

Graphes de Connaissances et Ontologie pour la Représentation de Données Immobilières Issues d'Annonces en Texte Libre

Lucie Cadorel, Fabien Gandon, Andrea G. B. Tettamanzi

lucie.cadorel@inria.fr



Ingénierie des Connaissances

3 juillet 2023, Strasbourg, France

Contexte



Utilisation des données immobilières

- Estimation des prix et tendances
- Recommandations de biens
- Données temporelles et spatiales
- Provenance variée (notaires, agents immobiliers, particulier,...)

Annonces immobilières

- Données exhaustives, à jour, massives et facilement accessible
- Caractéristiques du bien (prix, nb de pièces, surface) et de l'environnement (quartier, proximité aux services,...)
- Mais... données non-structurées donc difficile à interroger

➔ Utilisation d'une ontologie pour faciliter l'interopérabilité

Scénario 1 : Recherche d'un bien immobilier

- Quelles sont les caractéristiques du bien ?
- Dans quelle ville/quartier se trouve-t-il ?
- Quels lieux/services sont à proximité ?
- Dans quel environnement (ex. sonore) est le bien ?



Scénario 2 : Étude du marché immobilier

- Quelles sont les autres biens du même secteur ?
- Sont-ils similaires à celui à vendre ?
- Quel est le prix moyen du quartier ?
- Quels sont les services mentionnés dans ce secteur ?
- Quel est le volume de ventes par quartier ?



Scénario 3 : Analyse du territoire

- Comment est perçu un quartier (résidentiel, recherché, calme) ?
- Quels services mentionnés dans ce secteur ?
- Quels autres lieux sont mentionnés dans ce secteur ?
- Est-ce que les annonces qui mentionnent les transports en commun sont situées loin de certains lieux centraux ?
- Est-ce qu'un lieu est inclus dans un autre lieu ?



Etat de l'art et ses limites

Ontologie et Immobilier

- Selon le domaine d'étude (proDataMarket) : cadastre, droit et transactions
- NAREO : description du quartier et services à proximité

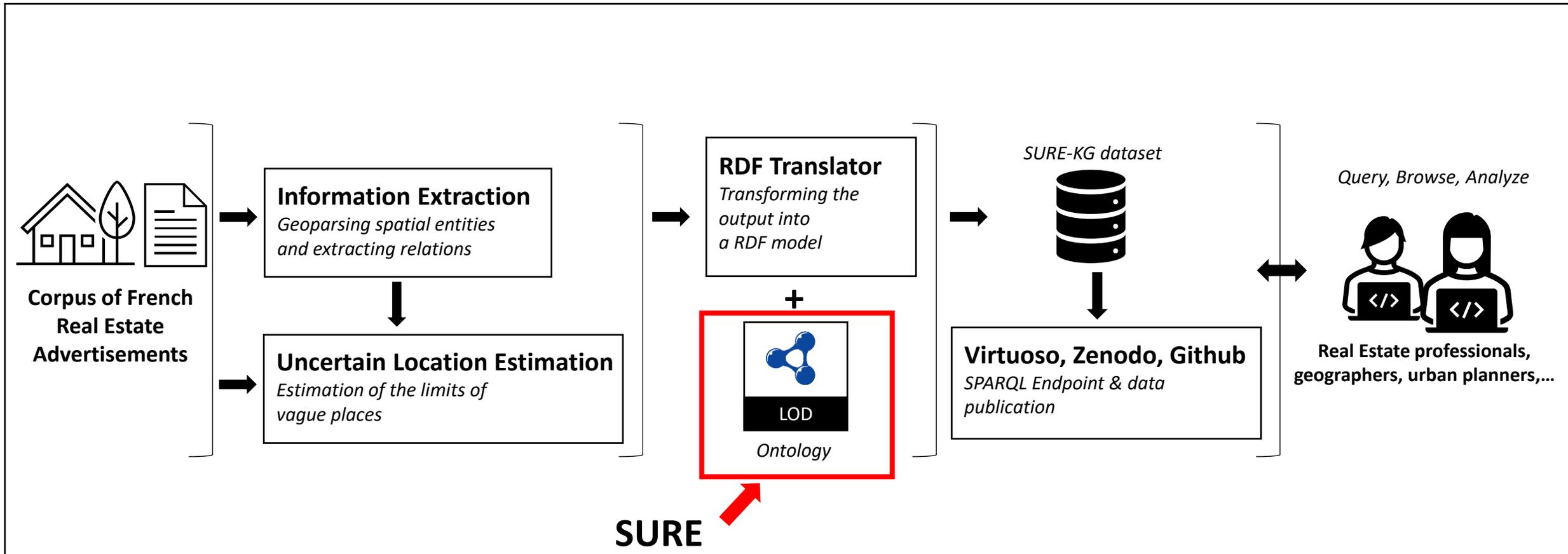
Ontologie et Géographie

- Représentation d'informations « générales » (DBpedia, Yago2Geo, ...)
- Dépend de l'application (KnowWhereGraph pour l'environnement)
- Ontologie selon la nature d'un lieu plutôt que de sa perception (Geonames)
- GeoSPARQL: très haut niveau mais limite pour requêter des relations vagues (proche)

