

Intégration et visualisation de données liées thématiques sur un référentiel géographique

Abdelfettah Feliachi*, Nathalie Abadie*
Fayçal Hamdi**,

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

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

Résumé. De nombreuses ressources publiées sur le Web des données sont décrites par une composante qui désigne d'une manière directe ou indirecte une localisation géographique. Comme toute autre propriété, cette information de localisation peut être mise à profit pour permettre l'interconnexion des données avec d'autres sources. Elle permet en outre leur représentation cartographique. Cependant, les informations de localisation utilisées dans les sources de données linked data peuvent parfois s'avérer imprécises ou hétérogènes d'une source à l'autre. Ceci rend donc leur exploitation pour réaliser une interconnexion difficile, voire impossible. Dans cet article, nous proposons de pallier ces difficultés en ancrant les données linked data thématiques aux objets d'un référentiel géographique. Nous mettons à profit le référentiel géographique afin de mettre en correspondance des données thématiques dotées d'indications de localisation hétérogènes. Nous exploitons enfin les relations de correspondance créées entre données thématiques et référentiel géographique dans une application de visualisation cartographique des données.

1 Contexte et objectifs

Le modèle *Linked Data* vise à favoriser l'interopérabilité de sources de données hétérogènes sur le Web, en proposant un ensemble de bonnes pratiques de publication de données qui s'appuient sur les technologies du Web sémantique. En particulier, il préconise de décrire les données selon le modèle RDF, de fournir les vocabulaires utilisés par les données sous la forme d'ontologies et d'interconnecter les différents jeux de données disponibles en identifiant les ressources équivalentes d'un jeu de données à l'autre.

Parmi les données publiées sous la forme de *Linked Data*, de nombreuses ressources font référence, de façon directe ou indirecte, à une localisation. Dans DBpedia¹, par exemple, 1373482 objets sont décrits par les propriétés `geo:long` et `geo:lat`. Geonames² contient plus de 8500000 entités spatiales nommées et environ 125000000 triplets RDF. La présence d'une

1. <http://dbpedia.org/>, consulté le 13/12/2013

2. <http://www.geonames.org/ontology/documentation.html>, consulté le 27/08/2013

information de localisation dans la description d'une ressource signifie que cette ressource est liée d'une manière ou d'une autre à une entité géographique du terrain réel. Deux ressources dotées d'informations de localisation identiques ou spatialement proches seront donc très susceptibles de présenter une relation sémantique. La prise en compte d'informations de localisation s'avère donc généralement importante pour l'interconnexion, et est le plus souvent réalisée au travers de mesures de distance géographiques entre les ressources à interconnecter. Cependant, les informations de localisation associées à des ressources *Linked Data* peuvent s'avérer très hétérogènes d'une source de données à l'autre, qu'il s'agisse d'hétérogénéités liées au type de localisation utilisé (directe via des coordonnées ou indirecte via une adresse par exemple), à l'origine de cette information, à sa précision ou au vocabulaire adopté pour la décrire. Cette hétérogénéité des informations de localisation peut rendre l'utilisation de mesures de distances géographiques pour l'interconnexion peu fiable voire impossible.

Dans cet article, nous proposons de mettre à profit des bases de données géographiques de référence pour l'interconnexion et l'exploitation de sources de données thématiques dotées d'informations de localisation hétérogènes. En effet, nous proposons d'ancrer les différentes ressources à interconnecter sur un référentiel géographique afin de favoriser la détection de relations d'interconnexion entre ces ressources. Parmi les applications découlant de cet ancrage de données thématiques à une base de données géographique de référence, la visualisation cartographique permet leur mise en valeur en fournissant aux utilisateurs un moyen convivial de découverte de ces données. Nous proposons donc une approche de visualisation cartographique des données thématiques ainsi interconnectées au référentiel géographique faisant porter les informations qu'elles renferment par les objets géographiques eux-mêmes et permettant leur visualisation à différentes échelles.

La section suivante dresse un état de l'art des approches proposées dans le domaine de l'interconnexion de données dotées d'informations de localisation, et dans le domaine de la visualisation cartographique de données *Linked Data*. La section 3 présente l'approche que nous proposons, pour l'ancrage de *Linked Data* sur un référentiel géographique et la visualisation des données. La section 4 décrit la mise en œuvre de cette approche sur des données décrivant des monuments historiques. La section 5 conclut cet article.

