

Que représentent les références spatiales des données du Web ? un vocabulaire pour la représentation de la sémantique des XY

Abdelfettah FELIACHI*, Nathalie ABADIE*
Fayçal HAMDI**,

*IGN-COGIT, Saint-Mandé, France.
abdelfettah.feliachi@ign.fr, nathalie-f.abadie@ign.fr

**CNAM-CEDRIC, Paris, France.
faycal.hamdi@cnam.fr

Résumé. La référence spatiale qui décrit la localisation des ressources Linked Data peut représenter un critère très utile pour l'interconnexion de ces ressources. Cependant, dans le contexte du Web de données cette référence spatiale peut être très hétérogène entre les différentes sources de données. Nous aborderons dans cet article la question de l'identification des causes de ces hétérogénéités et nous proposerons un modèle de représentation de ces causes afin de pouvoir les gérer de manière plus efficace dans le processus d'interconnexion.

1 Contexte et objectif

De plus en plus de ressources publiées dans le Web des données possèdent dans leur description une référence spatiale qui représente leur localisation dans le monde réel d'une manière directe ou indirecte. Dans le processus d'interconnexion des ressources des différentes sources du Web des données, ces références spatiales représentent l'un des critères privilégiés. En effet, deux ressources proches spatialement sont très susceptibles de représenter la même entité du monde réel. Ce principe est appliqué aussi bien dans le domaine de l'appariement de données géographiques que dans celui de l'interconnexion de Linked Data géoréférencées. Les références spatiales sont généralement comparées par le calcul d'une distance géographique dans le cas d'un référencement direct, ou par des distances de chaînes de caractères dans le cas d'un référencement indirect. Cependant la représentation des références spatiales peut être très hétérogène d'une source à une autre et parfois même parmi les ressources issues d'une même source. Ceci peut être compromettant pour l'utilisation de ce critère. Cette hétérogénéité des références spatiales peut être dû au fait que la référence spatiale dans le Web des données est souvent utilisée pour localiser la ressource qu'elle décrit d'une manière simple : par des adresses postales, des coordonnées de longitude et de latitude ou des géométries de type "point". Le cas des bases de données géographiques classiques est différent : leur référence spatiale est représentée de manière détaillée (point, ligne ou polygone par exemple). La référence spatiale n'a pas le même niveau d'importance, et donc pas la même qualité, entre

le Web des données et les bases de données géographiques classiques. Les causes d'hétérogénéité des références spatiales sont généralement peu prises en compte, et seulement d'une manière implicite, dans les processus d'interconnexion. Nous pensons qu'une identification de ces causes d'hétérogénéité dans le contexte du Web des données et une explicitation de celles-ci peuvent améliorer leur prise en compte dans les processus d'interconnexion. Donc, notre objectif est de proposer un modèle pour expliciter la sémantique des références spatiales afin de fournir des informations exploitables dans le processus d'interconnexion.

2 Un vocabulaire pour décrire les causes d'hétérogénéité entre références spatiales directes

En s'appuyant sur les causes d'hétérogénéité des données géographiques décrites dans la littérature, notamment les sources de conflits d'intégration de données géographiques décrites par Devogele (1997) et les sources d'imperfections de données géographiques décrites par Girres (2012), on distingue plusieurs types et causes d'hétérogénéité entre les références spatiales des différentes sources. Dans le contexte du Web de données, on peut identifier entre autres : l'utilisation de différents types de références spatiales (directe ou indirecte), l'utilisation de différents vocabulaires pour la représentation des références spatiales, la différence dans les sources d'information, le processus de saisie ou d'extraction, la modélisation géométrique et le niveau de précision géométrique, etc. Dans le cadre de l'interconnexion des données géoréférencées, nous nous intéressons plus particulièrement aux trois éléments suivants, que nous jugeons plus susceptibles d'influencer le processus d'interconnexion :

- La différence de précision géométrique des ressources.
- La différence dans la modélisation géométrique des références spatiales des ressources.
- L'aspect vague des entités géographiques que les ressources visent à représenter.

