

Cours DH UniNe

# Introduction à la cartographie

Simon Gabay

Neuchâtel, Lundi 27 avril 2020

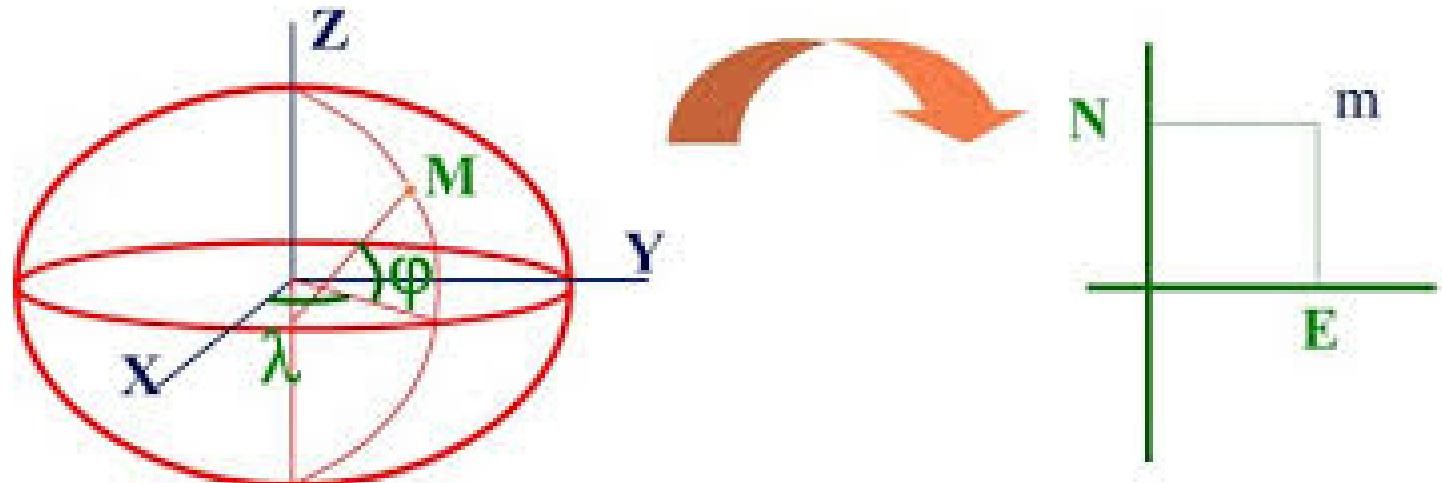
# Coordonnées et projection

# Coordonnées géographiques

- Système de trois coordonnées qui sont la latitude, la longitude et l'altitude (ou l'élévation).
- On utilise des points de références: l'équateur, le méridien de Greenwich, le niveau de la mer (qui n'a d'ailleurs pas la même élévation partout)
- Exprimé en degré, minute, seconde. Ainsi Genève à une latitude de  $46^{\circ} 12' 00''$  nord ("46 degrés, 12 minutes, 0 seconde nord") et une longitude  $6^{\circ} 09' 00''$  est ("6 degrés, 9 minutes, 0 seconde est")
- On peut exprimer ces coordonnées de degrés décimaux (par ex. pour Genève: 46.12 et 6.09)

# Coordonnées cartographiques

La projection est le passage de coordonnées géographiques exprimés en degrés ( $\lambda, \phi$ ) aux coordonnées cartographiques (E,N) exprimés en unités métriques via une fonction mathématique.



Source: [IGN](#)

Cette transformation induit nécessairement une distortion.