2 Interconnexion et visualisation de données thématiques dotées d'informations de localisation

2.1 Utilisation d'informations de localisation pour l'interconnexion

L'interconnexion désigne l'étape du processus de publication de données sur le Web des données qui vise à créer des liens d'équivalence entre les ressources des différentes sources de données qui représentent la même entité du monde réel. Cette tâche est généralement fondée sur l'évaluation du degré de similarité entre ces ressources par comparaison de leurs propriétés, sous l'hypothèse que plus la similarité entre les descriptions des différentes ressources est importante, plus la probabilité qu'elles représentent la même réalité l'est aussi Ferraram et al. (2013). L'automatisation de cette tâche d'interconnexion s'appuie donc sur des approches et des mesures de similarité proposées dans des domaines ayant des besoins similaires d'identification de relations de correspondance, comme la réconciliation de références ou l'alignement d'ontologies Heath et Bizer (2011).

Différentes approches ont été proposées afin de détecter automatiquement des relations de correspondance entre objets issus de bases de données géographiques hétérogènes qui représentent une même entité topographique du monde réel. Tout comme dans le domaine de l'interconnexion de *Linked Data*, l'appariement de données géographiques repose sur l'évaluation de la similarité des objets des différentes bases par comparaison de leurs propriétés. Le critère d'appariement privilégié est la géométrie des données qui représente la forme et la localisation des entités du monde réel via des primitives géométriques (point, ligne ou polygone). Différentes approches ont été proposées pour traiter chaque cas de figure : comparaison de données ponctuelles Minami (2000) Beeri et al. (2006) Safra et al. (2006), de données linéaires représentant des réseaux dotés du même de niveau de détail Walter et Fritsch (1999) ou dotés de niveaux de détail différents Mustière et Devogele (2008), ou encore comparaison de données surfaciques Bel Hadj Ali (2001) Samal et al. (2004). Dans chaque cas, la quantification du niveau de similarité entre les géométries est réalisée à l'aide d'une mesure spécifique. Des approches multicritères combinant critères géométriques, attributaires ou encore topologiques ont également été proposées Olteanu (2008).

A la croisée des deux domaines, des approches qui visent à trouver des équivalences entre des données *Linked Data* dotées d'informations de localisation ont également été proposées. Ces approches s'appuient sur les mesures de similarité entre géométries utilisées dans le domaine de l'appariement de données géographiques Salas et Harth (2011). Elles sont le plus souvent appliquées sur des données issues de bases de données géographiques représentées selon le modèle RDF et publiées sur le Web des données. L'approche proposée par Hahmann et Burghardt (2010) prend en considération des objets décrits par des géométries de type point et se fonde sur deux critères : le nom et la géométrie. La similarité entre deux points est estimée via un critère d'inclusion bilatérale d'un point candidat à l'appariement dans un rectangle englobant centré sur un point potentiellement homologue. L'approche proposée par Salas et Harth (2011) repose sur une projection cylindrique équidistante préalable des géométries, afin de pouvoir utiliser une mesure de distance de Hausdorff entre les géométries de type polygone que les auteurs souhaitent comparer pour détecter des cas de superposition entre géométries.

2.2 Visualisation cartographique des *Linked Data*

La plupart des approches de visualisation cartographique de données *Linked Data* visent à permettre l'affichage de données géographiques publiées selon le format RDF à l'aide d'outils de cartographie sur le Web. C'est le cas par exemple pour les données géographiques publiées par l'Ordnance Survey³. L'initiative GeoLinkedData.es propose une solution de visualisation fondée sur Google Maps⁴ permettant l'affichage de données ponctuelles sur un fond cartographique⁵. L'application de mash-up présentée par Shvaiko et al. (2012) permet de co-visualiser des données géographiques dotées de géométries diverses et publiées en RDF avec des données issues de LinkedGeoData⁶ dotées de géométries de type point et des images Flickr géoréférencées, sur un fond cartographique Google Maps.

L'approche proposée par Owusu-Banahene et Coetzee (2013) se distingue des précédentes dans la mesure où les auteurs proposent d'interconnecter des données thématiques issues de

3. <http://data.ordnancesurvey.co.uk>

4. <http://maps.google.com/>

5. <http://geo.linkeddata.es/browser>

6. <http://browser.linkedgeo.org/>

DBPedia⁷ avec des données géographiques en vue de créer des cartes thématiques. L'approche d'interconnexion réalisée en amont consiste à importer les données DBPedia souhaitées dans une base de données géographique PostGIS et à créer une jointure avec une table de données géographiques représentant des États en se fondant sur leurs noms. Les données ainsi liées sont ensuite mises à profit dans une application de visualisation cartographique.

