

Stockage et Traitement de Données à Grande Échelle

Partie IV – Introduction aux Systèmes d'Information Géographique

Sylvain Bouveret

sylvain.bouveret@grenoble-inp.fr

Mastère Spécialisé Big Data Ensimag / GEM – 2022



Introduction

1. Introduction et fondements de la gestion répartie des données
2. Stockage et traitements des masses de données
 - 2.1 SGFD & approches MapReduce
 - 2.2 Approches NoSQL
3. Exploitation des masses de données
4. Introduction aux Systèmes d'Information Géographique

1. Introduction et fondements de la gestion répartie des données ✓
2. Stockage et traitements des masses de données ✓
 - 2.1 SGFD & approches MapReduce ✓
 - 2.2 Approches NoSQL ✓
3. Exploitation des masses de données ✓
4. Introduction aux Systèmes d'Information Géographique

Introduction aux SIG

- **Pré-requis :**

- Bases de Données relationnelles (PSGDB) – SQL
- Programmation Java
- Notions d'applications Web

- **Organisation :**

- Cours d'introduction (1h30)
- TP OpenStreetMap, évalué (3h)
- Cours présentation de cas d'étude (1h30)

1 Première partie

Introduction aux SIG

Système d'Information Géographique

Système d'Information dédié à la gestion de données référencées spatialement (= géoréférencées).

Composantes habituelles d'un SIG :

- **Données** : primitives géométriques et données attributaires non géographiques
- **Logiciels** : Acquisition, Archivage, Analyse, Affichage, Abstraction, Anticipation.
- **Matériels** : serveurs, clients. . . .

- **Les débuts** (années 1990) : plans sur papier

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)
- **L'exploitation** (à partir de 2002)

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)
- **L'exploitation** (à partir de 2002)
- **L'intégration au SI** (années 2000) : premières cartouches spatiales – Oracle et PostGIS

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)
- **L'exploitation** (à partir de 2002)
- **L'intégration au SI** (années 2000) : premières cartouches spatiales – Oracle et PostGIS
- **Le Web** (2002) : ESRI, GeoServer, MapServer + interopérabilité : WMS, WFS, WFS-T.

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)
- **L'exploitation** (à partir de 2002)
- **L'intégration au SI** (années 2000) : premières cartouches spatiales – Oracle et PostGIS
- **Le Web** (2002) : ESRI, GeoServer, MapServer + interopérabilité : WMS, WFS, WFS-T.
- **La mobilité** (2005) : premiers dispositifs mobiles

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)
- **L'exploitation** (à partir de 2002)
- **L'intégration au SI** (années 2000) : premières cartouches spatiales – Oracle et PostGIS
- **Le Web** (2002) : ESRI, GeoServer, MapServer + interopérabilité : WMS, WFS, WFS-T.
- **La mobilité** (2005) : premiers dispositifs mobiles
- **L'externalisation** (2007) : services externalisés de données spatiales

- **Les débuts** (années 1990) : plans sur papier
- **La numérisation** (1990 → 2010)
- **L'exploitation** (à partir de 2002)
- **L'intégration au SI** (années 2000) : premières cartouches spatiales – Oracle et PostGIS
- **Le Web** (2002) : ESRI, GeoServer, MapServer + interopérabilité : WMS, WFS, WFS-T.
- **La mobilité** (2005) : premiers dispositifs mobiles
- **L'externalisation** (2007) : services externalisés de données spatiales
- **La VGI** (à partir de 2005, surtout après 2010) : *crowdsourcing* géographique

2 Deuxième partie

Systèmes de Référence Spatiale

L'information spatiale...

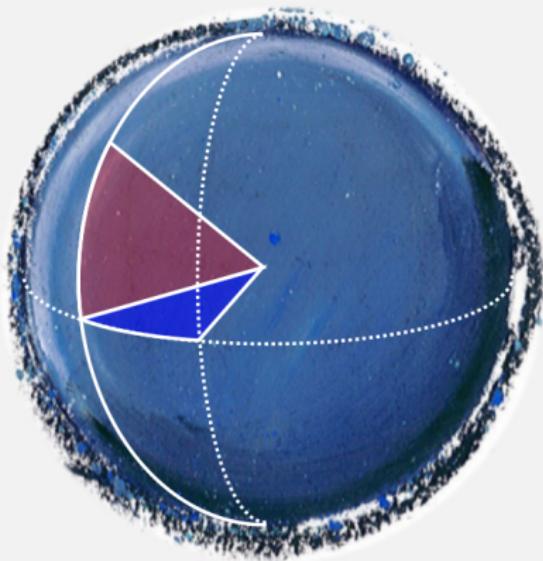




Le problème :

Comment repérer n'importe quel point sur la surface terrestre ?