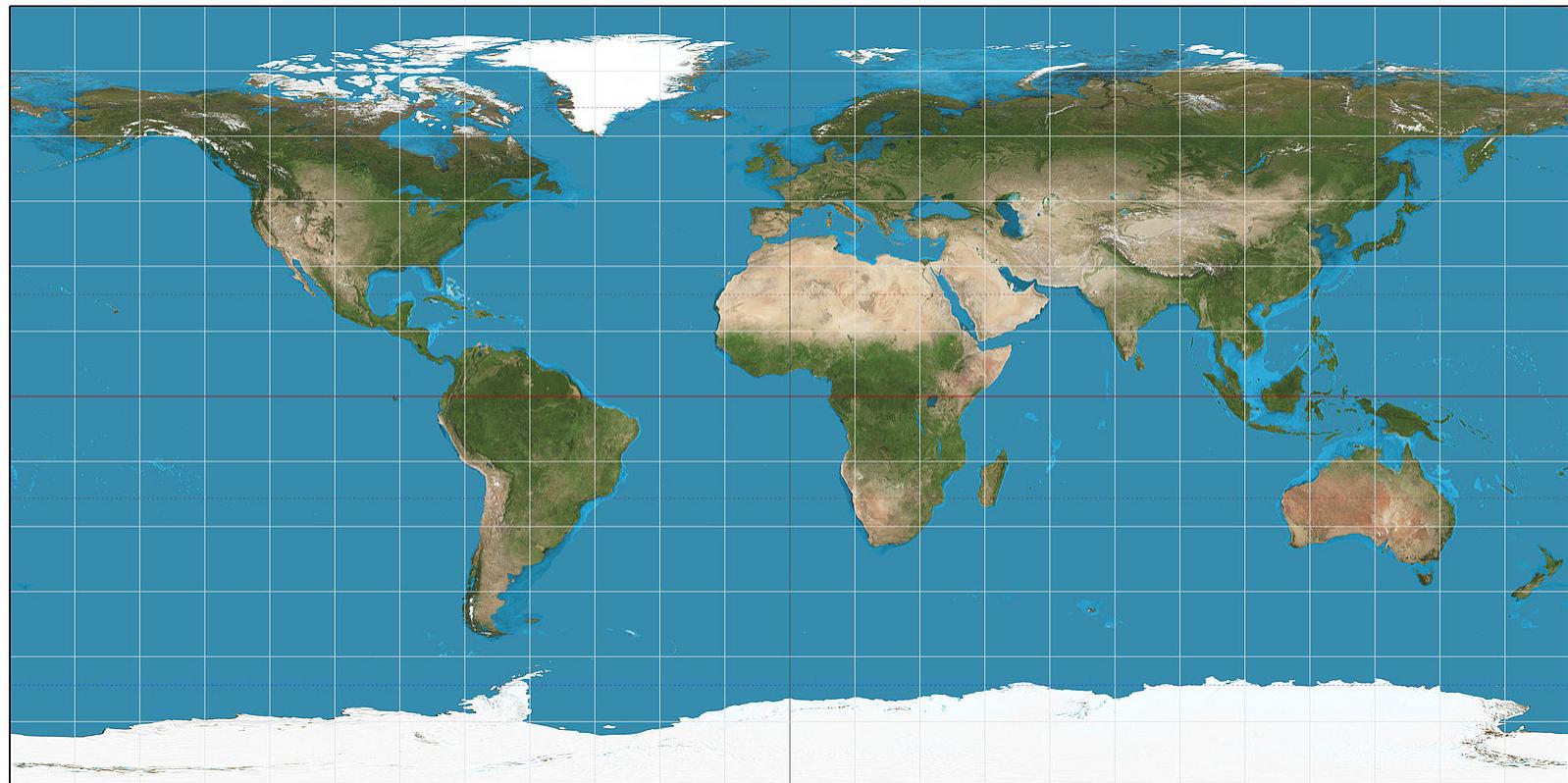
# Projection

Il existe une multitude de projections

- Certaines sont dites équivalentes: elles conservent les distances
- D'autres sont dites conformes: elles conservent les angles (c'est à dire les formes)
- La plus célèbre est probablement celle de Mercator

# Conservation des distances

Projection dite "Équirectangulaire", inventée par Marin de Tyr c. 120, conserve les distances le long des méridiens (distance d'un pôle à l'autre).



Source: [Wikipedia](#)

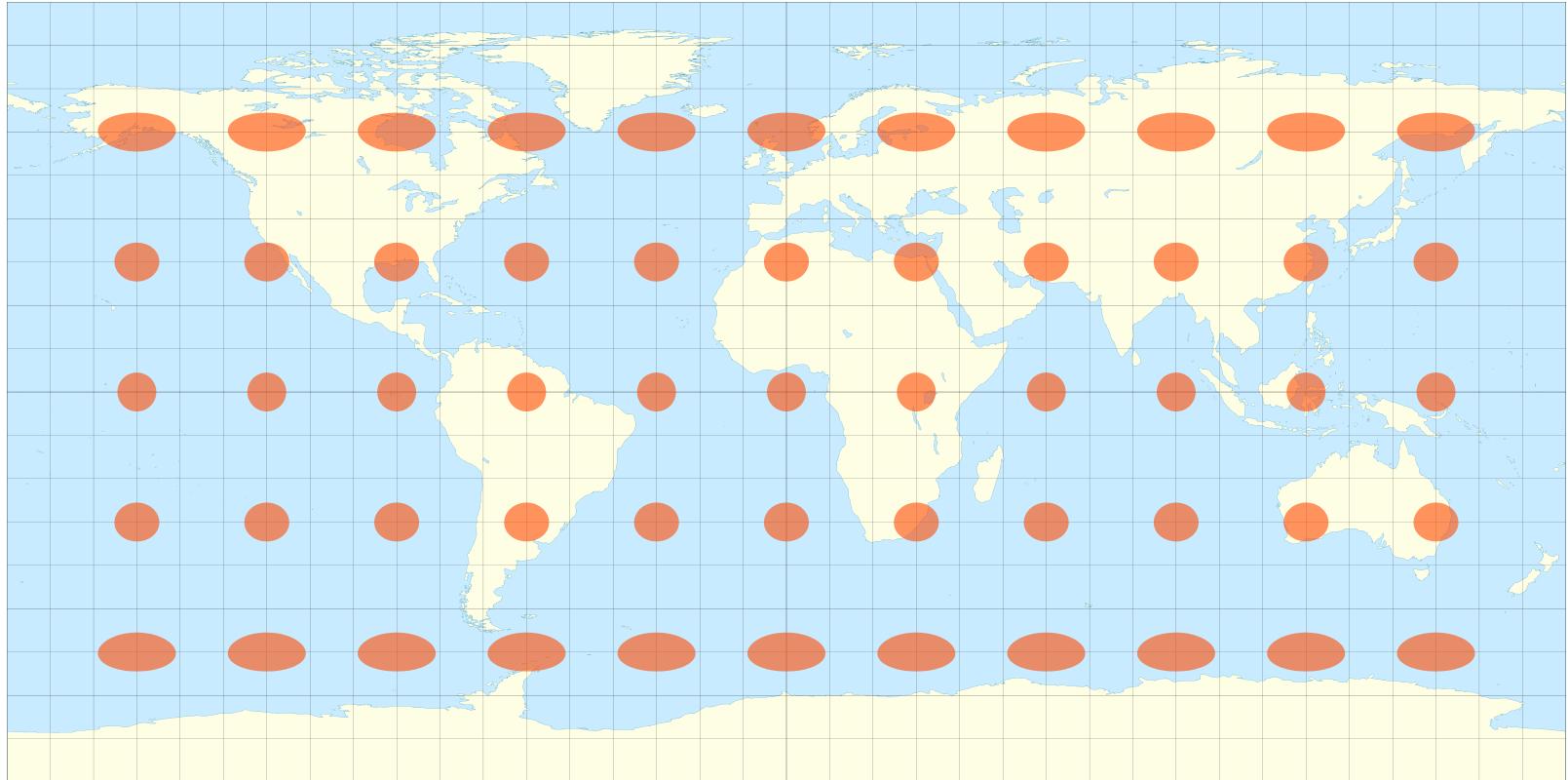
# Conservation des formes

Projection dite "de Mercator", inventée par Gerardus Mercator en 1569: la distorsion s'accroît au fur et à mesure de l'éloignement de l'équateur vers les pôles.



