

Cours *Distant Reading* : Visualisation

10. Introduction à la cartographie

Simon Gabay



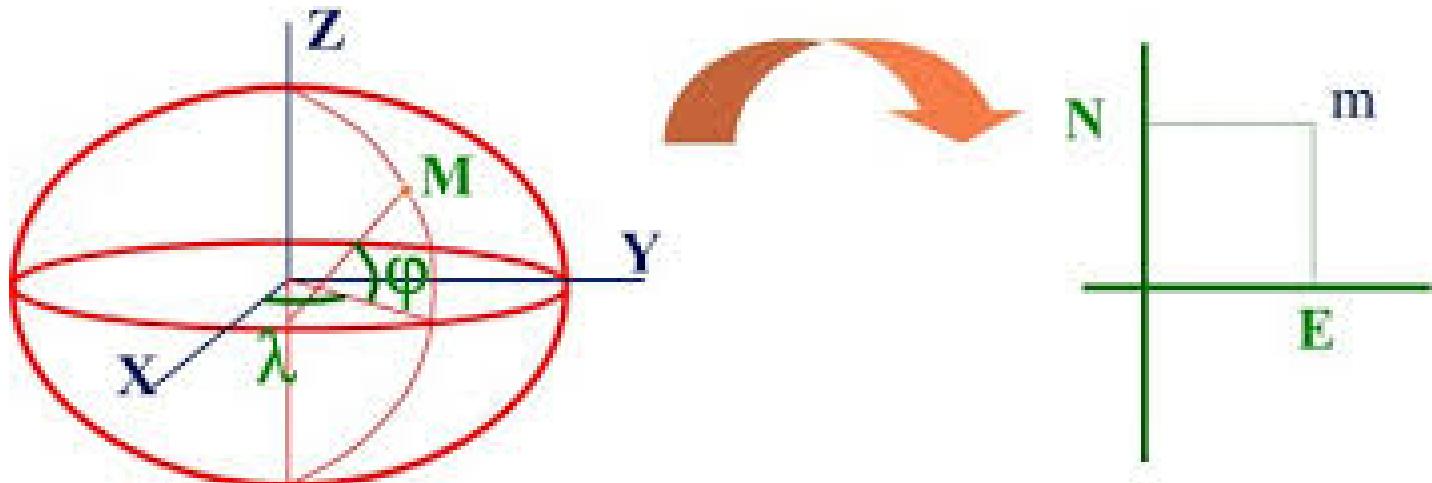
Coordonnées et projection

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), plus pratiques pour leur manipulation informatique.

Coordonnées cartographiques

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



Source: [IGN](#)

Cette transformation induit nécessairement une distortion.