Sur une sphère : (λ, φ) (longitude, latitude), par rapport à une origine.

Le problème...

Le problème...

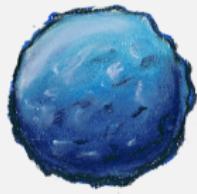


Le problème...



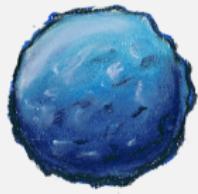
Il nous faut donc une **surface de référence**.

Il nous faut donc une **surface de référence**.



Géoïde

Il nous faut donc une **surface de référence**.



Géoïde



Ellipsoïde

Il nous faut donc une **surface de référence**.



Géoïde



Ellipsoïde



Sphéroïde

Important

Des coordonnées géographiques n'ont de sens que par rapport à un **Système de Référence Spatiale**.

Important

Des coordonnées géographiques n'ont de sens que par rapport à un **Système de Référence Spatiale**.

Exemple : *World Geodetic System 84*

```
GEOGCS ["WGS 84",
    DATUM ["WGS_1984",
        SPHEROID ["WGS 84", 6378137, 298.257223563,
            AUTHORITY ["EPSG", "7030"]], ,
        AUTHORITY ["EPSG", "6326"]], ,
    PRIMEM ["Greenwich", 0,
        AUTHORITY ["EPSG", "8901"]], ,
    UNIT ["degree", 0.01745329251994328,
        AUTHORITY ["EPSG", "9122"]], ,
    AUTHORITY ["EPSG", "4326"]]
```