3 Approche proposée

Les approches d'interconnexion de données *Linked Data* dotées d'informations de localisation proposées s'appuient sur des mesures permettant de comparer des informations de localisation du même type, qu'il s'agisse de calculer des distances entre géométries ou d'évaluer la similarité de chaînes de caractères représentant des adresses postales ou des toponymes. Dans le cas de ressources présentant des informations de localisation de différentes natures, il est donc nécessaire de procéder au géoréférencement des informations disponibles afin de disposer de coordonnées géographiques pour chaque source de données à comparer. De plus, dans le cas d'informations de localisation du même type, des différences de précision ou d'origine des données peuvent engendrer des valeurs de distance géographique très importantes entre les ressources à comparer, alors que leurs géométries se rapportent bien à une même entité topographique du monde réel. C'est pourquoi, nous proposons de mettre à profit des bases de données géographiques de référence comme ressources de support pour l'interconnexion de sources de données thématiques dotées d'informations de localisation hétérogènes. Suivant l'approche proposée par Aleksovski et al. (2006) dans le domaine de l'alignement d'ontologies, nous proposons d'ancrer les différentes ressources à interconnecter sur un référentiel géographique afin de favoriser la détection de relations d'interconnexion entre ces ressources. La figure 1 illustre cette approche sur un exemple d'interconnexion de ressources localisées dans le premier cas, par des coordonnées géographiques et dans le second, par des adresses. Le géocodage via la base de données d'adresses permet de disposer de coordonnées géographiques pour la seconde ressource. En revanche, ces coordonnées demeurent trop éloignées de celles de la première ressource pour permettre une interconnexion. Le fait d'ancrer les deux ressources sur le bâtiment auquel elles se rapportent permet finalement de les mettre en correspondance.

L'étape d'ancrage des sources d'informations thématiques au référentiel géographique nécessite de détecter des liens de correspondance entre ressources thématiques et objets géographiques auxquels elles se rapportent. Dans certains cas, il ne s'agira pas nécessairement de liens d'équivalence. Cependant, dans la mesure où les indications de localisation fournies avec les données thématiques font référence à des points de l'espace proches des objets géographiques auxquels on devra les ancrer, les approches proposées pour l'interconnexion de données dotées d'informations de localisation pourront être mises à profit lors de cette étape. Nous proposons donc pour cette étape d'ancrage de nous appuyer sur les outils d'interconnexion de données existants, et de les compléter à l'aide d'approches proposées dans le domaine de l'appariement de données géographiques.

Notre approche doit permettre la transformation des données géographiques dans le modèle RDF, afin de permettre leur traitement via un outil d'interconnexion de *Linked Data* disposant

7. <http://dbpedia.org/>

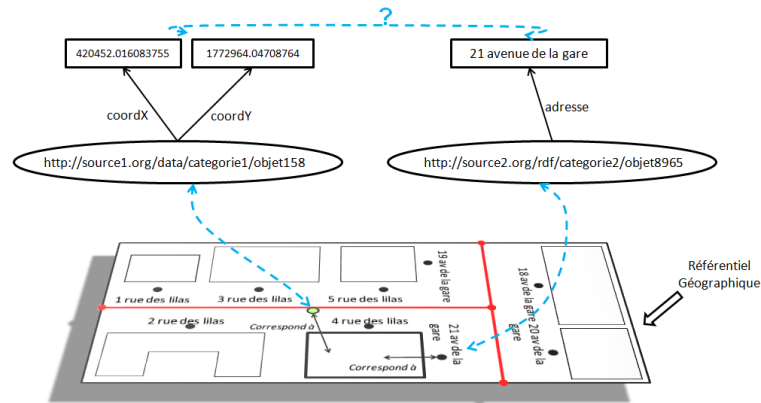


FIG. 1 – Utilisation du référentiel géographique pour l'interconnexion des Linked Data.

de mesures de distances adaptées aux données géographiques. Ces distances doivent prendre en considération les différences entre les types des géométries et les systèmes de coordonnées représentant les données dans les différentes sources. Les liens d'ancrage des données thématiques aux données géographiques peuvent être exploités pour diverses applications : analyse spatiale, croisement de données ou encore cartographie. Nous proposons ici de les mettre à profit dans une approche de visualisation cartographique.

