argileux soit un pH basique, mais pas en combinaison

A **NOT** B : permet de sélectionner toutes les unités ayant un sol argileux et un pH ≤ 7

Les opérateurs booléens peuvent également être appliqués aux propriétés spatiales, par exemple: chercher les unités qui présentent un sol argileux caractérisé par un pH > 7 et une superficie dépassant 10 hectares. La figure ci-dessous illustre l'utilisation de ces opérateurs dans les applications spatiales de la logique booléenne (superposition d'éléments zonaux dans un SIG).



Pour des raisons d'organisation et de stockage dans les bases de données géographiques, les données peuvent être considérées comme "spatiales" ou "non spatiales" (attributs). Tenant

compte de cette considération, on peut faire une distinction entre les opérations d'analyse selon qu'elles soient effectuées sur des données spatiales ou non spatiales.

V.4. ANALYSE DES DONNÉES NON SPATIALES

La manipulation des attributs dans une base de données non spatiales est beaucoup plus rapide et plus efficace que la manipulation des données spatiales dans une base de données géographiques. Il existe actuellement des systèmes de gestion de bases de données qui permettent de réaliser facilement des opérations d'extraction, de transformation, de manipulation et de sortie des données. Ces opérations peuvent être regroupées en trois catégories principales:

- des opérations d'interrogation ou **requêtes**
- des opérations de **généralisation**
- des opérations de **calcul**

V.4.1. Les requêtes

Les fonctions de requêtes consistent en l'extraction (sélective) des attributs *sans altérer les données existantes*. Ces extractions sont réalisées suivant des spécifications fixées par l'opérateur.

Exemples typiques de requêtes sur les attributs:

- *Quelle est la pluviométrie journalière ?* SELECT * FROM PLUIE

Cette requête retourne toutes les données concernant la pluviométrie journalière (* = toutes les données) stockées dans la base de données (table: PLUIE).

- *Quelle est la longueur et l'état de la chaussée de la section 11 de la route RN235 ?*

Souvent, on a besoin d'une extraction sélective. La requête est alors définie en utilisant des conditions spécifiées par l'opérateur. Cet exemple montre une extraction sélective de données attribut à partir d'une seule table (table: ROUTE).

```
SELECT
section#, route, longueur, état
FROM
route
WHERE
```

Section#	Route	Long. (km)	État
11	RN235	3	détérioré

section# = 11 AND route = RN235

- *Quel est le nom, l'adresse du propriétaire et la superficie de la parcelle 103 ?*

Cet exemple illustre l'utilisation d'une liaison relationnelle. L'extraction des données attribut est réalisée en reliant des données stockées dans des tables d'attributs différentes. La liaison entre les tables (PROPRIÉTAIRES et PARCELLES) est assurée à grâce à un champ commun, par exemple, le champ "Parcelle#".

Table: PARCELLES

Parcelle#	Superficie (en m ²)	Périmètre (en m)
101	10453.4	489.0
102	4105.9	284.7
103	11943.8	625.3
104	5457.4	361.0
105	14456.1	584.9

Table: PROPRIÉTAIRES

Parcelle#	Nom_prop	Adresse
101	Mansour	Kénitra
102	Benmalek	Casablanca
103	Moukhtar	Casablanca
104	Ijlal	Rabat
105	Farah	Temara

```
SELECTFROM      PARCELLE#, SUPERFICIE, NOM_PROP, ADRESSEPARCELLES,
WHERE           PROPRIETAIRES
               PARCELLE# = '103'
```

Parcelle#	Superficie (en m ²)	Nom_prop	Adresse
103	11943.8	Moukhtar	Casablanca

V.4.2. Généralisation

Les fonctions de généralisation consistent à une redistribution des données dans des classes. Les données existantes ne sont pas altérées. Les données identifiées par un ensemble de critères, formulés par l'opérateur, sont regroupées dans des classes. La généralisation se traduit par une réduction du niveau de détails dans la description du phénomène considéré

Exemple

Table: POPULATION

Commune	Province	Population	Superficie (km ²)
1	A	6,017	17
2	A	8,853	102
3	A	12,666	52
4	A	21,146	82
5	B	10,281	93
6	B	13,551	65

```
SELECT
FROM
PROVINCE,
SUM (POPULATION) "POPULATION2",SUM (SUPERFICIE) "SUPRFICIE2"
POPULATION
```

Province	Population2	Superficie2 (km ²)
A	48,682	253
B	23,832	158

V.4.3. Calcul

Dans les bases de données, les attributs sont enregistrés comme des valeurs fonctionnellement indépendantes. Cependant, une dépendance fonctionnelle peut exister dans le cas où une valeur est dérivée directement à partir d'une ou plusieurs autres valeurs. Par exemple, l'enregistrement d'attributs pour un certain nombre de communes, tels que le nombre de population, la superficie et la densité de population, engendre une dépendance fonctionnelle entre, d'une part, la densité de population et, d'autre part, le nombre de population et la superficie. En effet, La densité de population est facilement calculée à l'aide du rapport nombre de population / superficie. Les fonctions de calcul incluent des opérations arithmétiques, mathématiques et logiques.

```
SELECT
COMMUNE, PROVINCE,
SUPERFICIE / POPULATION "DENSITE"FROM
POPULATION
```

Commune	Province	Densité (hab./km ²)
1	A	353.9
2	A	86.8
3	A	243.6
4	A	257.9
5	B	110.5
6	B	208.5

V.5. ANALYSE SPATIALE INTÉGRANT LES DONNÉES GÉOGRAPHIQUES ET LEURS ATTRIBUTS

- La différence entre les systèmes de cartographie numérique et les systèmes d'information géographique réside principalement dans l'habilité des SIG d'analyser simultanément les données spatiales et les données non-spatiales.
- La diversité des fonctions d'analyse dépend largement de la puissance du SIG et de l'objectif pour lequel il a été conçu :

Un SIG conçu pour l'utilisation intégrée de l'imagerie satellitaire disposera plus de fonctions analytiques relatives au traitement de données en format raster;

Un SIG conçu pour servir les sociétés d'utilités publiques (réseaux de distribution d'eau, d'assainissement, d'électricité, etc.) aura des capacités largement axées sur le traitement de données en format vecteur.

- La manière selon laquelle l'analyse des données est réalisée dans un SIG dépend de l'architecture du logiciel utilisé:
 - Certains systèmes consistent en des programmes conçus pour réaliser des tâches spécifiques (efficacité);
 - D'autres systèmes fournissent une panoplie de programmes (ou modules) qui, combinés adéquatement, permettent de réaliser des tâches similaires (flexibilité).

Les SIG offrent une large variété de fonctions d'analyse. Ces dernières peuvent être regroupées en quatre groupes principaux:

- Opérations d'extraction, de (re)classification et de mesure
- Opérations de superposition (overlay)
- Opérations de surface
- Opérations de connectivité

Presentation de l'ArcGIS

ArcGIS est un ensemble de logiciels d'information géographique développés par la société américaine Esri (Environmental Systems Research Institute). Il est possible de créer, gérer et partager des données géographiques, cartes et modèles d'analyse via des applications bureautiques et serveurs.

Dans les faits, ArcGIS est composé de plusieurs « blocs » d'applications :

- **ArcMap**
- **ArcCatalog**
- **Arc Toolbox**
- ArcScene
- ArcGlobe



A- Présentation de l'interface : ArcMap

C'est l'application « cartographique » d'ArcGIS

ArcMap représente les informations géographiques sous la forme d'un ensemble de couches produisant *in fine* une « carte ». ArcMap présente donc une fenêtre géographique, ou bloc cartographique, dans laquelle vous pouvez afficher et utiliser les informations géographiques. Plus précisément, l'interface utilisateur d'ArcMap présente une fenêtre Windows qui se décompose de la manière suivante :

- 1- **La barre des menus** : située en haut de l'interface, et présentée sous la forme d'onglets, elle donne accès à des catégories de fonctionnalités permettant d'interagir avec toutes les fonctions d'ArcMAP ;
- 2- **la table des matières** : située sur la partie gauche de l'interface, elle permet de gérer l'affichage des couches ;
- 3- **Les barres d'outils** : Elles permettent d'accéder rapidement à des fonctions importantes d'ArcGIS, comme la gestion des sessions d'édition, les outils de sélection... Ces fenêtres peuvent être positionnées n'importe où, mais en règle générale les plus courantes sont « ancrées » sous le menu.

Travaux pratiques

4- L'espace carte : Facilement identifiable, il occupe le centre de l'interface et affiche les résultats.

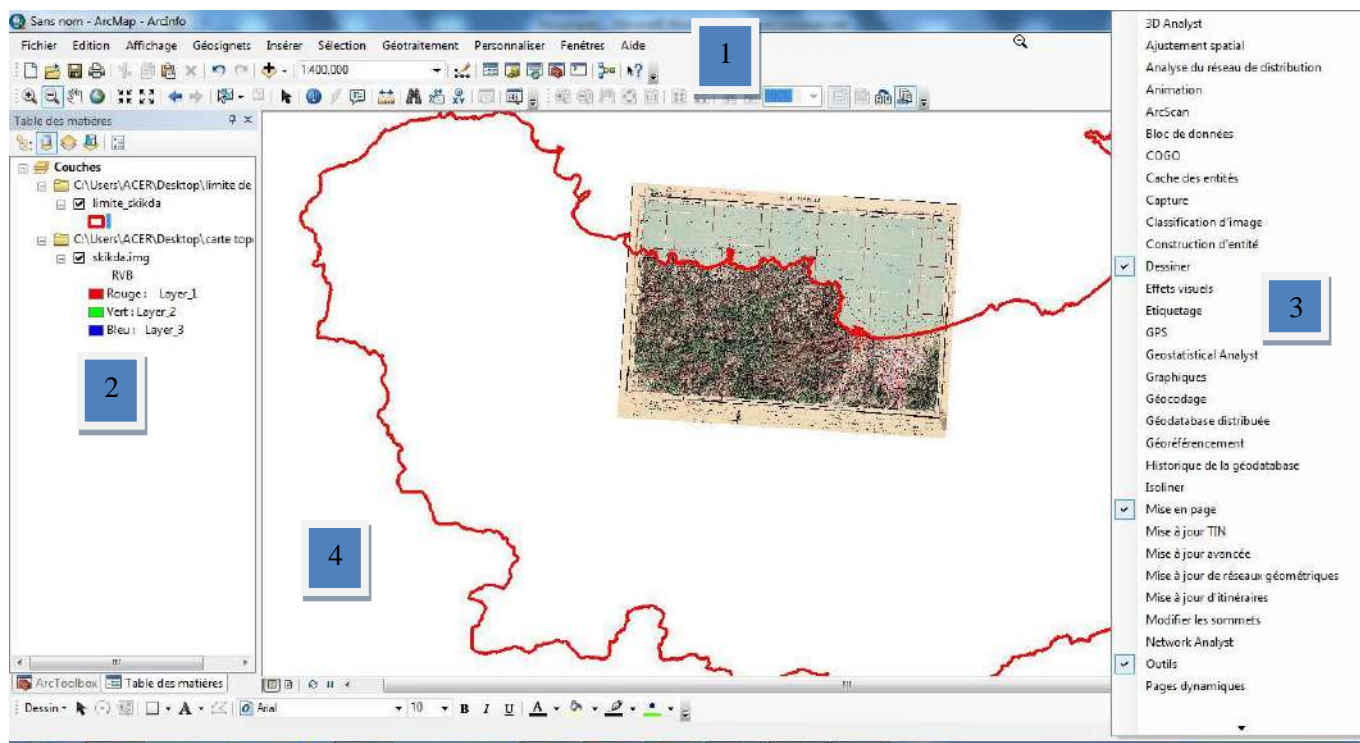


Figure 1 : L'interface ArcMap.