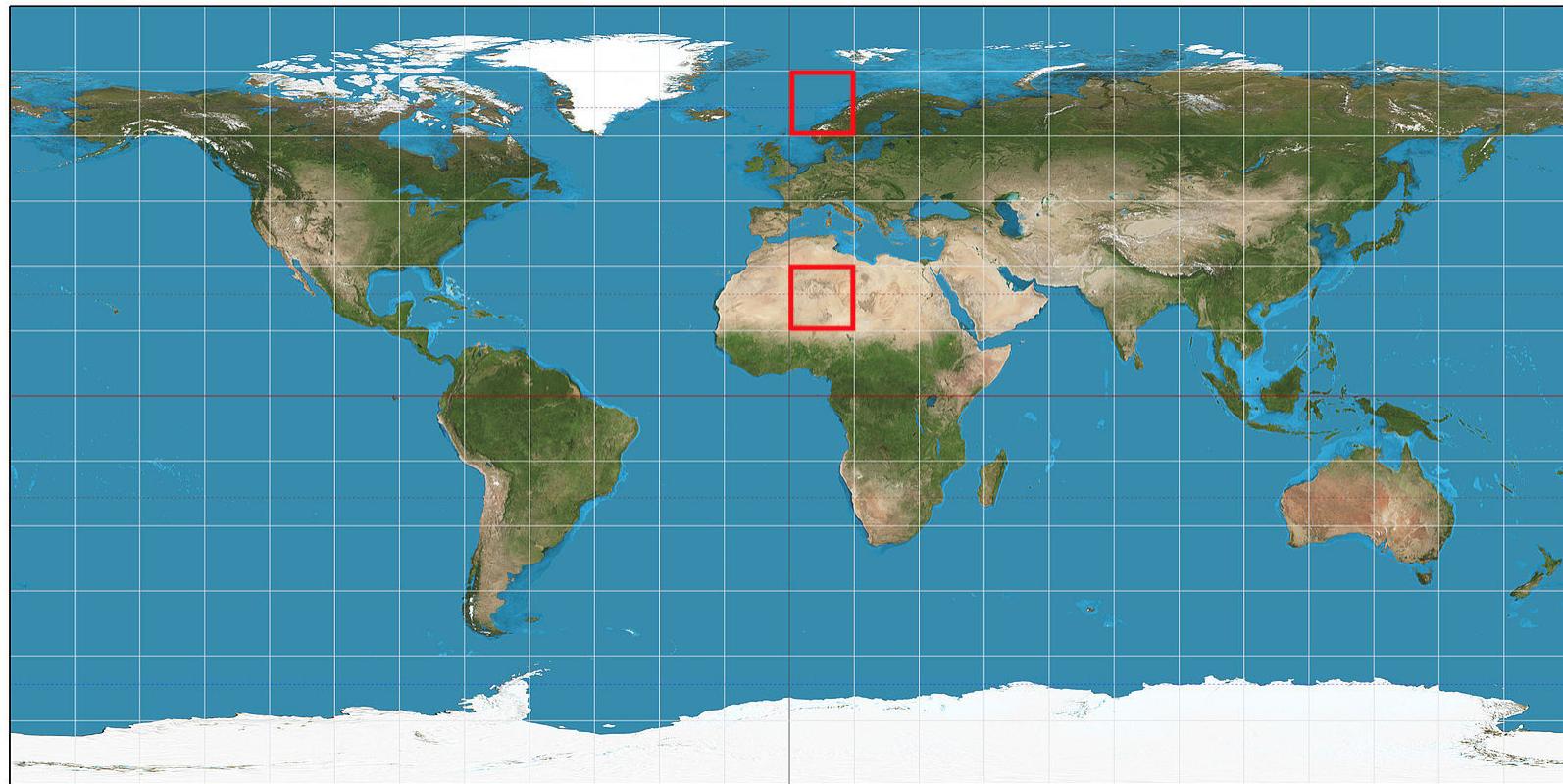
Projection

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, tout du moins de nom, est probablement celle de Mercator, mais il en existe plein d'autres

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](#)