4 Mise en œuvre

Afin de mettre en œuvre et tester l'approche que nous proposons, nous avons choisi d'ancrer des données sur les monuments historiques de la ville de Paris, issues de DBpedia en français⁸ et de la base Mérimée⁹ publiée sur le portail de données ouvertes data.gouv.fr, sur les données qui décrivent la topographie de la ville de Paris issues des bases BD TOPO® et BD ADRESSE® du Référentiel à Grande Échelle (RGE) produit par l'Institut National de l'Information Géographique et Forestière (IGN).

4.1 Implémentation de l'approche d'interconnexion

4.1.1 Architecture et outils

La figure 2 illustre l'architecture globale adoptée pour la mise en œuvre de notre approche d'interconnexion. Les données DBpedia, déjà en RDF, n'ont pas nécessité de conversion : nous avons effectué une requête SPARQL afin de sélectionner les ressources de type monument historique localisées par des coordonnées géographiques. La conversion des données Mérimée et des données géographiques de référence a été réalisée à l'aide de la plateforme Datalift¹⁰.

8. <http://fr.dbpedia.org/wiki/Accueil>

9. Ministère de la Culture et de la Communication, donnée publiée le 06-11-2012

10. <http://datalift.org/>

Intégration et visualisation de Linked Data sur un référentiel géographique

Celle-ci permet la transformation selon le modèle RDF de données représentées selon différents modèles, y compris des données géographiques sous format GML ou SHP (format vectoriel Shapefile). Les monuments historiques décrits par la base Mérimée sont localisés à l'aide d'adresses postales multiples. Nous avons donc dû procéder, après l'étape de conversion, à un nettoyage de ces données à l'aide de requêtes de type CONSTRUCT, afin de pouvoir traiter chaque adresse de façon individuelle lors de l'interconnexion. Dans le cas de la BD TOPO®, nous avons converti les données sur les bâtiments de Paris. Chaque bâtiment est décrit par une géométrie de type polygone. Dans le cas de la BD ADRESSE®, il s'agit de données décrivant des adresses de façon structurée, et associant à chaque adresse une géométrie de type point.

L'ensemble des données est ensuite stocké dans un triple store local (OpenRDF Sesame) pour faciliter l'étape d'interconnexion.

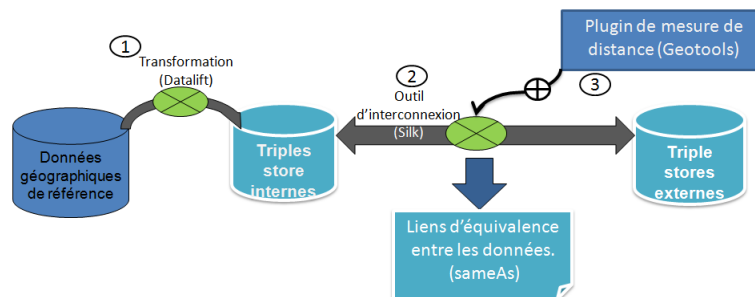


FIG. 2 – Architecture d'interconnexion.

L'étape d'interconnexion a été réalisée à l'aide de l'outil Silk Link Discovery Framework¹¹ que nous avons étendu en y ajoutant une mesure de distance spécifique aux géométries vectorielles. Ainsi nous avons pu bénéficier des mesures de distance et des méthodes d'agrégation et d'optimisation dont dispose Silk et mettre à profit la géométrie des données géographiques dans une approche d'interconnexion multicritère. Notre approche reste donc applicable à d'autres cas d'application à condition d'adapter la mesure de distance géographique utilisée et le type des relations à générer aux données en entrée.

4.1.2 Mesure de distance entre géométries vectorielles

Les indications de localisations généralement utilisées dans les sources de données *Linked Data*, sont le plus souvent des adresses, des géométries de type " point " ou des coordonnées de type longitude et latitude. Les géométries représentant les objets géographiques dans les bases de données géographiques décrivent la réalité d'une manière plus détaillée : outre des points, on rencontre fréquemment des lignes ou des polygones. La mesure de distance que nous avons choisie d'utiliser pour ancrer des données *Linked Data* sur le référentiel géographique calcule donc la distance minimale entre un point et une géométrie de type quelconque, ou retourne une valeur nulle si le point est inclus dans cette géométrie. Les coordonnées des géométries peuvent être exprimées dans différents systèmes de coordonnées. Nous commençons donc par

11. <https://www.assembla.com/spaces/silk/wiki/Home>

projeter toutes les coordonnées dans le même système de projection. Ainsi, nous pouvons faire reposer le calcul de notre mesure de distance sur le mesure de la distance euclidienne minimale entre deux objets de la bibliothèque Java Geotools ¹², et l'intégrer sous forme de plugin dans l'outil d'interconnexion Silk ((3) figure 2).