SURE : Spatial Uncertainty and Real Estate

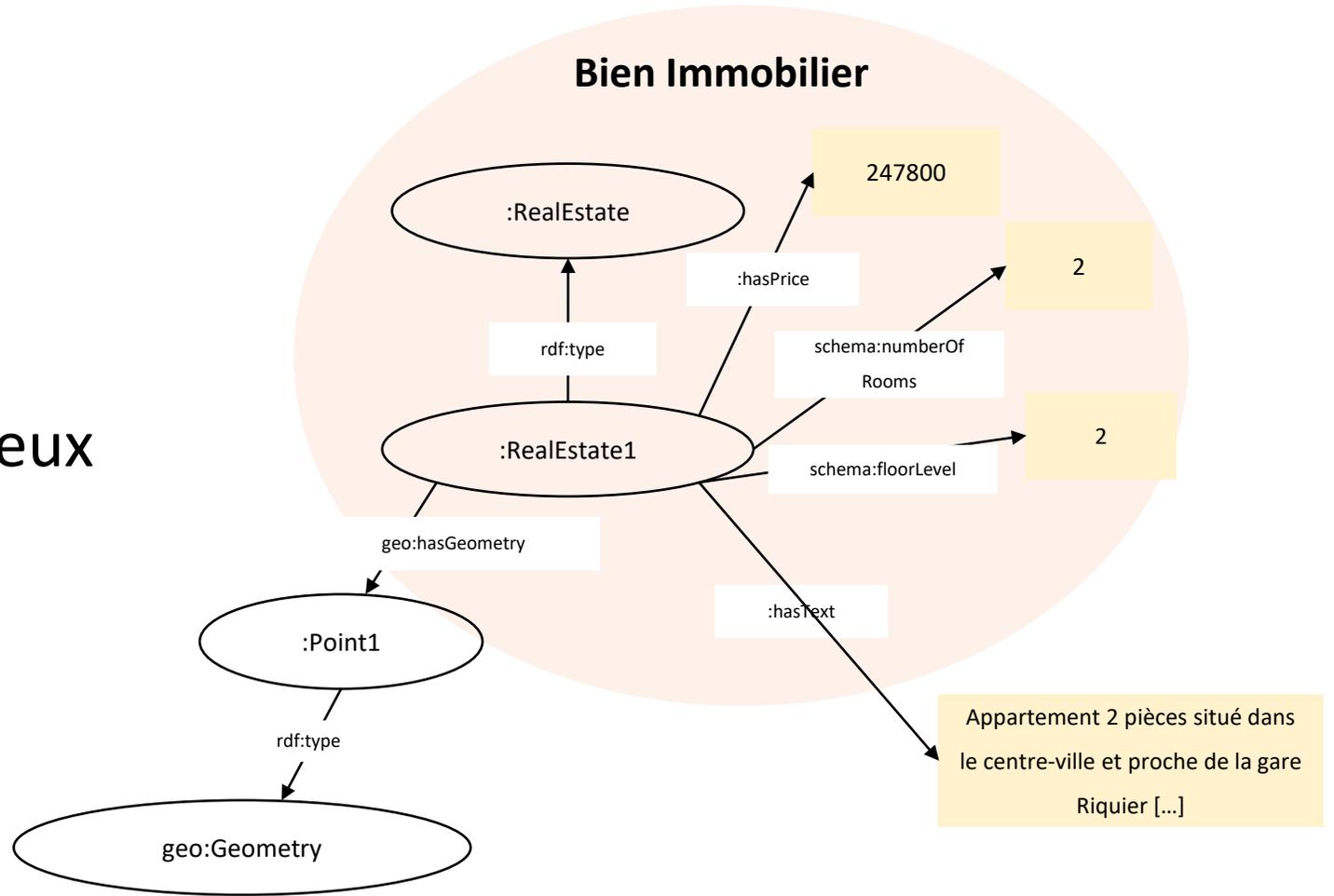
Namespace: <https://ns.inria.fr/sure/>

Préfixe: *sure:*



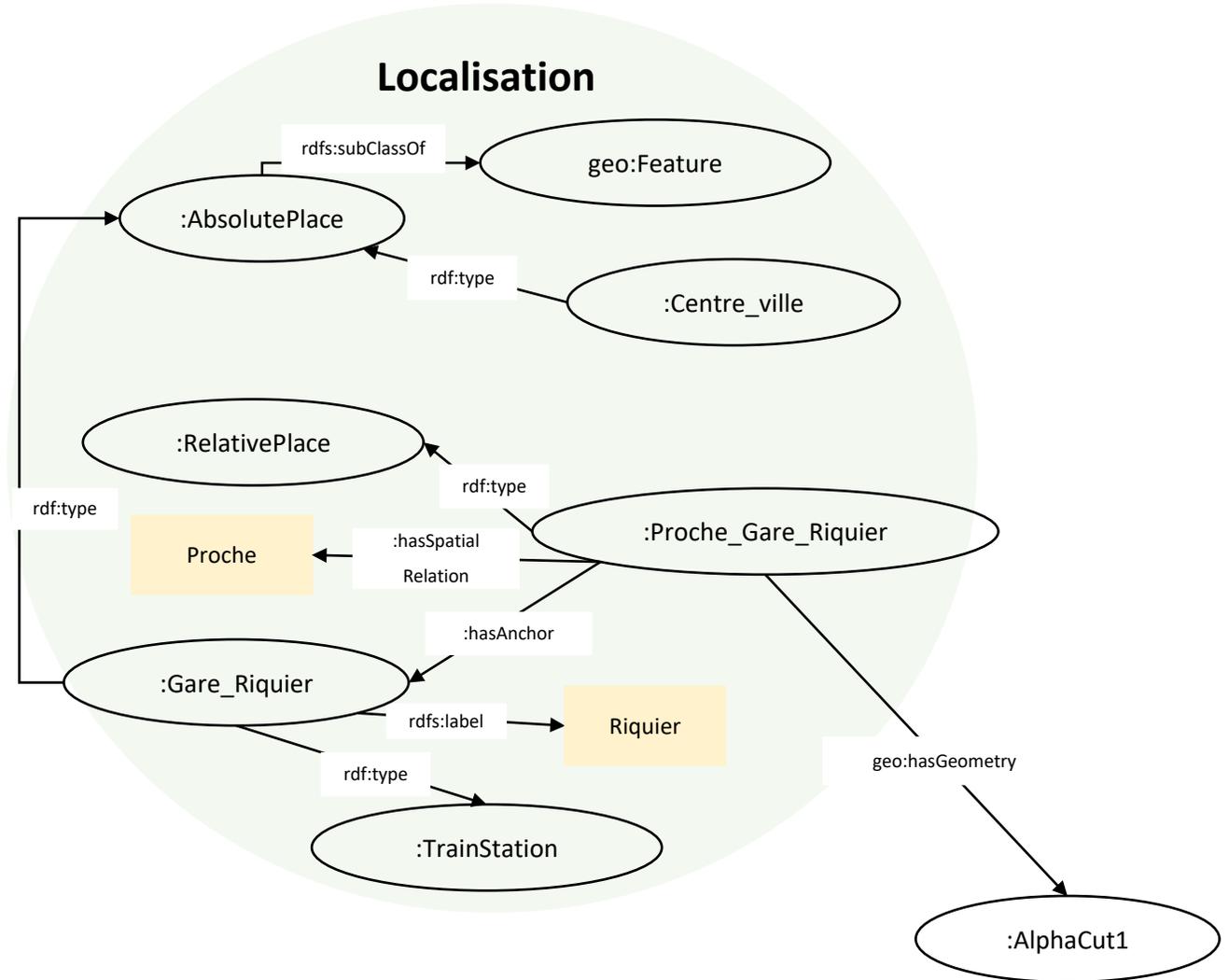
Représentation du bien immobilier

- Le bien et son type
- Caractéristiques (prix, étage, surface,...)
- Localisation (coordonnées, lieux mentionnés, ville)



