

Définition

La géolocalisation est l'action permettant d'associer les coordonnées géographiques correspondantes à une adresse précise.

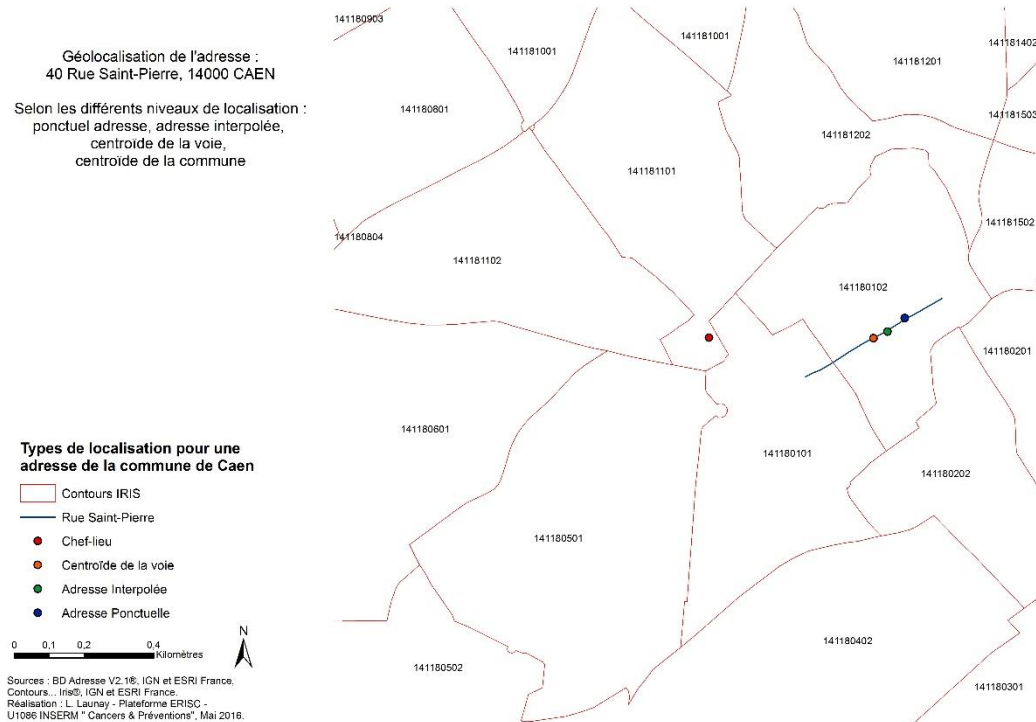


Principe

Elle peut être effectuée grâce à des Systèmes d'Information Géographique (SIG) tels que MapInfo®, ArcGIS®... Ces derniers permettent de rapprocher l'adresse issue d'une base de données à l'adresse présente dans une base de référence. Ces bases de référence proviennent notamment de la BD Adresse de l'IGN et font intervenir aussi bien les ponctuels des adresses que les linéaires des tronçons des voies. D'autres bases peuvent être utilisées telles que la Base Adresse Nationale (BAN), les données OpenStreetMap, les données Navstreets fournies par HERE.

La géolocalisation peut être effectuée à différentes échelles selon les données disponibles :
centroïde d'une unité administrative, centroïde d'une voie, adresse précise (carte).

Carte : Illustration des différents niveaux de localisation – exemple d'une adresse de la commune de Caen



Informations requises

Pour une géolocalisation précise (à l'adresse), il est nécessaire de disposer du :

- Numéro de voie
- Suffixe de la voie
- Type de voie
- Nom de la voie
- Complément d'adresse (si disponible)
- Nom de la commune
- Code postal de la commune
- Code insee de la commune