4.1.3 Paramétrage de l'interconnexion

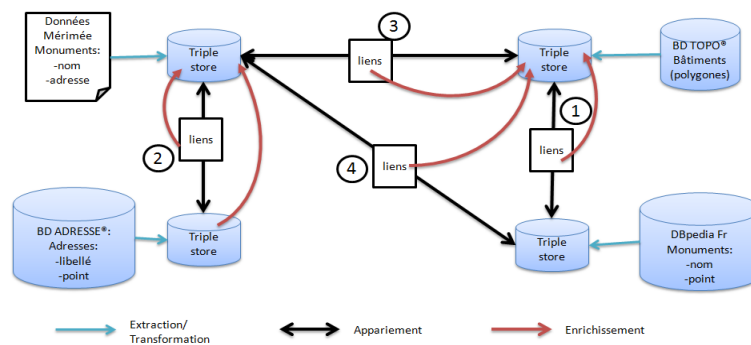


FIG. 3 – Interconnexions réalisées.

La première étape de notre démarche d'interconnexion est l'ancrage des données sur les monuments historiques issues de DBpedia sur les bâtiments de la base BD TOPO® ((1) figure 3). Dans ce cas, le seul critère d'interconnexion pouvant être utilisé est la localisation. Nous avons donc créé et exécuté un script Silk utilisant notre mesure de distance entre les positions ponctuelles de monuments de DBpedia et les polygones décrivant les bâtiments dans la BD TOPO®. Nous avons choisi un seuil maximal de 40 mètres, défini en nous appuyant sur la précision planimétrique connue de la BD TOPO® (10 mètres) et l'estimation empirique de la précision des indications de localisation dans les données DBpedia. Dans le cas des monuments de la base Mérimée, localisés par des adresses, nous avons tout d'abord procédé au géocodage des données. Nous avons donc réalisé une interconnexion entre les indications de localisation des monuments dans la base Mérimée et la base BD ADRESSE®. Pour ce faire, nous avons combiné les attributs qui constituent les adresses textuelles dans les deux bases et les avons comparés via une mesure de Levenshtein modifiée pour devenir insensible au désordre des mots. Ainsi, nous avons pu affecter à chaque monument interconnecté les coordonnées géographiques correspondant à son ou ses adresse(s) ((2) figure 3). Nous avons ensuite pu procéder à l'interconnexion des monuments Mérimée dotés de coordonnées avec les bâtiments BD TOPO® en réutilisant l'approche adoptée pour les monuments DBpedia ((3) figure 3). Enfin, nous avons procédé à une dernière interconnexion ((4) figure 3) entre les monuments issus de Mérimée et les monuments de Dpédia qui ne disposaient pas d'informations de localisation fiables. Pour ce faire, nous avons utilisé les noms attribués aux monuments comme critère d'interconnexion, et les avons comparés en nous servant de la distance de Levenshtein modifiée disponible dans Silk.

12. <http://www.geotools.org/>

4.1.4 Résultats et évaluation

Pour les deux sources de données, sur les 2462 ressources récupérées initialement, nous avons pu en interconnecter 1653 avec des bâtiments BD TOPO®. Ce résultat s'explique en partie par l'étape préalable de géocodage que nous avons dû appliquer aux données Mérimée. Lors de cette étape, nous avons paramétré Silk de sorte à obtenir le plus possible d'interconnexions justes, au risque de ne pouvoir géocoder certaines données Mérimée faute d'interconnexions. Pour cette étape, nous avons comparé les liens générés avec un mapping de référence créé manuellement sur la zone du premier arrondissement de Paris. Cette comparaison a révélé une précision de 100% et un rappel de 90,35%. Cette précision peut se justifier par le fait que les liens comparés ne concernent que le 1er arrondissement. Le rappel s'explique en partie par le paramétrage choisi et par la présence dans la base Mérimée d'indications de localisation non conformes à des adresses comme " face au 14 " ou " dans le cimetière ".



FIG. 4 – Interconnexion du Palais de la porte dorée avec les deux sources de données.

