

# Les DONNEES dans les SIG Ere de l'infobésité

# Volume des données

- Le volume des données: Le **type de données** ainsi que le **format de données** retenus pour la base de données ont pour conséquence une **très grande variation de son volume**. Les projets SIG nécessitent généralement un matériel permettant la gestion d'un grand nombre de données et d'objets.



## Les 5 A:

**-Abstraire:** revient à concevoir un modèle qui organise les données par composants géométriques et par attributs descriptifs ainsi qu'à établir des relations entre les objets.

**-Acquérir:** revient à alimenter le SIG en données. Les fonctions d'acquisition consistent à entrer d'une part la forme des objets géographiques et d'autre part leurs attributs et relations.

**-Archiver:** consiste à transférer les données de l'espace de travail vers l'espace d'archivage.

**-Analyser:** permet de répondre aux questions que l'on se pose.

**-Afficher:** pour produire des cartes de façon automatique, pour percevoir les relations spatiales entre les objets, pour visualiser les données sur les écrans des ordinateurs.

# Procédure SIG

## 1-Le dossier d'analyse des besoins des utilisateurs.

Une enquête approfondie est nécessaire pour identifier les besoins des utilisateurs métier. La difficulté est d'anticiper les fonctionnalités avant même que l'outil soit disponible.

Préparation du terrain

Saisie

Control

## 4-La mise en œuvre du produit.

Cette quatrième étape doit aboutir à la réalisation du projet, c'est-à-dire à l'intégration des outils, la structuration des données, la mise en place des procédures d'exploitation, etc... A ce stade on aboutit au **SIG en tant qu'outil**. Une phase de production permet d'enrichir le SIG avec tout type de données. Le SIG devient alors opérationnel.

## 2-Le dossier de conception.

Cette deuxième étape a pour objectif de spécifier les fonctions du produit ou du procédé, les normes de qualité, les procédures de contrôle et l'ensemble des fonctionnalités qui correspondent aux besoins. Cette phase aboutit à un dossier de conception, incluant la modélisation du phénomène ou du produit et le rapport de définition détaillé.

Exploitation

## 3-Le dossier de réalisation.

Cette troisième étape vise à préciser quels seront les outils, techniques, logiciels, matériels, formations, normes et tests qui permettront d'assurer les fonctions définies dans l'étape de conception. Cette phase se concrétise par un dossier de réalisation comprenant le rapport d'étude technique, les cahiers des charges et la proposition d'un planning.

Echange

# Provenance des données attributaires

1- Auprès d'**organismes officiels** nationaux ou internationaux producteurs ou revendeurs (*INCT Algérie, IGN France...*).

2- Auprès de **producteurs locaux**, cabinet de géomètres, sociétés de services, service de l'Etat, collectivités territoriales, concessionnaires de réseaux.

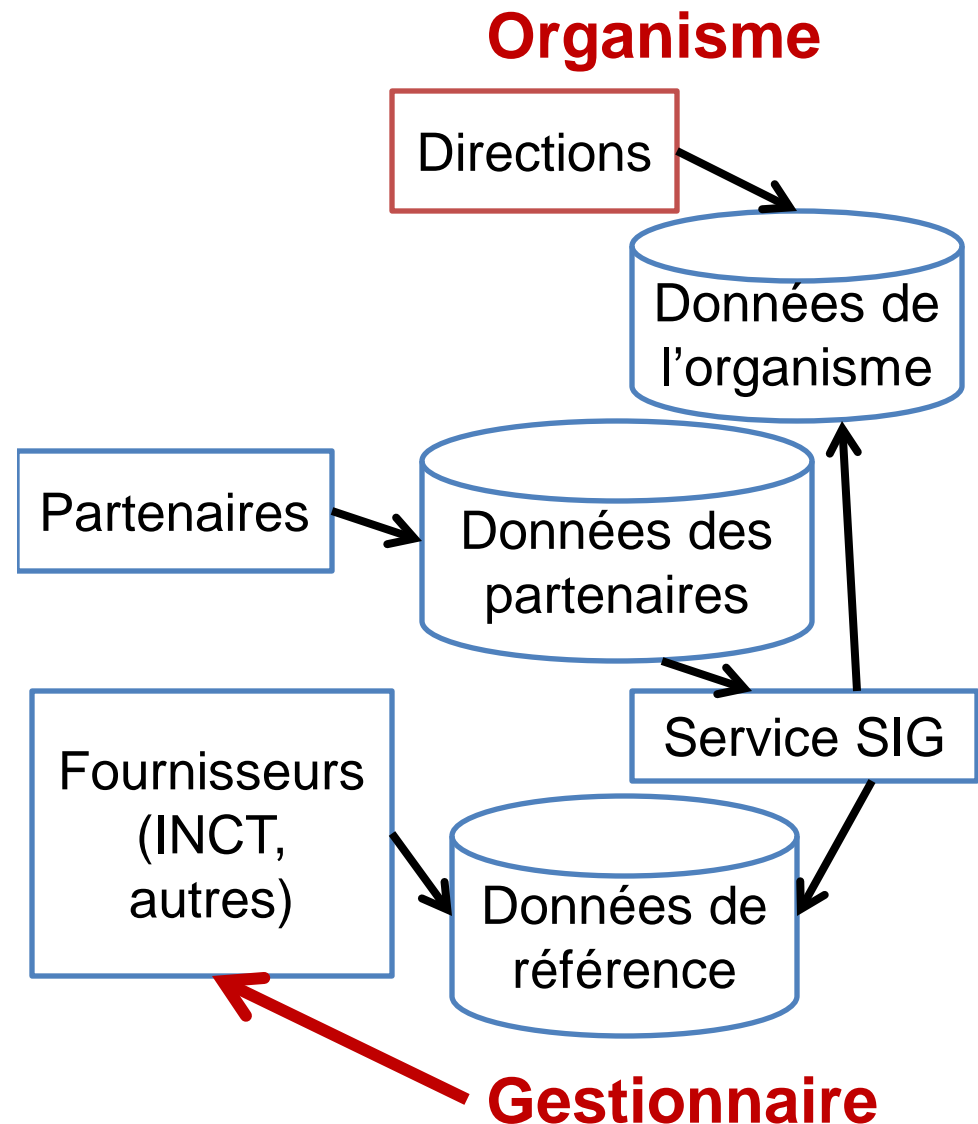
3- Acquisition à partir de **documents** existants.