Nous proposons donc un modèle qui permet de représenter ces informations dans un jeu de données au niveau des références spatiales des ressources, c'est-à-dire de leur géométrie. Nous choisissons de représenter ces informations au niveau le plus fin, à savoir la géométrie de chaque ressource. En effet ces informations peuvent changer d'une ressource à une autre, même au sein d'un seul jeu de données. De plus, une ressource peut avoir dans sa description plusieurs références spatiales nécessitant chacune des métadonnées différentes.

En l'absence de standard encore bien installé pour la représentation des géométries sur le Web des données, nous proposons d'étendre l'ontologie des géométries¹ qui présente l'avantage d'être compatible avec GeoSPARQL et de permettre la représentation de géométries structurées Hamdi et al. (2014). Nous présentons désormais les choix de modélisation effectués.

2.1 La précision géométrique

La précision planimétrique est le critère de qualité de données géographiques qui représente les écarts entre les positions des objets géographiques et les positions des entités réelles qu'elles représentent. Nous nous appuyons sur la description des éléments de qualité des données géographiques tels que définis dans les normes ISO19157 (2013) et ISO19115 (2003),

1. <http://data.ign.fr/def/geometrie>

afin de représenter l'élément de précision planimétrique. Cet élément est décrit principalement par le type de sa méthode d'évaluation et le résultat quantitatif de cette évaluation (Fig.1).

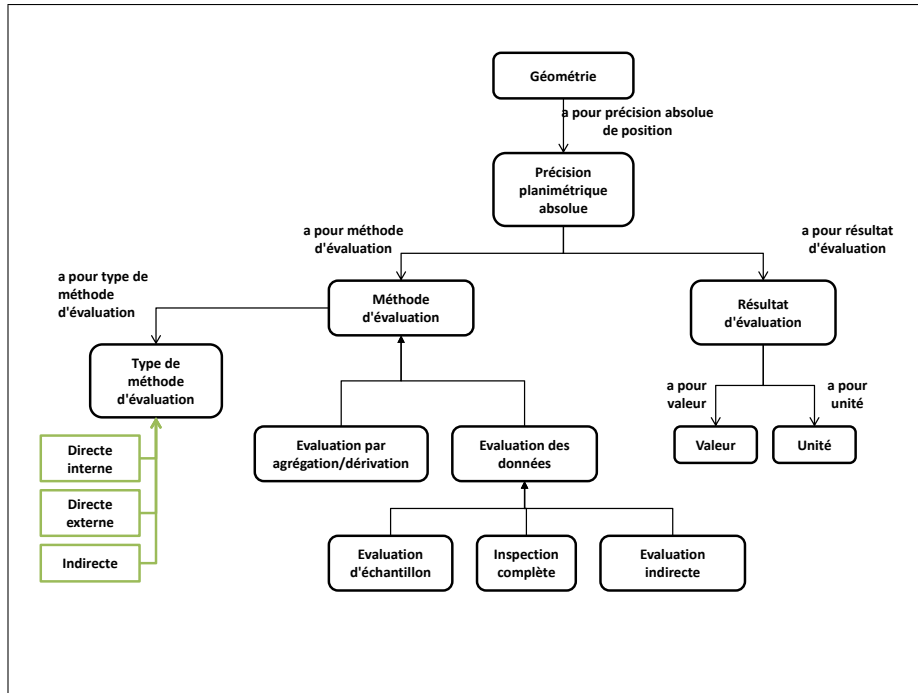


FIG. 1 – Extrait du modèle proposé de la représentation de l'élément de qualité qui décrit la précision planimétrique d'une référence spatiale.

Un exemple de l'utilisation possible de cette partie du modèle afin de décrire la précision planimétrique d'une référence spatiale est illustré dans la figure suivante (Listing 1). Dans cet exemple la précision planimétrique est de 100 mètres. Elle a été estimée par une vérification directe sur échantillon de données par rapport à la réalité du terrain.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix geom: <http://data.ign.fr/def/geometrie#>.
@prefix xysemantics: <http://data.ign.fr/def/xysemantics#>.

<http://data.ign.fr/id/xx/xx/xx> geom:geometry <http://data.ign.fr/id/xx/xx/
  Multipolygon_XX>.