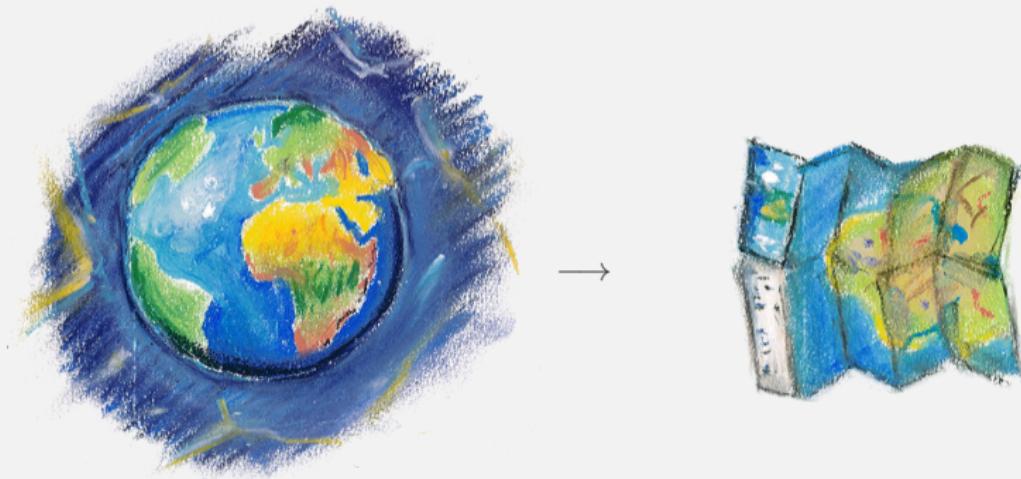
De la surface terrestre à la carte



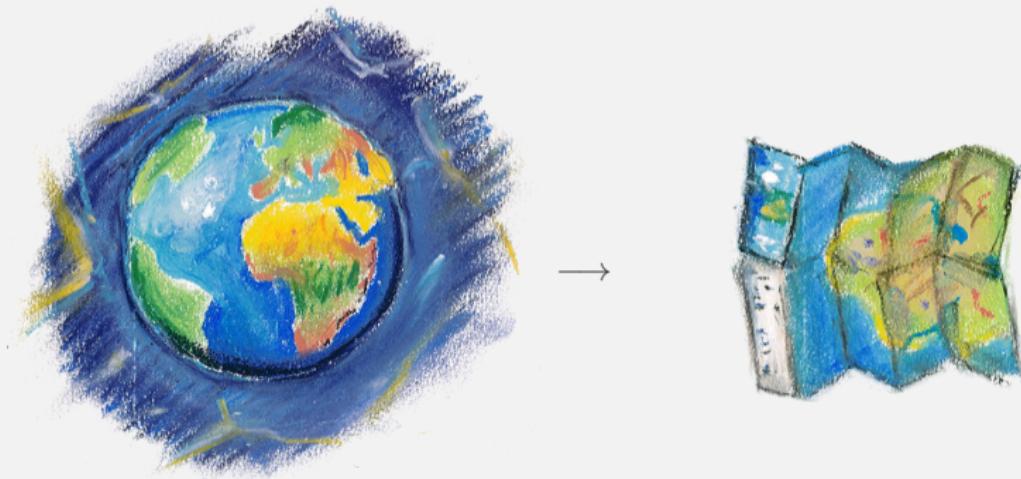
De la surface terrestre à la carte



De la surface terrestre à la carte

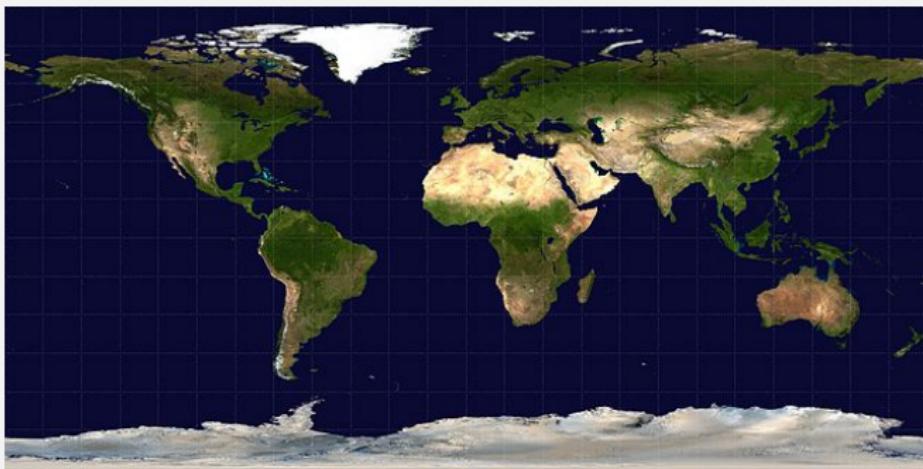


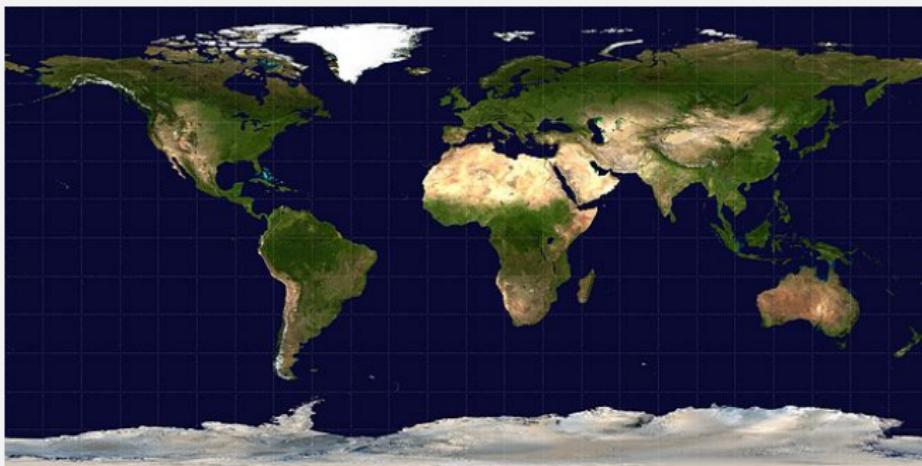
De la surface terrestre à la carte



Première idée : $(x, y) = (\lambda, \varphi) = (\text{longitude}, \text{latitude})$

Projection plate carrée





Déformations \leadsto carte inutilisable (pour la navigation par exemple)

Il nous faut un **Système de Projection**.