Nous n'avons pu procéder à une évaluation chiffrée des résultats d'interconnexion avec la BD TOPO®, faute de mapping de référence. Nous avons donc procédé par validation visuelle effectuée par une seule personne sur une interface cartographique. Ainsi, nous avons pu constater que l'approche d'ancrage sur un référentiel géographique permettait bien de mettre en correspondance des ressources qui n'auraient pas pu l'être par comparaison directe, en raison de localisations trop éloignées ou de dénominations différentes. C'est le cas pour " le palais de porte dorée " nommé ainsi dans DBpedia, et intitulé " Ancien Musée National des Arts Africains et Océaniens, devenu Cité nationale de l'histoire de l'Immigration " dans la base Mérimée avec une distance supérieure à 60m entre les deux localisations (figure 4).

Les interconnexions ont été réalisées sur un ordinateur bureautique sous Ubuntu12.04 LTS avec 3,8 Go de RAM et un processeur intel Core i5 2.40GHz*4, en utilisant la variante «Single Machine» de Silk. Nous avons utilisé les options de blocking disponibles pour optimiser le temps de calcul : chacune des interconnexions réalisées a pris moins de 3h30. Pour améliorer les temps de calcul d'interconnexion ou traiter plus de données, nous aurions pu utiliser la variante "Slik MapReduce". Cependant la validation demeure manuelle ; il est donc impossible pour l'instant de passer à l'échelle pour cette étape.

4.2 Implémentation de l'application cartographique

Afin de visualiser les liens d'ancrage des ressources thématiques sur les données géographiques de référence, nous avons implémenté une solution de visualisation cartographique. De plus, nous proposons une approche d'agrégation d'informations afin de permettre une meilleure visualisation des données à différentes échelles.

4.2.1 Architecture et outils

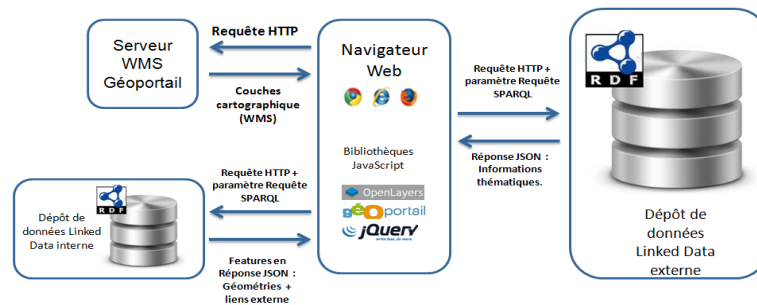


FIG. 5 – Architecture de visualisation cartographique des données.

L'architecture de notre application de visualisation cartographique est présentée en figure 5. L'interface proprement dite est un client léger développé à l'aide de la bibliothèque JavaScript OpenLayers¹³ et de l'API Géoportail¹⁴. La première permet l'affichage, sous forme de couches, de données stockées sur des serveurs géographiques et envoyées conformément aux protocoles décrits par les spécifications de l'OGC (Open Géospatial Consortium). La seconde nous fournit des fonds cartographiques et orthophotographiques produits par l'IGN.

OpenLayers permet également de créer d'autres couches vectorielles à partir de données récupérées par différents protocoles, notamment le protocole HTTP. Elle permet donc d'interroger un point d'accès SPARQL. De plus, elle offre la possibilité de mettre en forme les données ainsi récupérées avant de les visualiser. Ainsi, on peut créer des couches de données vectorielles en utilisant les géométries des données géographiques de référence que nous avons préalablement converties et stockées en RDF pour l'étape d'interconnexion. De plus, on peut mettre à profit les liens d'ancrage entre ressources externes et données géographiques pour récupérer des informations thématiques sur ces ressources externes. La figure 6 présente un exemple de co-visualisation des données géographiques de référence et des ressources thématiques ancrées sur ces données. Les bâtiments interconnectés avec des monuments historiques de la base Mérimée sont affichés dans des tons de bleu, la valeur du bleu variant selon le siècle de construction de monument en question fourni par la base Mérimée. Les bâtiments interconnectés avec des monuments historiques DBPedia sont dotés de contours noirs. Ceux interconnectés avec les deux sources sont à la fois en bleu et avec un contour noir. C'est le cas par exemple du bâtiment qui héberge la bourse du commerce de Paris.