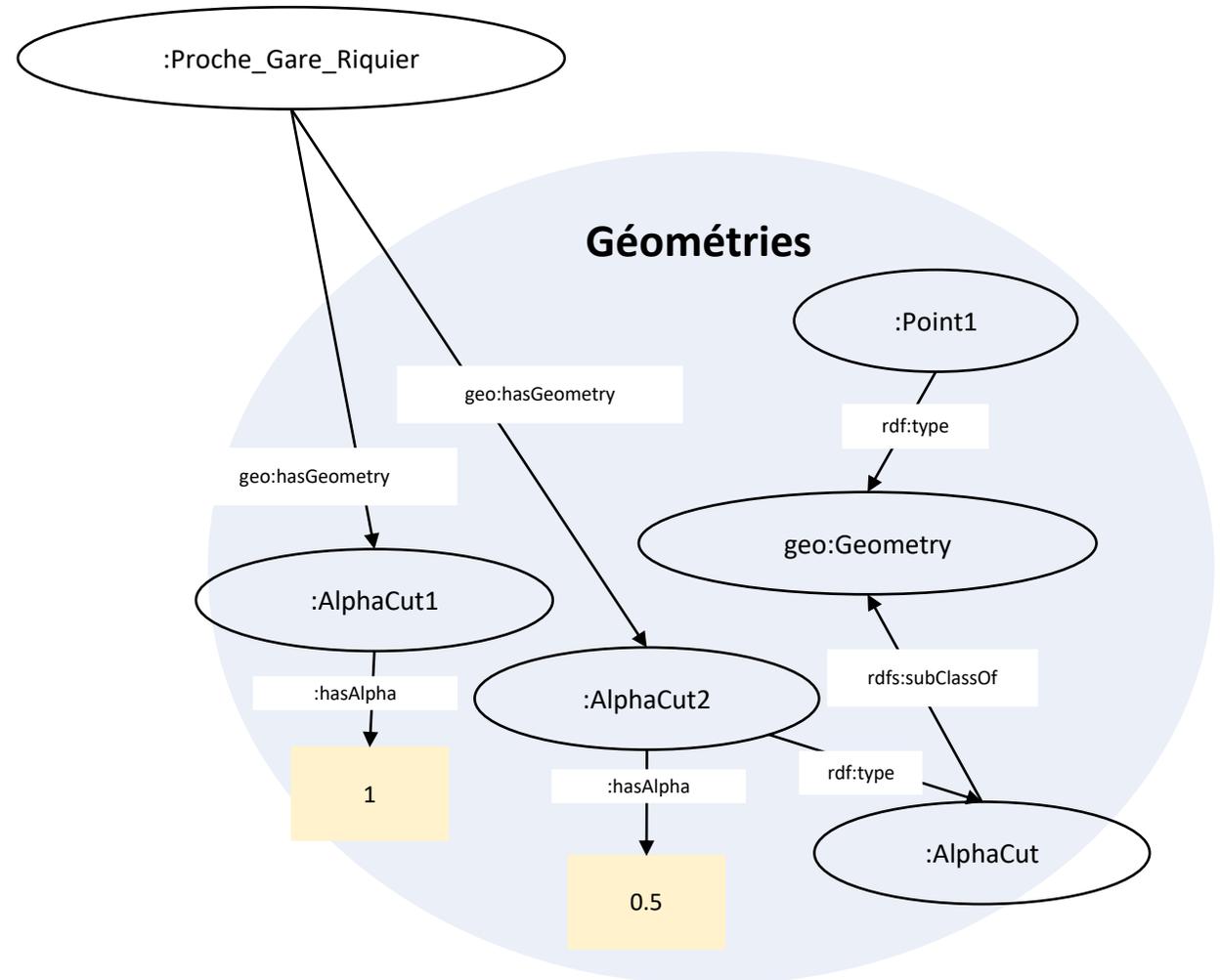
Représentation des lieux

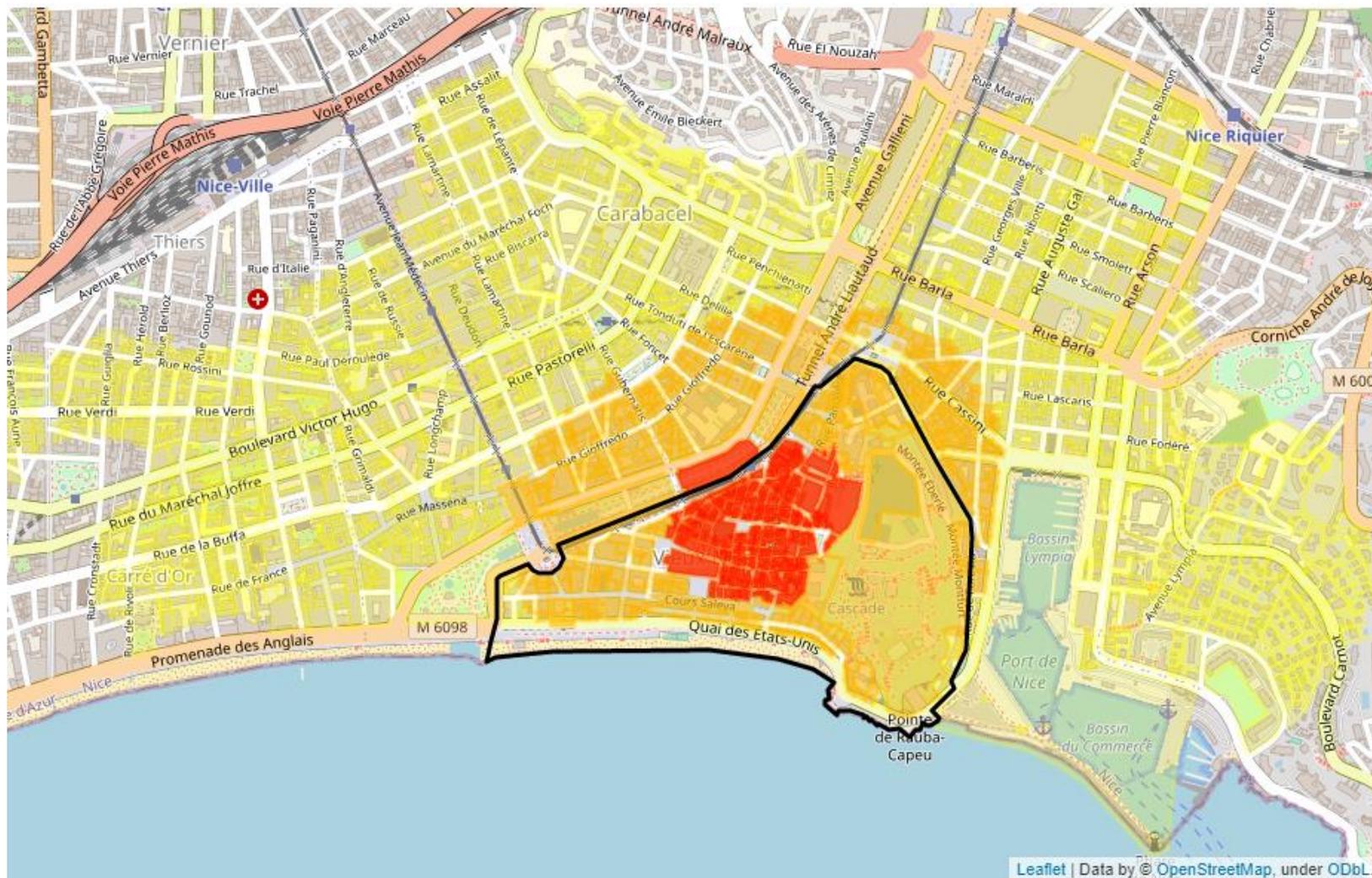
- Lieu **absolu** vs Lieu **relatif**
 - Promenade des Anglais
 - Place Masséna
 - Proche de la gare Riquier
 - Nice Ouest
- Lieu relatif: 2 **propriétés** supplémentaires (*:hasAnchor*, *:hasSpatialRelation*)
- Génération des classes à partir du texte (ex. gare, plage, école, etc.)



