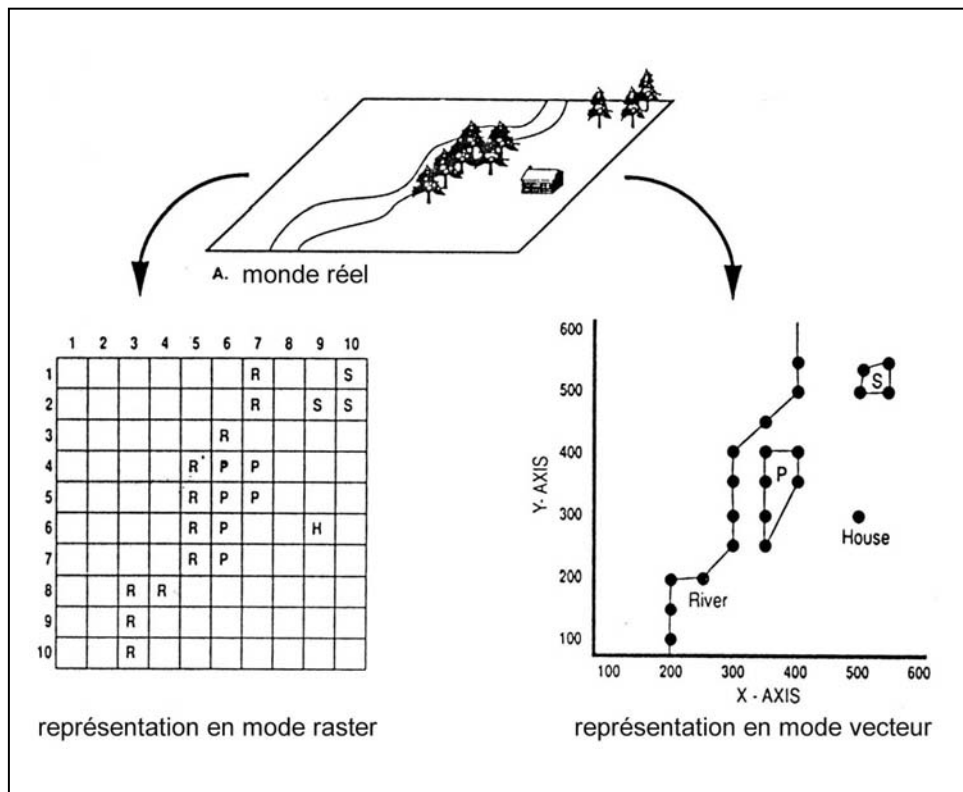


Comment représenter et structurer des données géographiques dans un environnement informatique ?

➤ Les différents modes / formats utilisés

Il existe **3** modes de base pour représenter, gérer et stocker des **données géographiques spatialisées** (cartes, plans, photographies aériennes, ...), dans un environnement **informatique** :

- le mode **raster** (représentation et gestion sous forme d'images),
- le mode **vecteur** (représentation et gestion sous forme de **points**, de **lignes** ou de **polygones**),
- le mode **TIN** (représentation et gestion sous forme de réseaux de triangles irréguliers - **Triangulated Irregular Network**).

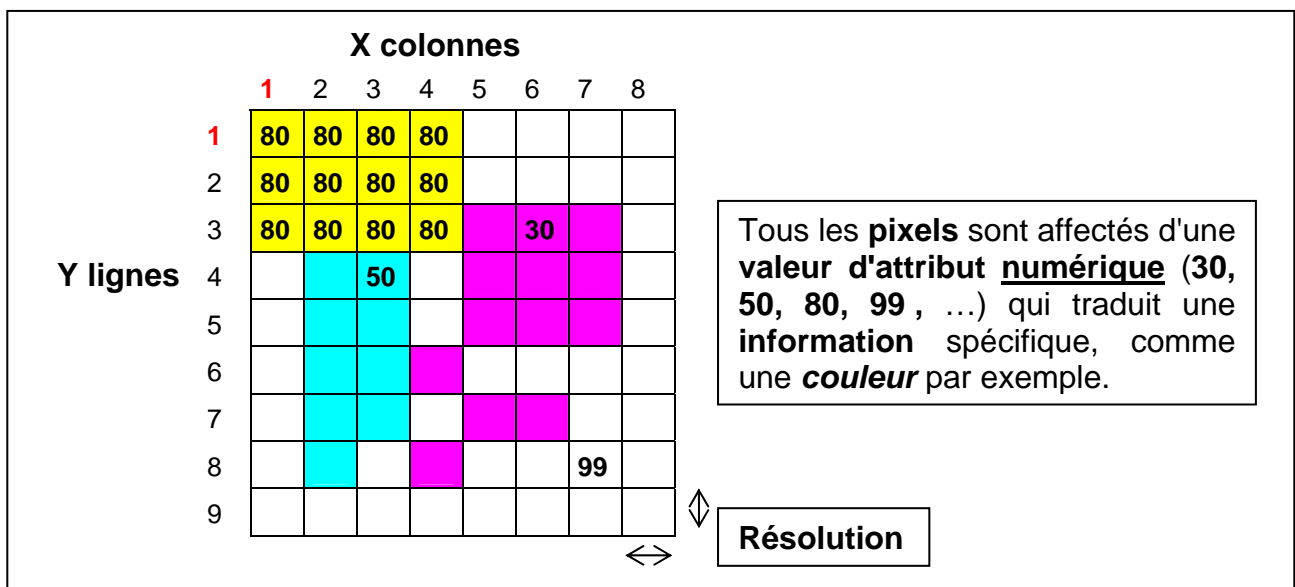


principe de représentation en formats raster et vecteur

1 – le mode raster ou « image »

Un **raster**, une **image numérique** c'est :

- une matrice de **cellules** organisées en un ensemble de **Y lignes** et de **X colonnes** : l'**origine géographique (0, 0)** est celle du **pixel Haut-Gauche (col 1, lig 1)** ;
 - les cellules, de même taille et carrées, sont appelées **pixels** (Picture Elements). Elles définissent autant de **surfaces élémentaires** de la scène observée permettant de transcrire la « réalité » en l'échantillonnant ;
 - ce mode d'**échantillonnage** du monde réel est le plus approprié pour analyser des phénomènes spatialement **continus** (la variation est continue dans l'espace, ex : variation de niveaux de gris, de couleurs ou d'altitude). La réalité va être **discrétisée** de manière régulière et de façon plus ou moins fine.
- (cf. les appareils photos numériques sont équipés de capteurs comptant de plus en plus de pixels (10M aujourd'hui) pour offrir des photos de plus en plus détaillées).