4- Acquisition à partir de **photos**

5- Acquisition à partir de **d'images satellites**

6- Acquisition à partir de **données alphanumériques**

7- Acquisition à partir du terrain Levé **G.P.S.**



# Données de références

## Données thématiques

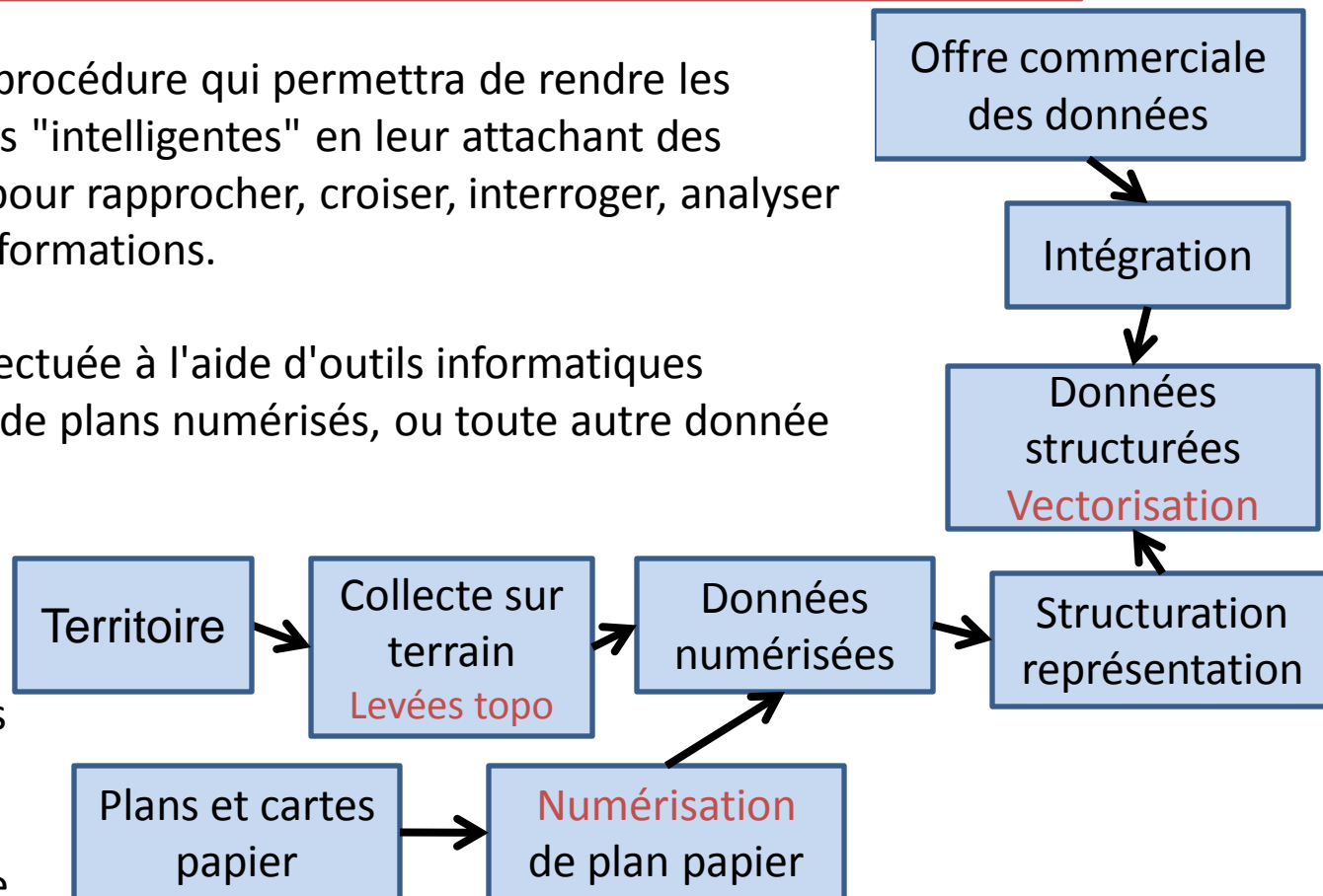
Type d'Organisme	Données de Références	Données thématiques
Wilaya	Plan des écoles, adresse postale, photo aérienne...	Fichier scolaire, entreprise, élections, éclairage public...
Commune	Carte communale, carte des quartiers	Voirie de la commune, espaces verts, routes
SNTR	Carte routière	Localisation entrepôts, clients, camions, itinéraires
Algérie Poste	Adresse postale	Clients, points de vente.

# Structuration des données spatiales

**Vectorisation** : c'est la procédure qui permettra de rendre les données géographiques "intelligentes" en leur attachant des données attributaires pour rapprocher, croiser, interroger, analyser toutes les couches d'informations.

La vectorisation est effectuée à l'aide d'outils informatiques spécifiques, et à partir de plans numérisés, ou toute autre donnée géographique.

**Levés topographiques** : Les matériels évoluent, et permettent aujourd'hui d'atteindre des niveaux de précisions importants. Cette technique d'acquisition constitue une solution pertinente si l'on considère le rapport qualité/prix.



**Numérisation** : c'est l'étape indispensable et préalable à l'intégration dans un Système d'Informations Géographiques (SIG). Elle consiste en un scan de documents papiers ou photographiques (*plans de réseaux, photographies aériennes*).

# Importance de la donnée

**-La nature des données:** L'information géographique est une association de données alphanumériques et de données spatiales, et doit permettre de répondre à un ensemble de questions comme par exemple:

**-Quel est cet objet ?**

*-Une ville d'Algérie*

**-Quels sont ses attributs ?**

- .....

**-Où se situe-t-il ?**

- .....

**-Quelle est sa forme?**

- .....

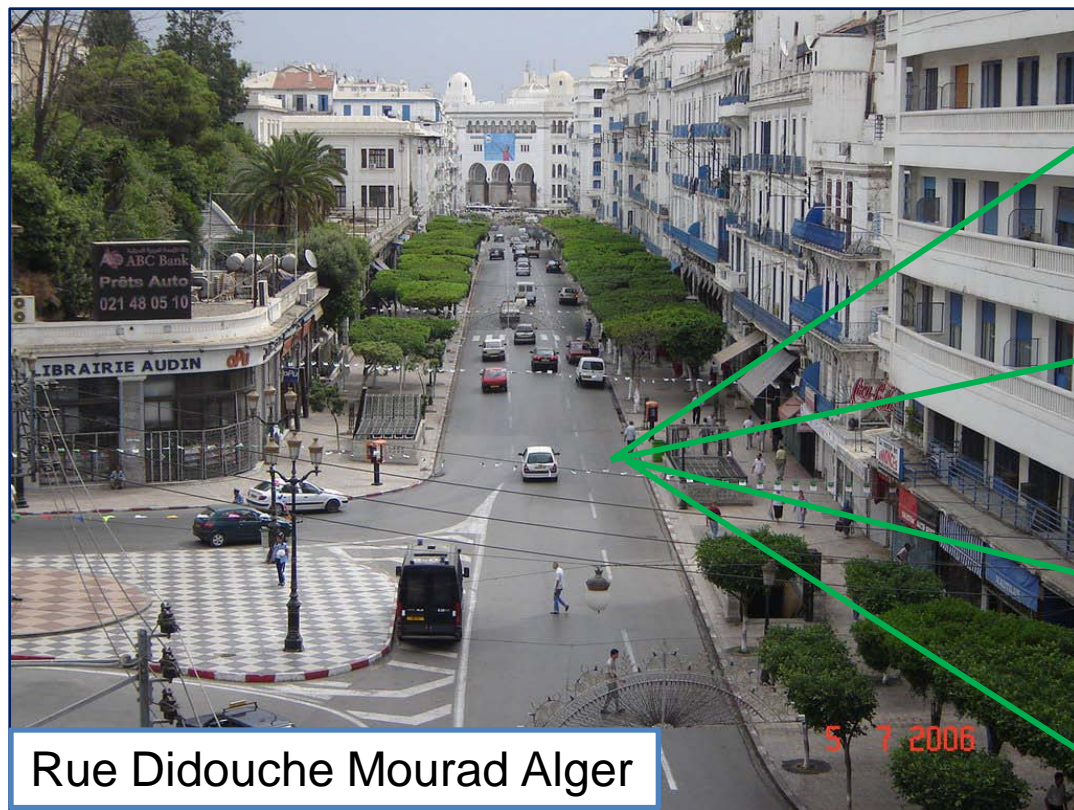
**-Quelle est sa relation avec les autres données?**