B- Présentation de l'interface : ArcCatalog



ArcCatalog : L'application ArcCatalog fournit une fenêtre de catalogues utilisés pour organiser et gérer différents types d'informations géographiques. Les informations qui peuvent être organisées et gérées dans ArcCatalog comprennent : des géodatabases ; des fichiers raster ; des fichiers vecteurs ; des cartes ; des « globes » ; des documents de « scènes » 3D et des fichiers de couches ; des métadonnées normalisées pour ces éléments d'informations SIG, etc.

ArcCatalog vous permet de passer en revue, organiser, distribuer et modifier les éléments de propriétés des données géographiques. L'interface permet d'employer des méthodes graphiques ou textuelles pour passer en revue, contrôler ou modifier l'ensemble des données.

Travaux pratiques

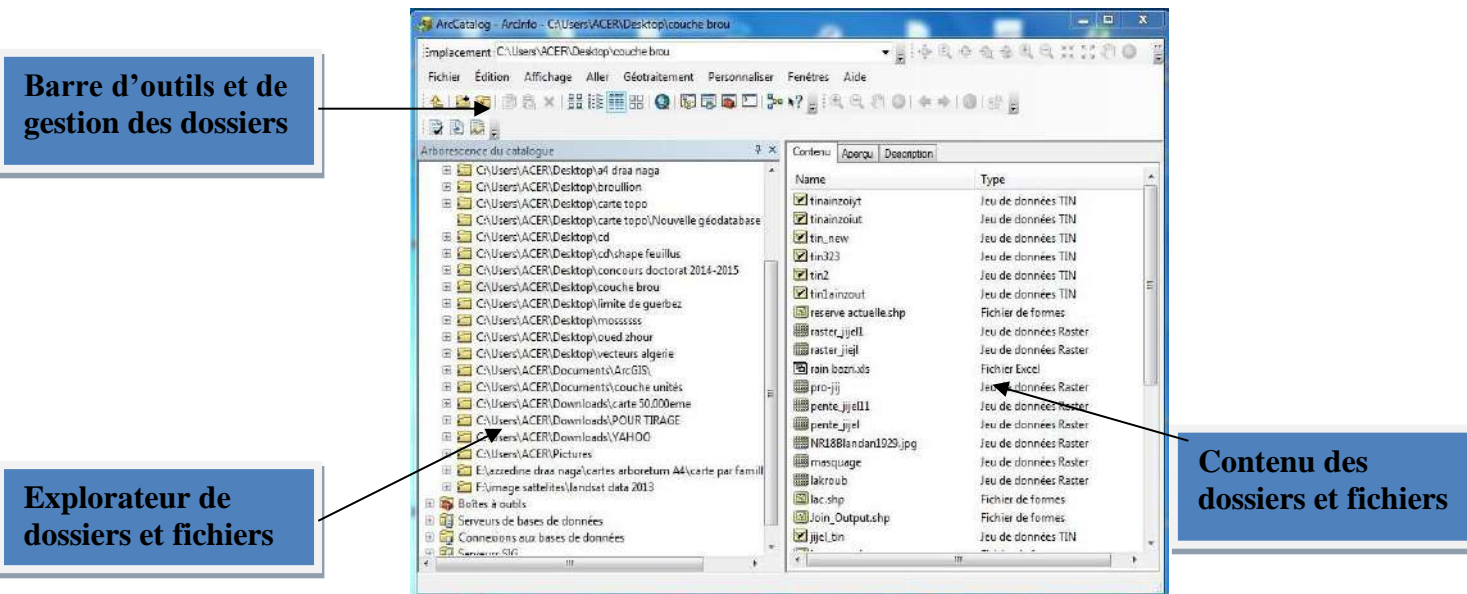



Figure 2 : L'interface ArcCatalog.

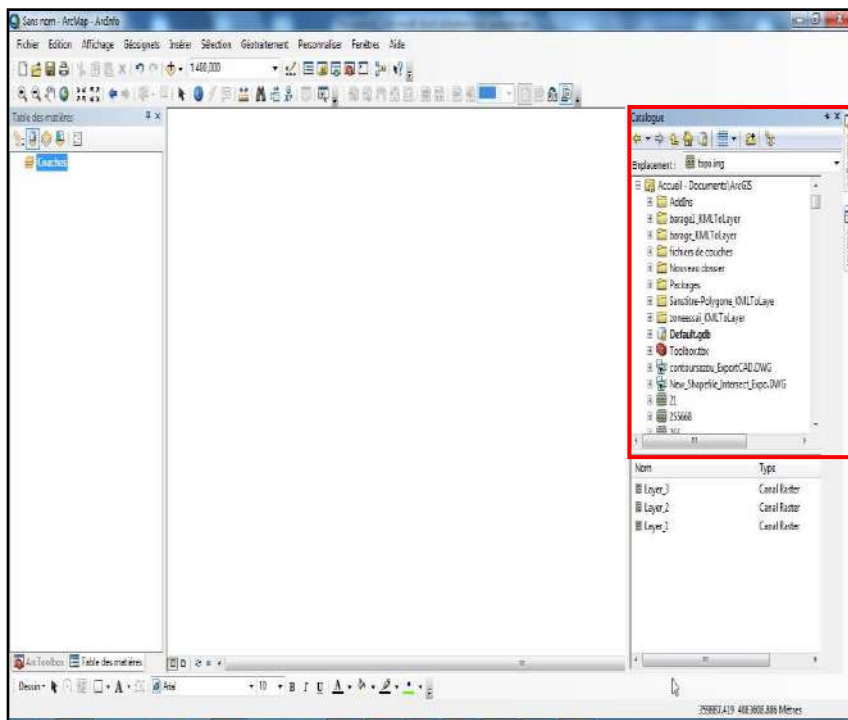
C'est idéalement dans cette interface que l'on exécutera les actions les plus courantes suivantes :

- Créer des nouvelles données (couches ou shapefiles, géodatabases...) ou supprimer une couche (soit l'ensemble des fichiers correspondants au shapefile),
- Attribuer un système de coordonnées lorsqu'il n'est pas reconnu par le logiciel, ou non renseigné,
- Avoir un aperçu géographique ou attributaire d'une couche donnée.

L'interface ArcCatalog est disponible au travers d'ArcMap, ce qui facilite grandement la gestion des fichiers.

Cet onglet se charge soit à partir de l'icône  dans la barre standard de ArcMap,


soit à l'aide de l'onglet Catalogue situé à droite de la fenêtre









Les formats de données visibles dans ArcCatalog

- Tables attributaires feuilles de calcul Excel et dBase (*.dbf)

Travaux pratiques

: Ce sont les fichiers généralement issus d'un fichier Excel que l'on intègre à l'étude cartographique. Ces données ne sont pas géoréférencées.

- Bases de données : Géodatabase (*.mdb)  , fichier Access .
- Fichiers de formes (*.shp) polygones  , lignes  , points  : Ce sont les entités graphiques qui sont géoréférencées. Un fichier de formes est associé à une table (*.dbf) qui peut contenir un nombre indéfini de données par entité.
- Fichiers couches (*.lyr)  : Ce sont les représentations graphiques issues des fichiers de formes.
- Fichiers raster  sont les fichiers images auxquels chaque couleur est rattachée une valeur numérique.

C- Présentation de l'interface : ArcToolbox

ArcToolbox est le module dans ArcGis 10 qui permet d'accéder à toutes les fonctionnalités puissantes de traitement et d'analyse. La boîte à outils fournit un ensemble très riche de fonctions de géotraitement. Il convertit aussi les formats de données pour l'importation ou l'exportation de fichiers, ainsi que les changements de projection.

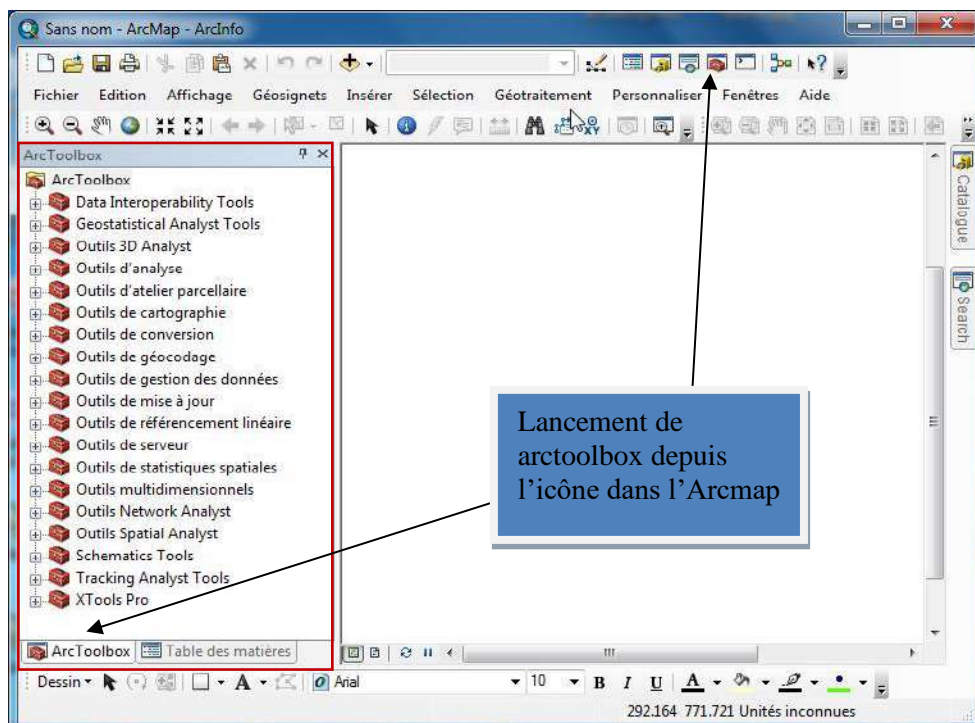


Figure 3 : L'interface ArcToolbox.