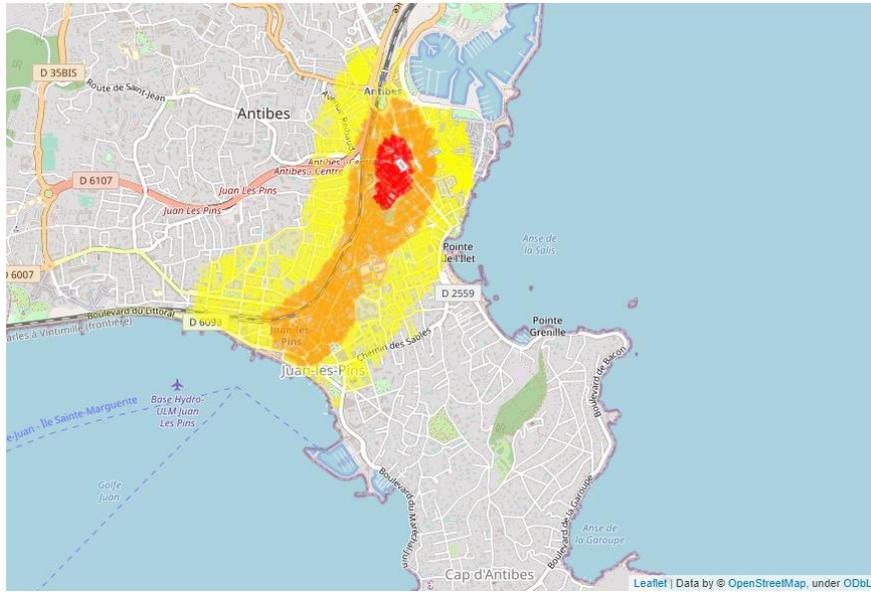
Représentation floue de la localisation

- Lieux vernaculaires : aucune géométrie officielle (centre-ville)
- Perception des lieux différentes (exagération)
- Représentation des lieux relatifs (ex. proche)
- Utilisation de la théorie des ensembles flous