**- de même type**.....

**- d'un autre type**.....



# Divergence des représentations Spatiales



**Trafic routier =**  
*Représentée par un graphe*

**Cadastre =**  
*Représentée par une ligne*

**Service de goudronnage =**  
*Représentée par un polygone*

**Service des évacuations d'eau de pluie =**  
*Représentée par un point (calcul de volume)*

# SGBD ou SIG ?

- Utilisation d'un SGBD pour les données géographiques

*En théorie OUI, mais des difficultés de gestion des entités spatiales*

- Utilisation d'un SIG pour toutes les données

*Pas de fonctions de gestion et d'administration*

- Combiner les 2 richesses:

1. Une BDD classique gérées par un SGBD pour les données descriptives.
2. Une BDD contenant les entités géographiques (géométrie, topologie) gérée par un outil SIG (GéodataBase)
3. Lien entre les 2 BDDs

A l'origine, un SGBD est un logiciel dédié à la gestion des **données alphanumériques** de type **attributaire** stockées dans des tables et organisées en bases de données structurées. Grâce à un logiciel standardisé de requêtes sur les attributs.

Un SGBD permet d'interroger les tables, de procéder à des jointures ou à des synthèses.

Les logiciels **SIG** proposent tous un module de **SGBD** en charge de la gestion des données attributaires.

Par contre, les SGBD doivent élargir le type de données qu'ils étaient capables de gérer.

Ce sont l'ensemble des données Géographiques, **Attributaires**, **Raster et Vecteur** que les **SGBDS**, peuvent gérer de manière intégrée et cohérente.

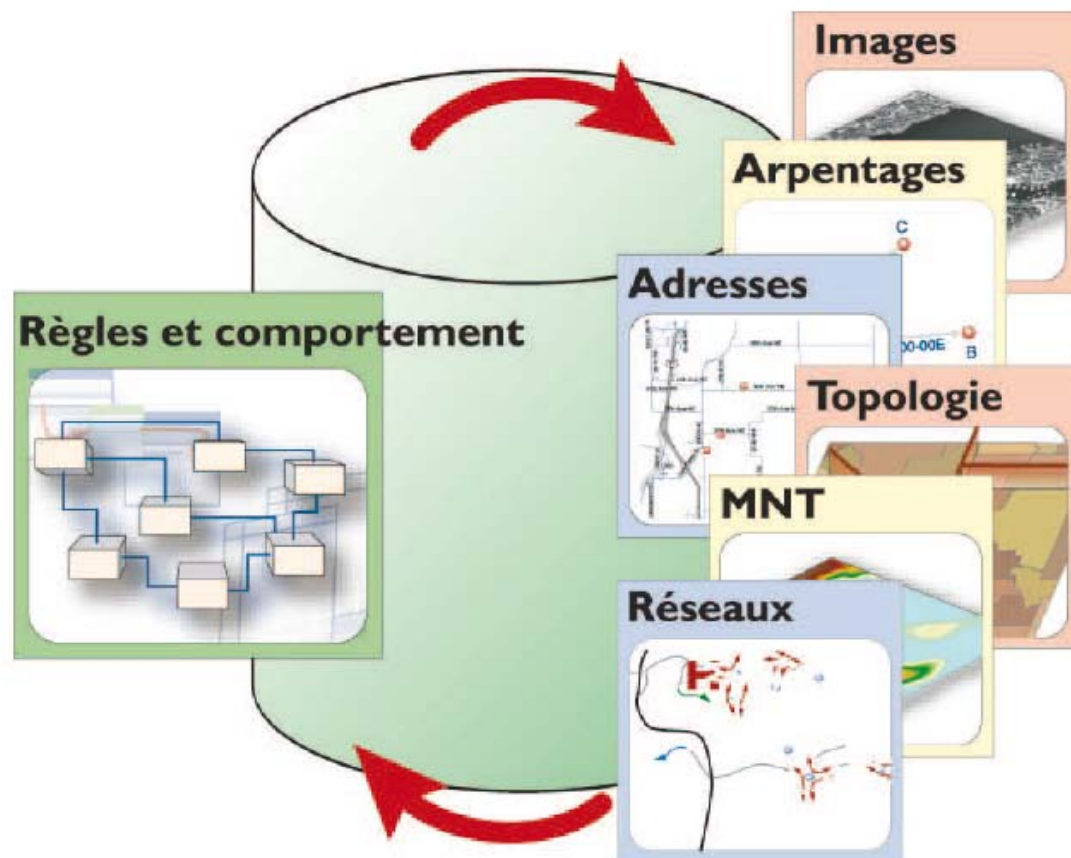
Toutefois le modèle classique associant une table à une couche d'entités reste encore dominant dans les systèmes simples.

Une Géodatabase est une base de données spatialisée contenant les couches de donnée d'information géographique.

L'organisation des données dans un SGBDS doit être stricte et formalisée par **un modèle conceptuel de données**.

La correspondance entre deux tables doit se faire via une seule colonne (*ou champ*) dans chacune des tables mises en relations.

Les relations se basent sur des **colonnes qui servent d'identifiant** pour être certain que la relation soit Stable dans le temps et se base bien sur des valeurs uniques.



# Rôle d'un SGBDS (Géodatabase)

## Fonctionnalités d'un **SGBDS (Géodatabase)**:

**-1-** C'est un SGBD standard

**-2-** Il offre un **type de donnée spatial dans son modèle de données** ainsi que son langage de requêtes;

- **Requêtes d'identification:** *de façon directe sur une carte, ou un graphique (pointage avec la souris)*

- **Requêtes de localisation, analyse spatiale, analyse temporelle...** *de façon indirecte par l'écriture de requêtes spatiales,*