Global Mapper



Global Mapper est un logiciel de traitement de données SIG (Système d'Information Géographique) avec toute sorte d'information cartographique et de cartes pour professionnels et initiés du monde de la cartographie.

Le logiciel peut s'exécuter de manière autonome ou bien intégré comme un module entre autres outils SIG, et il a la capacité d'accéder à une grande quantité de dépôts d'images, de données topographiques et de fichiers DEM/DSM.

Caractéristiques de Global Mapper

- Plus de 200 formats compatibles : données *raster*, de vecteurs, d'élévation...
- Importation et exportation depuis des bases de données spatiales.
- Accès aux données dans les nuages grâce à l'intégration en ligne.
- Support pour données KML et KMZ natives de Google Earth.
- Accès au matériel de DigitalGlobe, OpenStreetMap et TerraServer-USA/MSRMaps.



Figure 4 : L'interface du GlobalMapper.

Google Earth

Logiciel développé par Google, permettant de parcourir la Terre par des images satellite, en 2D ou en 3D. Il est également possible d'effectuer des recherches (restaurants, hôtels, etc.) et de consulter des itinéraires routiers.

Initialement développé par la société Keyhole (désormais rachetée par Google), Google Earth est un logiciel gratuit (dans sa version de base) proposé par Google. Google Earth permet à l'utilisateur de voir l'importe quel endroit du globe comme le font les satellites. La résolution n'est pas identique partout, mais les endroits stratégiques (grandes villes) sont souvent très bien couverts, à tel point que l'on distingue facilement les voitures dans les rues.

Pour l'utiliser, il suffit de taper un nom de lieu (en précisant parfois le pays) pour s'y rendre. Le logiciel télécharge en temps réel toutes les informations depuis Internet : images satellite, noms des rues, des routes, des villes, situation des hôtels, restaurants, etc.

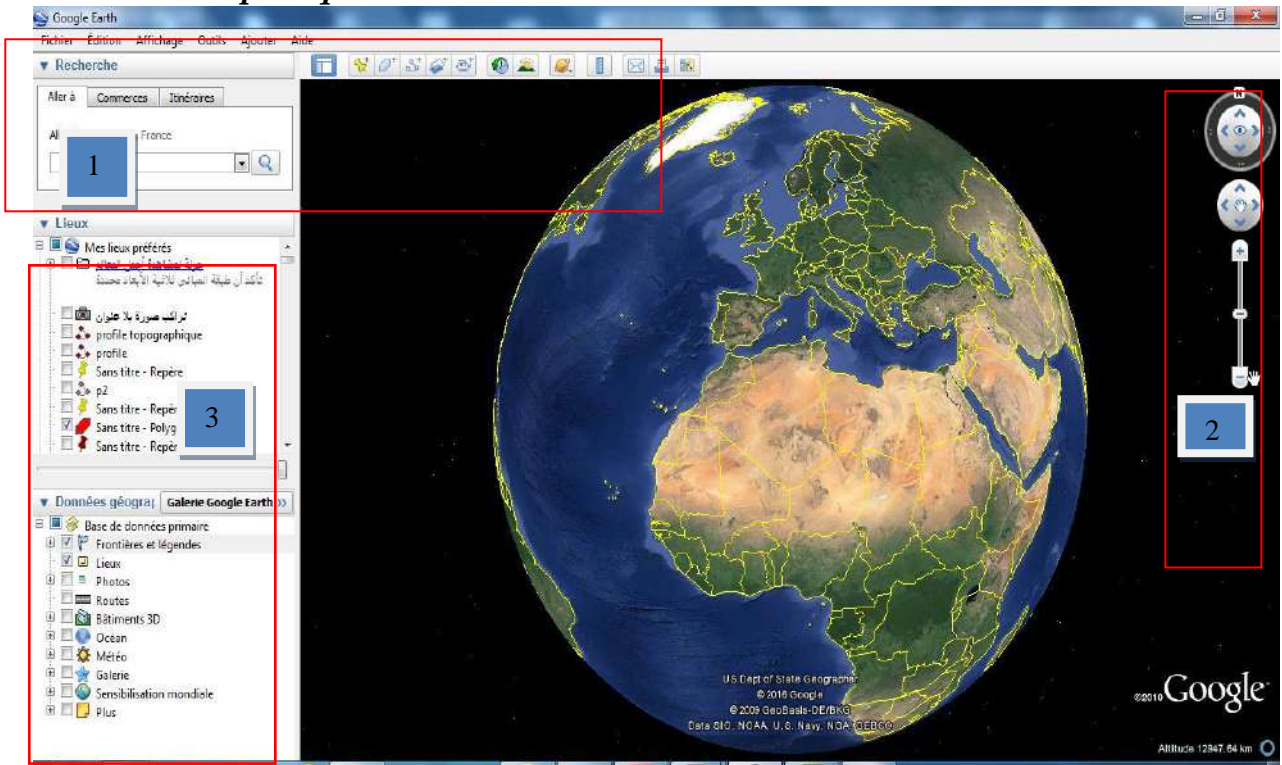
Outre la vue du-dessus classique, il est possible d'incliner l'angle de vue et de voir alors le paysage en relief. La hauteur de vue (altitude) est également réglable.

Bien entendu, il est possible de faire pivoter la carte. L'utilisateur peut enregistrer des lieux en mémoire en y adjoignant des commentaires. Google Earth propose également des itinéraires pour se rendre d'un point quelconque du globe à un autre.

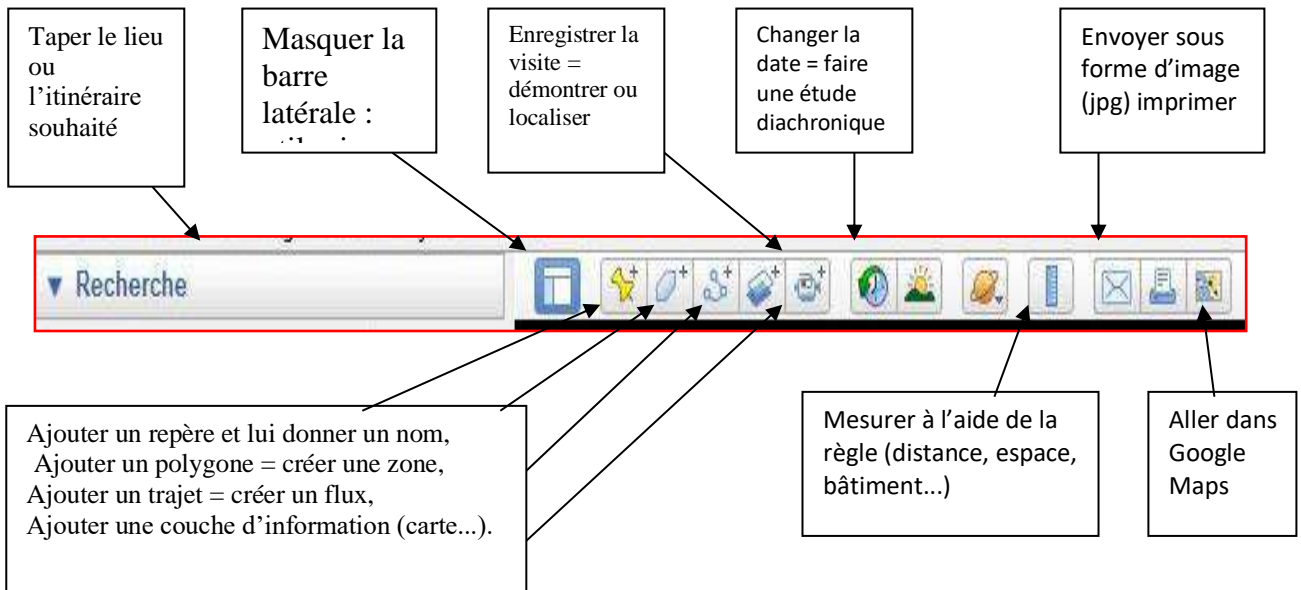
Le logiciel existe en plusieurs versions :

- Google Earth est gratuit
- Google Earth Plus coûte 20\$ par an ; il intègre le support des appareils GPS, l'import de données sous forme de feuilles de table, des outils de dessin et de meilleures fonctions d'impression
- Google Earth Pro coûte 400\$ par an ; destiné aux professionnels pour un usage commercial, il comprend les améliorations de Google Earth Plus ainsi que le calcul de surfaces et d'autres options (générateur de vidéos, impression haut de gamme, import de données GIS, etc.)
- Google Earth Entreprise est un ensemble de solutions pour les entreprises, qui permettent de déployer des solutions commerciales en mode ASP ou en mode hébergé (il s'agit de Google Earth Fusion et de Google Earth Server).