schématisation d'une image

2 éléments caractérisent une image :

- sa **résolution** (taille du pixel)
- le **type d'information** = **attribut** lié à la **valeur numérique** du pixel

Table de couleurs = LUT (Look-Up Table)

Valeurs numériques des pixels = **attribut** : ici codage couleur

The screenshot displays two windows from the TNTmips software. On the left is the 'Éditeur de palette de couleurs' (Color palette editor) window, which includes a grid of color swatches and a section for defining RGB values: Rouge: 36, Vert: 15, Bleu: 9. On the right is the 'Inspection de l'image' (Image inspection) window, which provides metadata for the image: 'Image: F:\SAUVEGARDE_FAC_180706\data_D\Projets télédec\LANGUEDOC_ROUSS', 'Lignes: 4200 Colonnes: 4980 Type: 8-bit non signé(s)'. It also shows a grid of pixel values, with a green crosshair highlighting a specific pixel with a value of 200. The main image area shows a false-color satellite image of a landscape with a green crosshair over a specific area.

Composition colorée d'une image aster à 15 m de résolution RVB 8 bits (non signé)

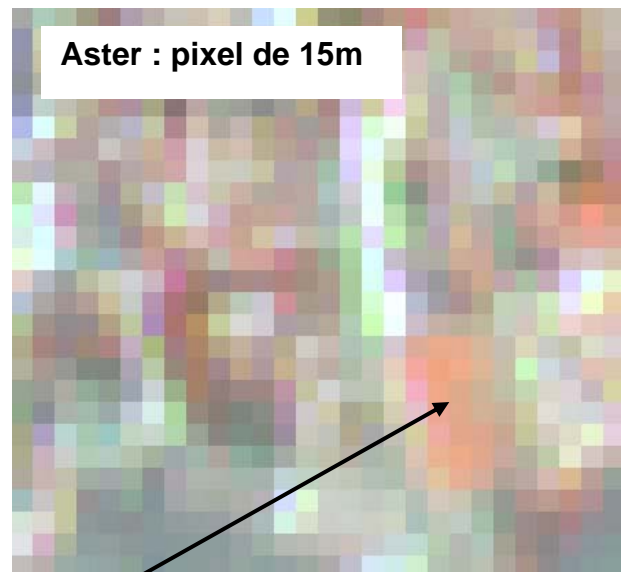
Visualisation sous logiciel de traitement d'image TNTmips

1 ⇒ La résolution :

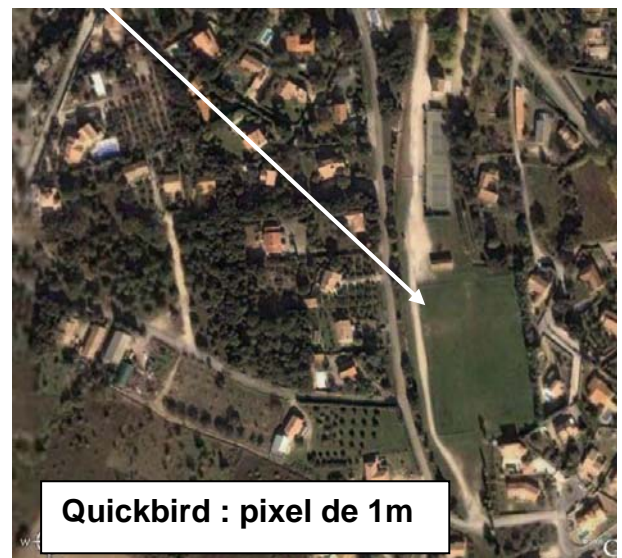
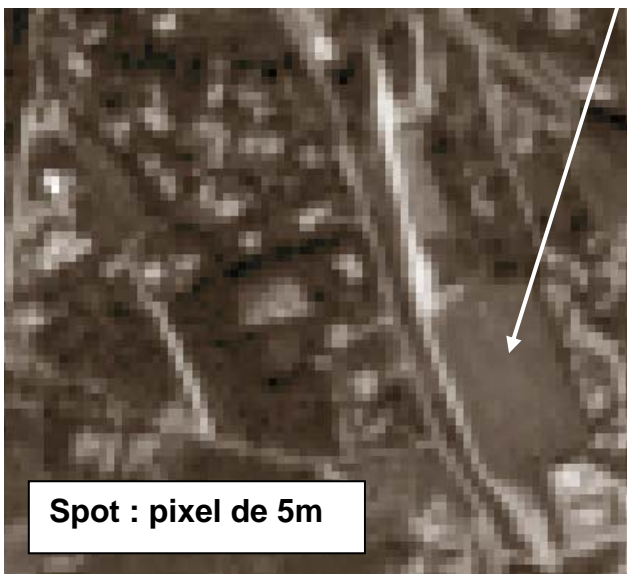
1-1 pour une image provenant d'un **capteur de télédétection** (ex image satellitale), la résolution exprime une **résolution spatiale mesurée au niveau du sol**.

On parle alors de pixel de 1 km (de coté), de 10 m, de 1m, ...

⇒ **Réalité-terrain échantillonnée** sous forme de surfaces de 1 km², 100 m², 1 m², ...



stade



Comparaison d'un même quartier observé par 4 satellites et à 4 résolutions différentes

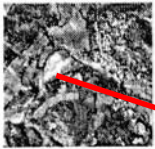


Visualisation de la ville de **Mauguio** sur **Google Earth** :

- à gauche résolution de **15 m (Landsat rééchantillonné)**
- à droite résolution de **1m (QuickBird rééchantillonné)**

1-2 pour une image provenant du **scan** d'un document papier (carte topographique ou photographie aérienne par exemple), la **résolution** choisie lors du scan correspond à un **nombre de pixels par unité de longueur**. La résolution s'exprime ici en **dpi** (**dots per inch** ou « points par pouce »), avec **1 pouce = 2.54 cm**

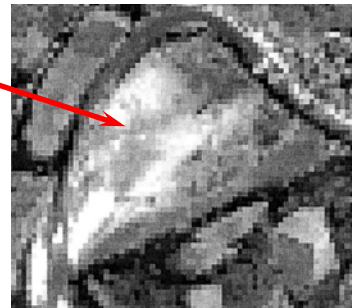
⇒ exemple de scans à **haute résolution** (à **400 dpi**) ou **basse résolution** (**50 dpi**)



scan à basse résolution : 50 dpi

50 pts / 2.54 cm => 1 pt = 0.5 mm
Image de 207 colonnes sur 200 lignes
- 228 Ko en format Tiff
- 130 Ko en format Jpeg