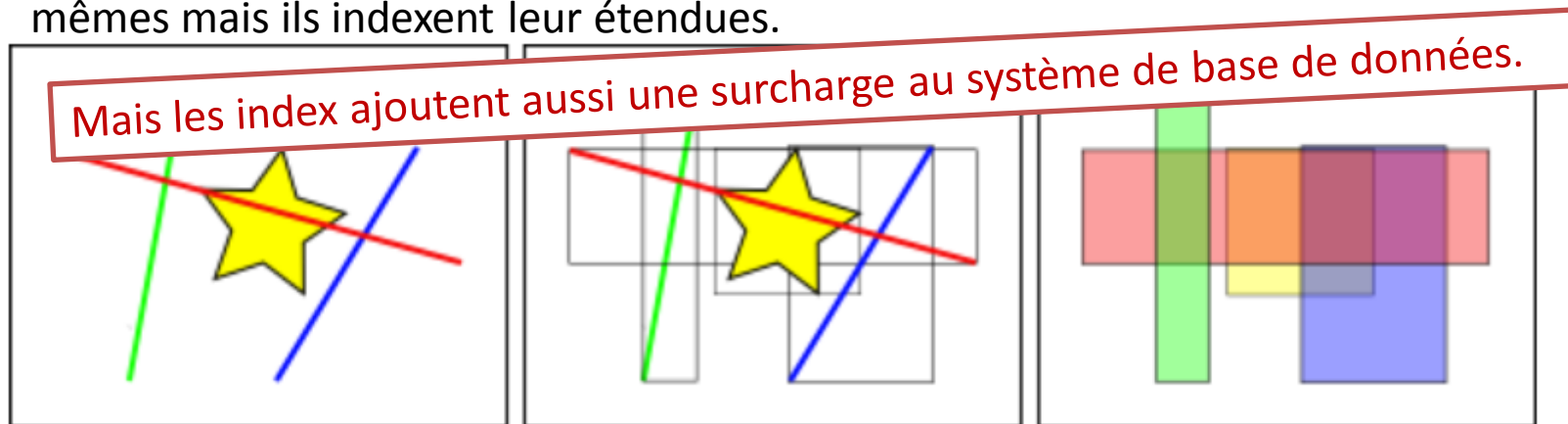
- ✓ *soit en utilisant SQL qui est langage de requête des BD et enrichi avec des opérateurs spatiaux (Spatial SQL ou GéoSQL qui calcul des distances, utilise des relations topologiques DEDANS, TOUCHE, etc., ou des relations de cardinalités).*

- ✓ *soit avec un Langage de requête visuel utilisant des icônes...*

# Les indexations spatiales

-3- Il implémente ce type de données et ses opérateurs, fournissant **l'indexation spatiale**.

- ✓ L'indexation spatiale accélère les recherches en organisant les données dans des arbres de recherche qui peuvent être parcourus efficacement pour retrouver une entité particulière.
- ✓ Les index spatiaux ne sont pas capables d'indexer les entités géométriques elles-mêmes mais ils indexent leur étendues.



Dans la figure ci-dessus, le nombre de lignes qui intersectent l'étoile jaune est seulement la ligne rouge. Mais l'étendue des entités qui intersectent la boîte jaune sont *deux*, la boîte rouge et la boîte bleue.

La manière dont les bases de données répondent rapidement à la question "Quelles lignes intersectent l'étoile jaune ?" correspond au début à la question "Quelle étendue intersecte l'étendue jaune" en utilisant les index (*ce qui est très rapide*) puis à calculer le résultat exact de la question ensuite "Quelles lignes intersectent l'étoile jaune ?"



# Les Jointures spatiales

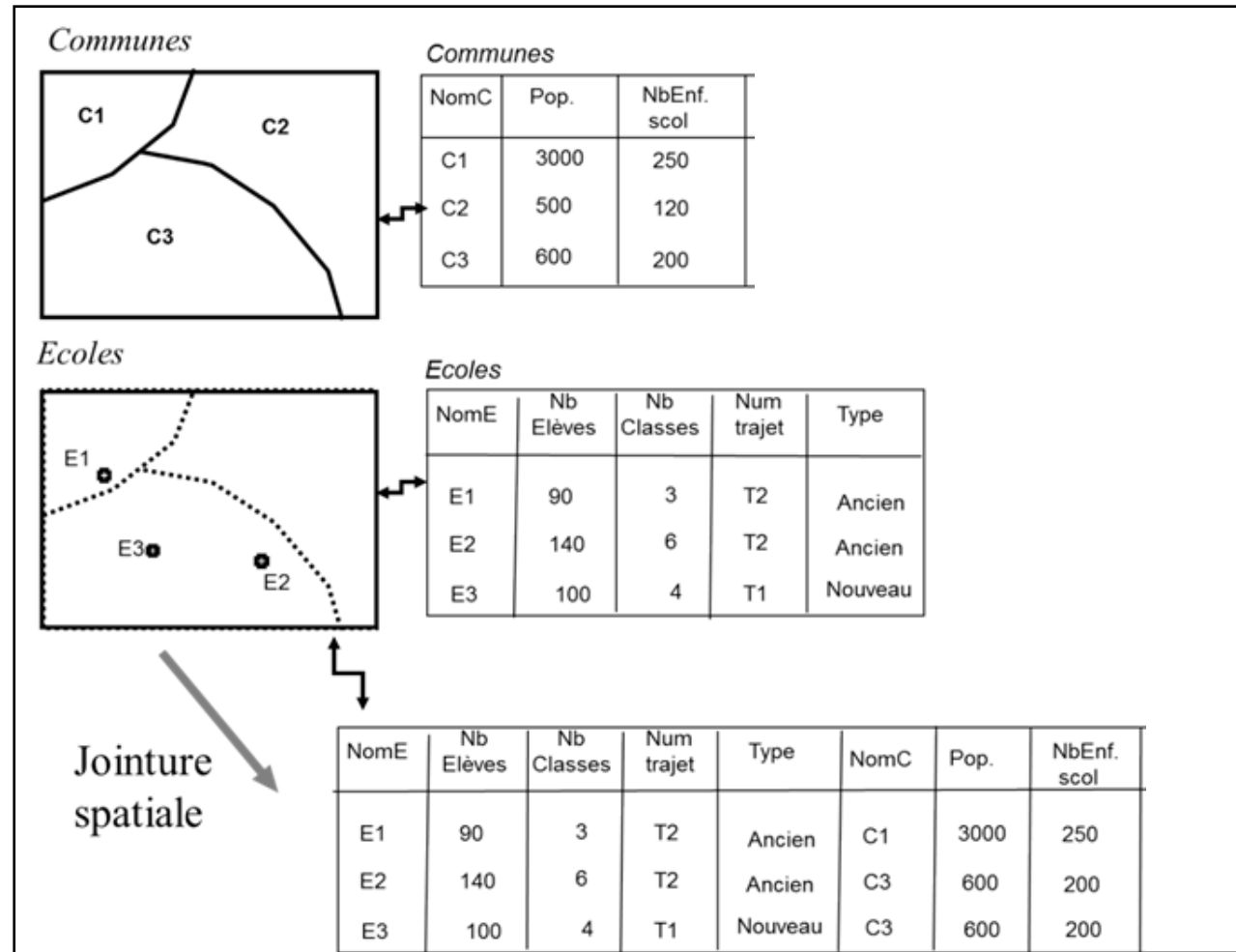
## -4- Il permet la **Jointure spatiale** :

Au lieu que la jointure se fasse entre les tables et s'appuie sur une relation logique (*le partage d'un même champ*) c'est une **relation spatiale**, c'est à dire une caractéristique de positionnement relatif des deux **entités** (*point, ligne ou polygone*).