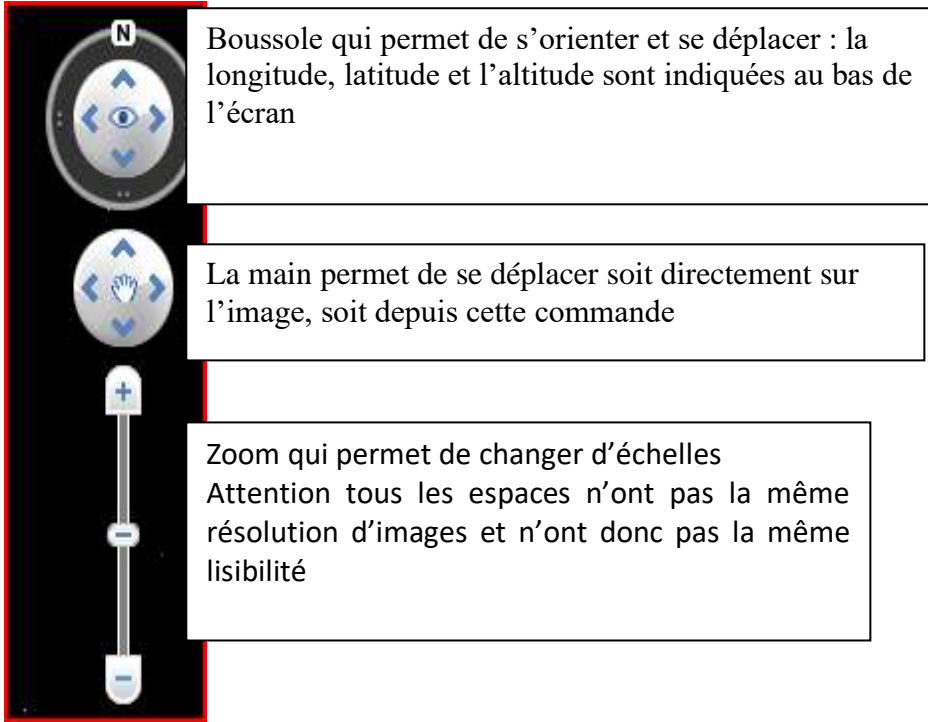
Travaux pratiques



Les fonctionnalités de Google Earth sont réparties sur ces trois parties de l'écran.



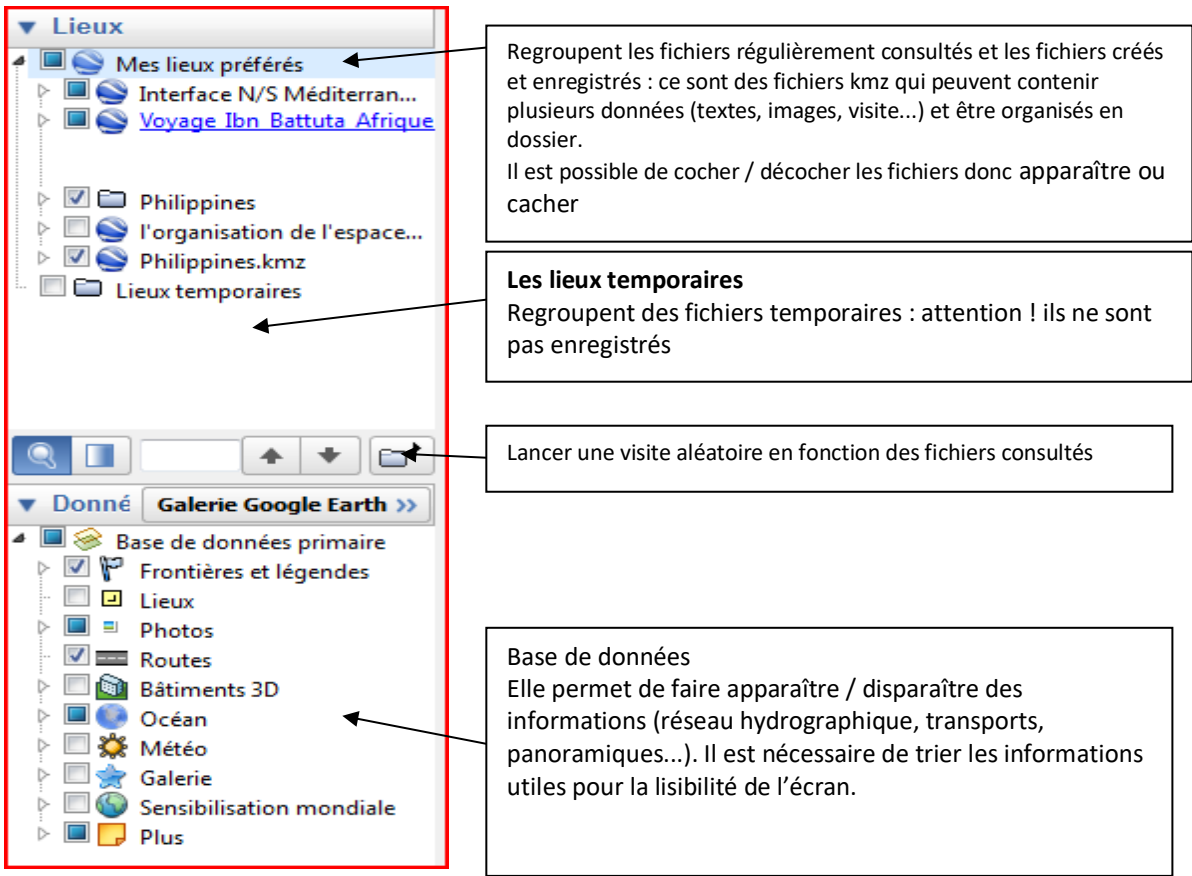
Travaux pratiques



Boussole qui permet de s'orienter et se déplacer : la longitude, latitude et l'altitude sont indiquées au bas de l'écran

La main permet de se déplacer soit directement sur l'image, soit depuis cette commande

Zoom qui permet de changer d'échelles
Attention tous les espaces n'ont pas la même résolution d'images et n'ont donc pas la même lisibilité



Regroupent les fichiers régulièrement consultés et les fichiers créés et enregistrés : ce sont des fichiers kmz qui peuvent contenir plusieurs données (textes, images, visite...) et être organisés en dossier.
Il est possible de cocher / décocher les fichiers donc apparaître ou cacher

Les lieux temporaires
Regroupent des fichiers temporaires : attention ! ils ne sont pas enregistrés

Lancer une visite aléatoire en fonction des fichiers consultés

Base de données
Elle permet de faire apparaître / disparaître des informations (réseau hydrographique, transports, panoramiques...). Il est nécessaire de trier les informations utiles pour la lisibilité de l'écran.

Travaux pratiques

Application 1 : Les différents types des données et leur organisation dans le système d'information géographique

1- Présentation du logiciel ArcGis

ArcGIS Desktop est un ensemble complet d'applications SIG professionnelles permettant de résoudre des problèmes, remplir une mission, augmenter l'efficacité, prendre de meilleures décisions et communiquer, visualiser et comprendre une idée, un plan, un conflit, un problème ou une situation.

ArcGis est en fait composé de plusieurs applications accessibles depuis ce menu.

Depuis le bouton démarrer (en bas à gauche) dans le Menu : **ArcGis**

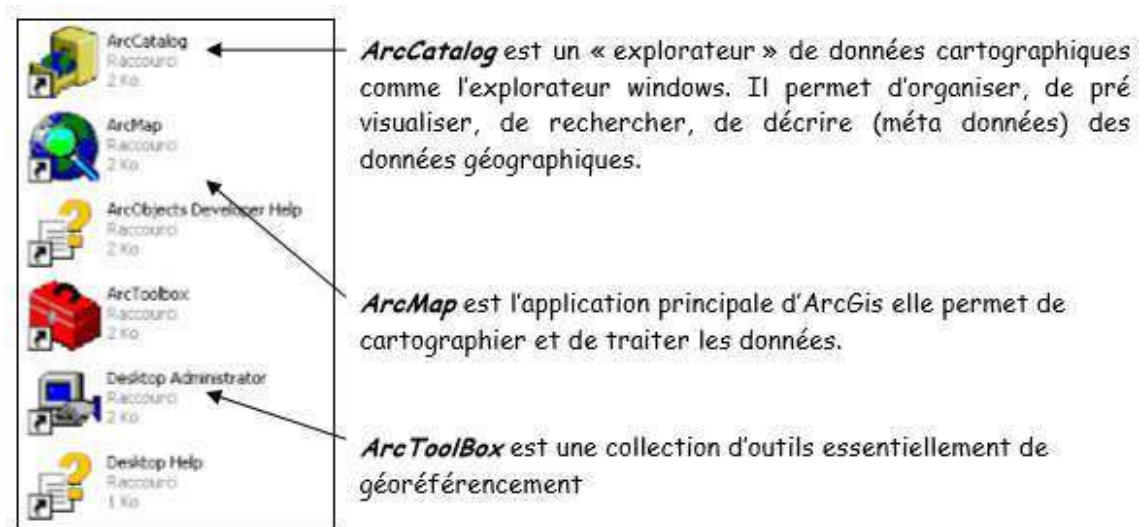


Figure 1 : Démarrage de l'ArcGis

Pour commencer à explorer les possibilités d'ArcGis, vous allez utiliser une « carte » sur un jeu de données préparé : **Ouvrir la carte « projet_map.mxd » Menu : Fichier -> Ouvrir**

Travaux pratiques

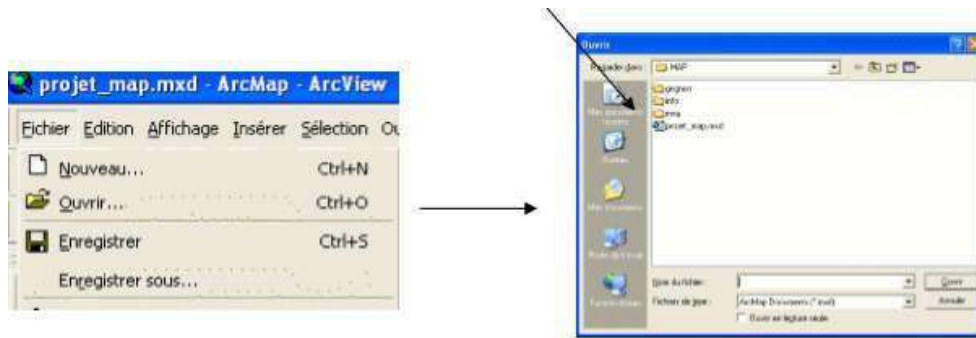


Figure 2 : Ouverture d'un fichier

2. Données et organisation des données



Les données manipulées peuvent être :

- Sémantique : table de données
- Sémantique et cartographique : table et cartographie

Sur la couche routes, cliquez avec le bouton droit de la souris et choisir dans le menu « Ouvrir la table attributaire » pour visualiser une table de données d'une couche d'information :



L'outil d'identification permet de faire afficher dans une boîte de dialogue les champs d'information d'un objet cliqué sur la carte.