Système de Projection : $(\lambda, \varphi) \mapsto (x, y)$

Système de Projection : $(\lambda, \varphi) \mapsto (x, y)$

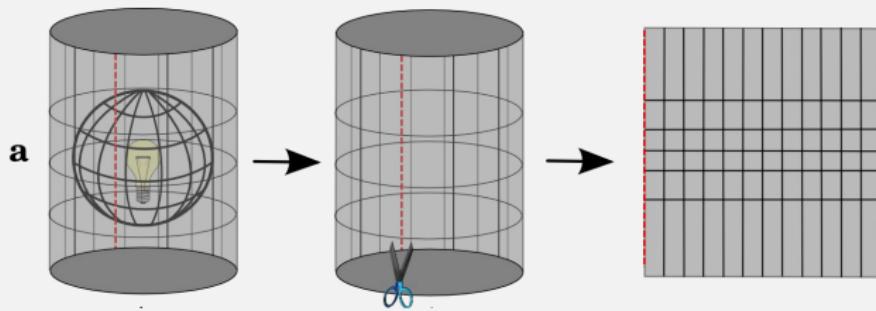
Toute projection induit forcément une déformation.

Système de Projection : $(\lambda, \varphi) \mapsto (x, y)$

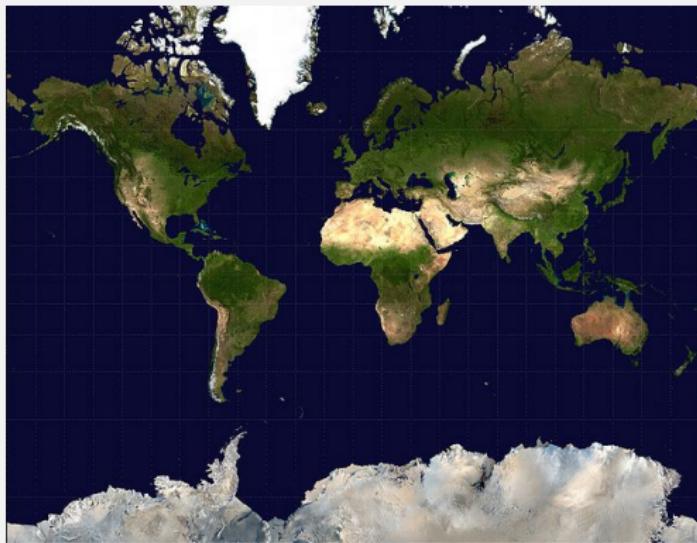
Toute projection induit forcément une déformation.

- **projection équivalente** : conserve localement les surfaces
- **projection conforme** : conserve localement les angles, donc les formes
- **projection aphylactique** : les projections équidistantes conservent les distances à partir d'un point donné
- **projection gnomonique** : tout grand cercle (= cercle coupant la sphère en deux hémisphères égaux) est représenté comme une ligne droite.

Projections courantes : cylindrique

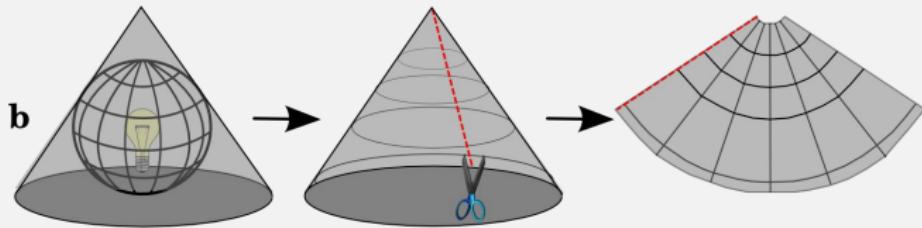


Projections courantes : cylindrique



Exemples : Mercator, Peters, Robinson, UTM...

Projections courantes : conique

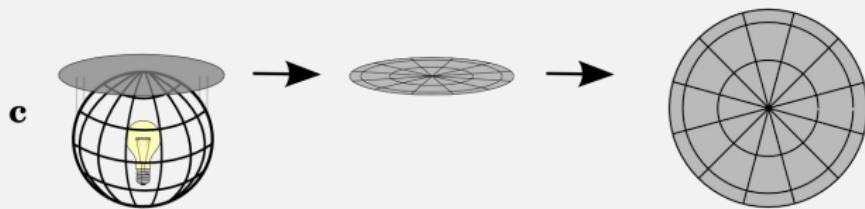


b

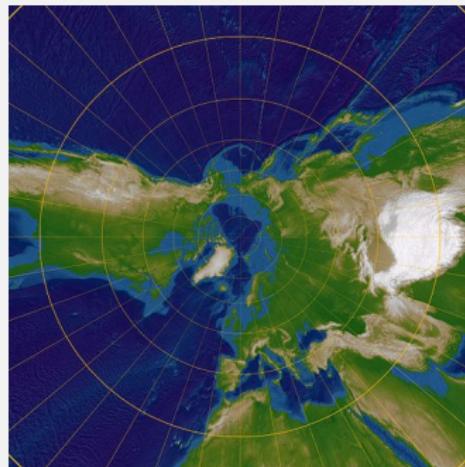


Exemples : Conique conforme de Lambert

Projections courantes : azimutale



Projections courantes : azimutale



Exemples : gnomonique, orthographique. . .