Zoom de l'image à 50 dpi



scan à haute résolution d'image : 400 dpi (400 dots per inch /

400 pts / 2.54 cm => 1 pt = 0.06 mm

Image de 1661 colonnes sur 1600 lignes

- 5 Mo en format Tiff (pas de compression)

- 1 Mo en format Jpeg (image compressée)

points par pouce (ppp)

pixel par pouce

2 ⇒ type d'information lié à la valeur numérique du pixel

Selon la **nature de l'image** la valeur numérique liée à chaque pixel permet de représenter et de moduler différents **attributs** tels que :

- un **code** :
 - **code couleur** pour une photo ou une carte scannée avec correspondance pour assigner les couleurs d'une **table de couleur (couleurs indexées)**,
 - **code de classe** sur une **image classée** (ex **10** = vignes, **20** = vergers, **30** = bâti), la **légende** étant alors contenue dans une table alphanumérique associée,
- une **valeur de luminance** (mesure d'une quantité d'énergie **réfléchi**e par des objets de la surface) ou une **T° de surface** (quantité d'énergie **émise**) ⇒ cas des **images-satellite**
- une **altitude** (pour les **Modèles Numériques de Terrain** ou MNT),
- une **valeur de pente**, une **orientation**, ...

Le **nombre de valeurs numériques différentes** qui peuvent être attribuées à un pixel dépend du **type codage** de l'image, les images numériques étant codées sous forme **binaire** (suite de **0** et de **1**) :

- le codage sur **1 octet (8-bits)**, le plus utilisé en traitement d'image, permet **256** (2^8) gradations possibles (**256 niveaux de gris, 256 couleurs ou 256 valeurs de luminance**),

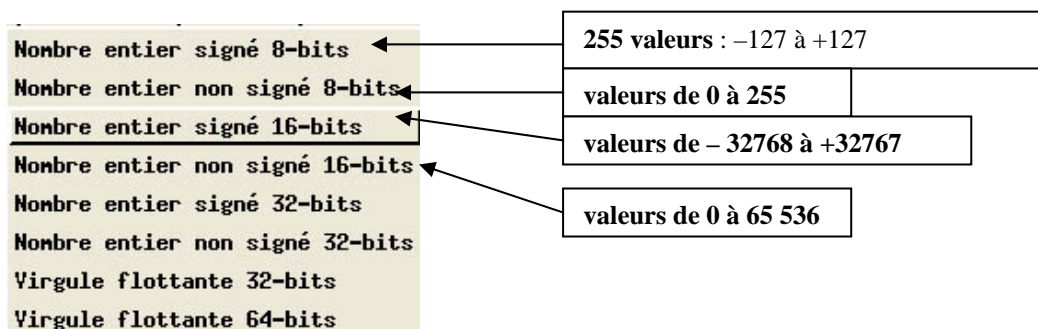
exemple d'écriture en binaire sur 1 octet :

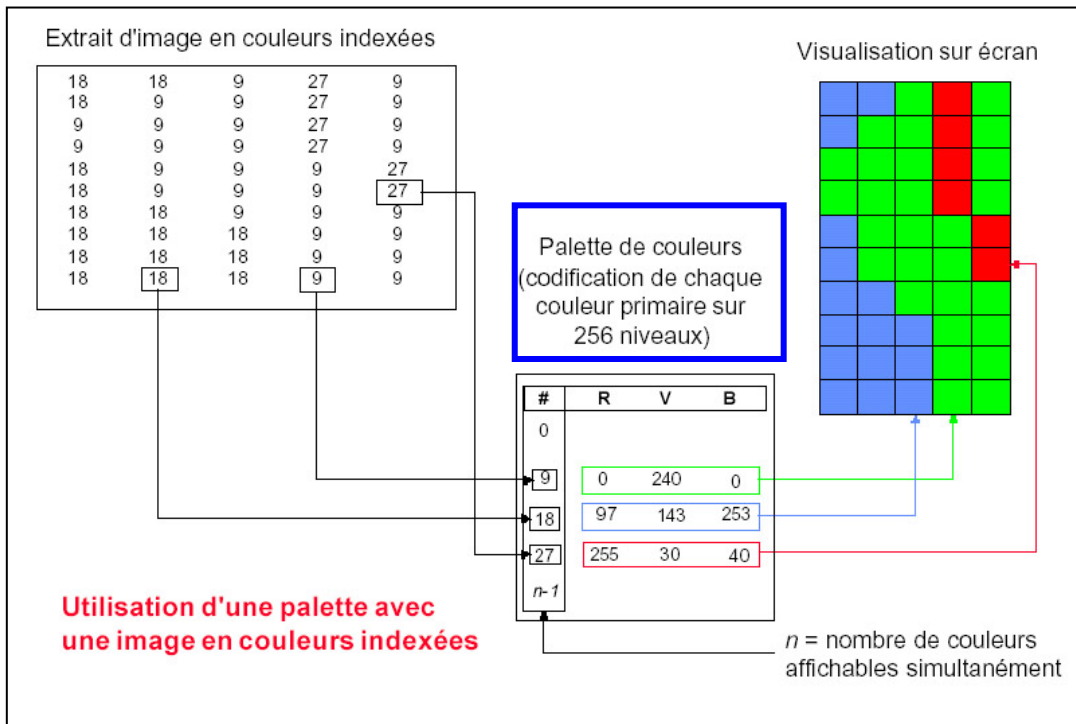
$$2 = 00000010$$

$$252 = 11111100$$

- le codage sur **2 octets (16-bits)** offre **65 356 niveaux** (2^{16}) (codage approprié pour les MNT, les images-radar),
- le codage sur **3 octets (24-bits)** offre **16 millions de valeurs** (pour les images en *vraies couleurs*)
- le codage sur 1 bit (**1/8 d'octet**) = **2 valeurs** : 0 ou 1 (pour les images binaires utilisées pour les masques en traitement d'image).

Exemple : type d'image proposée en sortie pour la réalisation d'un MNT (altitudes)





Le **nombre total de pixels** (= nbre col x nbre lig) et le **type de codage** déterminent la **taille** du fichier ou la **quantité de données** contenues dans l'image : Il faut parfois choisir entre la **taille** d'une image et sa **qualité**,

⇒ Possibilité de **compression** d'image, avec ou sans perte d'information.

⇒ **formats raster communs** : **TIFF, JPEG**, (permettent les **échanges** standards entre logiciels) :

- **Tiff (Tagged Image File Format)** : autorise **pleine résolution** mais grosses images ;
- **Jpeg (Joint Photographic Experts Group)** : format permettant de fortement réduire le poids d'une image en la compressant. Cette **compression** entraîne une **perte de qualité** (+ ou – selon le choix du niveau de compression). Supporte **16 millions de couleurs**.