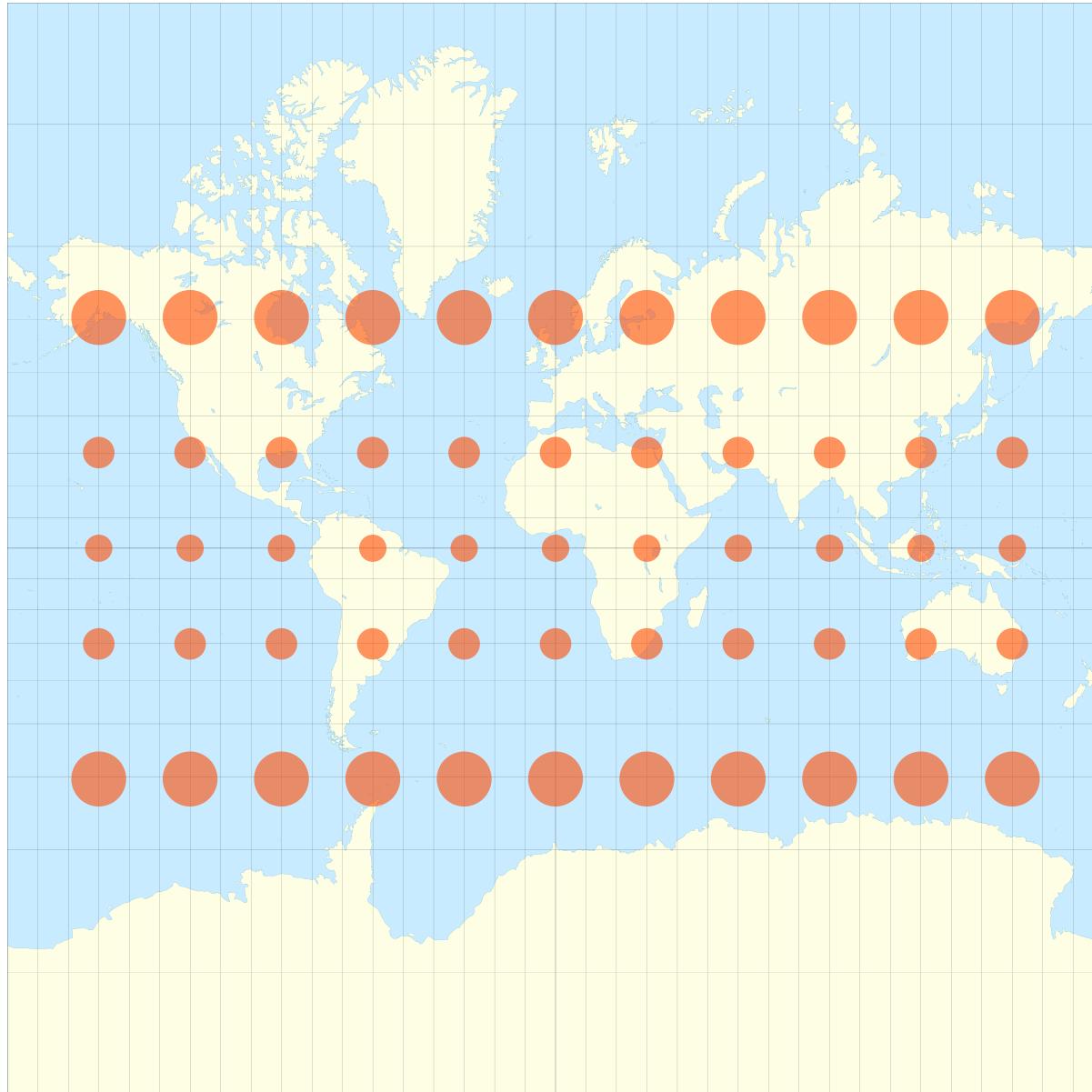
Source: [Wikipedia](#)

# La distortion



Source: [Wikipedia](#)

On peut évaluer la déformation avec des indicatrices de Tissot (ici pour la projection "Équirectangulaire")



Source: [Wikipedia](#)

