
Modélisation sémantique et publication d'observations environnementales du nord canadien.

Philippe Genoud¹, Marianne Douglas², Danielle Ziébelin¹

1. *Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes
LIG-STEAMER, 681 rue de la passerelle 38041 Grenoble CEDEX, France
Philippe.Genoud@imag.fr, Danielle.Ziebelin@imag.fr*

2. *Queen's University, Department of Biology Kingston, ON, Canada K7L 3N6
Cette recherche a été faite au CCI University of Alberta
msvdouglas@gmail.com*

RESUME. Ce travail issu d'une collaboration entre le département de biologie de l'Université de Queen's, le Canadian Circumpolar Institute (CCI) de l'Université d'Alberta et le Laboratoire d'Informatique de Grenoble (LIG) pour la mise en place d'une plate-forme de publication et d'analyse de données environnementales s'appuyant sur les standards du web sémantique. Les données concernent la chimie et la biologie de l'eau des régions arctiques et subarctiques. Elles sont structurées en des séries de mesures effectuées lors de campagnes de prélèvements estivales.

Le système présenté est un premier prototype qui démontre l'intérêt de publier et exploiter les données en s'appuyant sur les standards du web sémantique. Nous verrons ainsi : comment des données initialement tabulaires peuvent être représentées sous la forme d'un graphe RDF en s'appuyant sur des vocabulaires issus d'ontologies de domaine permettant de leur associer une sémantique plus riche, comment elles peuvent être liées à d'autres données disponibles sur le web des données ouvertes et liées et comment ainsi enrichies elles peuvent être interrogées, via une interface cartographique, au travers de requêtes exploitant leur sémantique.

ABSTRACT. This work, stemming from a collaboration between the departments of biology of the University of Queen's, the Canadian Circumpolar Institute (CCI) of Alberta University and the Laboratory of Computing of Grenoble (LIG), is dedicated to the implementation of a platform for publishing and analyzing environmental data by using semantic Web standards. The data concern the chemistry and the biology of the water in Arctic and subarctic regions. These data sets are structured in a series of measurements made during summer sampling collections.

This platform is a first prototype which demonstrates the interest in publishing and exploiting the data using semantic Web standards. We shall demonstrate how initial tabular data can be represented by an RDF graph based on vocabularies according to domain ontologies which allow the data a richer semantics; how these data can be linked to other data available on the Web; and how these enriched data then can be queried, via a cartographic interface, through requests which exploit their semantics.

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

MOTS-CLES : Web sémantique, Ontologies, OWL, RDF, SPARQL, GeoSPARQL, Données Ouvertes et Liées, Observations environnementales.

KEYWORDS: Semantic Web, Ontologies, OWL, RDF, SPARQL, GeoSPARQL, Linked Open Data, Environmental Observations.

1. Introduction

Dans le domaine de l'environnement, l'étude de l'évolution de données d'observation est une tâche centrale, ainsi au-delà de l'analyse de leurs propres données, il est primordial pour les chercheurs de les croiser avec d'autres données. L'évolution d'internet facilite grandement ces possibilités d'accès et d'échange de données, comme en témoigne en géomatique la mise en place d'infrastructures de données spatiales. Cependant, l'hétérogénéité des modèles, données, métadonnées et des formats, leur changement au cours du temps, demeure une difficulté majeure lors de l'intégration de sources différentes. Actuellement, après le formidable essor du web 2.0, nous assistons à une évolution du world wide web vers ce que le W3C désigne sous le terme de web des données : un modèle de données simple, flexible et puissant, le Resource Description Framework (RDF) (Cyganiak *et al.*, 2014), qui en s'appuyant sur l'infrastructure du web, facilite la publication et l'échange de données à l'échelle du web ; des modèles de représentation, en particulier RDFS (Brickley et Guha 2004) et OWL (Hitzler *et al.*, 2012), permettant de définir sous forme d'ontologies la connaissance permettant de donner une sémantique aux données, sémantique qui pourra être exploitée par des programmes informatiques. C'est dans cette lignée que s'inscrit l'initiative des données ouvertes et liées (Linked Open Data) (Bizer *et al.*, 2009) dont les principes ont été énoncés par Tim Berners-Lee (Berners-Lee 2006) : (1) utiliser des URIs (Uniform Resource Identifiers) pour nommer (identifier) les choses, (2) utiliser des URIs HTTP de sorte de pouvoir consulter ces adresses, (3) quand une URI est consultée, fournir une information utile en utilisant des standards ouverts (RDF, SPARQL, ...), (4) inclure des liens vers d'autres URI afin de pouvoir découvrir plus de choses.