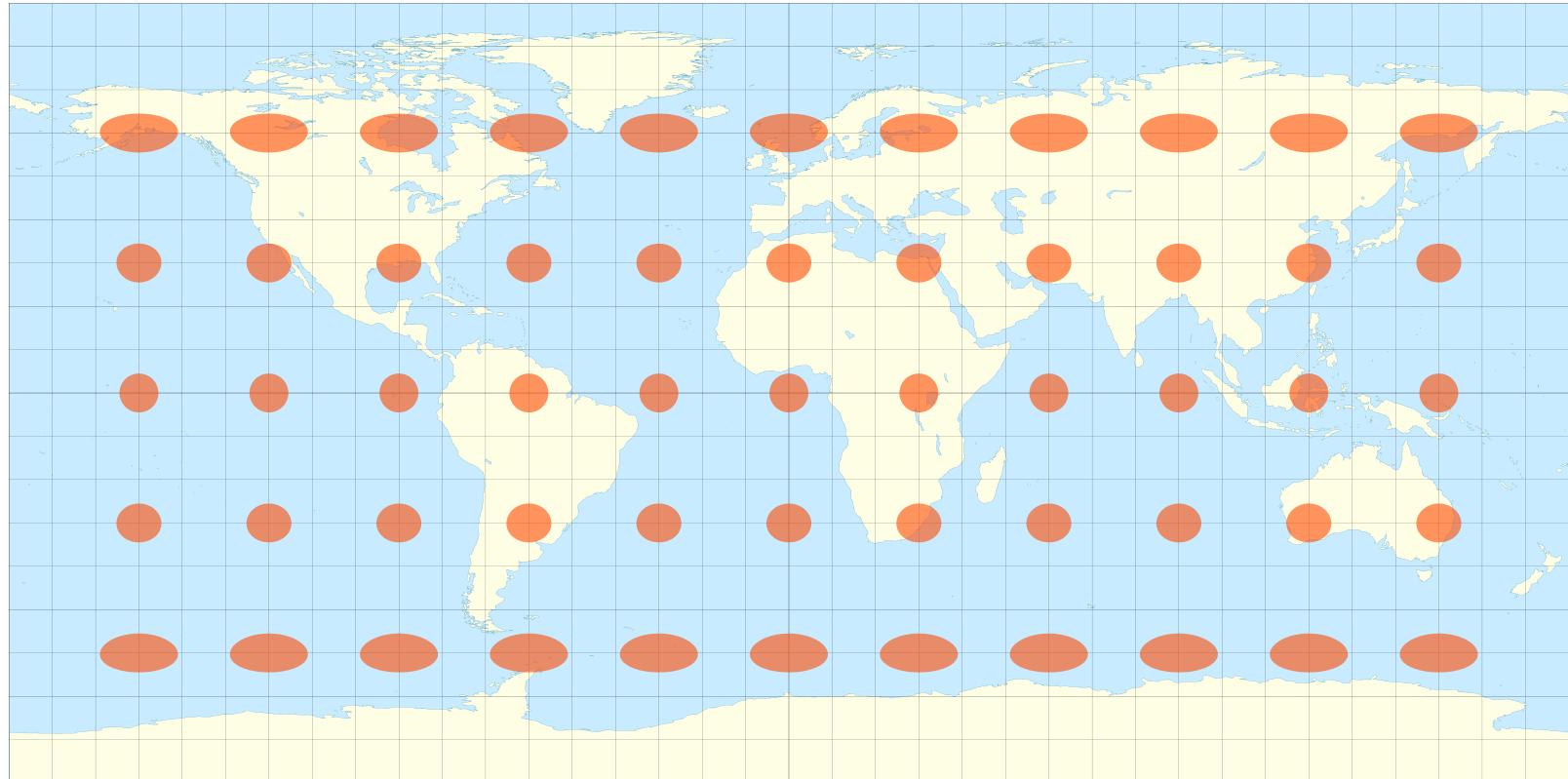
Conservation des formes

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



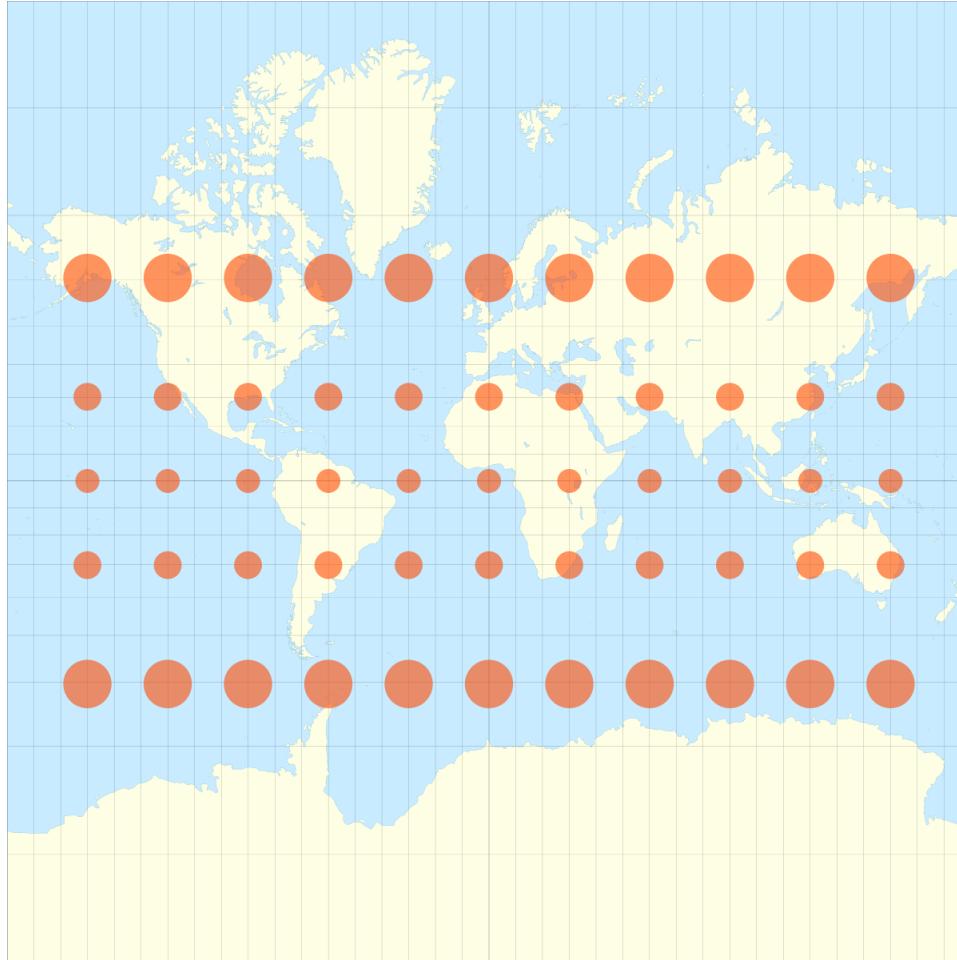
La distortion

On peut évaluer la déformation avec l'indicatrice de Tissot



Source: [Wikipedia](#)

Pour la projection "Équirectangulaire": la hauteur des cercles varie peu, car les distances sont conservées le long des méridiens.