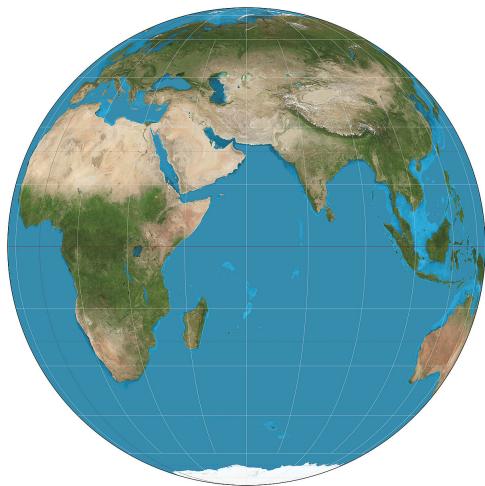
# Autre projections

Il n'existe pas que des projections cylindriques



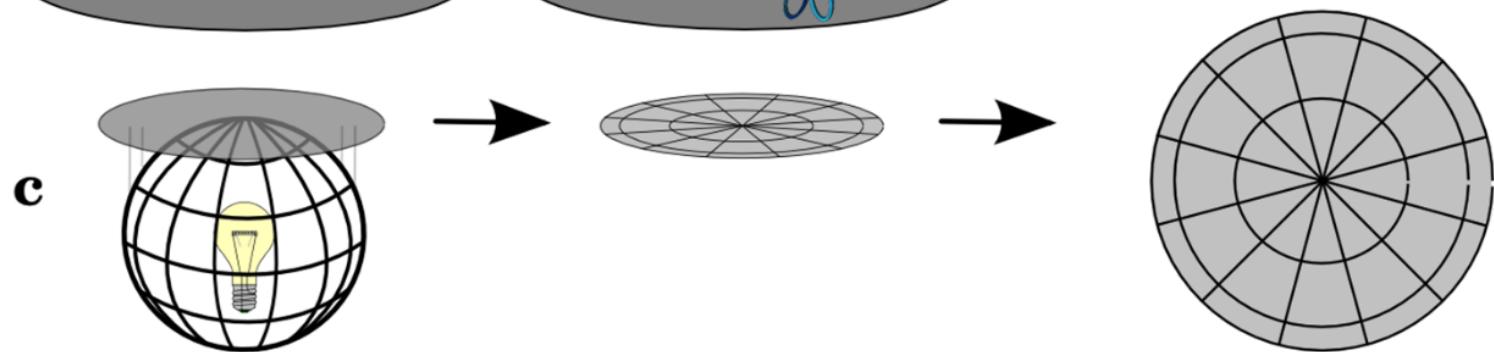
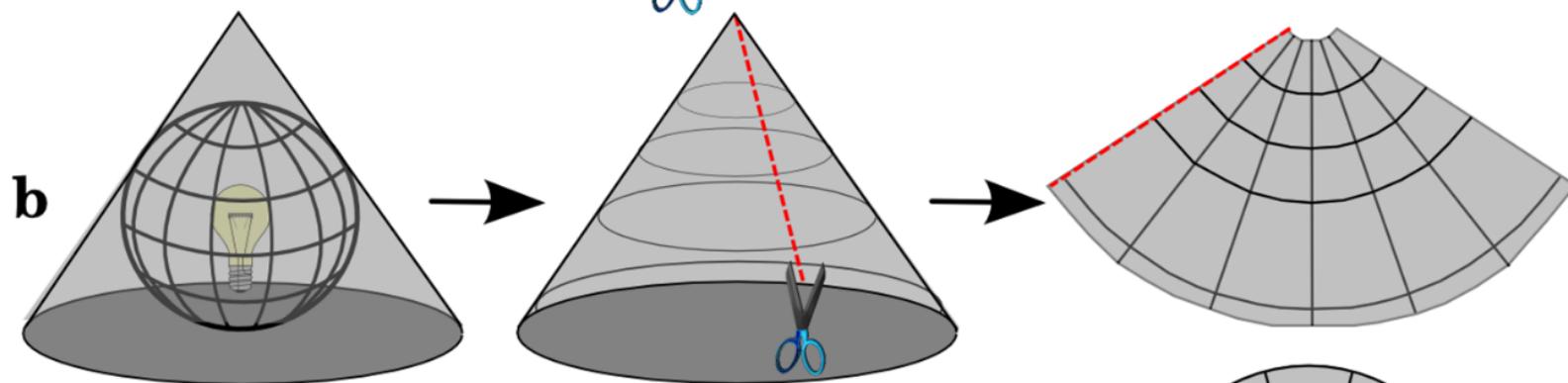
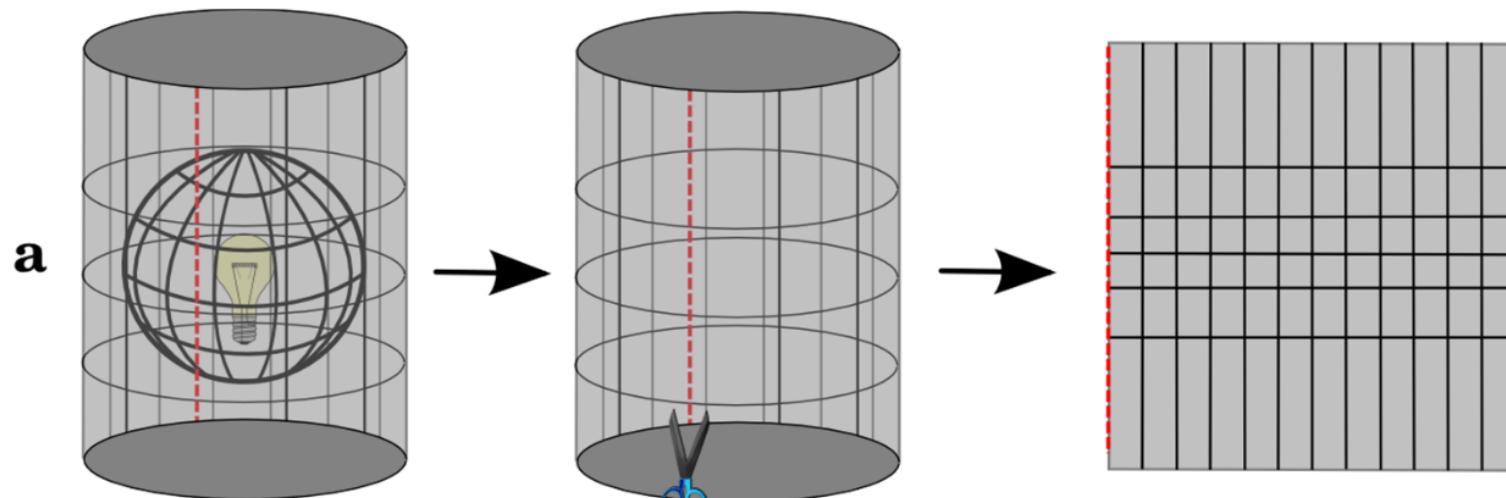
Projection conique (de Lambert)

Source: [Wikipedia](#)



Projection azimutale (orthographique)

Source: [Wikipedia](#)