Les couches sont représentées cartographiquement selon deux modes :



Travaux pratiques

Application 2 : Gestion des « couches » du SIG

1- L'affichage des couches

La carte affichée est composée de plusieurs calques transparents superposés : les couches.

La table des matières sert à contrôler l'affichage des couches :

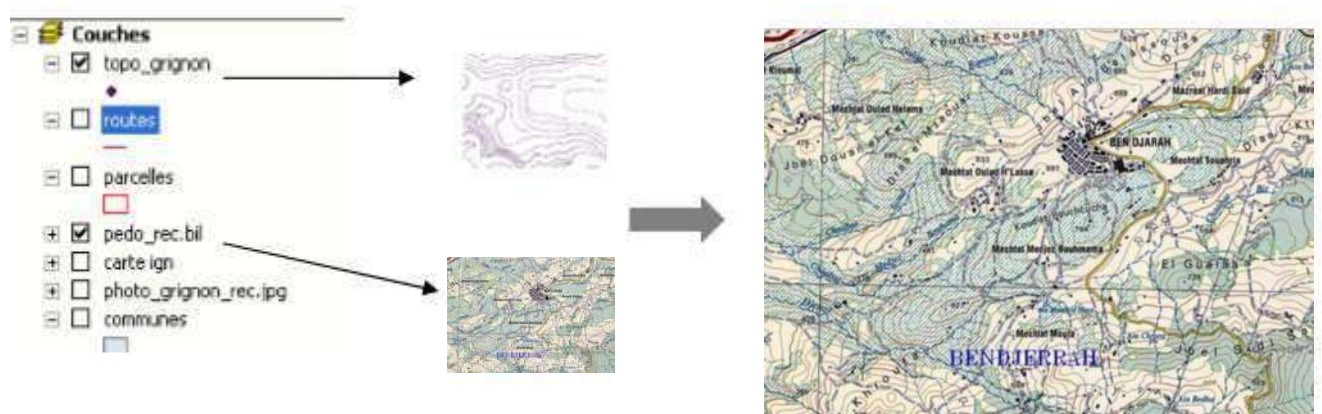


Figure 1 : L'affichage des couches

2- Les types de couches du SIG

a/ Les couches vectorielles

Ponctuelles :



Travaux pratiques

Linéaires :

routes



Polygonales

communes



Travaux pratiques

b/ Les couches raster

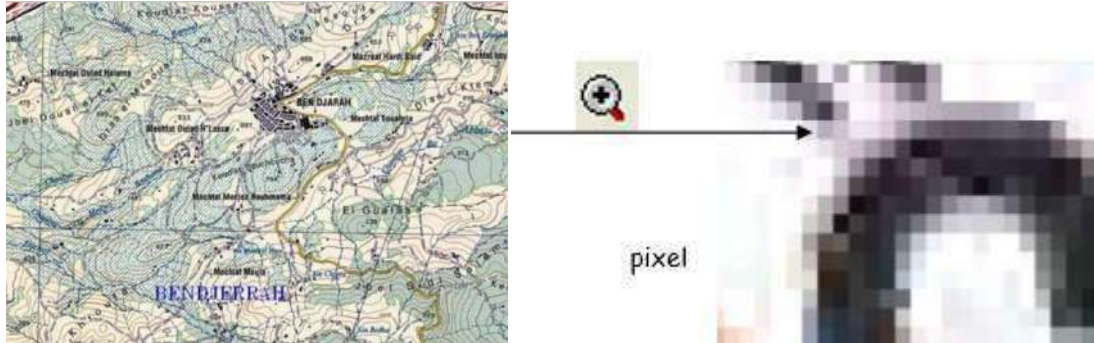


Figure 2 : Image raster

Exercice 1 :

Soit un écosystème représenté par une image raster d'une résolution de 30m. Cette image contient 1000 pixels.

Calculer la superficie de cet écosystème en ha ?

Solution :

Un pixel à une résolution de 30 m c.a.d la taille de l'unité est de 30×30

Donc la superficie de cette unité = 900m^2

La superficie de la zone = $1000 \times 900 = 900000 \text{m}^2$

1 ha correspond à $100 \times 100 = 10000 \text{m}^2$

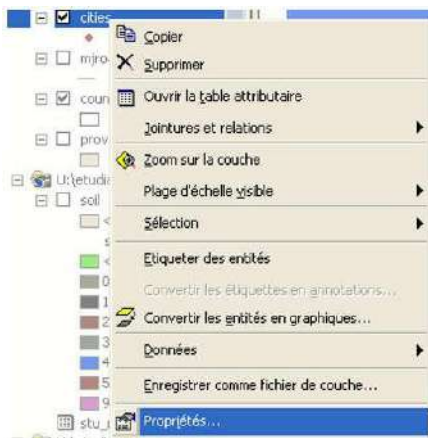
La superficie de la zone = $900000 / 10000 = 90 \text{ ha}$.

La superficie de cet écosystème est de 90 ha.

Travaux pratiques

Application 3 : Projections cartographiques

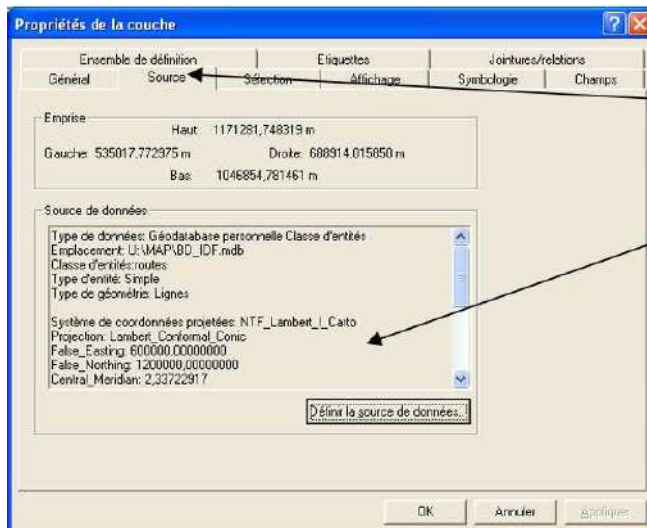
1. Systèmes de coordonnées des couches



Aller dans les propriétés de la couche routes

(Clique bouton droit de la souris sur la couche)

Item Propriétés



Aller sur l'onglet « Source »

Le système de coordonnées de la couche est affiché

La couche est en Lambert I

Dans certains cas le Système de coordonnées n'est pas défini



Propriétés de la couche parcelles



Système de coordonnées:
<Non défini>

Travaux pratiques



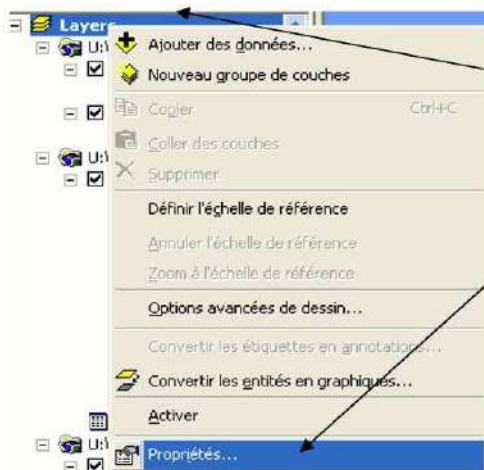
Ouvrir la couche département

```
Système de coordonnées:  
Lambert_Conformal_Conic  
False_Easting: 600000,000000  
False_Northing: 2200000,000000  
Central_Meridian: 2,337229  
Standard_Parallel_1: 45,898919  
Standard_Parallel_2: 47,696014  
Scale_Factor: 1,000000  
Latitude_Of_Origin: 46,800000
```

Système de coordonnées en Lambert II étendu

2. Système de coordonnées de la carte

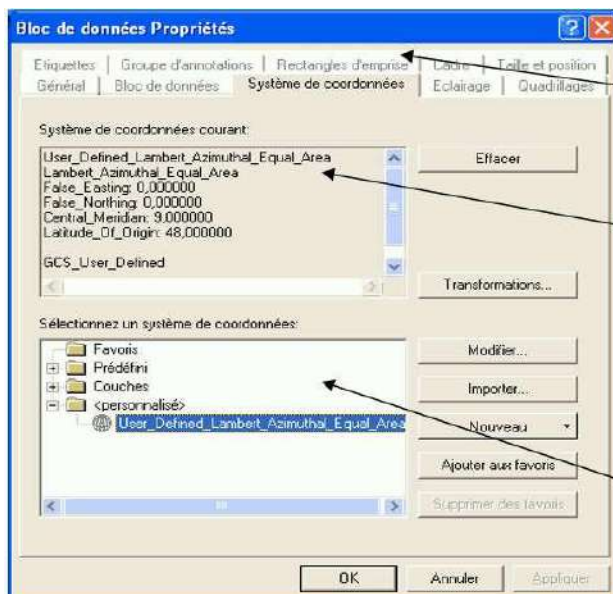
Ajouter les couches (en latitude/longitude WGS84) : continents et latitude_meridien (grille des méridiens et parallèles du globe tout les 30° d'arc)



Aller voir dans les propriétés de la carte

Cliquer (bouton droit) sur Couches

Propriétés



Aller dans l'onglet :

Système de coordonnées

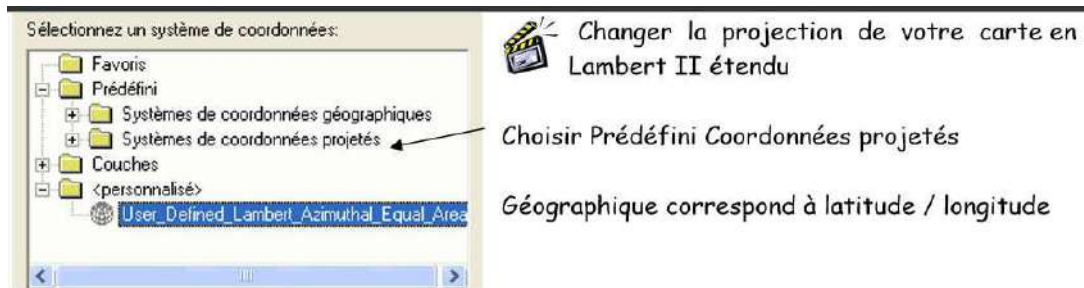
Le système qui est actuellement utilisé pour toute la carte

Ici vous pouvez choisir un nouveau système de coordonnées

Travaux pratiques

Remarque : Quand une couche n'a pas la même projection de la carte elle est alors reprojétée « à la volée » pour être affichée dans la même projection que la carte.