⇒ **apparition des nouveaux formats ECW, MrSid, JPEG2000**. Ces formats autorisent des **taux de compression très importants** (*techniques de compression par ondelettes*) tout en préservant la **qualité** de l'information.

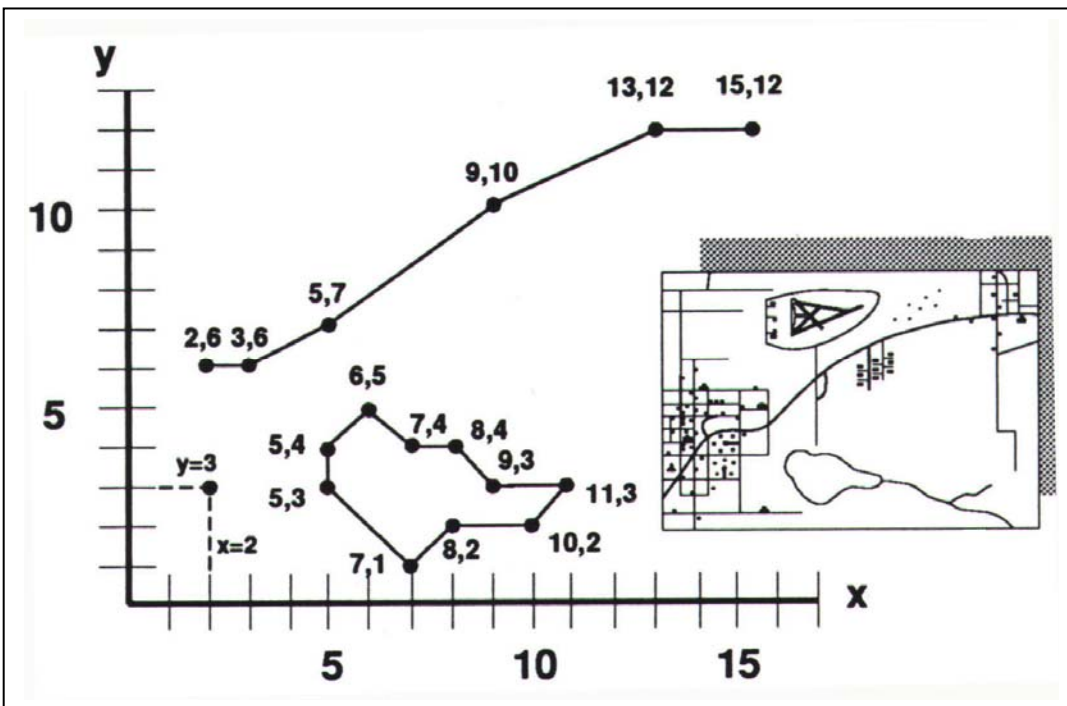
⇒ Certains logiciels subdivisent le mode **raster** en différenciant les **images** proprement dites et les **grilles (grides)** :

- on parle, dans ce cas, d'**image** pour tout ce qui concerne les photographies (aériennes), les images-satellite, les cartes numérisées, ... ,
- on parle de **grilles** pour désigner des données dérivées souvent utilisées pour l'analyse ou la modélisation. Les Modèles Numériques de Terrain (**MNT**) sont ainsi considérés comme des grilles car ils représentent un semis régulier de points côtés.

2 – le mode vecteur

Les données spatialisées sont représentées en mode vecteur sous la forme de **points**, de **lignes** ou de **polygones** : **composante géométrique** (/ **composante attributive**).

C'est le mode de représentation et de stockage privilégié des données de type **discret** : bâti, parcellaire, type d'occupation du sol, ...



Il existe 3 primitives graphiques :

- **point** ⇒ **entité ponctuelle** (suivant l'échelle) : ville à petite échelle (**attributs** : nom, nbre d'hab., taux de croissance, etc...), puit, ... ;
- **polyligne** ⇒ **entité linéaire** : route (**attributs** : trafic, revêtement, ...), faille, ...
- **polygone** ⇒ **entité surfacique** : quartiers d'une ville à grande échelle (**attributs** : noms, ...), parcellaire cadastrale, ...

Ces points, lignes ou polygones sont définis dans un **système cartésien** (représentation sur un **plan**) par un ensemble de **coordonnées x, y** (et **z** si on ajoute l'altitude). Un jeu de **coordonnées x,y** représente un **sommet** (ou un **nœud**).

Les valeurs des coordonnées varient selon le système de coordonnées dans lequel sont gérées les données.

Les **attributs** associés aux entités spatiales sont stockés dans des **tables** de données descriptives : **composante attributive**

Fond carto vectoriel avec tables associées

Tables associées aux polygones de l'objet vectoriel

The screenshot displays a GIS application window titled 'Vue de l'éditeur de données spatiales 1'. The main map area shows a vector cartographic background of Europe. A pink polygon highlights a specific region. Two callout boxes point to the map: one to the background and another to the pink polygon.

The 'Éditeur des données spatiales' window is open, showing a toolbar and a list of layers. The 'POLYSTATS' layer is selected, and its properties are displayed in a separate window titled 'Frontière_3 / PolyData / POLYSTATS'. This window shows a table of statistics for the selected polygon.

The 'Frontière_3 / PolyData / Pays' window is also open, showing a table of countries associated with the polygon. The table has two columns: 'NUM' and 'PAYS'. The country 'Belgique' (row 74) is selected, indicated by a red dot in the 'NUM' column.

NUM	PAYS
69	Ukraine
70	Moldavie
71	Biélorussie
72	Pologne
73	Allemagne
74	Belgique
75	Pays Bas
76	Lituanie
77	Lettonie
78	Éthiopie
79	Estonie
80	Yougoslavie
81	Tchécoslovaquie
82	Autriche
83	Suisse
84	Croatie
85	Roumanie
86	Bulgarie
87	Grèce
88	Zimbabwe
89	Swaziland
90	République Centrafricain
91	Koweït
92	Sahara Occidental
93	Italie
94	Hongrie

Attaché	Enregistrement 927 de 3008 (1/1 attaché)	
Area:	32794,43812066	km ²
BoundLen:	1309,91505636	kn
CentX:	-1659835,59724124	
CentY:	5392826,76388564	
AreaIncI:	32794,43812066	km ²
BoundNotIncI:	1309,91505636	kn
CentXNotIncI:	-1659835,59724124	
CentYNotIncI:	5392826,76388564	
PointInPolyX:	-1645459,26062619	
PointInPolyY:	5378171,50642580	
CompactRatio:	2,04050707	
CompactRatioI:	2,04050707	

⇒ Attributs, Tables, Base de Données : notions

Les informations descriptives que l'on va pouvoir relier à un objet géographique sont contenues dans une **base de données** qui représente une collection d'informations hétérogènes mais structurées.