Ce sont ces principes que nous nous proposons d'appliquer à des données environnementales concernant la qualité de l'eau dans le nord canadien et utilisées entre autres pour l'étude de l'impact du changement climatique dans ces régions arctiques et subarctiques (Lim *et al.*, 2008). Dans cet article nous présentons la démarche qui a été la nôtre, de la transformation de données tabulaires en RDF en leur associant un modèle sémantique, leur enrichissement en les liants à d'autres données, à leur accès au travers d'une interface web cartographique. Ce travail nous a permis d'appréhender les différentes étapes nécessaires à une telle démarche et d'approcher différentes technologies nécessaires à sa mise en œuvre.

2. Représentation des données

Notre cas d'étude porte sur les écosystèmes arctiques et subarctiques, les données biologiques et chimiques l'eau proviennent soit de stations de monitoring, soit de campagnes de prélèvement. Le but de cet ensemble de données est d'aider le chercheur à répondre à la question : quel est l'état de la biodiversité dans les zones arctiques et subarctiques et établir une ligne d'évolution temporelle (Lim *et al.*, 2005). A titre d'exemple, la Figure 1 présente les données de chimie de l'eau relevées sur l'île de Banks durant l'été 2012.



Figure 1: Les données relevées sur différents sites d'observation sur l'île de Banks (Territoires du Nord-Ouest, Canada) lors de la campagne 2012 et enregistrées sous forme de tableaux excel. Chaque site de prélèvement (un lac ou une mare (pond)) est repéré par ses coordonnées géographiques (latitude et longitude) et désigné par un identifiant unique (par exemple BKAF). Chaque site est associé à une région (par exemple, ici l'île de Banks). Les observations correspondent aux colonnes de la table. Le titre de la colonne décrit la nature de la mesure : par exemple la température de l'eau en degrés Celsius, sa teneur en Sodium (Na) en mg par litre.... L'ensemble des mesures (mesures set) effectuées à une date donnée pour un site donné correspondent à une ligne de la feuille de calcul.

Au cours des années (25 ans) un certain nombre de données ont ainsi été collectées, parfois de nature et de formes différentes, enregistrées dans fichiers de format divers. Pour faciliter leur exploitation, nous proposons un modèle conceptuel (ontologie) qui permettra de les représenter dans le format RDF et ensuite de les relier à d'autres sources de données externes.

2.1. Ontologie de l'application

Pour modéliser ces données sous la forme d'un graphe RDF, une ontologie OWL a été définie de manière spécifique pour cette application. Cette ontologie comprend un certain nombre de classes et de propriétés permettant de représenter les données d'observations. Pour représenter l'information spatiale (coordonnées des sites de prélèvements, géométrie des régions) nous nous sommes appuyés sur le standard GeoSPARQL proposé par l'OGC (OGC, 2012) (Battle et Kolas, 2012). La Figure 2 présente sous la forme d'un diagramme UML les différentes classes, et relations (*owl:ObjectProperties*) les liant, définies pour notre application.