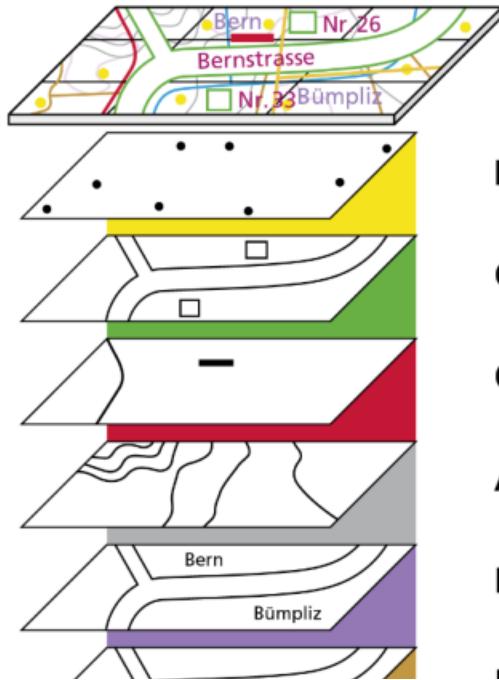
# Modèles, formats

## Vers la géomatique

Géomatique: ensemble des outils et méthodes permettant d'acquérir, de représenter, d'analyser et d'intégrer des données géographiques.  
Elle utilise des SIG.

SIG: Systèmes d'information géographique (*geographic information system* ou GIS), qui permet de représenter graphiquement le contenu des BDG

BDG: base de données géographiques contiennent des informations géographiques



**Points fixes**

**Couverture du sol**

**Objets divers**

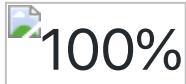
**Altimétrie**

**Nomenclature**

**Pisces fonds**

## Les rasters

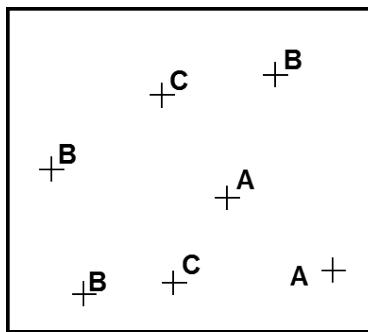
Pour représenter ces couches, on va pouvoir utiliser un modèle raster (ou matriciel), fait de mailles régulières (ou matrice de points) représentées sous la forme d'une grille.



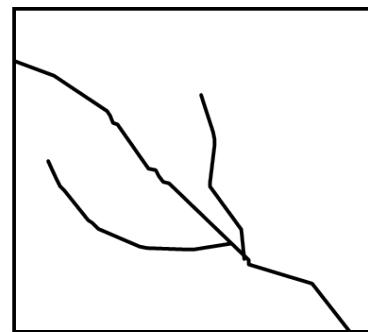
Source: [courses.washington.edu](https://courses.washington.edu)

# Mode matriciel vs mode vectoriel

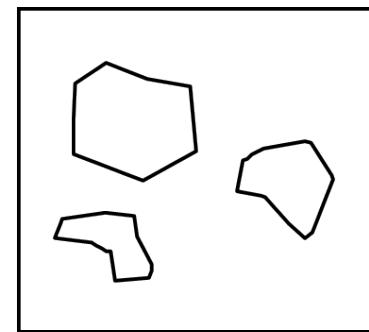
Le modèle raster s'oppose au modèle vecteur. Ce dernier s'appuie sur un découpage a priori du monde en entités (route, région administrative...). En pratique, plutôt qu'une grille qui recouvre toute la carte, y compris les espaces vides, on va faire ressortir seulement quelques points, que l'on va relier.



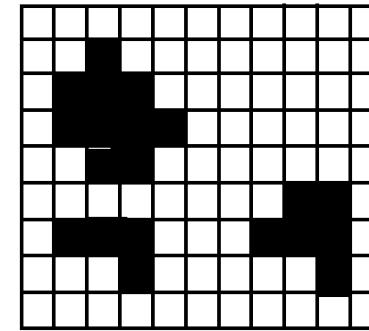
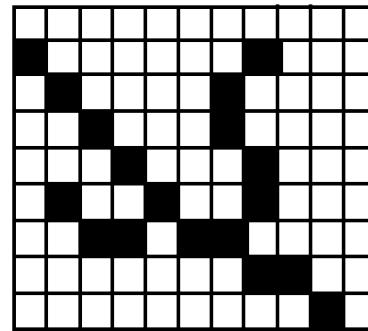
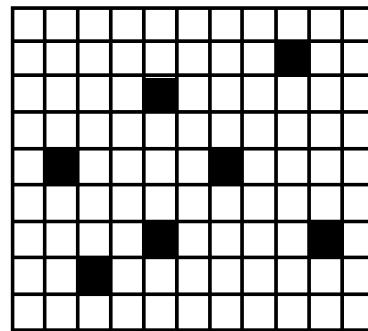
Points



Lignes



Polygones



## *pros and cons*

- Taille des fichiers: vecteur
- Réalisme: raster
- Précision: vecteur (surtout si mauvaise qualité d'image)
- Continuité géographique: vecteur
- Rapidité de constitution des données: raster

Comme nous n'allons pas nous lancer dans les prises de vue satellite, regardons les fichiers nécessaires aux représentations vectorielles.

# Shapefile

Le shapefile, ou fichier de formes, format d'informations géographiques ouvert.

Fichier	Contenu	Présence
.shp	Formes géométriques	obligatoire
.dbf	Base de données (attributs aux formes)	obligatoire
.shx	Lien entre .dbf et .shp	obligatoire
.prj	Info. sur le syst. de coordonnées	facultatif
.xml	Métadonnées	facultatif

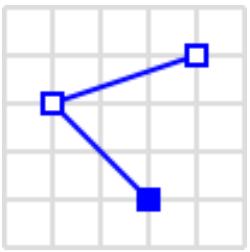
## **GeoJSON**

Il permet de décrire des données de type point, ligne, chaîne de caractères, polygone, ainsi que des ensembles et sous-ensembles de ces types de données et d'y ajouter des attributs d'information qui ne sont pas spatiales.