Une base de données est aujourd'hui gérée à l'aide d'un **Système de Gestion de Base de Données (SGBD)**.

Actuellement, la structuration et l'interrogation des données se font essentiellement dans le cadre des **Systèmes de Gestion de Base de Données Relationnelle (SGBDR)**.

Vocabulaire

- Relation = table :

Une relation, une table, se définit par un nom, une liste d'attributs ainsi que par les "individus" qu'elle contient. C'est un ensemble de données relatives à un même concept.

Les individus sont généralement représentés en lignes et les attributs et en colonnes.

Exemple : la relation = table *Exploitations*

5 colonnes = 5 **champs** correspondant à 5 **attributs** différents (**un identifiant**, info thématiques, ...)

Exploitations				
NumExp	Nom_Exploita	INSEE	SAU	Quota_lait
1	Nom1	95183	93	0
2	Nom2	95535	335	0
.
29	Nom29	95213	87	0
31	Nom31	78317	108	0

31 lignes d'enregistrements

- Règle d'unicité :

une table ne peut pas avoir deux enregistrements identiques, c'est à dire qui auraient la même valeur pour chacun des attributs. Le rôle d'un SGBD est de garantir cette unicité des enregistrements, grâce à la notion de **clé**.

- Clé primaire d'une table (ou clé) :

c'est le plus petit ensemble d'attributs (un ou plus) pour lequel la règle d'unicité est respectée : notion d'**identifiant**, chaque enregistrement étant identifié sans aucune ambiguïté.

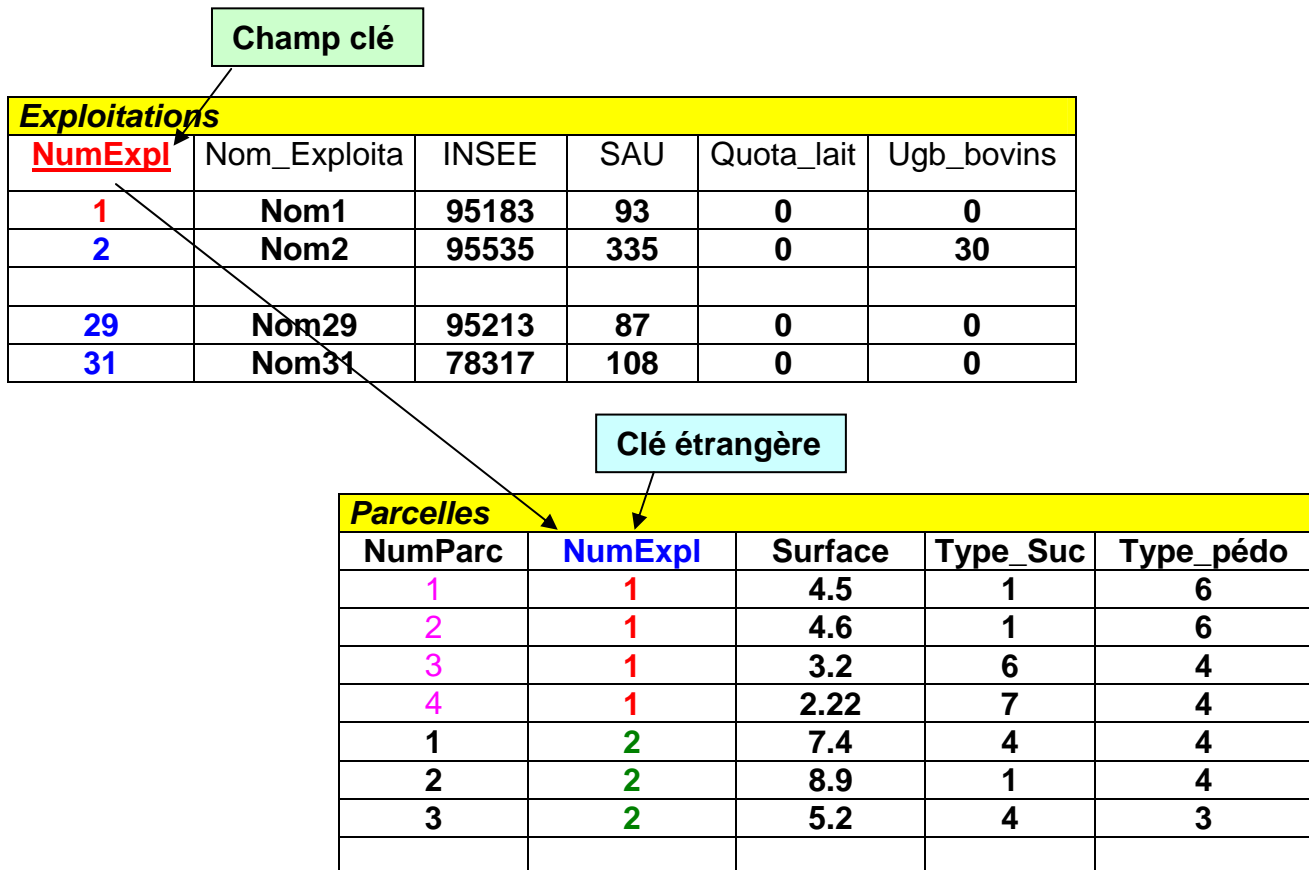
Par convention, les attributs clé sont soulignés et il faut toujours porter une grande attention à la définition de la clé d'une relation.

- Clé étrangère (ou clé externe) d'une table :

c'est un sous-ensemble d'attributs qui renvoie à la clé primaire d'une autre table. **La clé primaire et la clé étrangère doivent contenir la même information** même si le nom des champs peut être différent.

⇒ **concept fondamental des bases de données relationnelles** qui autorisent la mise à disposition de grandes quantités de données réparties sur des tables différentes que l'on peut **relier** entre-elles

exemple :



NumExpl correspond à une **clé étrangère** dans la table **Parcelles** car elle est liée à l'attribut **NumExpl** dans la table **Exploitations**.