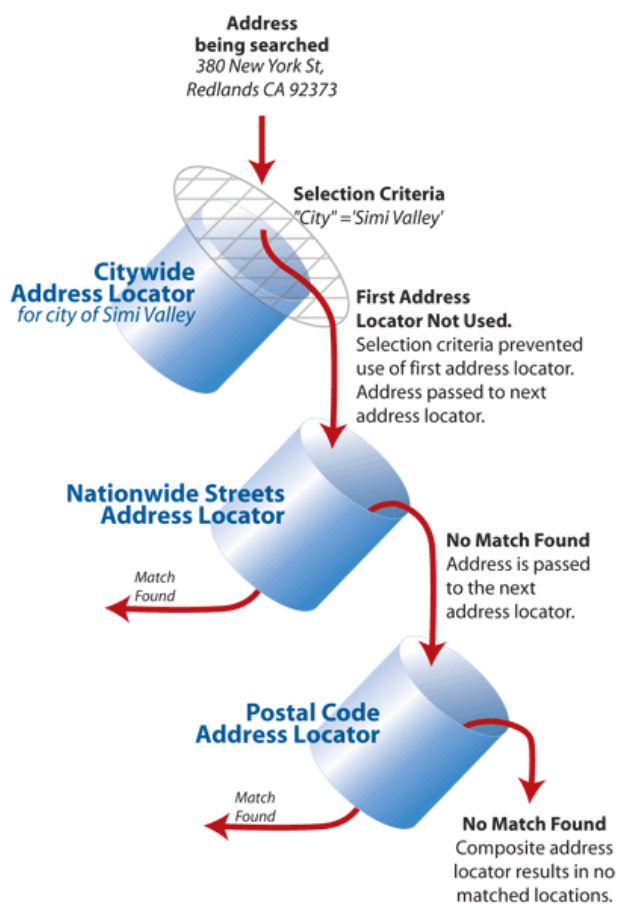
Etape du géocodage

La reconnaissance de l'adresse comprend deux phases :

- Une phase automatique réalisée par le SIG
- Une phase manuelle lorsque le SIG n'a pas trouvé de correspondance (type de voie incorrect, fautes d'orthographe...)

Le géocodage sous ArcGIS est réalisé en cascade (illustration) grâce à différents localisateurs d'adresses regroupés dans un localisateur d'adresses composite.

Illustration : les localisateurs sous ArcGIS



Source :

<http://help.arcgis.com/fr/arcgisdesktop/10.0/help/index.html#//00250000003r000000>

Ces localisateurs permettent une géolocalisation selon les différents niveaux évoqués ci-dessus. Ces localisateurs sont définis par différents critères dont les 3 principaux concernent :

- La sensibilité orthographique : seuil à partir duquel une adresse avec faute d'orthographe est considérée comme candidate,
- Le score de candidat : score à partir duquel une adresse est jugée comme pouvant correspondre à celle de la base,
- Le score d'appariement : score à partir duquel l'adresse est considérée comme "identique".

A l'issue du traitement en automatique, différents cas sont retrouvés :

- L'adresse a été appariée sans ambiguïté,
- L'adresse n'a pas pu être appariée et aucun candidat n'a été trouvé,
- L'adresse n'a pas été appariée en raison de candidats multiples entre lesquels l'utilisateur devra choisir.



La part de données appariées en automatique - et par là-même le temps nécessaire à la réalisation de ce travail - dépend de la taille de la base et de la qualité des données. Privilégiez si possible des adresses de qualité : adresse complète avec le complément d'adresse contenu dans un champ séparé, attention aux fautes d'orthographe et aux articles.

Une fois les coordonnées (x,y) obtenues, il est possible de changer d'échelle au profit de l'IRIS afin d'ajouter des données socioéconomiques à l'échelle agrégée. Il est à noter que les communes de moins de 5 000 habitants ne sont pas découpées en IRIS aussi la valeur de celui-ci est la même quelle que soit la localisation du point. Pour les communes de plus de 5 000 habitants, un découpage est réalisé en unité géographique homogène d'environ 2 000 habitants. Par conséquent, la valeur de l'IRIS dépend des coordonnées géographiques et par là-même de la qualité de la géolocalisation comme illustré ci-dessus (carte).