Quand la projection d'une couche n'est pas définie dans son fichier, elle sera supposée être dans le même système de coordonnées que la carte, si ce n'est pas le cas attention erreur.



Changer la projection de votre carte en Lambert II étendu

Choisir Prédéfini Coordonnées projetés

Géographique correspond à latitude / longitude

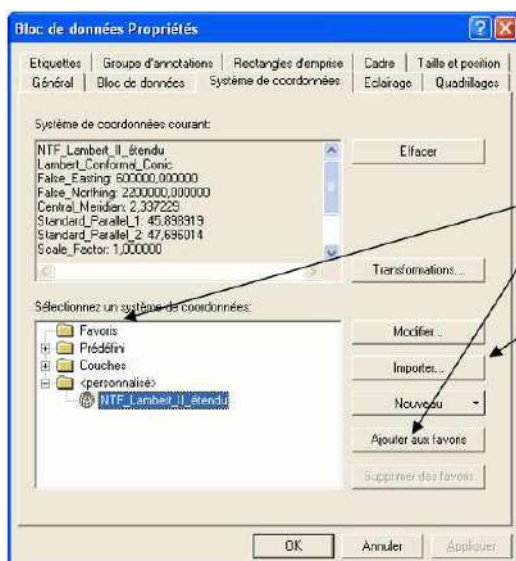


Aller dans les Grilles nationales



Aller dans le dossier « France »

Choisir NTF Lambert II étendu



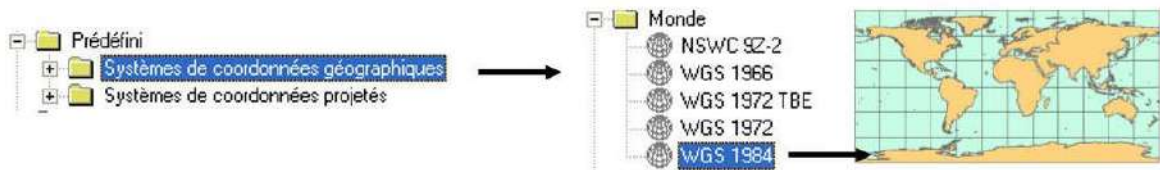
Vous pouvez conserver dans une liste de favoris les projections dont vous vous servez le plus souvent

Vous pouvez importer la projection de la carte depuis un fichier de couche (shape ou autre)

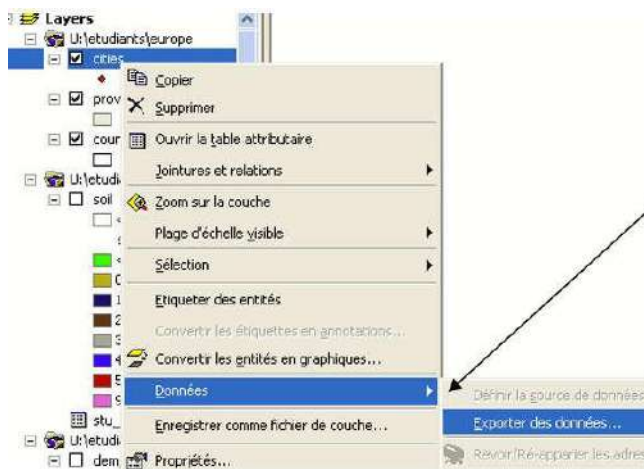
Travaux pratiques



Mettre votre carte en coordonnées géographiques latitude / longitude avec le datum mondial WGS84 :



3. Changer la projection d'une couche vectorielle sous ArcMap



Une fois la projection de la carte définie en Lambert II étendu, exporter la couche des continents



Choisir Même système de coordonnées que le bloc de données (c'est à dire la carte)

Vous allez ainsi, enregistrer une copie de la couche **reprojetée** dans un nouveau système de coordonnées.



ArcToolBox

On dispose aussi dans la « boîte à outils » ArcToolBox de fonctions pour la gestion des projections cartographiques.

Travaux pratiques

Application 4 : Description topographique

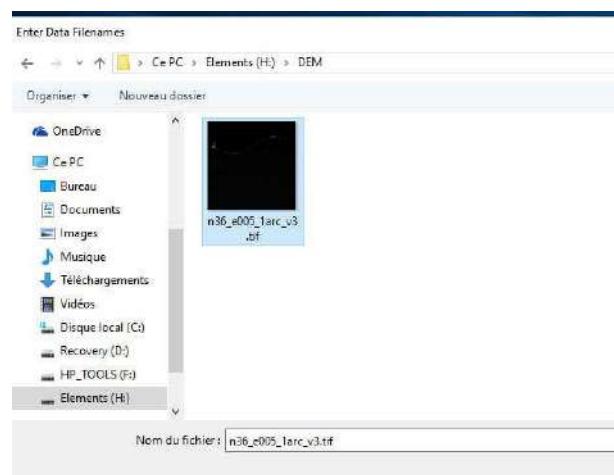
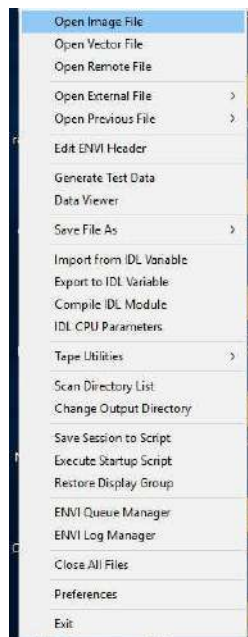
1- Le modèle numérique de terrain

Il fournit une information altimétrique, c'est une représentation numérique simplifiée de la surface du territoire, intégrée dans le SIG cette information joue un rôle très important dans les méthodes de l'analyse spatiale en particulier pour la prise en compte de la morphologie du terrain.

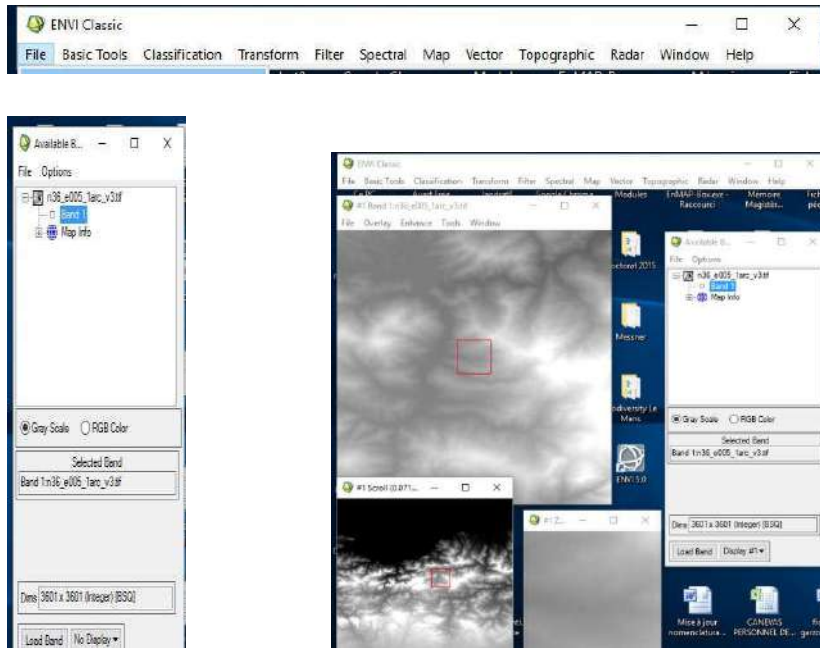
A partir de l'information altimétrique du MNT des cartes dites dérivées vont pouvoir être calculées, en particulier dans un objectif d'analyse morphologique du milieu: Carte des pentes, des orientations, d'ensoleillement, d'inter visibilité, bassins versants, profils en long et en travers, coupes de terrain.

2- Réalisation de la carte des classes des pentes :

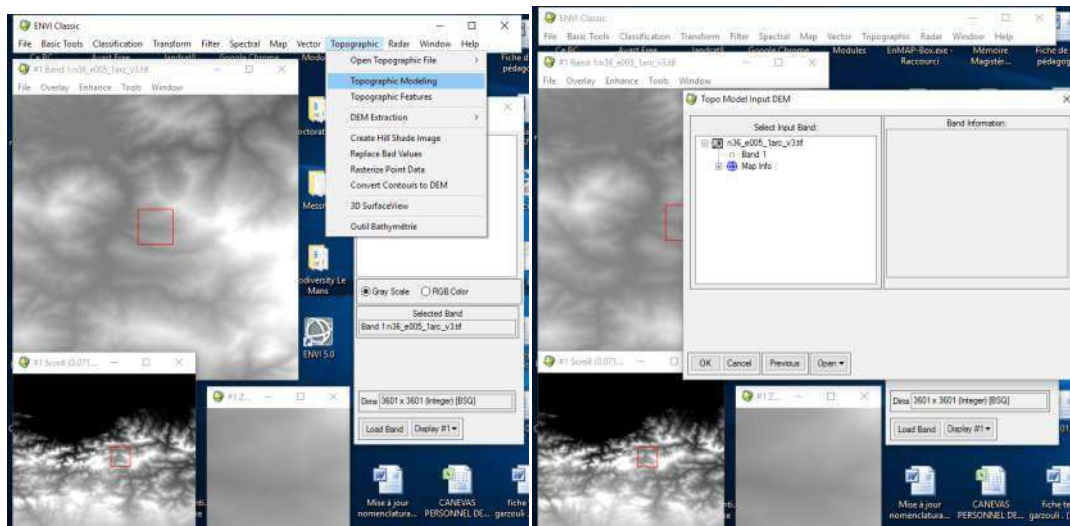
Etape 1 : Lancer le logiciel ENVI et ouvrir le MNT