<http://data.ign.fr/id/xx/xx/Multipolygon_xx> a geom:MultiPolygon;
xysemantics:absolutePositionalAccuaracy [
  a xysemantics:AbsolutePositionalAccuaracy;
  xysemantics:evaluationResult [
    a xysemantics:Result;
    xysemantics:value "100"^^xsd:float;
    <http://qudt.org/schema/qudt#unit> <http://qudt.org/vocab/unit#Meter>;
  xysemantics:evaluationMethod [
    a xysemantics:SampleBasedInspection;
  ]
]

```

```
xysemantics:evaluationMethodTypeCode <http://data.ign.fr/id/codes/geomtrie/
methodeevaluation/DirectExternal>]].
```

Listing 1 – Exemple de triplets RDF qui représentent la précision planimétrique

2.2 La modélisation géométrique

Elle représente le choix de représentation géométrique par rapport à l'entité du monde réel référencée par l'objet géographique. Gesbert (2005) et Abadie (2012) se sont intéressés à la description formelle de la modélisation géométrique dans le cadre de leurs travaux de modélisation des règles de saisie des données géographiques, afin d'améliorer leur intégration. Dans ce cadre, nous notons principalement la proposition de Abadie (2012) basée sur les travaux de Sahade (2010) et Smith et Mark (1998), qui définit un ensemble d'éléments représentatifs de la forme des entités géographiques du monde réel, tels que perçus dans un contexte cartographique. Ceci est concrétisé au niveau de notre modèle par un « élément caractéristique de la forme » décrit par son type, qui fait référence à une taxonomie des types d'éléments caractéristiques de la forme (Fig.2). Cette taxonomie peut éventuellement être étendue. Ces éléments de forme sont des éléments parasites qui ne peuvent exister sans un objet hôte, qui est l'entité du monde réel qui les porte (Fig.2).

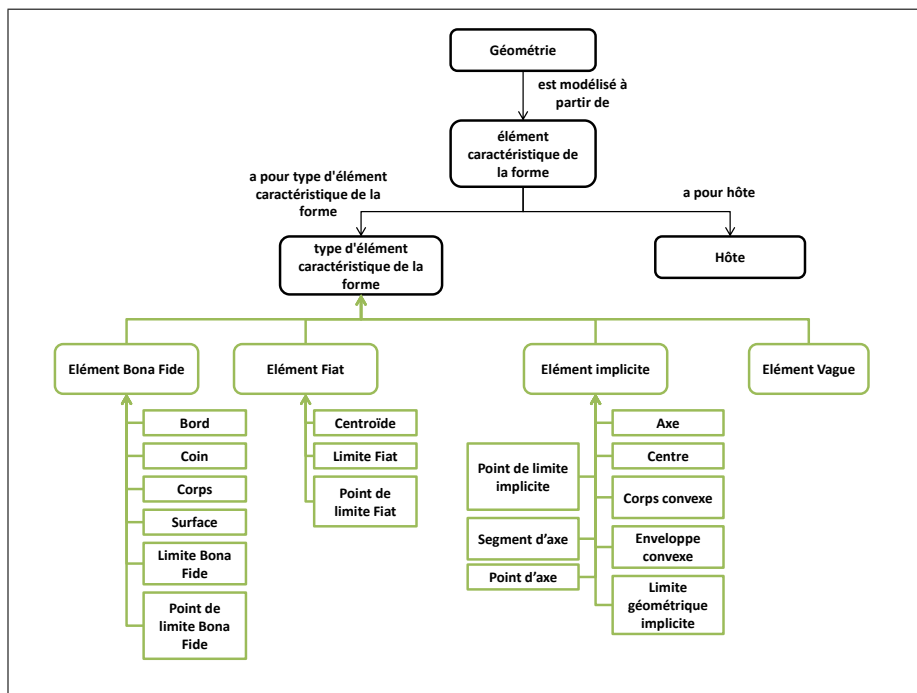


FIG. 2 – Extrait du modèle proposé qui décrit l'élément de représentation géométrique d'une ressource spatiale.

Un exemple possible de l'utilisation de cette partie de modèle afin de décrire la modélisation géométrique d'une référence spatiale est illustré dans la figure suivante (Listing 2). Dans cet exemple la géométrie représente une limite de type Bona Fide saisie à partir des bâtiment.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix geom:<http://data.ign.fr/def/geometrie#>.
@prefix xysemantics:<http://data.ign.fr/def/xysemantics#>.

<http://data.ign.fr/id/xx/xx/xx> geom:geometry <http://data.ign.fr/id/xx/xx/
  Multipolygon_XX>.

<http://data.ign.fr/id/xx/xx/Multipolygon_xx> a geom:MultiPolygon;
xysemantics:sModeledFrom [
  a xysemantics:ShapeCharacteristicElement;
  xysemantics:shapeCharacteristicElementType <http://data.ign.fr/id/codes/geometrie/
    elementcaracteristique/BonaFideBoundary>;
  <http://www.loa-cnr.it/ontologies/DOLCE-Lite.owl#host> <http://data.ign.fr/def/topo#
    Bati> ].
```