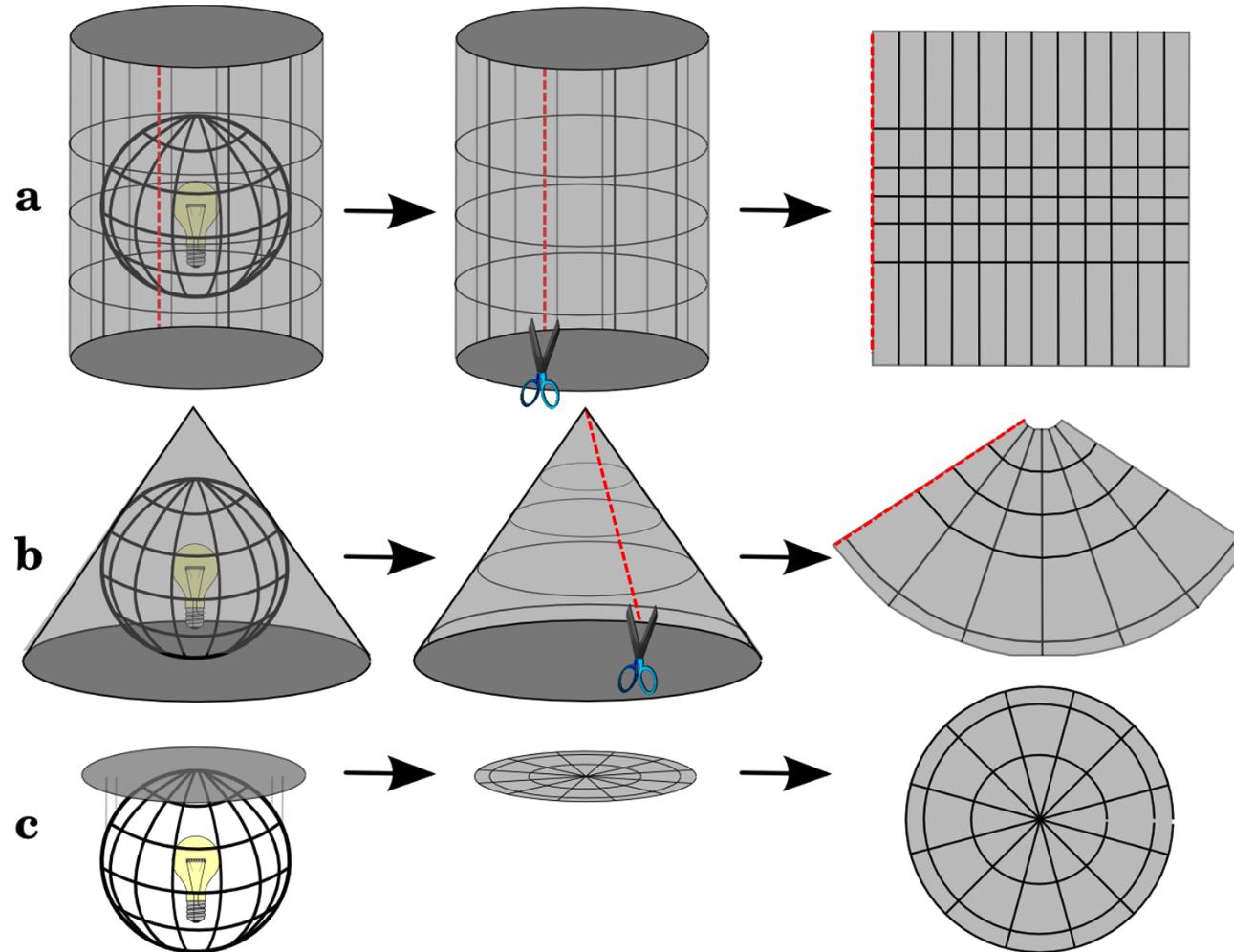


Source: [Wikipedia](#)

Pour la projection de Mercator, les indicatrices de Tissot montre une grande variation dans le diamètre des cercles mais pas de déformation ovaloïde: la forme est conservée, mais pas les distances.

Autre projections

Les représentations que nous avons vues sont cylindriques, mais il existe d'autres formes, notamment conique et azimutale.

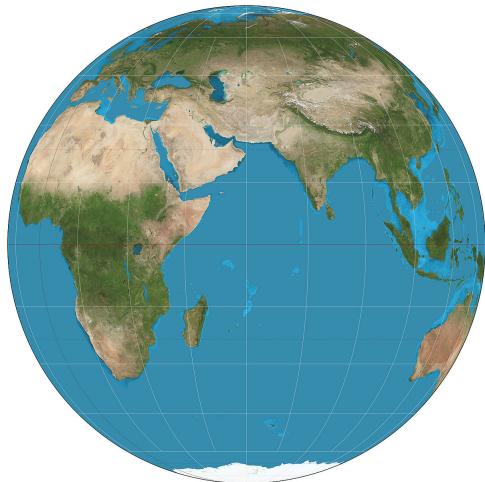


Projections conique et azimutale



Projection conique (de Lambert)

Source: [Wikipedia](#)



Projection azimutale (orthographique)

Source: [Wikipedia](#)

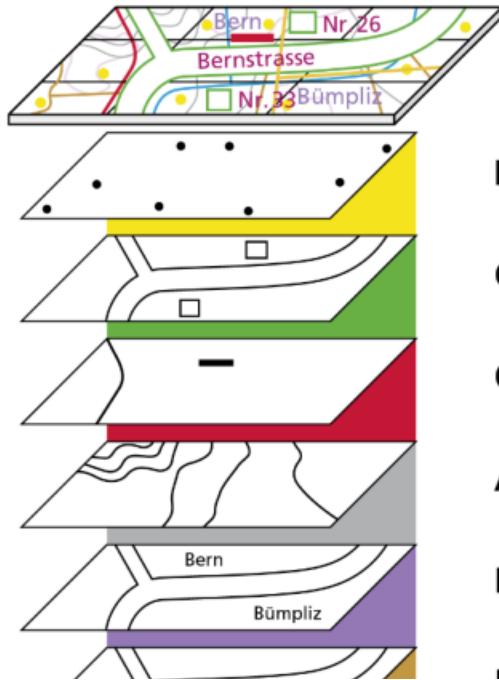
Modèles, formats, logiciels

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 permettant des représentations cartographiques.