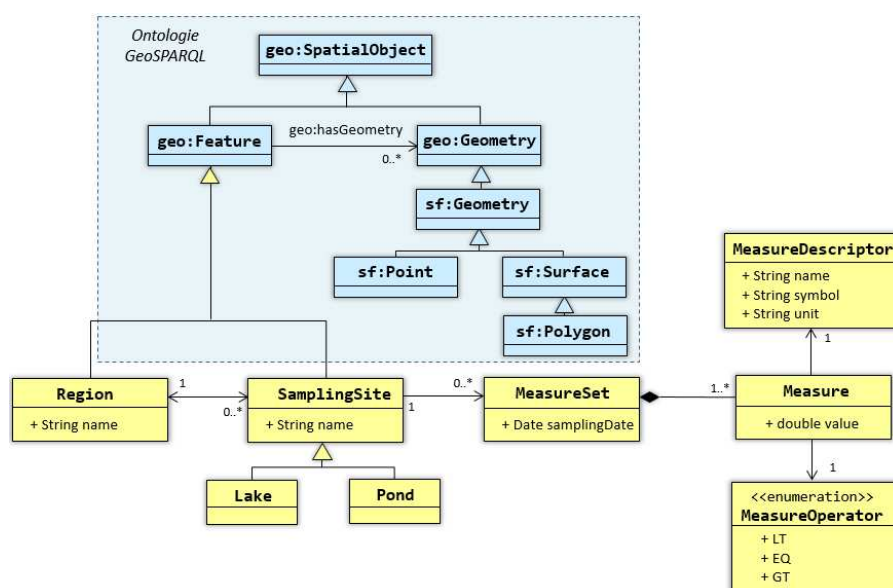


Figure 2 : l'ontologie de l'application et ses liens avec l'ontologie GeoSPARQL.

Les classes Region et SamplingSite sont définies comme sous-classes de la classe geo:Feature de l'ontologie GeoSPARQL. La représentation spatiale de ces entités géographiques est réalisée en utilisant les types géométriques GeoSPARQL basés sur l'OGC Simple Features model (OGC, 2011) et définis comme sous classes de la classe geo:Geometry, dans notre cas: sf:Polygon pour les régions et sf:Point pour les sites de prélèvement.

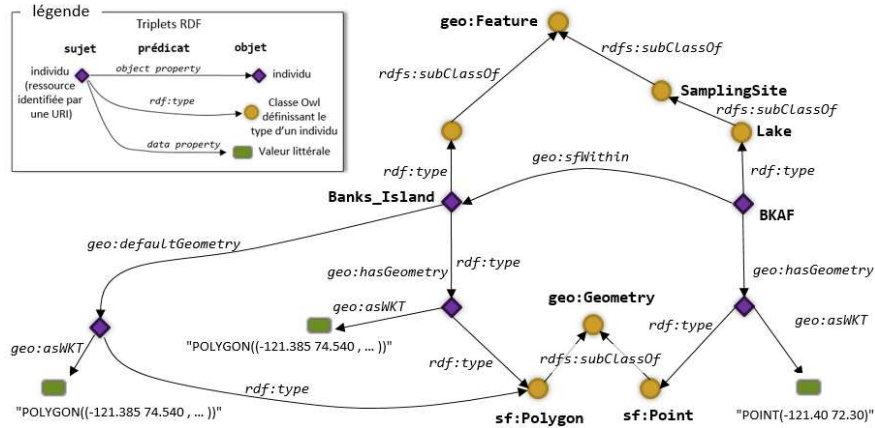


Figure 3 : Extrait du graphe RDF représentant, à partir des données extraites du tableau excel de la Figure 1, le site BKAF et sa région d'appartenance l'île de Banks.

Les coordonnées spatiales sont quant à elles définies sous la forme de littéraux au format WKT (*Well Known Text*) associés à l'objet géométrie par la propriété de données (*owl:DataProperty*) `geo:asWKT` (Figure 3).

2.2. Ontologie de domaine

Nous avons étendu notre ontologie d'application par une ontologie générale issue du domaine de l'hydrologie proposée par le Consortium of Universities for the Advancement of Hydrologic Science¹ (CUAHSI) (Couch *et al.*, 2014).