# JSON

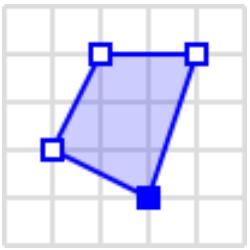
C'est un dérivé du format Json (*JavaScript Object Notation*), qui fonctionne avec des noms associés à des valeurs pour structurer de l'information.

```
{  
    "objet_1": "truc",  
    "id": "0001",  
    "info_a": "chose",  
    "info_b": "machin",  
    "objet_2": {  
        "id": "0002",  
        "info_a": "chose",  
        "info_b": {  
            "liste": [  
                { "item_1.1": "machin", "item_1.2": "bidule" },  
                { "item_2": "chouette" }],  
            ]  
        }  
    }  
}
```



```
{ "type": "Point",  
  "coordinates": [30, 10]  
}
```

Source: [Wikipedia](#)



```
{ "type": "Polygon",  
  "coordinates": [  
    [[30, 10], [40, 40], [20, 40], [10, 20], [30, 10]]  
  ]  
}
```

Source: [Wikipedia](#)

## Entrée complète avec métadonnées

```
{  
  "type": "Feature",  
  "geometry": {  
    "type": "Polygon",  
    "coordinates": [  
      [  
        [100.0, 0.0], [101.0, 0.0], [101.0, 1.0],  
        [100.0, 1.0], [100.0, 0.0]  
      ]  
    ]  
  },  
  "properties": {  
    "Postal code": "2000",  
    "Name": "Neuchâtel"  
  }  
}
```

## ***Leaflet***

Bibliothèque javascript, depuis développée en *R* (package *leaflet*) et en python (librairie *Folium*). Comme l'indique son langage d'origine (javascript, donc), elle est particulièrement utilisée par les développeurs web.

*Leaflet* est au cœur du projet *openStreetMap* (<https://www.openstreetmap.org>), alternative libre et gratuite à *Google maps*.

# Géoréférencement

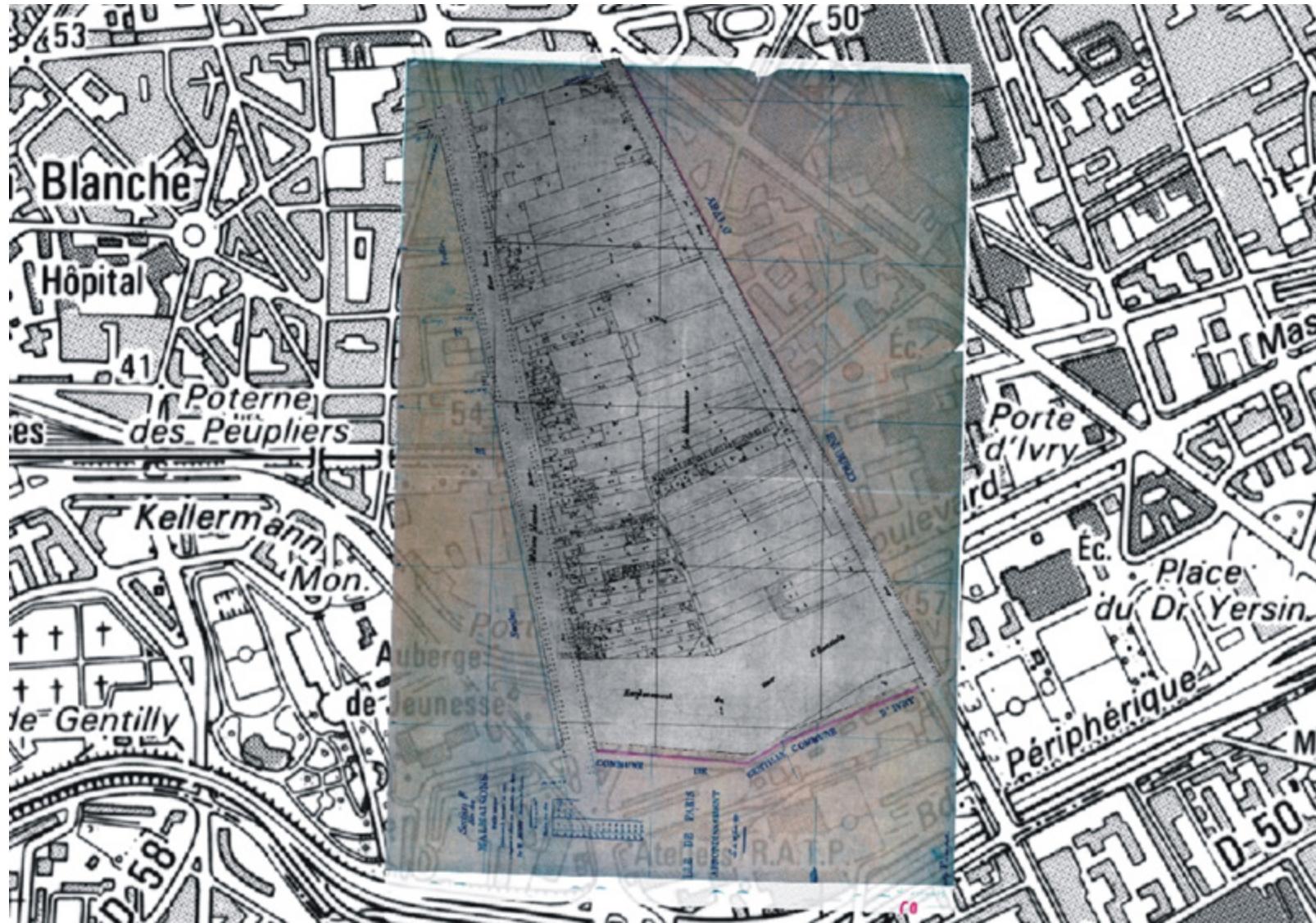
## Définition

Le géoréférencement est le processus dans lequel on applique à une entité cartographique un emplacement spatial en lui donnant des coordonnées géographiques et en appliquant une transformation.