Par exemple,:

Une table des écoles pourra être jointe à une table des communes en fonction d'une relation spatiale de type inclusion géométrique.

Dans ce cas la jointure ne peut donc être effectuée qu'entre des données géographiques, et non directement en utilisant les tables d'attributs.



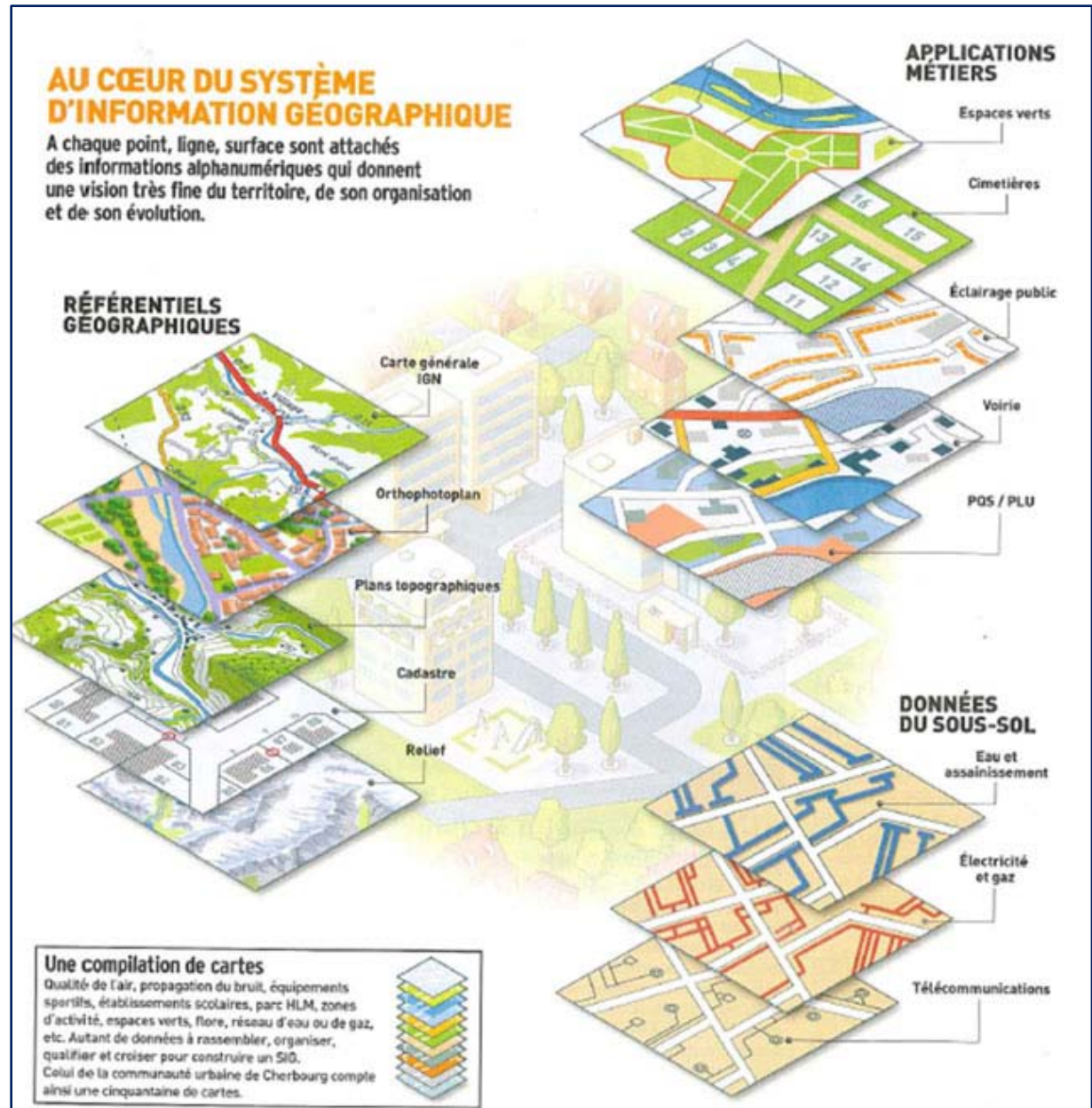
# Relation spatiales Inter couches

-5- Il permet des **Relations spatiales** essentielles car « tout ce qui se passe à un endroit est lié à ce qui se passe ou se passera au voisinage et ce lien va décroître avec l'éloignement.  
(1<sup>ère</sup> loi en Géographie TOBLER 79).

Les relations sont de 2 types:

– **Inter-couches** :  
relations verticales

– **Intra-couche** :  
relations horizontales





# Extension Spatiale des SGBDS Relationnels

	Relationnel	Spatial
<b>Données</b>	Entier, Réel, Texte,...	Plus complexe: Point, ligne, Polygone,...
<b>Prédicats et calculs</b>	Test: =, >, <...	Géométriques et topologiques: <i>intersecte, adjacent à ...</i> Fonction géométrique: <i>calcul de surface, de périmètre,...</i>
<b>Manipulation</b>	Opérateur de l'algèbre, sélection, protection, jointure... Agrégats: add, mul,...	Manipulation par thème ou inter-thèmes. Section et jointure sur critère spatial Agrégats: fusion d'objets adjacents
<b>Liens entre objets</b>	Par clé de jointure	Liens spatiaux ( <i>souvent implicites</i> )

# Composants d'un SIG d'entreprise standard

En plus de l'environnement bureautique, les logiciels SIG peuvent être centralisés sur des **serveurs d'applications** et **des serveurs Web** pour fournir des fonctions SIG à un nombre infini d'utilisateurs par l'intermédiaire de réseaux.

Des ensembles spécifiques de **logique SIG** peuvent être incorporés et déployés dans des **applications personnalisées**.

De même, le SIG est de plus en plus déployé sur des **périphériques nomades** pour des applications **de terrain**.

Dans un environnement d'**entreprise**, les utilisateurs se connectent à des **serveurs SIG centraux** par l'intermédiaire de SIG bureautiques traditionnels, de navigateurs Web, d'applications spécialisées, de matériels nomades et d'appareils numériques.

Le concept de **plate-forme SIG** progresse.