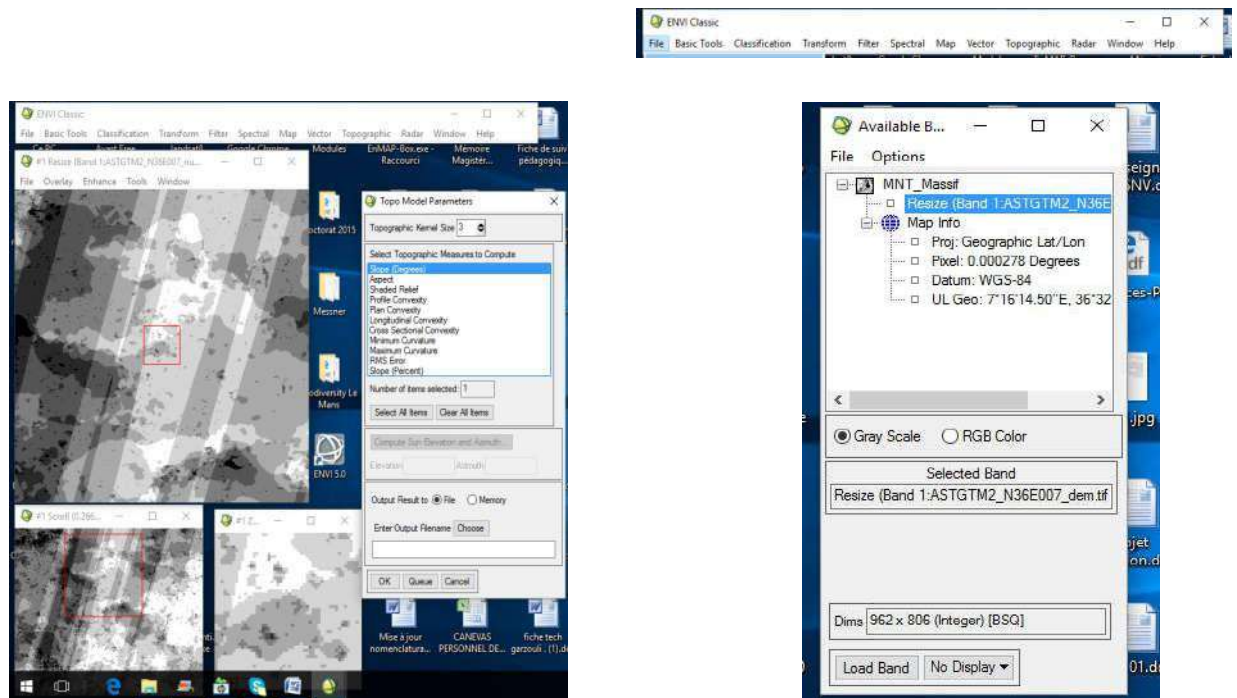
Travaux pratiques



Etape 2 : Aller a Topographic/topographic modeling/ slope

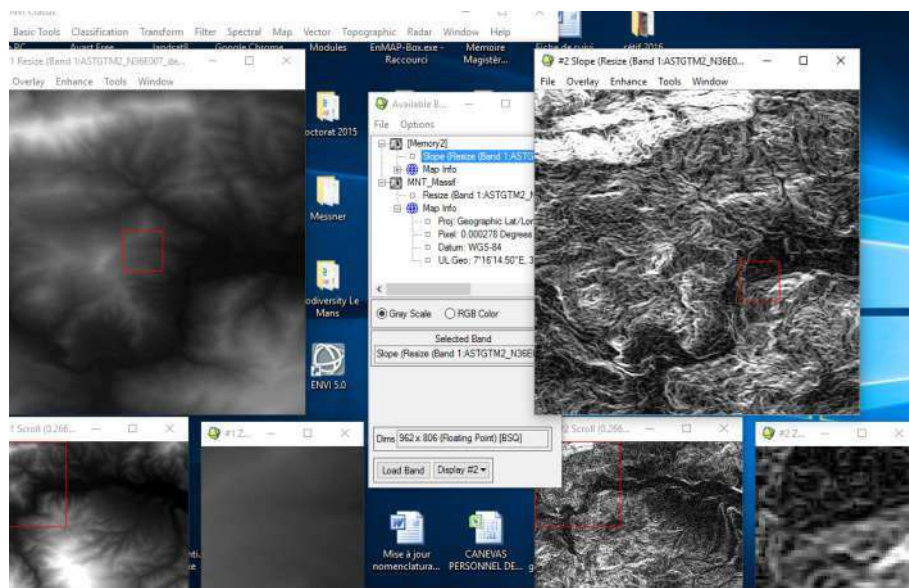


Travaux pratiques



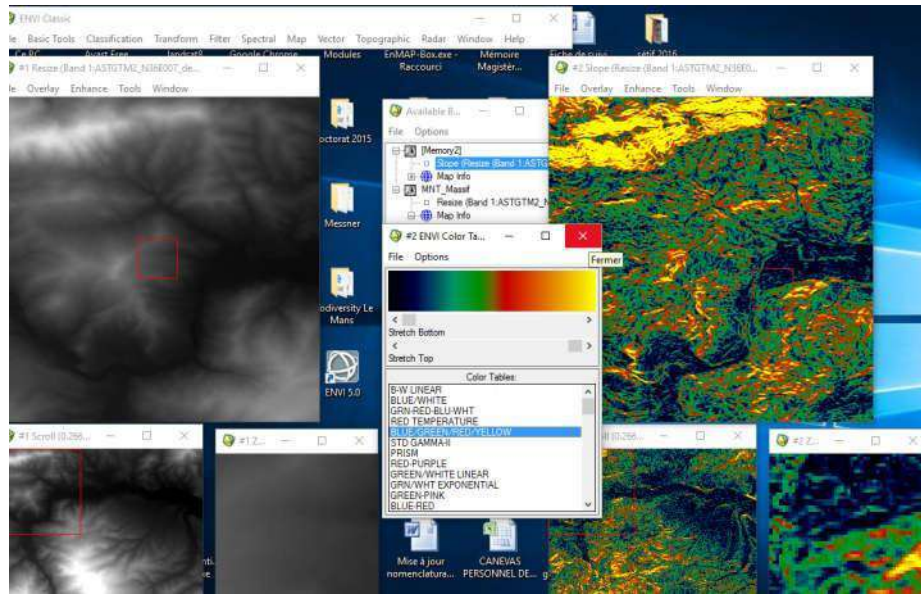
Résultat : Enregistrer le fichier sous le nom pente.

Etape 3 : Pour l'affichage de résultat : sélectionner le fichier/ load band

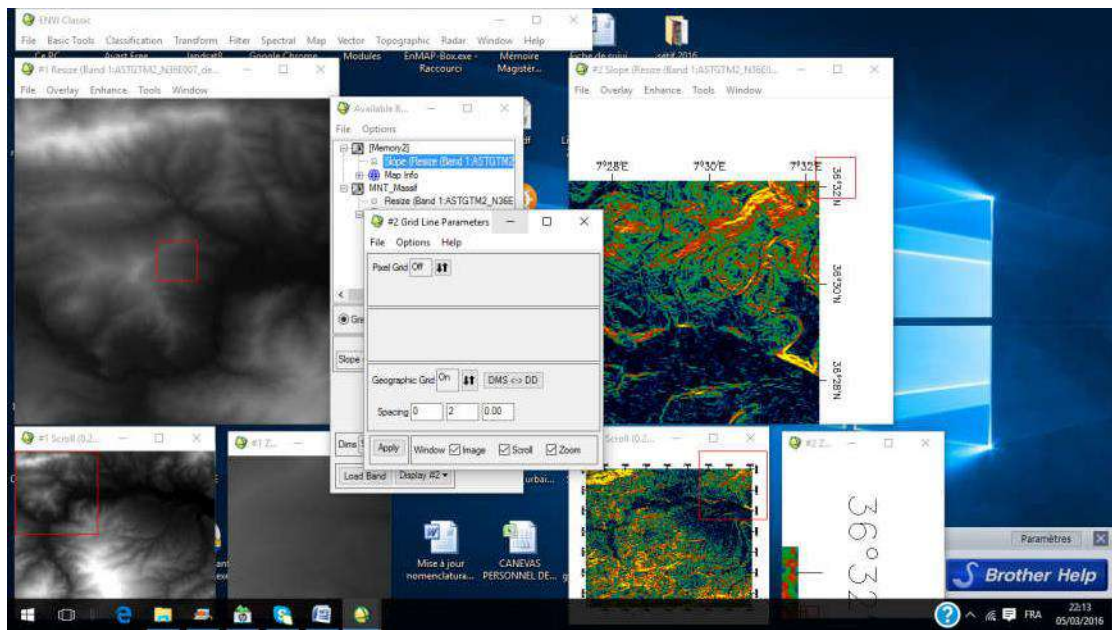


Etape 4 : Pour changement des couleurs de la carte : aller à Tools/ENVI Color Table/ sélectionner les couleurs

Travaux pratiques



Etape 5 : la mise en page de la carte : faire avec l'ENVI ou ArcGis.