# Problèmes

- Comment définir quelles sont les bonnes coordonnées géographiques? (où est précisément New York?)
- Comment retrouver des lieux disparus? (où est Troie?)
- Comment faire la différence entre un même lieu dans le temps? (Paris médiéval vs le Grand Paris)
- Comment reconnaître un même lieu avec différent noms? (Constantinople vs Istanbul)
- Comment harmoniser les données pour améliorer l'interopérabilité? (ID021432 vs city\_035gr134)

# Géoréférencement: l'image



# Géoréférencement: le texte

**EDMOND LUDLOW** (vers 1617-1692) est un parlementaire **ANGLAIS**, plus connu pour son implication dans l'exécution de **CHARLES IER**, et pour ses mémoires, publiés à titre posthume et qui sont devenus une source importante pour les historiens des **GUERRES DES TROIS ROYAUMES**. Après avoir servi dans les **GUERRES CIVILES ANGLAISES**, Ludlow a été élu membre du **LONG PARLEMENT**. Après la création du **COMMONWEALTH** en 1649, il est nommé adjoint de Ireton, commandant des forces du Parlement en **IRLANDE**, avant de rompre avec **OLIVER CROMWELL** lors de la création du **PROTECTORAT**. Après la Restauration, **LUDLOW** part en exil en **SUISSE**, où il passe une grande partie du reste de sa vie.

## Interopérabilité

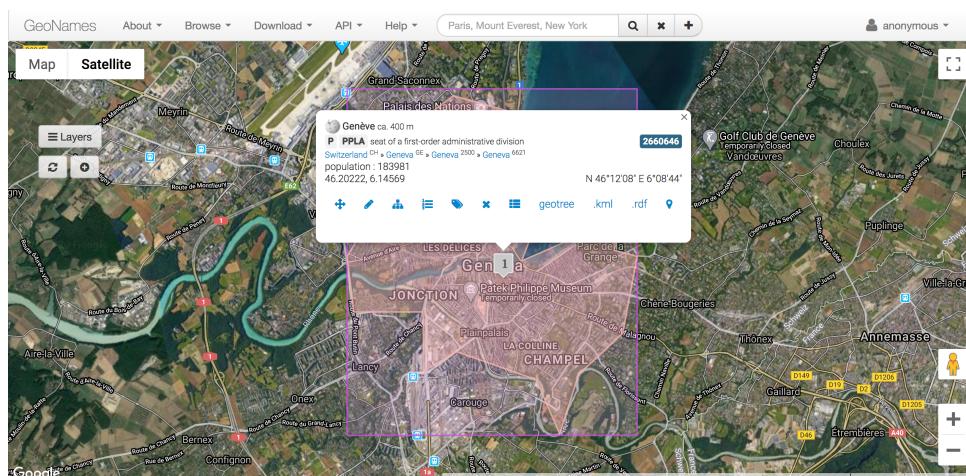
Il existe des référentiel des noms géographiques – comme pour les personnes physiques au demeurant. Il s'agit d'une liste qui fait autorité: chaque entité (nom de lieu, de personne) possède un identifiant unique qui permet d'harmoniser l'information.

Ces référentiels peuvent être assortis d'informations supplémentaires. On trouve souvent le nom de l'entité en différente langues, ou avec différentes orthographies. Parmi ces informations, on trouve la longitude et la latitude pour les noms de lieux: on parle alors de gazette (*gazetteer*).

# GeoNames

GeoNames (<https://www.geonames.org>) permet de faire des recherches dans les informations géographiques d'aujourd'hui.

The screenshot shows a web browser displaying the GeoNames search results for "Genève". The URL in the address bar is [www.geonames.org/search.html?q=Genève&country=CH](https://www.geonames.org/search.html?q=Genève&country=CH). The page title is "GeoNames Home | Postal Codes | Download / Webservice | About". A search bar at the top contains "Genève" with a dropdown menu showing "all countries" and "search [advanced search]". Below the search bar is a table with 1315 records found for "Genève". The columns are "Name", "Country", "Feature class", "Latitude", and "Longitude". The first record is "Geneva" with "Switzerland, Geneva" as the country and "seat of a first-order administrative division" as the feature class. The latitude is N 46° 12' 8" and the longitude is E 6° 8' 44". Other entries include "Geneva International Airport" (CH, FR, airport), "Carouge" (Switzerland, Geneva, populated place), "Thônex" (Switzerland, Geneva, populated place), "Versoix" (Switzerland, Geneva, populated place), "Vernier" (Switzerland, Geneva, populated place), "Plan-les-Ouates" (Switzerland, Geneva, populated place), "Onex" (Switzerland, Geneva, populated place), "Meyrin" (Switzerland, Geneva, populated place), "Geneva" (Switzerland, Geneva, second-order administrative division), "Anières" (Switzerland, Geneva, populated place), and "Jussy" (Switzerland, Geneva, populated place). The table has a green header row and rows for each record.