Listing 2 – Exemple de triplets RDF qui représentent la modélisation géométrique

2.3 L'aspect vague des entités géographiques

Certaines entités géographiques sont difficiles à définir ou délimiter par essence, comme c'est le cas d'une vallée par exemple. Dans ce cas, l'élément caractéristique de la forme géométrique doit être de type « élément vague » (Fig.2).

3 Conclusion

Dans cet article nous nous sommes intéressés aux causes des hétérogénéités entre les ressources géoréférencées qui peuvent affecter la fiabilité du critère spatial dans le processus d'interconnexion. Nous identifions un ensemble d'éléments permettant une représentation plus riche de la précision planimétrique et de la modélisation géométrique de chaque référence spatiale. Nous avons organisé ces éléments dans un modèle sous forme d'ontologie² OWL. La suite de notre travail consistera à mettre au point une approche exploitant ce modèle afin d'améliorer la prise en compte du critère spatial dans un processus d'interconnexion.

Références

- Abadie, N. (2012). *Formalisation, acquisition et mise en œuvre de connaissances pour l'intégration virtuelle de bases de données géographiques : Les spécifications au cœur du processus d'intégration*. Thèse de doctorat, Université Paris-Est.
- Devoegele, T. (1997). *Processus d'intégration et d'appariement des bases de données géographiques. Application à une base de données routière multi-échelles*. Thèse de doctorat, Université Paris-Est.

2. <http://data.ign.fr/def/xysemantics>

Un vocabulaire pour la représentation de la sémantique des XY

- Gesbert, N. (2005). *Formalisation des spécifications de bases de données géographiques en vue de leur intégration*. Thèse de doctorat, Université Paris-Est.
- Girres, J.-F. (2012). *Modèle d'estimation de l'imprécision des mesures géométriques de données géographiques. Application aux mesures de longueur et de surface*. Thèse de doctorat, Université Paris-Est.
- Hamdi, F., N. Abadie, B. Bucher, et A. Feliachi (2014). Geomrdf: A geodata converter with a fine-grained structured representation of geometry in the web. *The 1st International Workshop on Geospatial Linked Data - (GeoLD 2014) - SEMANTiCS 2014*, pp.12.
- ISO19115 (2003). Geographic information – metadata. Standard, International Organization for Standardization (TC 211).
- ISO19157 (2013). Geographic information – data quality. Standard, International Organization for Standardization (TC 211).
- Sahade, S. (2010). Computer-tractable translation of geospatial data. *International Journal of Spatial Data Infrastructures Research, Revue en ligne publiée par le Joint Research Centre (European Commission) 5*.
- Smith, B. et D. Mark (1998). Ontology and geographic kinds. *8th International Symposium on Spatial Data Handling (SDH'98)*, 308–320.

Summary

The spatial reference that describes the location of Linked Data resources can be a very useful criterion for the interlinking of these resources. However, in the context of Web of data this spatial reference can be very heterogeneous between different data sources. We discuss in this article the question of identifying the causes of these heterogeneities and we propose a model of representation of these causes in order to better manage them in the interlinking process.