Points fixes

Couverture du sol

Objets divers

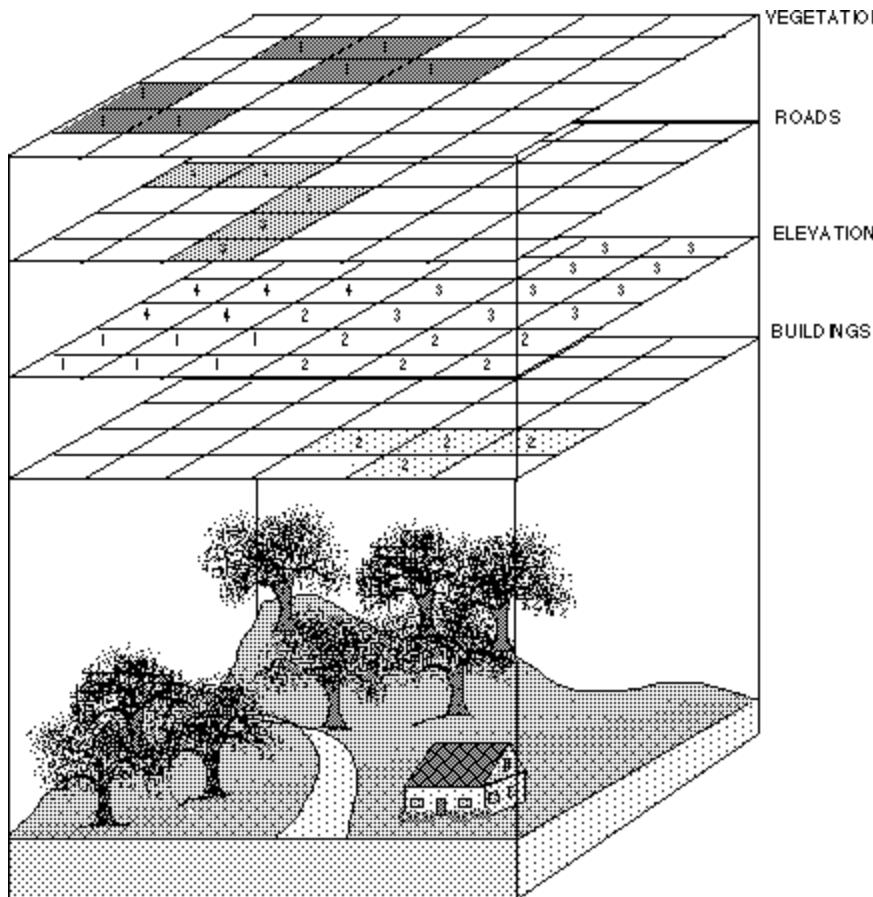
Altimétrie

Nomenclature

Pisces fonds

Les rasters

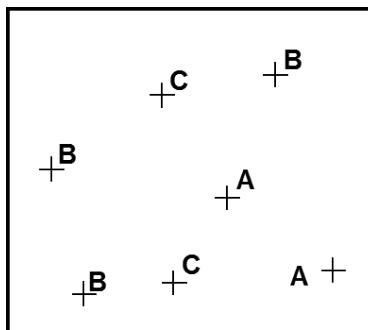
Pour représenter ces couches, on va pouvoir utiliser un modèle raster (ou matriciel), fait de mailles régulières (ou matrice de points) représentées sous la forme d'une grille.



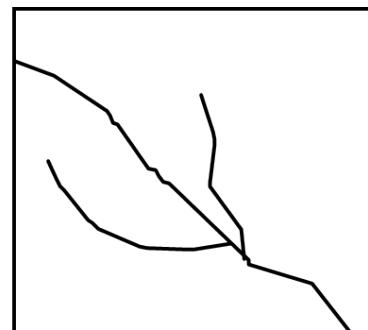
Source: courses.washington.edu

Mode matriciel vs mode vectoriel

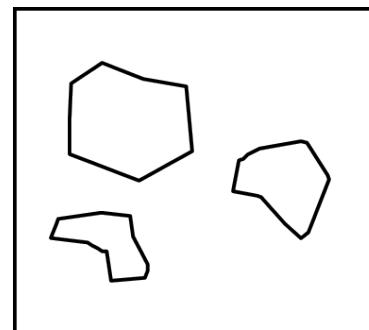
Le modèle raster s'oppose au modèle vecteur. Ce dernier s'appuie sur un découpage *a priori* du monde en entités (route, région administrative...). En pratique, plutôt qu'une grille qui recouvre toute la carte, y compris les espaces vides, on va faire ressortir seulement quelques points, que l'on va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



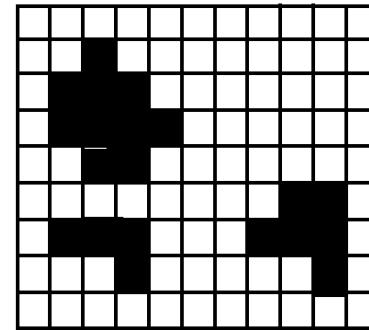
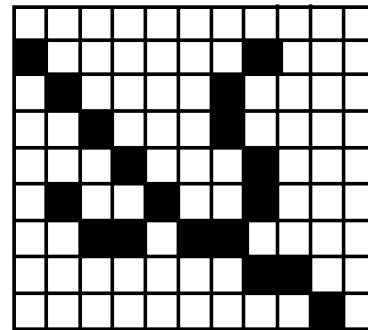
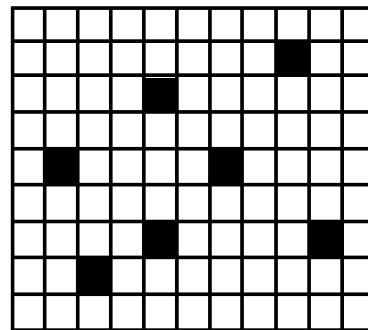
Points