- Requêtes SQL :

Le langage **SQL** (Structured Query Language = *Langage de requêtes structuré*) est un langage de recherche et de mise à jour d'une base de donnée. Il permet notamment de rechercher des entités en fonction de leurs attributs

Exemple de **requête SQL** permettant une **sélection** sur la table **Exploitations** :

```
SELECT *
FROM Exploitations
WHERE SAU > 250
```

On veut **SELECTIONNER** à **PARTIR** de la table **Exploitations** tous les enregistrements **où** la SAU est > à 250 Ha

Type d'opérations possibles sur des tables au travers de requêtes SQL

➤ Projection (résultat : relation/table)

- Extraction des données d'une table relatives à un ou plusieurs attributs (colonnes) : le nom et l'adresse d'un client

➤ Sélection

- Sélection des lignes (enregistrements) qui respectent certaines conditions : les emprunts dont la date de retour est dépassée

➤ Jointure

- Rapprochement de deux tables A et B pour créer une table C selon des critères qui associent des lignes de A et de B: tous les livres d'un auteur particulier

➤ Union

- Combiner deux tables A et B pour créer une table C qui contient toutes les lignes appartenant à l'une ou l'autre des tables de départ : tous les clients et tous les auteurs

➤ Différence

- Combiner deux tables A et B pour créer une table C qui contient toutes les lignes de A non contenues dans B: les auteurs qui ne sont pas des clients

➤ Intersection

- Combiner deux tables A et B pour créer une table C qui contient toutes les lignes appartenant à A et B : les auteurs qui sont clients

➤ Produit

- Combiner chaque ligne d'une table A avec l'ensemble des lignes d'une table B: les livres empruntés par chaque client

➤ Division

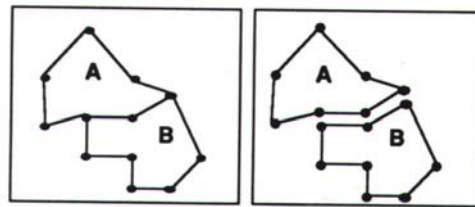
- Récupérer les lignes d'une table A qui sont mises en correspondance avec toutes les lignes d'une table B : les livres empruntés par tous les clients

La **gestion** et l'**analyse** des entités vectorielles peuvent se faire en gérant ou non les **relations topologiques** pouvant exister entre les différentes entités d'une couche vectorielle (ex les différents polygones représentant un type d'occupation du sol) :

- 1) mode non topologique = **modèle spaghetti** (le plus simple) ;
- 2) modes **topologiques de réseau, planaire** ou **de surface** (plus complexes à mettre en œuvre mais permettant des analyses sophistiquées)

1 - le mode spaghetti (mode non topologique)

⇒ Format utilisé principalement dans les logiciels de **DAO** (plans, cadastres, ...) et format natif des entités vectorielles gérées par les logiciels de SIG **MapInfo** ou **ArcView** (niveau de base de **ArcGis**) .



2 polygones adjacents représentés en mode spaghetti

Ce format permet de construire, cartographier et analyser des couches vectorielles peu complexes (nécessite une construction "propre").

En revanche, il présente plusieurs inconvénients majeurs :

- la redondance d'information (ex bordures de polygones communes) rend difficile les **opérations de modification et de mise à jour** ;
- l'absence de relations topologiques ne garantit pas la **cohérence des données** et rend plus difficile certaines **analyses spatiales** (calcul d'adjacence, de connectivité, agrégation ou intersection de couches).

→ Format des fichiers de forme (*shape*) dans ArcView (ArcGis)

Utilisations :

- Cadastre
- Plans
- ...

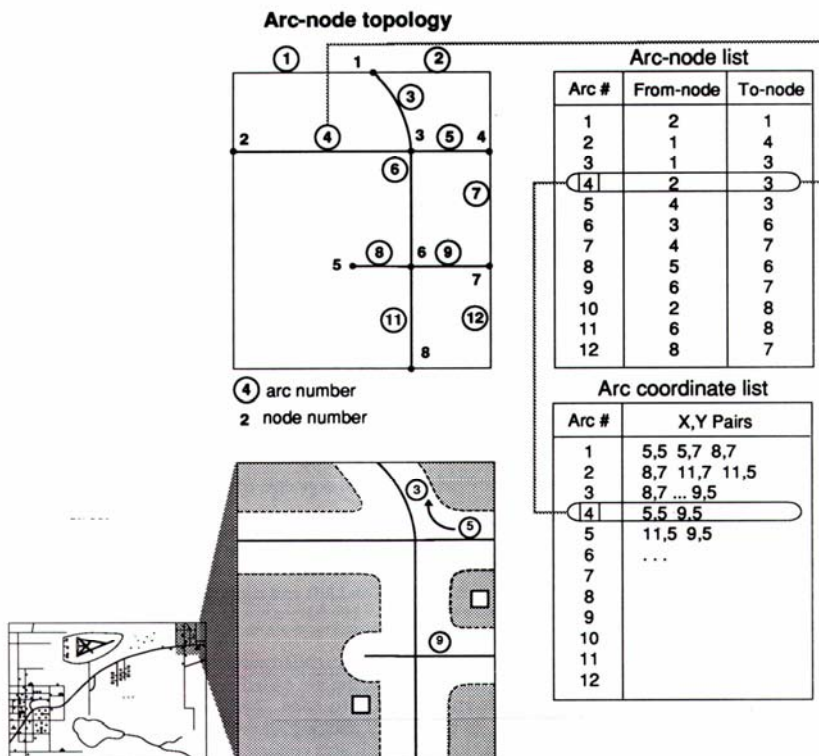
2 - les modèles topologiques

Topologie : la topologie ou « algèbre de carte » décrit et structure les **relations** existant entre les unités spatiales d'un **graphe**, la construction et l'analyse d'un graphe reposant avant tout sur 2 primitives graphiques indissociables : le **nœud** et l'**arc**.

- l'**arc** : **ligne brisée (polyligne)** dont la forme est définie par les **coordonnées x_i, y_i** de ses **sommets** ;
- le **nœud** : **sommet** de **début** et de **fin** d'un **arc**. Plusieurs arcs peuvent se rejoindre ou se croiser à un même nœud.

Exemple de **relations topologiques** entre **entités spatiales** :

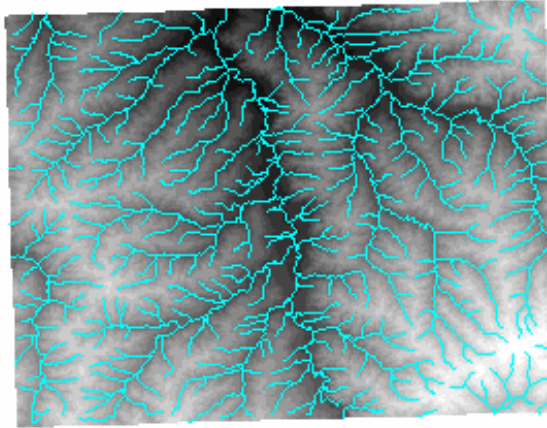
- un arc est **orienté** du nœud N1 vers le nœud N2 ;
- les arcs A1 et A2 sont **connexes** (ils partagent un même **nœud**) ;
- un arc limite à sa droite le polygone P1, à sa gauche le polygone P2 (les polygones sont formés d'une succession d'arcs orientés) ;
- les polygone P1 et P2 sont **adjacents** (ils partagent un même arc) ;
- un polygone est inclus dans un autre polygone (**gestion des îles**).



modèle de topologie planaire (topologie arc-nœud)
(les arcs 3, 4, 5 et 6 se rejoignent au niveau du nœud 3).