<http://www.geonames.org/2660646>

# Pleiades

*Pleiades* (<https://pleiades.stoa.org>)

The screenshot shows the Pleiades homepage with a grid of ancient place entries. The first row contains three entries: 'Rome - 423025' (a map of Italy with a blue marker), 'Romeyer - 167886' (a map of a mountainous region with a blue marker), and 'Romechium - 456119' (a blank page with the note 'An ancient place, cited: BAtlas 46 unlocated Romechium'). The second row contains one entry: 'Rometta - 462471' (a map of Sicily with a blue marker). Below the grid is a detailed view for the place 'Rome'. The view includes a header with the place name, a list of creators and contributors, copyright information, and a history link. It also shows a green bar with tags: 'dare:major=1', 'dare:feature=major settlement', 'UWHS', 'extant remains', and 'dare:ancient=1'. Below this is a brief description: 'The capital of the Roman Republic and Empire.' A 'Canonical URI for this page:' link is provided. The 'Representative Point (Latitude, Longitude):' is listed as 41.891775, 12.486137. A 'Locations:' section lists 'DARMC location 30635 (750 BC - AD 640)'. The 'Names:' section lists several names in various languages, including 'Ῥώμη' (Rōmē) in Ancient Greek and 'Roma' in Latin and English. The 'Connections:' section lists connections to 'Tiberis (river)', 'Via Aemilia Scauri', and 'Via Appia'. To the right of the detailed view is a large map of Italy and surrounding regions, with numerous green markers indicating other ancient sites.

<https://pleiades.stoa.org/places/423025> (ajoutez /json et observez ce que vous obtenez)

# World-Historical Gazetteer



World Historical Gazetteer

BETA 0.3

[site tour](#) | [feedback](#)

[API](#) [Tutorials](#) [About](#) ▾

[Register](#) [Login](#)

## Istanbul

Attestations

Related

Related: [WWII in Istanbul, Turkiye, Asia, World](#)

**Istanbul**

dataset: [black](#) (14273)

Variants: [Byzantium](#) ; [Istanbul](#)

Types: [settlement](#)

Links: [dbp:Istanbul](#) (closeMatch) ; [gn:745042](#) (closeMatch) ; [gn:745044](#) (closeMatch) ; [loc:n79034985](#) (closeMatch) [tgn:7002473](#) (closeMatch) ; [viaf:135931454](#) (closeMatch) [wd:Q406](#) (closeMatch)

1914/1918, 1939/1942, 1942/1945, 1945/1952, 1947/1968

**Constantinople**

dataset: [black](#) (12203)

Variants: [Byzantium](#) ; [Constantinople](#)

Types: [settlement](#) ; [battle](#)

Links: [dbp:Constantinople](#) (closeMatch) ; [gn:745042](#) (closeMatch) ; [gn:745044](#) (closeMatch) ; [loc:n79034985](#) (closeMatch) [tgn:7002473](#) (closeMatch) ; [viaf:135931454](#) (closeMatch) [wd:Q406](#) (closeMatch)

-700/120, 300/500, 313/642, 400/650, 400/800, 500/500, 500/1500, 600/800, 622/627, 750/809, 762/800, 789/1100, 800/1200, 800/1000, 1040/1100, 1090/1090, 1096/1270, 1100/1300, 1100/1400, 1200/1700, 1200/1400, 1204/1300, 1225/1350, 1250/1300, 1300/1500, 1310/1360, 1325/1346, 1379/1405, 1400/1860, 1453/1571, 1512/1639, 1792/1815, 1794/1815, 1800/1913, 1815/1852, 1840/1910, 1860/1920, 1870/1910, 1914/1918, 1914/1918, 1918/1923

**Byzantium**

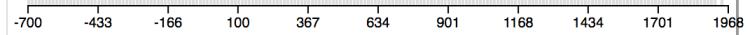
dataset: [black](#) (11491)

Variants: [Byzantium](#) ; [Istanbul](#)

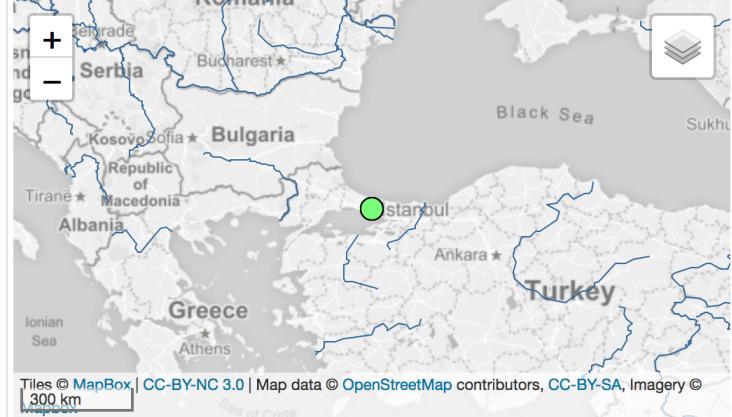
Types: [state](#)

Links: [dbp:Istanbul](#) (closeMatch) ; [gn:745042](#) (closeMatch) ; [gn:745044](#) (closeMatch)

### TEMPORAL ATTESTATIONS



### GEOGRAPHY



### TRACES (related persons, events, works)

None yet

<http://dev.whgazetteer.org/places/12345979/portal>

## Extraction d'information

En anglais on parle d'*information retrieval* (IR). Le but est d'arriver à retrouver de l'information, donc des noms de lieux, automatiquement dans une source (la plupart du temps écrite).

## Format *BIO*

Avec le développement de l'apprentissage profond (*deep learning*), l'entraînement d'outil d'IR devient crucial. Pour cela un standard d'annotation a été élaboré: le format *BIO* (*beginning, inside, outside*) développé lors de la ConLL shared task 2003.

Le principe est d'arriver à annoter/étiqueter des séquences de token qui renvoient à des entités nommées (noms de personne, de lieu, d'organisation...) pour constituer des corpus qui permettent d'entraîner des modèles pour la reconnaissance d'entités nommées (*Named entities recognition* ou NER)

TOKEN	POS (CATTEX)	BIO (ConLL)
Monsieur	NOMcom	PER-B
de	PRE	PER-I
La	DETdef	PER-I
Rochefoucauld	NOMpro	PER-I
est	VERcjg	O
un	DETndf	O
écrivain	NOMcom	O
né	VERppe	O
à	PRE	O
Paris	NOMpro	GEO-B
en	PRE	O
France	NOMpro	GEO-B