Système de référence spatiale : système permettant de représenter des éléments dans l'espace, de manière non ambiguë,

- à l'aide de coordonnées géographiques $(\lambda, \varphi) \rightsquigarrow$ coordonnées **non projetées**
- à l'aide de coordonnées planes $(x, y) \rightsquigarrow$ coordonnées **projétées**

Système de référence spatiale : système permettant de représenter des éléments dans l'espace, de manière non ambiguë,

- à l'aide de coordonnées géographiques $(\lambda, \varphi) \rightsquigarrow$ coordonnées **non projetées**
- à l'aide de coordonnées planes $(x, y) \rightsquigarrow$ coordonnées **projétées**

Un SRS est défini par :

- un ellipsoïde de référence
- une origine
- un système de projection (si SRS projeté)
- un **identifiant unique** \rightsquigarrow nomenclature EPSG (*European Petroleum Survey Group*) ; SRID (*Spatial Reference System Identifier*) propres aux SIG.

SRS : exemples

Exemples :

- WGS84 (Système de coordonnées non projetées) → EPSG 4326
- Lambert CC42 → CC50 (Systèmes de coordonnées projetées utilisés en France, s'appuyant sur une projection conique conforme à partir de coordonnées exprimées selon un sphéroïde de référence GRS 1980) : EPSG3942 → EPSG3950.

Pour s'amuser un peu :

http://imgs.xkcd.com/comics/map_projections.png

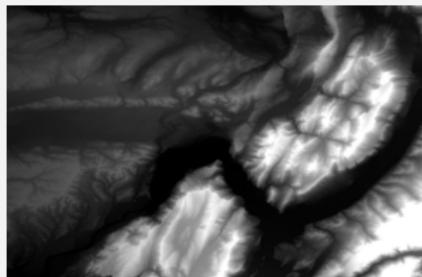
3 Troisième partie

Données vectorielles et matricielles

2. Données vectorielles

3. Données matricielles

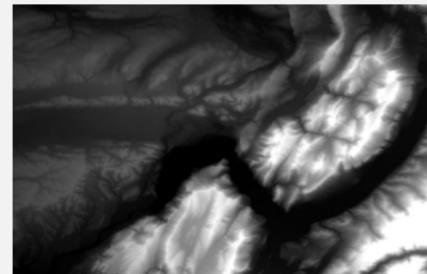
Deux types de données...



Deux types de données...



Données vectorielles



Données matricielles

Données vectorielles

Données vectorielles = ensemble d'**entités géographiques**

Données vectorielles = ensemble d'**entités géographiques**

Entité géographique = **géométrie + attributs**

Données vectorielles = ensemble d'**entités géographiques**

Entité géographique = **géométrie + attributs**

Différents types de géométries possibles :

- **Point (0D)** : 2 coordonnées (λ, φ) ou (x, y) ou 3 s'il y a en plus l'altitude
- **Ligne / Polyligne (1D)** : suite de points
- **Polygone (2D)** : suite de points dont le dernier est identique au premier





Attributs

Les attributs servent à **caractériser** les entités géographiques

- Attributs classifiants (exemple : type = 'rivière')
- Autres attributs (exemple : nom = 'Isère')

Attributs

Les attributs servent à **caractériser** les entités géographiques

- Attributs classifiants (exemple : type = 'rivière')
 - Autres attributs (exemple : nom = 'Isère')
-
- En général attributs au sens modèle relationnel
 - D'autres modèles plus flexibles sont possibles. **Exemple** : clefs-valeurs (OpenStreetMap)

Exemple :

- **Bâtiment D ENSIMAG dans OpenStreetMap**

Données matricielles

Données matricielles (*raster*) = grille de cellules

Données matricielles (*raster*) = grille de cellules

Cellule = une valeur ou un vecteur de valeurs

Données matricielles (*raster*) = grille de cellules

Cellule = une valeur ou un vecteur de valeurs

Caractérise en général des phénomènes continus tels que :

- images aériennes ou satellites ;
- données d'élévation (ex : NASA SRTM) ;
- températures... .