2.1 le mode topologique planaire

Le vecteur en mode planaire intègre la notion de nœud et des relations topologiques entre les lignes et les nœuds. Il y a donc orientation des lignes ce qui permet de construire un réseau hydrographique par exemple, où les bras sont orientés et reliés entre eux.



Exemple d'une topologie de type planaire

Caractéristiques du vecteur en mode planaire :

- Présence de nœuds
- **Intersection automatique des lignes**
- **Pas de notion de polygones**
- Relations lignes / nœuds
- Attributs possibles sur lignes / nœuds / points
- **Notion d'orientation des lignes**
- Possibilité d'avoir des nœuds intermédiaires

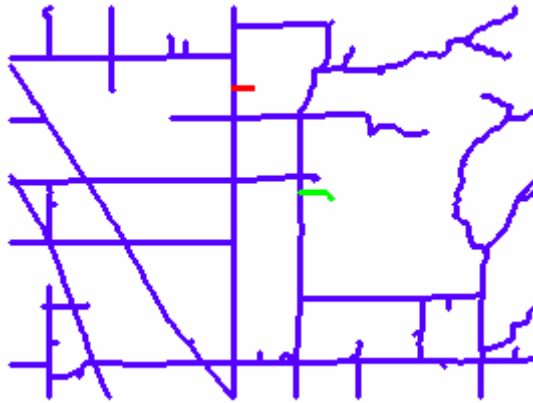
Utilisations :

- **Modélisation de réseau hydrographique**
- **Courbes de niveaux**
- ...

2.2 le mode topologique réseau

Le vecteur en mode réseau ressemble beaucoup au mode planaire, car il intègre la notion de nœud et des relations topologiques entre les lignes et les nœuds. En revanche, il permet de croiser les lignes sans obligatoirement créer un nœud.

C'est pourquoi il est typiquement utilisé pour les réseaux routiers : deux routes qui se croisent via un pont ne permettent pas de passer de l'une à l'autre, il ne faut donc pas de nœud entre les lignes qui les matérialisent.



Exemple d'une topologie de type réseau

Caractéristiques du vecteur en mode réseau :

- Présence de nœuds
- Intersection **non automatique** des lignes
- Les lignes **peuvent se couper ou passer au-dessous / au-dessus d'une autre**
- **Pas de notion de polygones**
- Relations lignes / nœuds
- Attributs possibles sur les lignes / nœuds
- Notion d'orientation des lignes
- Possibilité d'avoir des nœuds intermédiaires

Utilisations :

- Réseau routier
- Réseau électrique
- Lignes aériennes
- ...

3) le modèle topologique de surface

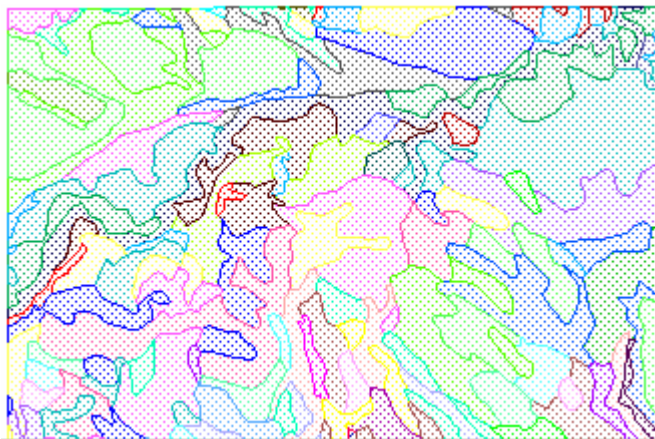
Le vecteur en mode polygonal intègre la notion de nœud et des relations topologiques entre les lignes, les nœuds, et les polygones. Il y a donc orientation des lignes et notion de polygone à droite, et polygone à gauche d'une ligne. Ce mode est le plus complexe des modes topologiques, mais aussi celui qui permet le plus de possibilités d'analyse.

Caractéristiques du vecteur en mode polygonal :

- Présence de nœuds
- Intersection automatique des lignes
- Relations polygones / lignes / nœuds
- Attributs possibles sur polygones / lignes / nœuds / points
- Notion d'orientation des lignes
- Possibilité d'avoir des nœuds intermédiaires

Utilisations :

- Occupation du sol
- Géologie
- ...



Exemple d'une topologie de type polygonal

→ Topologie des polygones = Format des couvertures dans ArcInfo (ArcGis)

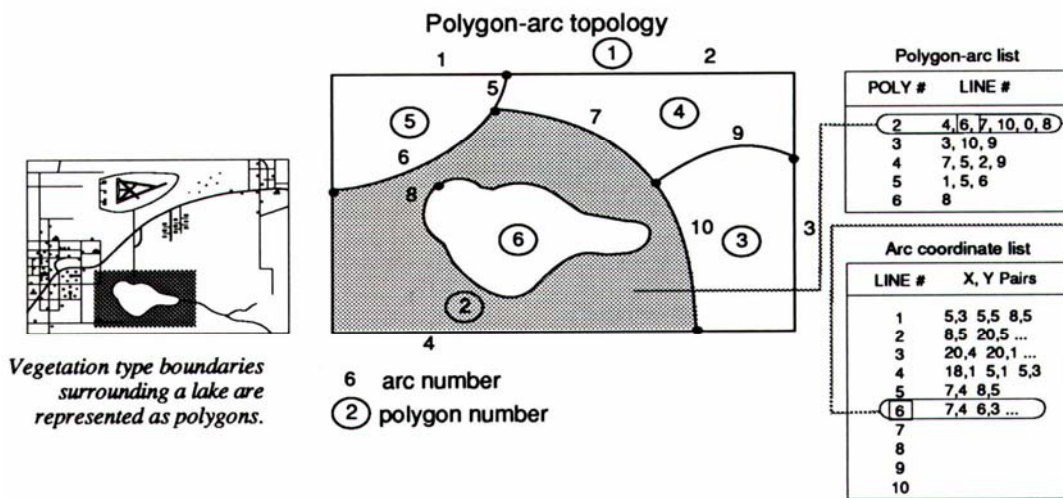
exemple de gestion de la topologie des polygones sous ArcInfo

- chaque **polygone** est constitué d'une série d'arcs orientés et connectés (nœuds de départ et d'arrivée confondus) => **topologie polygone-arc (1)**
- chaque **arc orienté**, délimite 2 faces (faces gauche et droite *RPOLY* et *LPOLY*) = **Adjacence** (par arc commun) => **topologie gauche-droite (2)**.