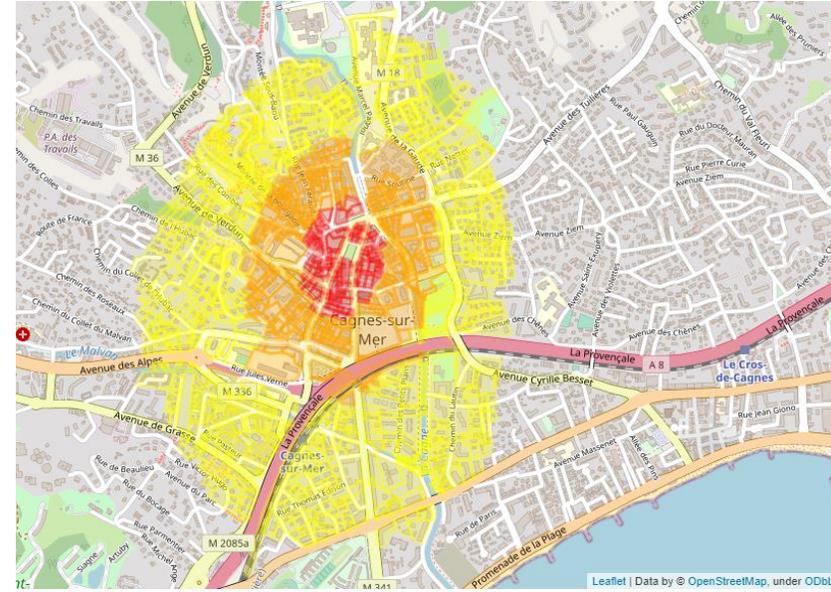




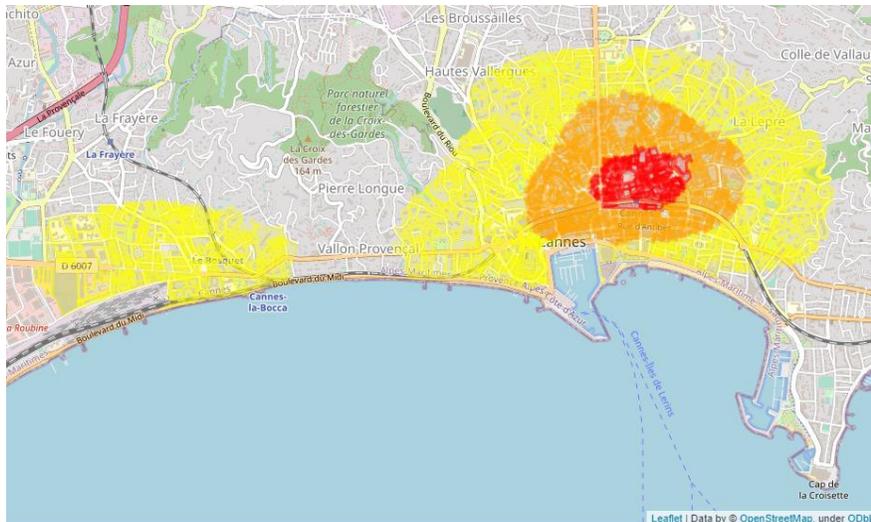
Quartier du Vieux Nice, France



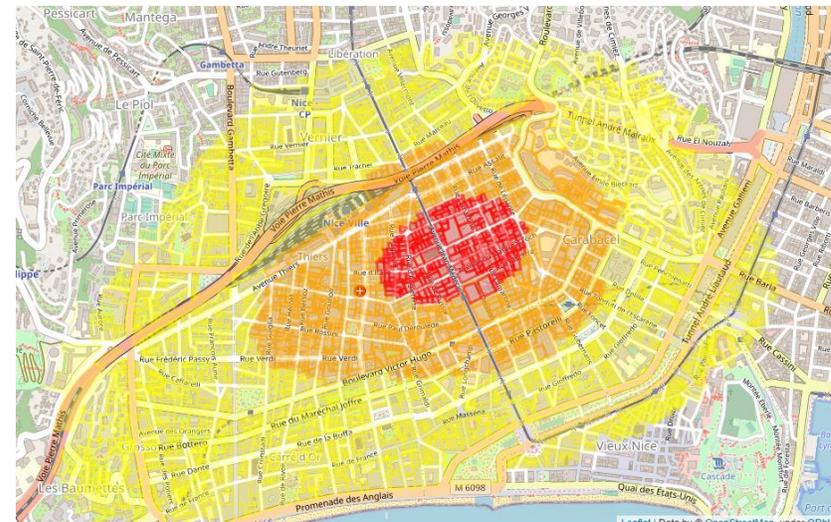
Centre-ville Antibes, France



Centre-ville Cagnes-sur-mer, France



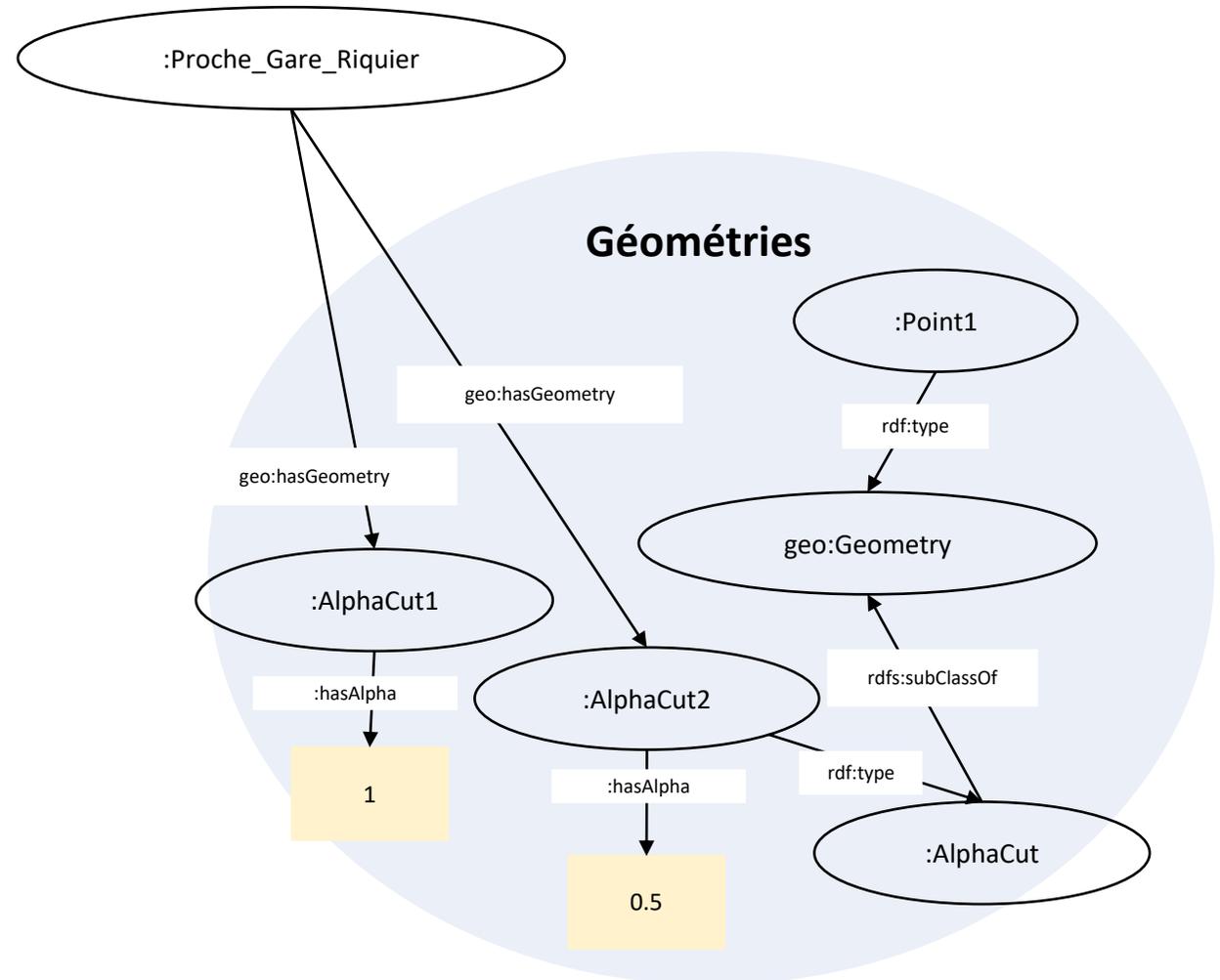
Centre-ville Cannes, France



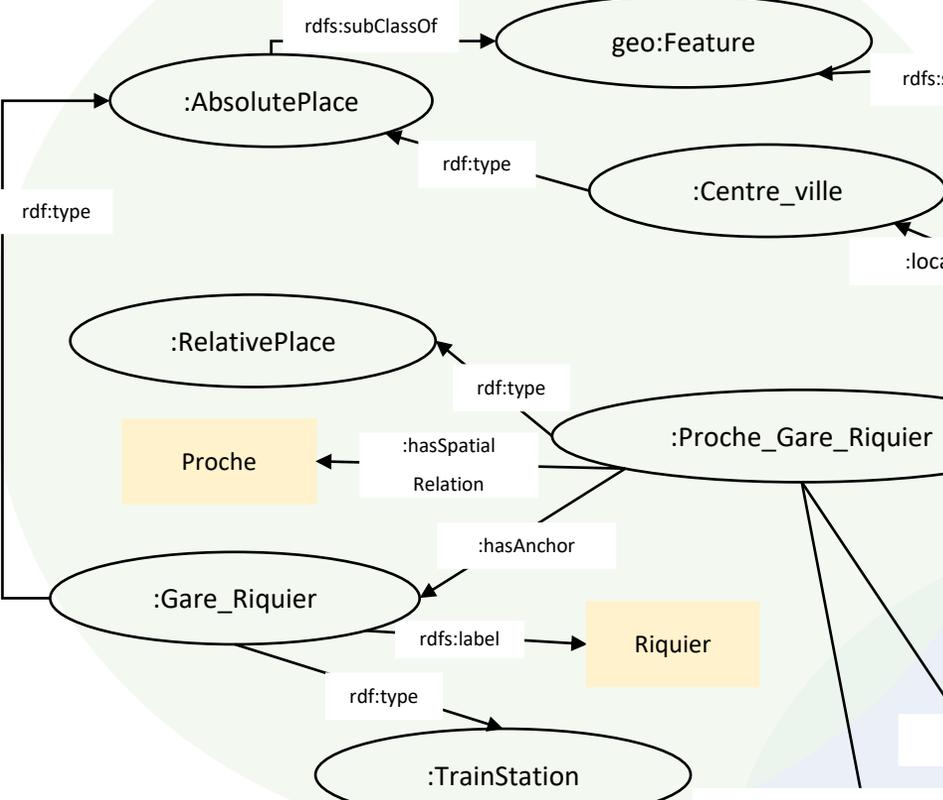