# Organisation des données sémantiques

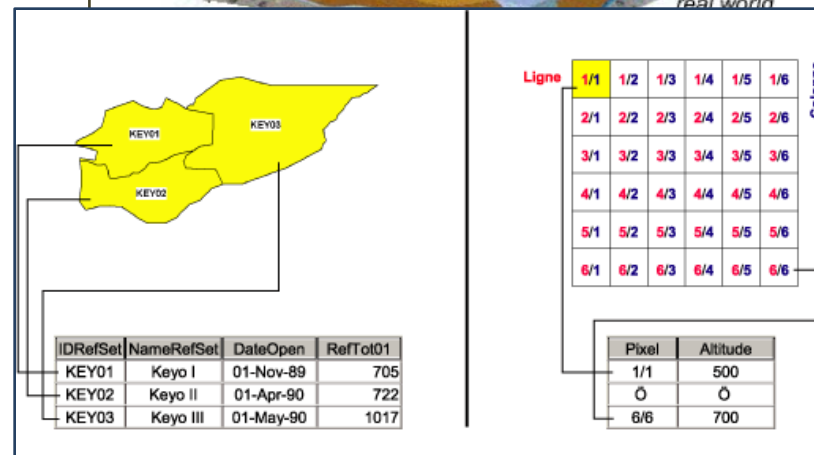
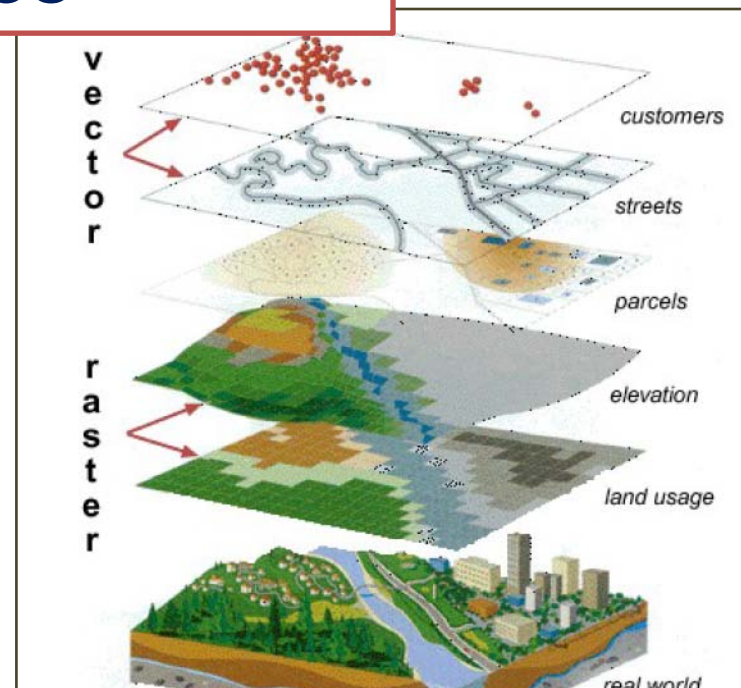
Les données **attributaires** sont celles qui vont **décrire les objets graphiques** de la carte.  
Il existe donc un lien dynamique dans le logiciel SIG entre les données graphiques, et les données alphanumériques (*données quantitatives et qualitatives*).

Ex: nom du propriétaire de la parcelle

Toutes ces données sont stockées dans des tables dites attributaires.

**Une table attributaire est identique à une table d'une base de données:**

- Chaque ligne (*ou enregistrement*) représente un objet graphique du plan vecteur (*ponctuel, linéaire ou surfacique*)
- Chaque colonne (*ou champ*) représente une information (*attribut*)

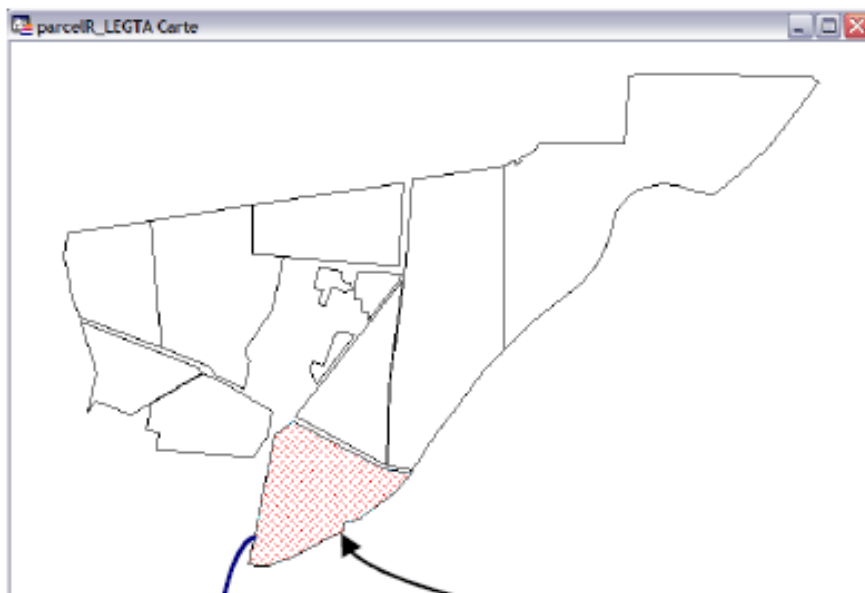


# Organisation des données attributaires

## Lien entre objet et table attributaire

### Création d'objets graphiques

La numérisation des objets se fait à l'écran sur la fenêtre carte



Objet sélectionné

### Saisie des champs attributaires

La saisie des champs se fait en fonction des éléments graphiques



Ligne sélectionnée

# Renseignement des tables

Attributs de bati\_industriel

FID	Shape *	ID	PREC PLANI	PREC ALTI	ORIGIN BAT	NATURE	HAUTEUR
0	Polygone ZM	BATIMENT0000000040492185	2,5	2,5	BDTopo	Bâtiment commercial	11
1	Polygone ZM	BATIMENT0000000040479298	1,5	1	Autre	Bâtiment industriel	3
2	Polygone ZM	BATIMENT0000000040479329	1,5	1	Autre	Bâtiment industriel	4
3	Polygone ZM	BATIMENT0000000040481425	1,5	1	Autre	Bâtiment industriel	6
4	Polygone ZM	BATIMENT0000000040483886	1,5	1	Autre	Bâtiment industriel	8
5	Polygone ZM	BATIMENT0000000040484009	1,5	1	Autre	Bâtiment industriel	2
6	Polygone ZM	BATIMENT0000000040484017	1,5	1	Autre	Bâtiment industriel	3

## Le renseignement manuel

**Saisie manuelle des attributs**, associer la table à un plan, ainsi si on sélectionne un objet (= *une ligne*) dans la table, elle est aussi sélectionnée sur le plan, il existe donc un lien dynamique entre les deux.

Nous noterons que la superficie et la longueur peuvent être calculés automatiquement, ce qui est pratique pour calculer la longueur d'une route par exemple.

## Le renseignement automatique

il est possible d'**importer** (et c'est souvent comme cela que l'on procède) **des données** issues d'une table provenant d'un tableur (comme EXCEL), d'un fichier texte ou d'une base de données (ACCESS, Dbase, SQL, Oracle..).

Pour cela, il suffit d'avoir un attribut commun appelé « **Attribut de jointure** » .