13. <http://openlayers.org/>

14. <http://www.geoportail.gouv.fr/accueil>

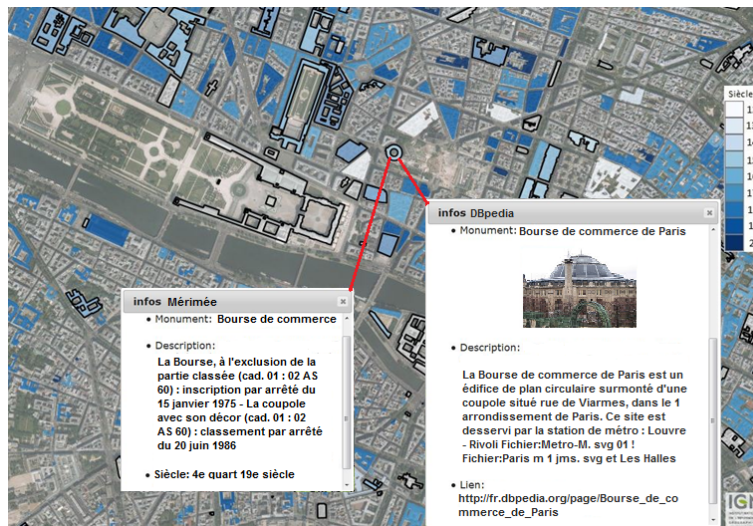


FIG. 6 – Visualisation des données géographiques interconnectées à des ressources Mérimée et DBPedia, affichage au niveau des bâtiments (exemple : bourse du commerce de Paris).

4.2.2 Visualisation à différentes échelles

Afin de fournir un affichage cartographique lisible à différentes échelles, on a généralement recours à des techniques dites de généralisation. La généralisation est un processus de synthèse d'information, qu'on peut comparer à un processus de résumé de texte qui réduit le nombre de mots mais garde les idées principales Ruas (2002). Elle vise à simplifier l'information en passant d'un niveau d'échelle à un autre en gardant le même sens pour la carte. Dans notre cas, nous avons eu recours à une approche de simplification fondée sur une classification sous contrainte de contiguïté spatiale Carvalho et al. (2009).

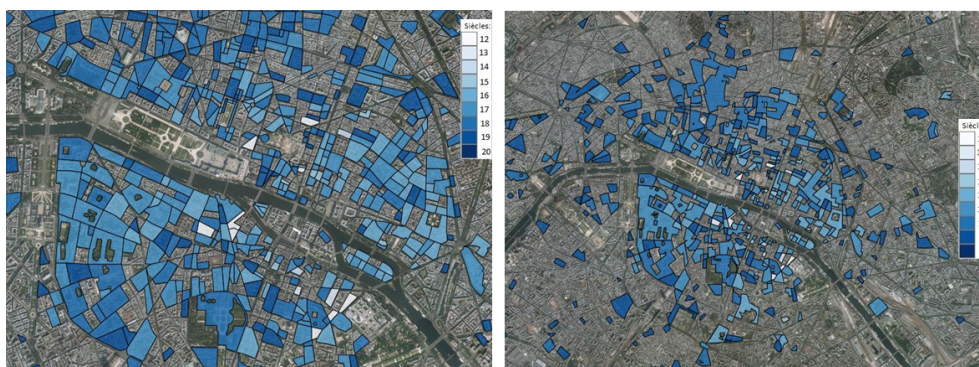


FIG. 7 – Visualisation des îlots et des îlots agrégés à différentes échelles

Cette méthode consiste à unir les objets voisins qui présentent des propriétés similaires. Nous regroupons les objets en fonction du siècle de construction des monuments historiques. Cette technique ne peut pas être appliquée directement sur les bâtiments, compte tenu de la non-contiguïté entre ces derniers. Nous avons donc appliqué cette méthode sur les îlots surfaciques créés à partir du réseau routier, en leur affectant le siècle de construction le plus fréquemment rencontré parmi les bâtiments correspondant à des monuments classés qu'ils contiennent (partie gauche figure 7). Ceci nous a permis de générer des représentations plus agrégées pour un affichage à petite échelle (partie droite figure 7).