En général obtenues par un processus d'acquisition de données systématique, elles sont caractérisées par :

- une emprise, qui détermine exactement la portion de la surface terrestre concernée \leadsto géoréférencement ;
- une résolution spatiale, définie par la taille d'une cellule ;
- éventuellement d'autres résolutions (spectrale \leadsto nombre de bandes stockées – notion d'image multi-spectrale, altitude, etc.).

- des images raster peuvent être utilisées comme sources de données pour extraire (de manière automatique ou manuelle) des données vectorielles (routes, lignes de niveau, etc.) ;
- réciproquement, il peut être utile de convertir des données vectorielles au format matriciel, par exemple pour des questions d'affichage (cf tuiles OpenStreetMap).

4

Quatrième partie

Solutions techniques

- 4. Formats standards
- 5. Protocoles d'échange
- 6. Logiciels
- 7. Autres problématiques

Formats standards

- **ESRI Shapefile** : géométries (.shp), attributs (.dbf) et index (.shx)
- **Geography Markup Language (GML)** : standard XML (OGC)
- **Keyhole Markup Language (KML)** : équivalent Google de GML
- **GeoJSON** : JSON + géométries
- **Mapinfo TAB format (MIF / MID)**
- **Well-Known-Text (WKT)**

- **GeoTIFF** : format TIFF enrichi avec des informations SIG
- **JPEG2000**
- **ESRI grid...**

Protocoles d'échange

Protocoles standards développés par l'OGC (*Open Geospatial Consortium*) s'appuyant sur HTTP (GET).

Les deux principaux :

- **Web Map Service (WMS)** : Obtenir une image matricielle sur un serveur
- **Web Feature Service (WFS)** : Obtenir des données vectorielles

Logiciels

- SGBD PostgreSQL (l'un des plus puissants SGBD relationnels libres), avec une cartouche spatiale PostGIS \leadsto solution la plus répandue
- SGBD Oracle et sa cartouche spatiale

- Intermédiaire côté serveur entre des SGBD et des clients
- Fondés sur les services webs de l'OGC (WMS-WFS)
- Exemples :
 - GeoServer
 - MapServer

Il existe trois bibliothèques Javascript principales permettant d'afficher des informations géographiques sous forme de couches matricielles (tuiles) ou vectorielles :

- OpenLayers 2
- OpenLayers 3
- LeafLet

- ESRI ArcGIS (précurseur et référence)
- QGIS (libre).

Autres problématiques

Métadonnées

Métadonnées ≈ données sur les données

Exemples :

- Description générale (nature des données, emprise géographique, système de projection, organisme producteur, etc.)
 - Qualité des données (résolution, date de validité, etc.)
 - Informations légales (responsable légal, licence d'utilisation, etc.)
-
- Permet l'**interopérabilité** des systèmes (ISO19115)
 - Construction de **géocatalogues**

Plusieurs manières de représenter une donnée vectorielle.

Exemple, pour un point :

- .
- +
- 

On peut jouer sur tous les attributs graphiques : symbole, épaisseur de ligne, couleur, etc.

Plusieurs manières de représenter une donnée vectorielle.

Exemple, pour un point :

- .
- +
- 

On peut jouer sur tous les attributs graphiques : symbole, épaisseur de ligne, couleur, etc.

La discipline chargée d'analyser et de formaliser les signes et les règles de représentation est la **sémiologie graphique**.

Observons la même carte à différentes échelles...



Observons la même carte à différentes échelles...



Les entités géographiques ne sont pas représentées de la même manière à différents niveaux de zoom.

Règles de représentation multi-échelles → **généralisation cartographique**

Dans de nombreuses applications l'emplacement spatial d'une entité (utilisateur par exemple) n'est pas donné explicitement sous la forme de coordonnées, mais est fourni par exemple sous la forme d'une adresse, d'un numéro de parcelle, etc.

L'opération qui consiste à associer, à une telle information imprécise, une coordonnée spatiale précise, est appelée **géocodage**. Ce n'est pas une opération triviale, et même si un traitement automatique est possible, une validation manuelle est souvent nécessaire.

Informations de licence

- Map Projection Families - CC-by-SA, source : https://docs.qgis.org/2.6/fr/docs/gentle_gis_introduction/coordinate_reference_systems.html
- Images des différentes projections - CC-by-SA ou domaine public. Source Wikimedia.
- Images vectorielles et raster - CC-by-SA, source : travail personnel à partir de données OpenStreetMap et Nasa SRTM.
- Autres images d'illustration - CC-by-SA, source : travail personnel.