# Bases de données : aspect sécuritaire

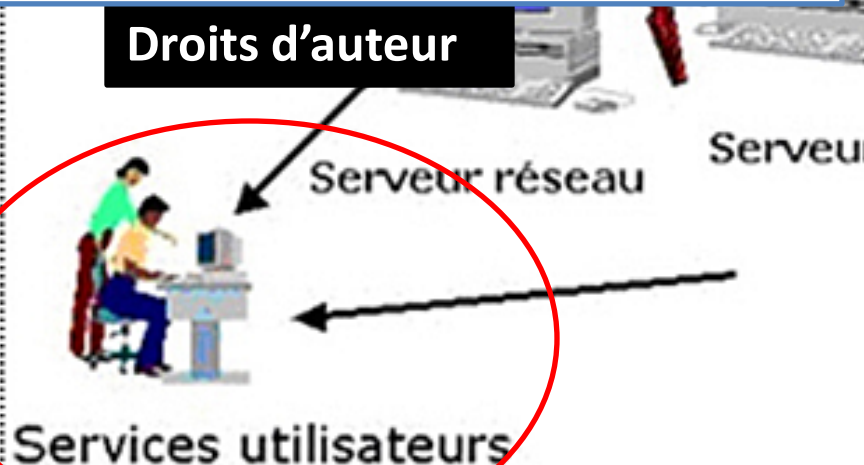
## -le droit de la propriété

### intellectuelle (droit d'auteur),:

La mise en forme des données (*structure*) doit présenter un caractère d'originalité.

Le chef de projet devra s'assurer auprès de l'auteur qu'il est autorisé à reproduire tout ou partie de l'œuvre de celui-ci.

### Droits d'auteur



### Intégrité des données

**-le droit économique** (droit "sui generis"): Permet de protéger les données en raison de l'investissement substantiel qui a permis de les produire.

Toute concession de droits d'utilisation de données géographiques accordées par un fournisseur à un tiers doit faire l'objet d'un écrit (*contrat ou licence*) et ce quel que soit le mode de protection s'appliquant aux données. Le contrat ou la licence doit prévoir:

- la nature des droits cédés (*exemple: possibilité de traitement des données*),
- l'étendue des droits (*exemple : la zone géographique*),
- la destination des données (*exemple : interne, externe, sur serveur,etc ...*),
- la durée de la concession et la rémunération de l'auteur.

# Normes ISO pour les SIG

A partir des années 1995, l'**ISO** et l'**OGC** ont commencé à développer des normes pour les SIG.

Ces initiatives ont pour but de gagner la confiance et la reconnaissance de la communauté de la cartographie et de la géographie afin de montrer la nécessité d'avoir un standard international.

L'**ISO** (**Organisation internationale de normalisation**) est le plus grand producteur et éditeur mondial de Normes internationales. L'ISO est un réseau d'instituts nationaux de normalisation de 157 pays, selon le principe d'un membre par pays, dont le Secrétariat central, est situé à Genève, Suisse.

L'**Open Geospatial Consortium (OGC)** est un consortium international industriel qui se compose de plus de 270 entreprises. Avec les spécifications de l'OGC, les développeurs devraient être capables de rendre des géoinformations complexes et les services se basant sur celles-ci accessibles à une variété d'utilisateurs et d'applications.

La série ISO 19100 concerne l'information géographique. Il existe environ 50 normes différentes chacune ayant un thème bien défini concernant la représentation de données et métadonnées, ainsi que des services.



# Contraintes d'intégrité spatiale

- Certaines contraintes **géométriques ou topologiques** devront toujours être vérifiées (*par exemple, 2 zones doivent toujours être tangentes*), car elles dépendent du modèle logique ou du modèle interne de description
- D'autres dépendent de la **définition sémantique** de la collection (*un réseau routier doit toujours être connexe, mais un réseau téléphonique peut ne pas l'être*)
- Il s'agit de vérifier:
  - **la généalogie** (*information sur les sources des données*).
  - **l'actualité** (*validité temporelle*), *la cohérence logique* (*contrôle de l'intégrité des données*)
  - **la précision géométrique, la précision sémantique, l'exhaustivité**



La composante sémantique représente l'information relative à **la nature**, l'**aspect** et les **propriétés descriptives** d'un objet ou à un phénomène du monde terrestre:

Exemple une Wilaya décrite par son nom (**Tlemcen**), sa population (949 132 **hab**), pour une surface de (10 182 km<sup>2</sup>) et une autre de (**Tamanrasset**) sa population (176 636 hab) pour une surface de (558 310 km<sup>2</sup>).

Les méthodes d'analyse spatiale ont été classifiées selon 5 types:

- **Les méthodes d'interrogation et raisonnement** :représentent les formes les plus simples d'analyse spatiale. Le SIG est utilisé pour **répondre aux questions** sous la forme de requêtes qui utilisent des prédicats spatiaux ou alphanumériques pour sélectionner un ensemble d'objets géographiques. **Aucun changement** n'intervient dans les bases de données spatiales.

## Suite...

# Composante sémantique

- **Les méthodes de mesure:** permettent d'obtenir des valeurs numériques qui décrivent des objets géographiques. Elles incluent des mesures comme le périmètre, la surface pour un seul objet spatial et la distance entre deux objets spatiaux.
- **Les méthodes de transformation:** sont des opérations spatiales qui changent les données en le combinant entre elles pour obtenir de nouvelles données géographiques. Ces transformations utilisent de simples règles géométriques, logiques et/ou arithmétiques, par exemple l'overlay, la fusion, etc.
- **Les méthodes de synthèse:** synthétisent un ensemble de données en une ou plusieurs valeurs numériques. Elles sont les équivalents spatiaux des descripteurs statistiques classiques et elles incluent des méthodes d'analyse statistiques spatiale, comme le barycentre.

# Précision et qualité (SIG)

**Erreur =** *"la non concordance entre les objets encodés dans la base de données géographiques et les entités du monde réel que ces objets sont supposés représenter »*  
(VEREGIN, 1995)

17 types d'erreurs possibles:

- **Exactitude**: relation entre la mesure et la réalité
- **Précision**: degré de détail dans le stockage ou la manipulation lors des calculs
- **Résolution**: Taille réelle du plus petit objet qui peut être discerné sur une carte
- **Echelle**: Limite physique du dessin (*0,5 mm correspond à 12,5m pour 1/25 000*)
- **Justesse du contenu**: pour les données thématique (*parcelle non cultivée ou pas*)
- **Erreur de mesure**: mauvaises observation, fausse altitude
- **Facteur humain**: lors d'une interprétation (*mauvaise analyse de photos aériennes*)

# Précision et qualité (SIG)

- **Erreurs de laboratoire**: erreur de manipulation de l'information cartographique
- **Justesse de localisation**: primordiale au sein des SIG (déformation lors d'un géoréf.)
- **Variabilité spatiale de l'information**: selon le thème le contenu peut changer
- **Age des données**: en dehors de la géologie la validité de l'info diminue avec le temps
- **Densité de l'observation**: plan d'échantillonnage pertinent (variabilité du milieu étudié)
- **Pertinence**: l'information étant indirecte, il faut juger de la pertinence du support
- **Format des données**: l'intégrité des données dépend de la projection, de l'échelle,...
- **Accessibilité**: secret militaire ou inversement multiplicité des données
- **Coût et propriétaire**: acquisition des données achetées et les copyrights
- **Erreur numérique**: en relation avec le logiciel (erreurs d'arrondis, compactage,...)

# Les Métadonnées: Définition

**Définition** : Données qui **renseignent** sur la nature des **données** et qui permettent ainsi leur utilisation pertinente.