5 Conclusion et perspectives

Dans cet article, nous avons proposé une approche visant à tirer parti des informations de localisation associées à des ressources *Linked Data* dans un processus d'interconnexion. Cette approche consiste à interconnecter les ressources non pas directement entre elles, mais avec un référentiel géographique afin de pallier les difficultés d'interconnexion liées à l'hétérogénéité des indications de localisation. Les interconnexions générées entre données thématiques et données géographiques ont été ensuite mise à profit dans une application de visualisation cartographique des données à différentes échelles reliant les informations thématiques directement aux objets géographiques. Les liens d'ancrages générés entre données thématiques et données géographiques sont des relations d'équivalence. Pour plus de cohérence sémantique, nous envisageons d'étudier les différents types de liens sémantiques qui peuvent exister entre une ressource thématique et un objet géographique.

Remerciements

Ce travail a été réalisé dans le cadre du projet Datalift, en partie financé par l'Agence Nationale de la Recherche (ANR-10-CORD-009).

Références

- Aleksovski, Z., W. ten Kate, et F. van Harmelen (2006). Exploiting the structure of background knowledge used in ontology matching. In *Ontology Matching Workshop, 5th International Semantic Web Conference*, Athens, Georgia, USA, pp. 13–24.
- Beeri, C., Y. Kanza, E. Safra, et Y. Sagiv (2006). Object fusion in geographic information systems. In *VLDB*, pp. 816–827.
- Bel Hadj Ali, A. (2001). *Qualité géométrique des entités géographiques surfaciques - Application à l'appariement et définition d'une typologie des écarts géométriques*. Ph. D. thesis, Université de Marne-la-Vallée.
- Carvalho, A. X. Y., P. H. M. Albuquerque, G. R. de Almeida Junior, et R. D. Guimarães (2009). Spatial hierarchical clustering. *Rev. Bras. Biom* 27(3), 411–442.
- Ferraram, A., A. Nikolov, et F. Scharffe (2013). Data linking for the semantic web. *Semantic Web : Ontology and Knowledge Base Enabled Tools, Services, and Applications*, 169.

- Hahmann, S. et D. Burghardt (2010). Connecting linkedgeodata and geonames in the spatial semantic web. In *Proc. of GIScience 2010 Extended Abstracts*, Purves, R. and Weibel, R.(eds.), pp. 28–34.
- Heath, T. et C. Bizer (2011). Linked data : Evolving the web into a global data space. *Synthesis lectures on the semantic web : theory and technology* 1(1), 1–136.
- Minami, M. (2000). *Using ArcMap*. ESRI Press.
- Mustière, S. et T. Devogele (2008). Matching networks with different levels of detail. *Geoinformatica* 12(4), 435–453.
- Olteanu, A.-M. (2008). *Appariement de données spatiales par prise en compte de connaissances imprécises*. Ph. D. thesis, Université de Marne-la-Vallée.
- Owusu-Banahene, W. et S. Coetzee (2013). Integrating linked open data into open source web mapping. In *International Cartographic Conference*. International Cartographic Association.
- Ruas, A. (Ed.) (2002). *Généralisation et représentation multiple*. Information Géographique et Aménagement du Territoire. Hermes Lavoisier.
- Safra, E., Y. Kanza, Y. Sagiv, et Y. Doytsher (2006). Efficient integration of road maps. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, GIS '06, New York, NY, USA, pp. 59–66. ACM.
- Salas, J. et A. Harth (2011). Finding spatial equivalences accross multiple rdf datasets. In *Terra Cognita Workshop*, pp. 114–126.
- Samal, A., S. Seth, et K. Cueto1 (2004). A feature-based approach to conflation of geospatial sources. *International Journal of Geographical Information Science* 18(5), 459–489.
- Shvaiko, P., F. Farazi, V. Maltese, A. Ivanyukovich, V. Rizzi, D. Ferrari, et G. Ucelli (2012). Trentino government linked open geo-data : a case study. In *The Semantic Web–ISWC 2012*, pp. 196–211. Springer.
- Walter, V. et D. Fritsch (1999). Matching spatial data sets : a statistical approach. Volume 13, pp. 445–473.

Summary

Many resources published on the Web of Data are described by predicates that refer directly or indirectly to locations on the Earth. Like any type of property, this information about location can be used with benefits as data linking criterion. Besides, resources associated to a location can be displayed on maps. However, location information used in linked data sources may be inaccurate or very heterogeneous from one source to another, which makes the use of this criterion for data linking difficult. In this article, we propose to anchor linked data sources to a reference geographic database to support the discovery of relationships between resources that are related to the same real world geographic entity but that are described with heterogeneous location information. Finally, we use the links generated between thematic linked data and reference geographic data in a Web mapping interface for data visualization.