Lignes



Polygones



pros and cons

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

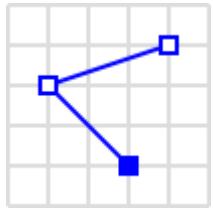
C'est un dérivé du format Json (*JavaScript Object Notation*), qui fonctionne avec des noms associés à des valeurs pour structurer de l'information. C'est un format d'échange d'information, qui a pris dans nombre de cas la suite du XML, avec lequel il partage nombre de points communs

```
<objet>
  <id>0001</id>
  <info n="a">chose</info>
  <info n="b">machin</info>
</objet>
```

```
{
  "objet": {
    "id": "0001",
    "info_a": "chose",
    "info_b": "machin",
  }
}
```

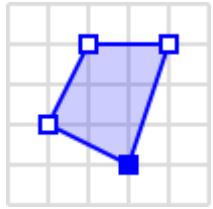
Le format JSON offre une grande flexibilité dans la structuration de l'information.

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



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

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*.

Il existe d'autres bibliothèques très utiles (comme *Gdal*). Nous y reviendrons lors des travaux pratiques.

*GIS

En plus des langages que nous venons de mentionner, il existe plusieurs outils de cartographie numérique. Chez les spécialistes de sciences humaines, les deux plus célèbres sont:

- *ArcGIS*, payant, développé par l'ESRI (*Environmental Systems Research Institute*): <https://www.arcgis.com>
- *QGIS*, gratuit et *open source*: <https://www.qgis.org>

Géoréférencement

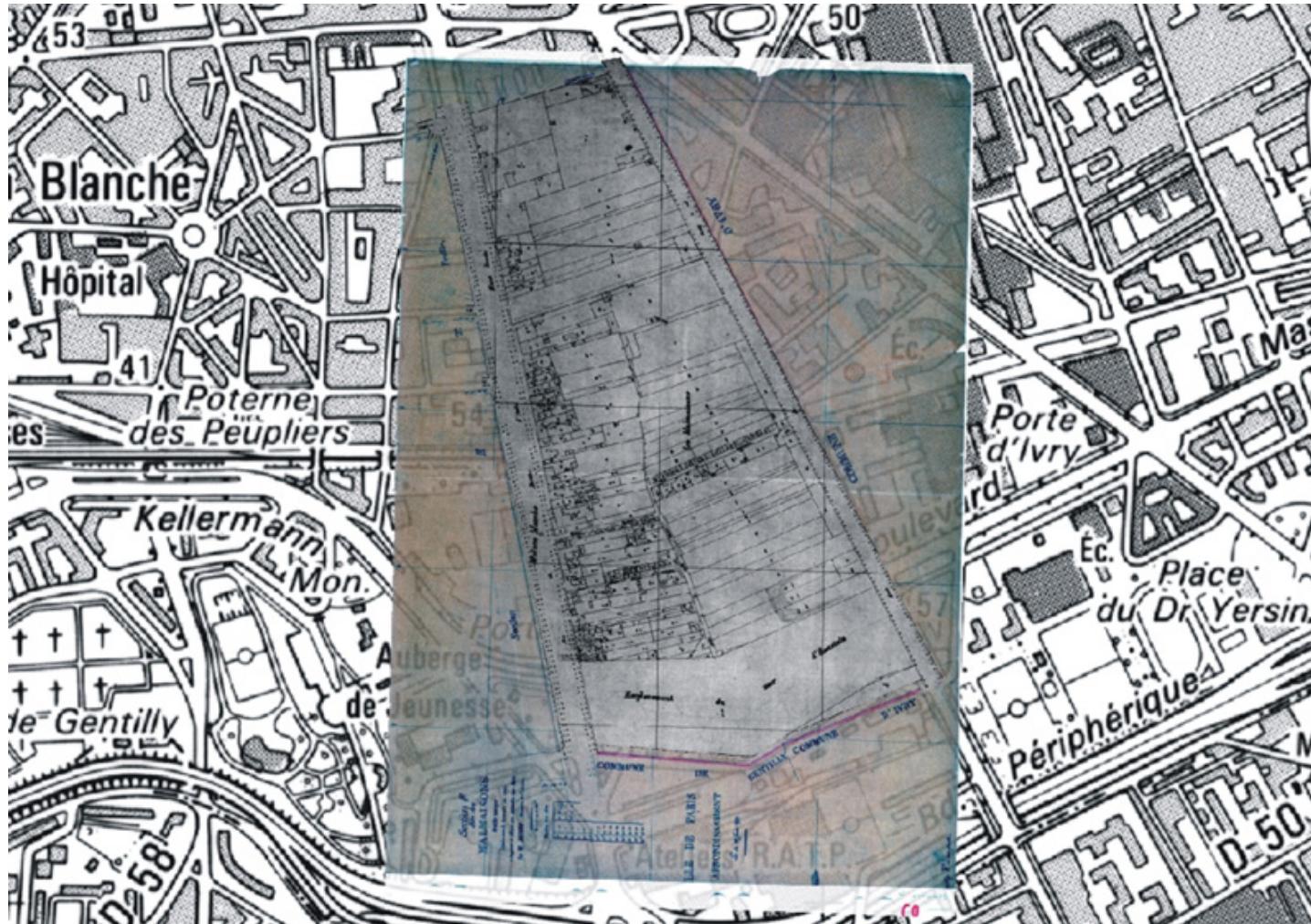
Définition

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



Source: Capizzi 2004

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*).