Elle permettent de:

- 1- Connaître l'**origine** et la **nature** des données
- 2 -Comprendre leur **structure**
- 3- Savoir comment **accéder aux données** et comment les interpréter
- 4- Connaître les modèles de données et leur **règle de gestion**

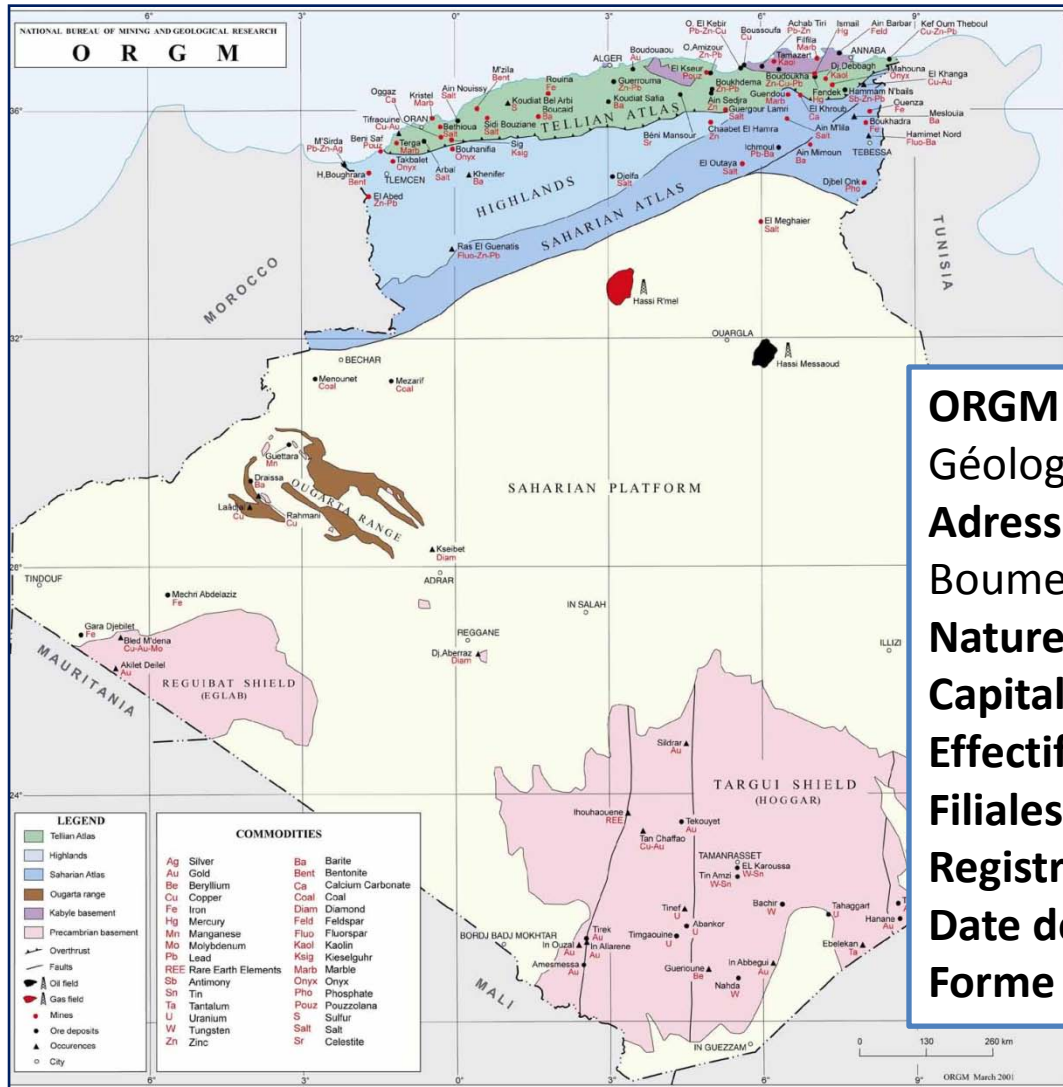
**Intérêts** : L'utilisation de standards permet aux utilisateurs d'avoir **une terminologie commune** permettant la réalisation de recherche efficace pour la découverte des données dans les catalogues. Les métadonnées reposant sur les standards qui permettent **d'éviter la perte de connaissance** sur les données.

Les données géographiques sont souvent produites par des **organisations** ou des **indépendants** et peuvent répondre aux besoins de différents types d'utilisateurs (opérateurs SIG, analyse d'image, politiques, ...).

# Métadonnées Géographiques

## Métadonnées géographiques

– Echelle, Emprise, Référentiel géographique (Système de projection), Qualité (précision de localisation et des attributs), Datation, ...



**ORGM / Office National de Recherche  
Géologique et Minière**

**Adresse:** Cité Ibn Khaldoun.Bp 102  
Boumerdes, Algérie

**Nature de l'entreprise:** Prestataire

**Capital Social (en DA):** 1.500.000.000 DA

**Effectif:** 497

**Filiales et groupe:** Filiale Du Groupe Manal

**Registre de commerce:** 16/00-00B0723061

**Date de début d'activité:** 06/05/1966

**Forme juridique:** SPA

# Les Métadonnées: Types

On distingue trois types de métadonnées :

**Métadonnées pour la découverte:** offrent le minimum d'informations juste suffisantes pour **identifier un lot de données** afin d'extraire celles qui seront peut être utiles. La question à laquelle elles répondent « *Qui a des données sur...?* »

**Métadonnées pour le catalogage:** **caractérisent** avec plus de détail **les ressources**, et donne des détails permettent le choix réel des ressources utiles qui répondent aux besoins « *Parmi les données lesquelles sont les plus utiles pour mes besoins?* »

**Métadonnées pour l'exploitation:** servent à configurer l'interface du SI d'accueil en fonction du jeu de données livré + **informations sémantiques** + **informations sur les limites d'utilisation** « *Comment vais-je intégrer ses données?* »

<b>Données Descriptives</b>	Titre, auteur, dates, mots clés, couverture géographique,...
<b>Données Administratives</b>	Date et lieu de création, droits d'accès, sources, liens d'accès, format, ...

# Les Métadonnées: Rôle

<b>But de la mise en place</b>	<p>1- Améliorer la connaissance des données au sein de la structure.</p> <p>2- Anticiper les futurs besoins et/ou obligations de catalogage</p>
<b>Objectifs</b>	<p>1- Eviter la création de données inutiles</p> <p>2- Normaliser les données afin de pouvoir échanger les données et métadonnées avec les partenaires</p>

## Exemples de Métadonnées conformes à la norme ISO 19115

**Titre:**.....REFERENCES\_SIG  
**Date de dernière m.à.j:** .....23-05-2021  
**Type de date:**.....Révision  
**Code:**.....Référence\_SIG.tab  
**Langue:**.....Français  
**Jeu de caractères:** .....utf8  
**Thématique:**.....Société  
**Ouest:**.....-5,79023  
**Est:** .....9,56222  
**Nord:**.....51,09111  
**Sud:**.....41,36493.....