Mise en page cartographique par ArcMap

Travaux pratiques

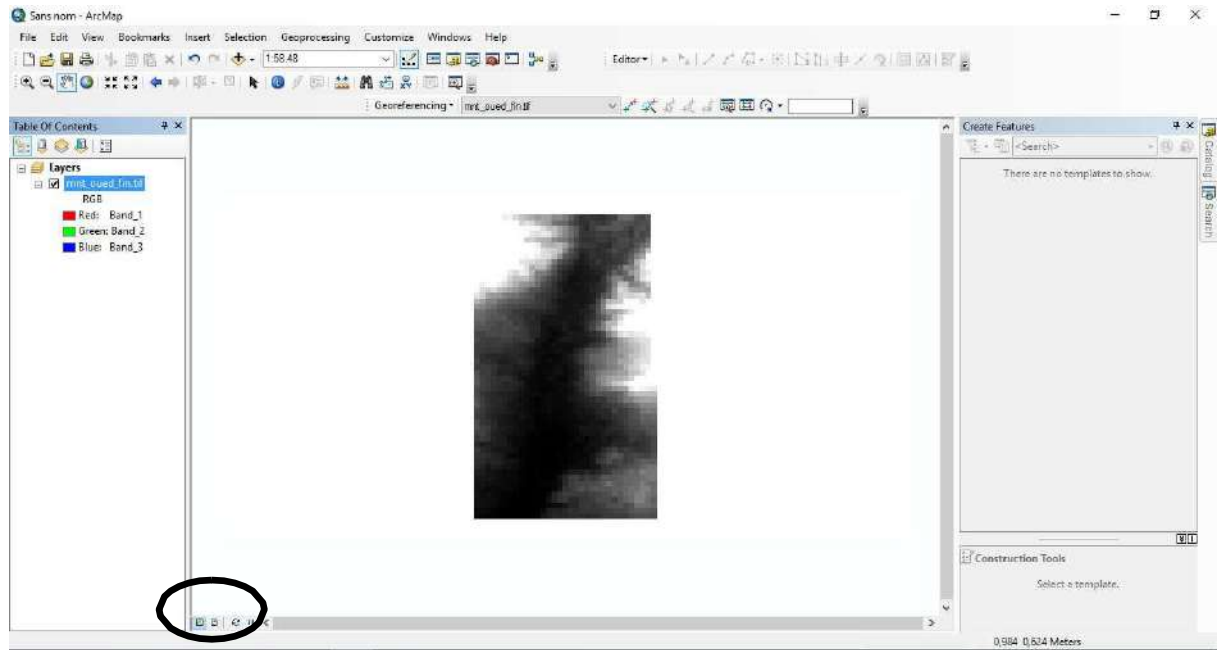


Figure 1 : Préparation d'une mise en page

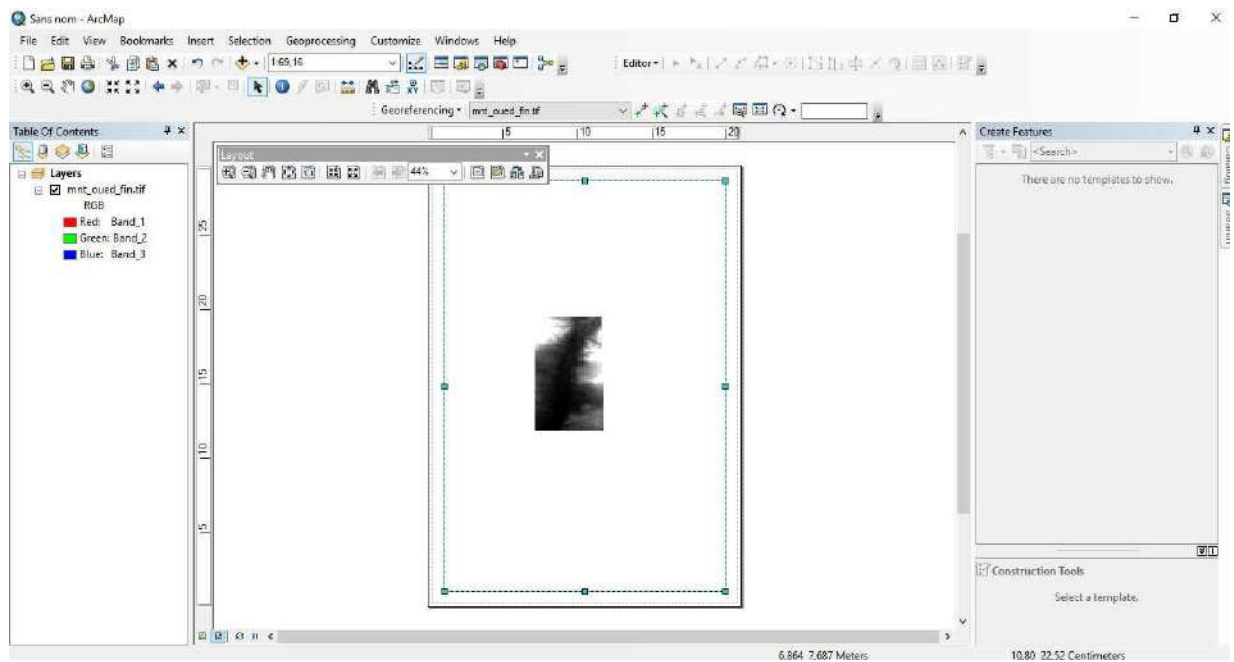


Figure 2 : cadre de la mise en page

Travaux pratiques

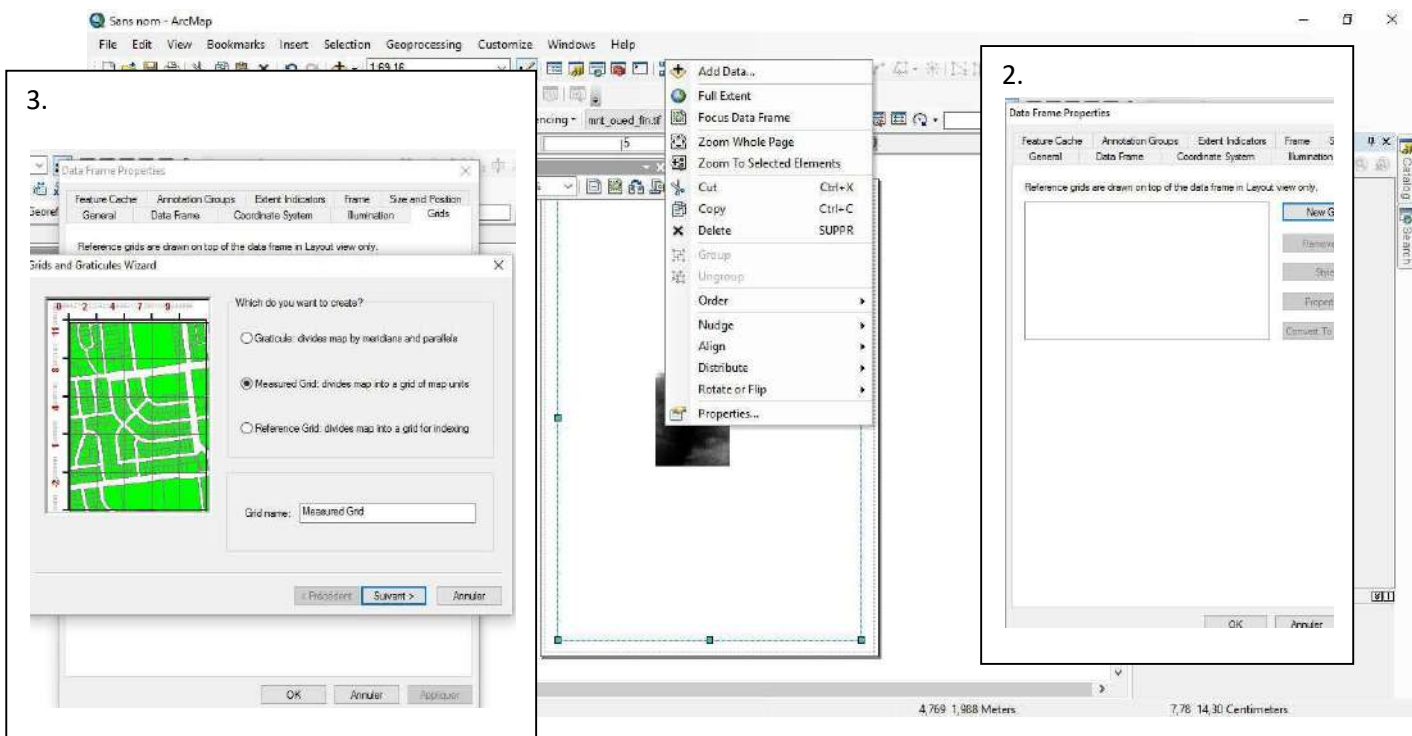


Figure 3 : Eléments de la mise en page

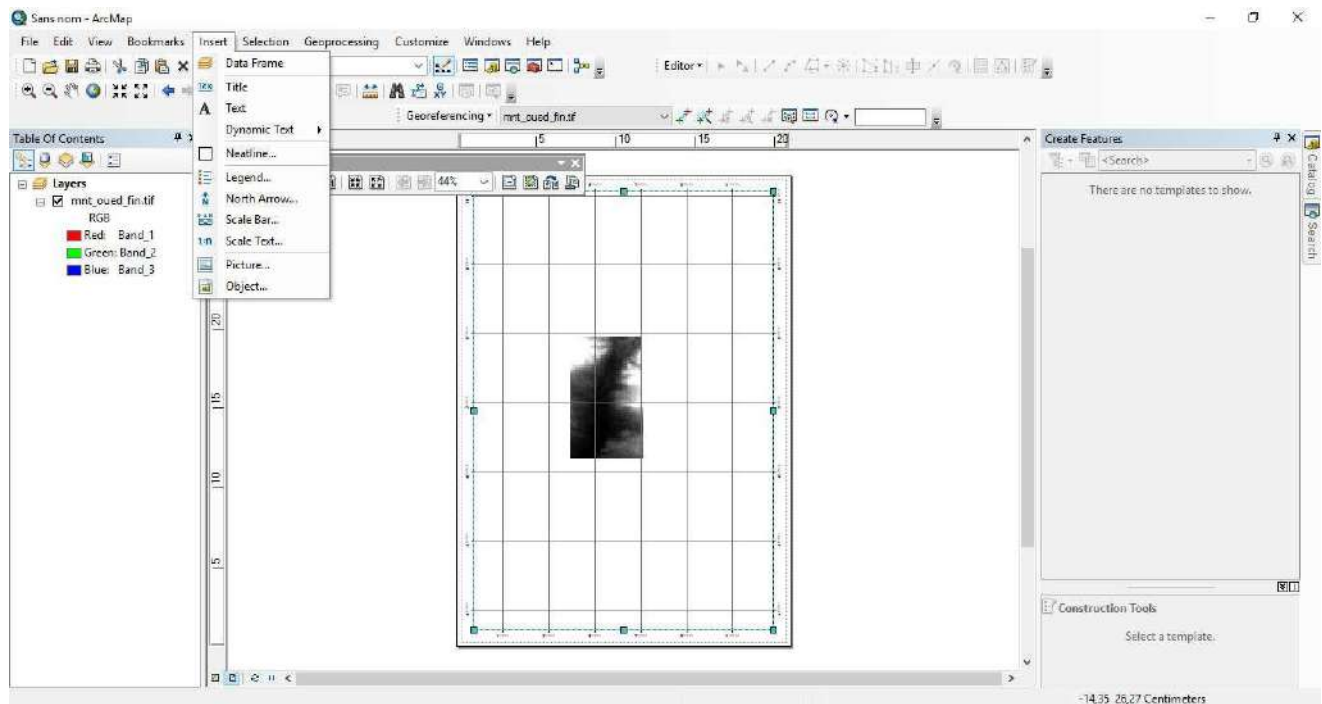


Figure 4