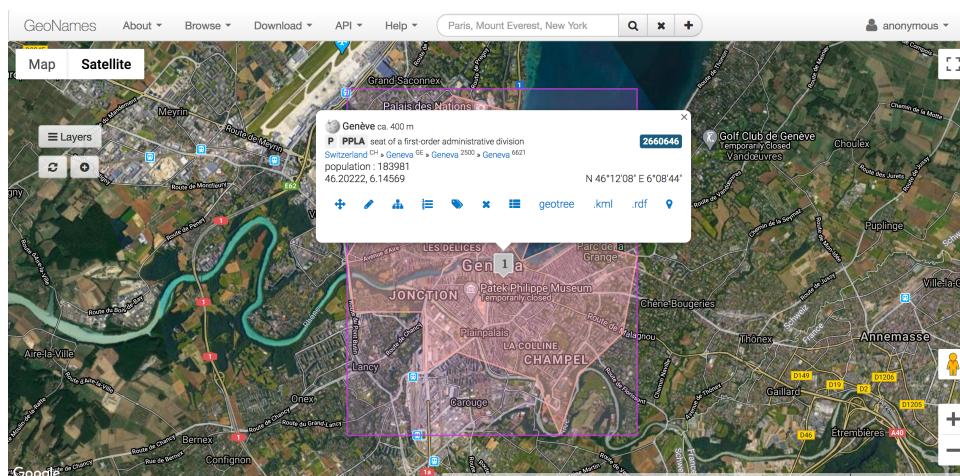
GeoNames

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

Screenshot of the GeoNames search results for "Genève".

The search bar shows "Genève" and the results page displays 1315 records found for "Genève".

Name	Country	Feature class	Latitude	Longitude
Geneva	Switzerland, Geneva	seat of a first-order administrative division Geneva > Geneva	N 46° 12' 8"	E 6° 8' 44"
Geneva	Switzerland, Geneva	first-order administrative division	N 46° 11' 48"	E 6° 6' 37"
Geneva International Airport	CH,FR	airport elevation 430m	N 46° 10' 15"	E 6° 6' 33"
Carouge	Switzerland, Geneva	populated place Geneva > Carouge (GE)	N 46° 10' 51"	E 6° 8' 21"
Thônex	Switzerland, Geneva	populated place Geneva > Thônex	N 46° 11' 17"	E 6° 11' 56"
Versoix	Switzerland, Geneva	populated place Geneva > Versoix	N 46° 17' 1"	E 6° 9' 43"
Vernier	Switzerland, Geneva	populated place Geneva > Vernier	N 46° 13' 1"	E 6° 5' 5"
Plan-les-Ouates	Switzerland, Geneva	populated place Geneva > Plan-les-Ouates	N 46° 10' 4"	E 6° 6' 59"
Onex	Switzerland, Geneva	populated place Geneva > Onex	N 46° 11' 2"	E 6° 6' 8"
Meyrin	Switzerland, Geneva	populated place Geneva > Meyrin	N 46° 14' 3"	E 6° 4' 48"
Geneva	Switzerland, Geneva	second-order administrative division Geneva > Geneva	N 46° 12' 28"	E 6° 7' 38"
Anières	Switzerland, Geneva	populated place Geneva > Anières	N 46° 16' 36"	E 6° 13' 19"
Jussy	Switzerland, Geneva	populated place Geneva > Jussy	N 46° 14' 9"	E 6° 16' 1"