Centre-ville Nice, France

Représentation floue de la localisation

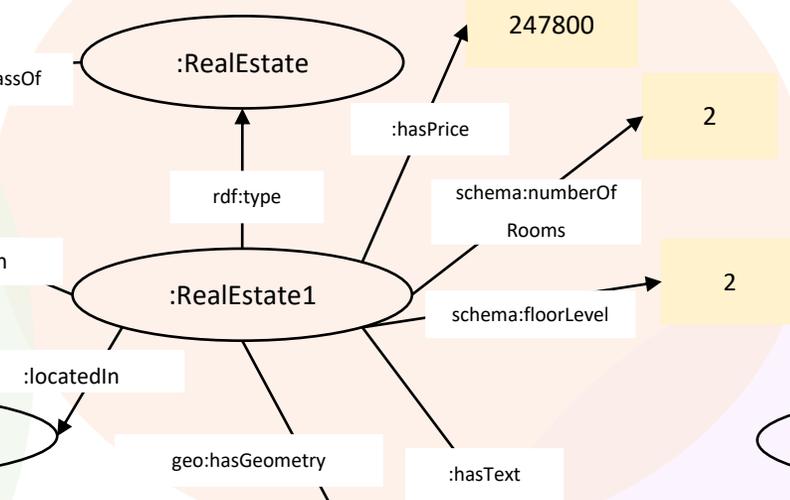
- Lieux vernaculaires : aucune géométrie officielle (centre-ville)
- Perception des lieux différentes (exagération)
- Représentation des lieux relatifs (ex. proche)
- Utilisation de la théorie des ensembles flous