Stockage des informations de **connectivité** et **contiguïté**

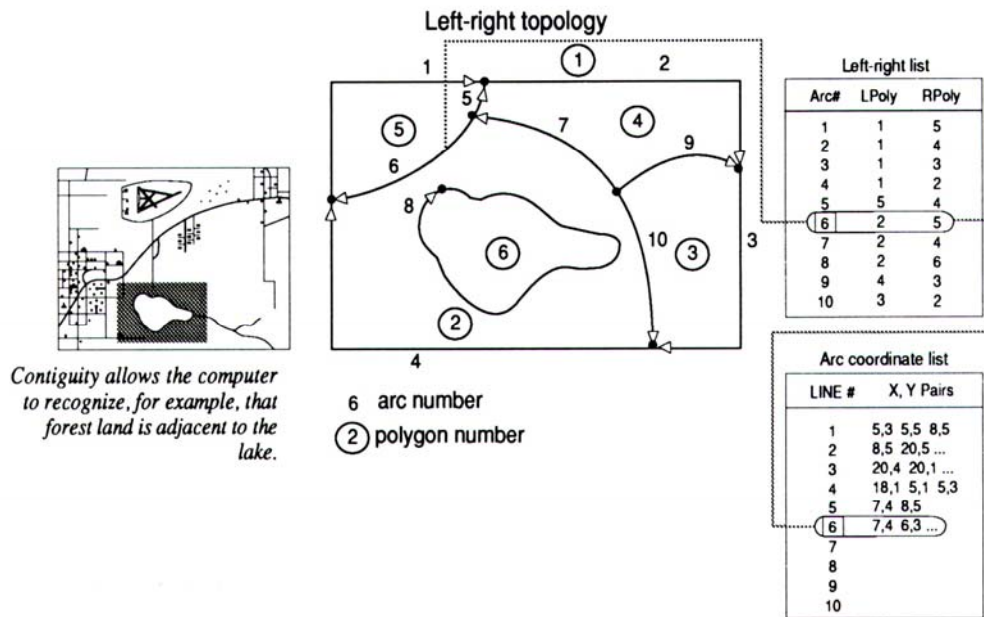
2 polygones ne peuvent se superposer mais ils peuvent s'imbriquer (gestion des îles)

(1) topologie polygone-arc



Les arcs 4, 6, 7, 10 et 8 délimitent le **polygone 2** (arc 8 crée un îlot = **polygone 6**, d'où le 0 avant 8 ds polygon-arc list)

(2) topologie gauche-droite



3 - le mode TIN (Triangulated Irregular Network)

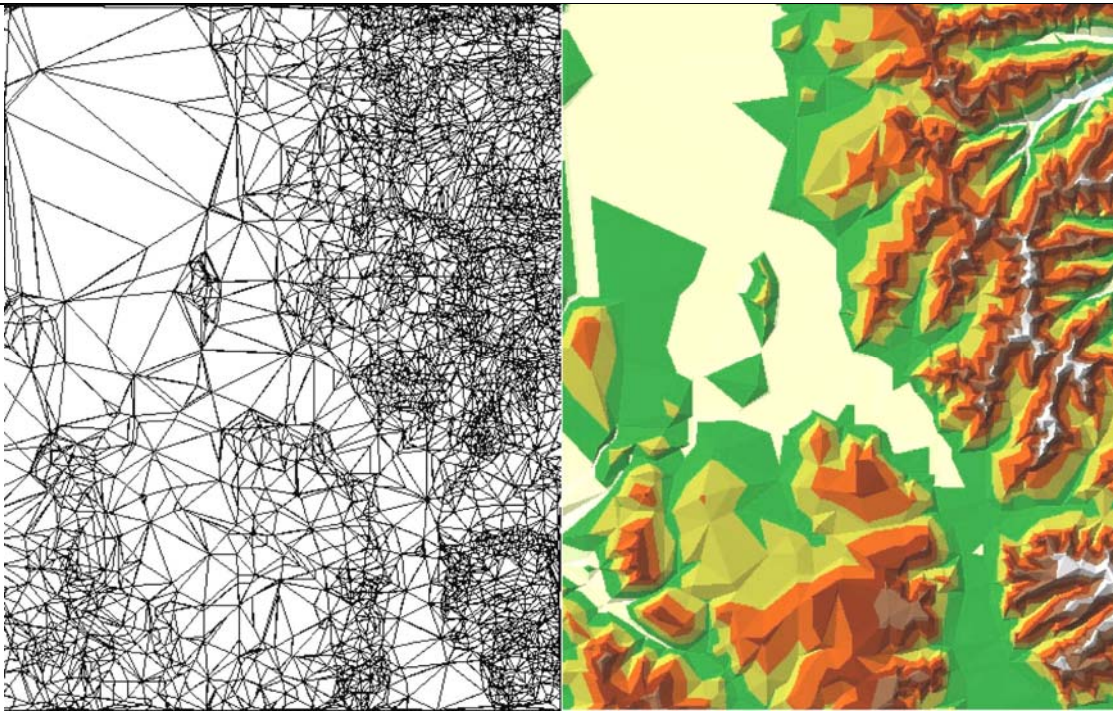
Dans un réseau de triangulation irrégulier, on représente une **surface** par un réseau de triangles liés et tracés entre un **semis de points** - de coordonnées (x, y, z) - situés à des distances variables les uns des autres.

A chaque triangle constitué (ou **facette**) sont associés les attributs suivant :

- sa **pente** ;
- son **azimut** (angle de la face avec le nord géographique)
- pour chaque arête du triangle, le **triangle à gauche** et le **triangle à droite**.

Un TIN est souvent une étape intermédiaire entre un semis de point irrégulier et la construction d'un MNT.

Les surfaces topographiques hétérogènes qui varient beaucoup dans certaines zones et moins dans d'autres sont modélisées avec plus de précision, pour un certain volume de données, dans un **TIN** que dans une **grille** (avantage d'un semi de points irrégulier sur une maille régulière)



exemple de représentation d'un relief par un TIN

Un **TIN** permet : calcul de pente, exposition, écoulement, intervisibilité, profil le long d'itinéraires, drapage de plans d'information