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

Pleiades

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

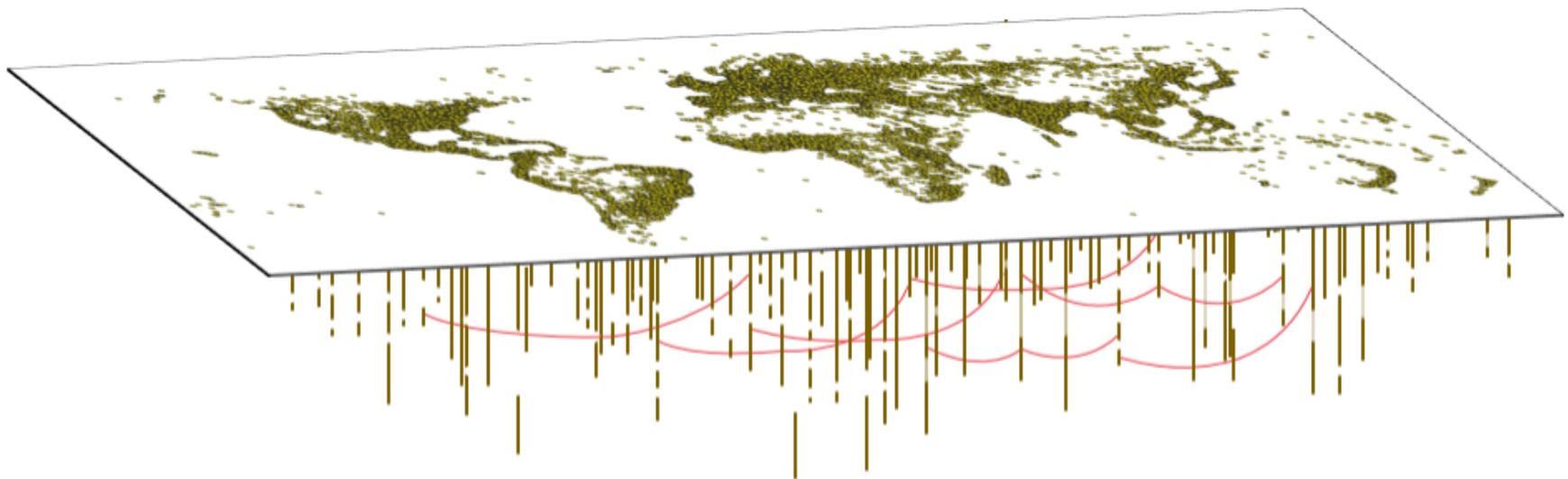
The screenshot shows the Pleiades interface. At the top, there are four small place cards: 'Rome - 423025' (map of Italy with a blue marker), 'Romeyer - 167886' (map of a coastal area with a blue marker), 'Romechium - 456119' (empty placeholder), and 'Rometta - 462471' (map of Sicily with a blue marker). Below these is a larger, detailed view of the 'Roma - 423025' page. The page title is 'Roma' with a small icon, followed by 'a Pleiades place resource'. Below this are sections for 'Creators', 'Contributors', 'Copyright', and 'Last modified'. A green bar at the bottom contains the tag 'tags: dare:major=1, dare:feature=major settlement, UWHs, extant remains, dare:ancient=1'. The main content area describes 'The capital of the Roman Republic and Empire.' It includes a 'Canonical URI for this page:' link, a 'Representative Point (Latitude, Longitude):' coordinate, a 'Locations:' section listing 'DARMC location 30635 (750 BC - AD 640)', a 'Names:' section listing 'Póλη (Rome: Ancient Greek, 30 BC - AD 300)', 'Roma (Latin, 750 BC - AD 640)', and 'Rome (English, modern)', and a 'Connections:' section listing links to 'Tiberis (river)', 'Via Aemilia Scauri', and 'Via Appia'. To the right of the text is a large map of the Tiber River area in ancient Rome, showing the river, hills, and surrounding regions.

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

World-Historical Gazetteer

Projet *World-Historical Gazetteer* à l'université de Pittsburgh

1. Ajouter de la profondeur diachronique à la carte
2. Tisser des liens entre les différents points (né ici, mort là-bas, passé par cet endroit...): cf. projets *Linked places* et *Linked traces*.



L'exemple de Byzance/Constantinople/Istanbul

World Historical Gazetteer BETA 0.3 site tour | feedback API Tutorials About Register Login

Istanbul

Attestations Related

Related: [Byzantium](#), [Istanbul](#), [Turkey](#), [Asia](#), [World](#)

Istanbul [highlight](#) [location](#) 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

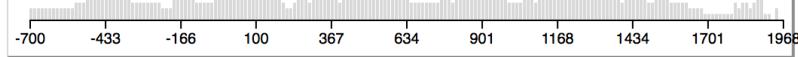
Constantinople [highlight](#) [location](#) dataset: [black \(12203\)](#)

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

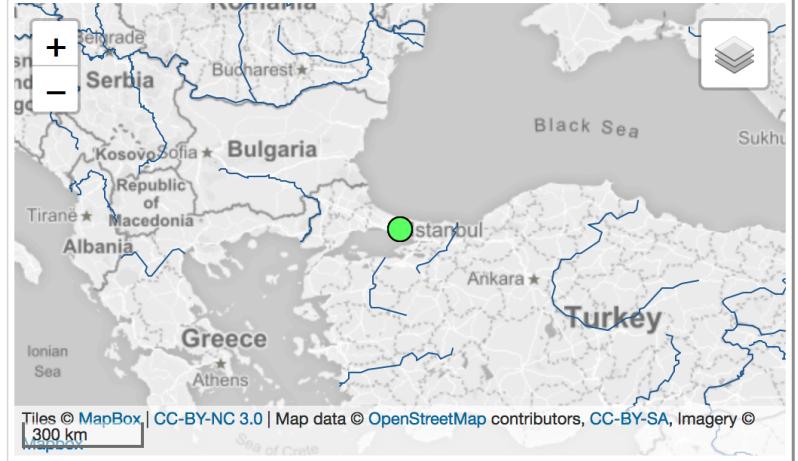
Byzantium [highlight](#) [location](#) dataset: [black \(11491\)](#)

Variants: [Byzantium](#) ; [Istanbul](#)
Types: [state](#)
Links: [dbp:Istanbul](#) (closeMatch) ; [gn:745042](#) (closeMatch) ; [gn:745044](#) (closeMatch)

TEMPORAL ATTESTATIONS



GEOGRAPHY



nearby places (300 max)

Tiles © MapBox | CC-BY-NC 3.0 | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © 300 km

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)	Entités (ConLL)
Monsieur	POS(CATTEX)	Entités (ConLL)
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
Saint	NOMpro	GEO-B
-	PUNfbl	GEO-I
Martin	NOMpro	GEO-I