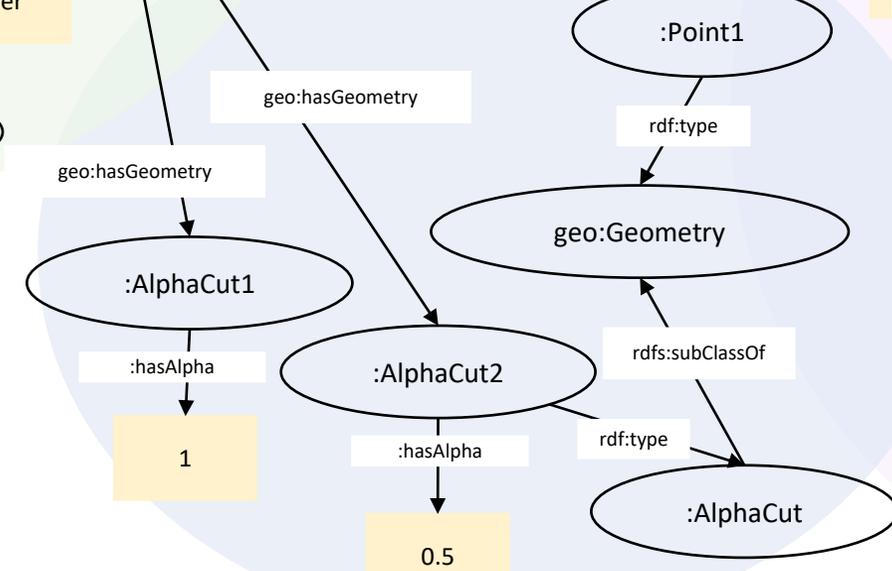
Localisation



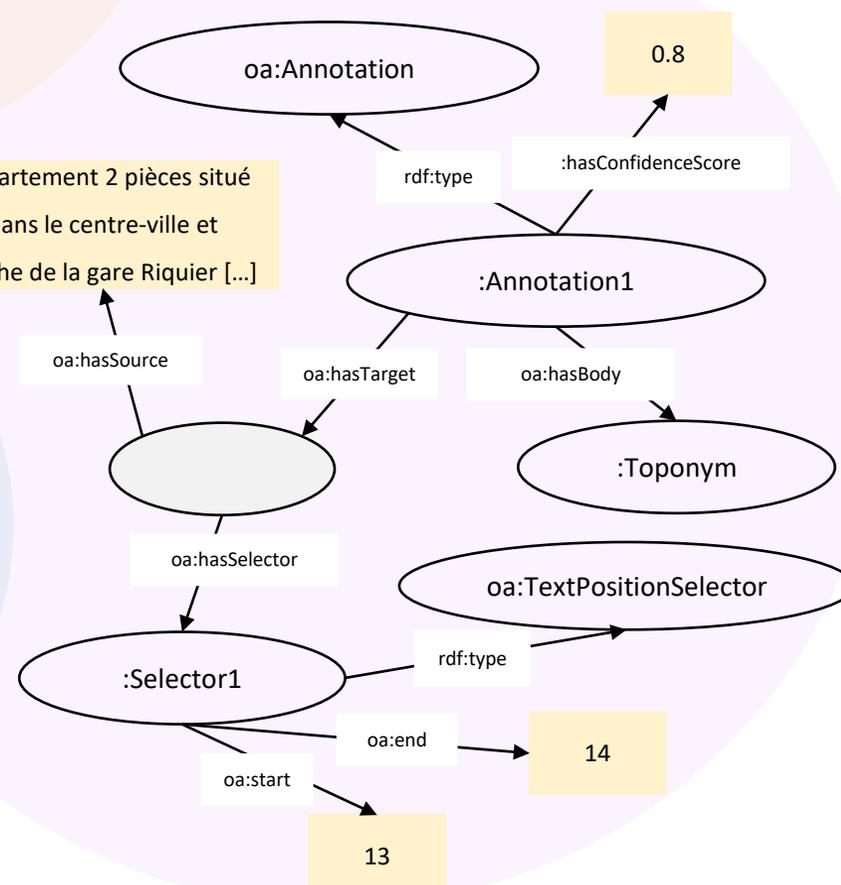
Bien Immobilier



Géométries



Annotations



SURE-KG : A Knowledge Graph to Represent Real Estate and Uncertain Spatial Data from Advertisements

Chiffres clés

- 7M triples
- 100K Annonces localisées dans les **Alpes-Maritimes**
- 6000 lieux extraits

Accessibilité

- Zenodo: <http://zenodo.org/record/7885757>
- SPARQL endpoint: <http://erebe-vm2.i3s.unice.fr:5000/sparql/>
- URIs Namespace: <https://ns.inria.fr/sure#>

Reproductibilité

- Code et Documentation: <http://github.com/Wimmics/sure>
- Licence: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License



Conclusion

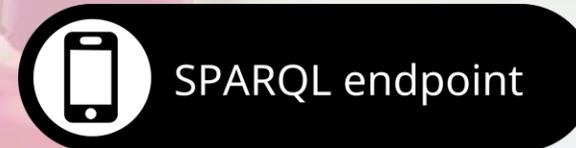
Création d'une ontologie et d'un graphe de connaissances

- Représentation des données spatiales incertaines
- Représentation des informations issues des annonces
- Utilisation des ensembles flous pour représenter les limites spatiales
- Graphe de connaissances sur les Alpes-Maritimes

Perspectives

- Etendre à la France entière
- Evaluation
- Relier les lieux extraits à d'autres bases de données

Merci !



Lucie Cadorel
lucie.cadorel@inria.fr