¹ <https://www.cuahsi.org/>

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

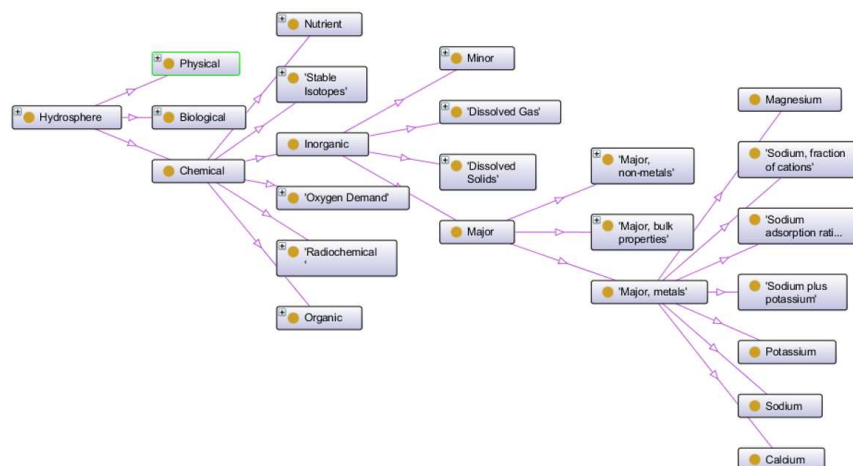


Figure 4 : un extrait de l'ontologie hydrosphère du CUAHSI. Seuls sont développés sur cette hiérarchie, les classes correspondant aux concepts de métaux lourds.

Cette ontologie définit une taxonomie qui permet de structurer de manière hiérarchique plus de 4000 termes décrivant des mesures physiques, chimiques et biologiques liées à l'eau. Elle est utilisée par le Système d'Information du CUAHSI (CUAHSI-HIS Hydrologic Information System) composé d'un ensemble de serveurs et bases de données connectés à des applications clientes sous la forme de services web pour faciliter la découverte de séries temporelles de données collectées en un point donné. Nous avons repris cette ontologie définie sous forme tabulaire² pour la traduire sous la forme d'une hiérarchie de classes OWL (Figure 4).

L'utilisation de cette ontologie dans notre modèle, s'effectue en associant aux descripteurs des mesures relevées lors des campagnes de prélèvements le type de la mesure correspondante dans la terminologie du CUAHSI (Figure 5).

² <http://his.cuahsi.org/ontologyfiles.html>

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEomatics conference, SAGEO 2015.

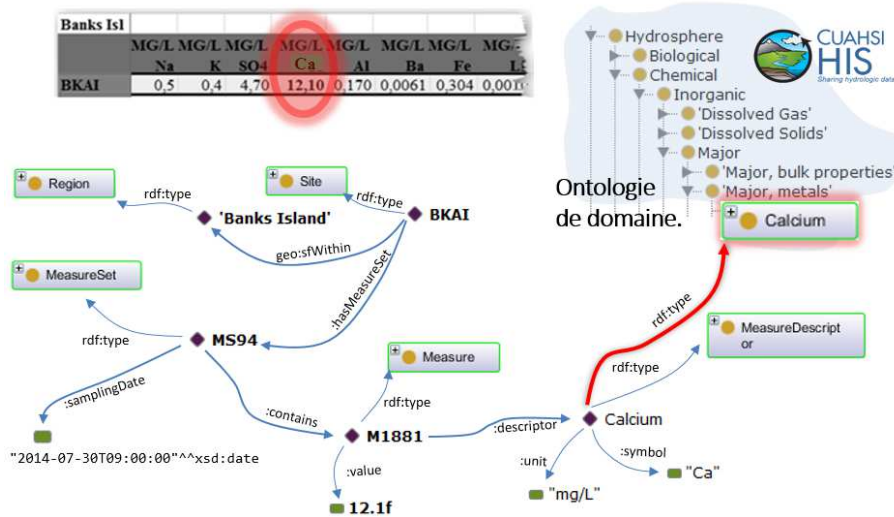


Figure 5 : sur le site BKAI de Banks Island un ensemble de mesures MS94 a été effectué. La mesure M1881 de valeur 12.10 est associée à un descripteur qui indique qu'il s'agit d'une mesure du Calcium et que son unité est mg/l. Ce descripteur est relié à l'ontologie de domaine du CUAHSI par une relation rdf:type.

2.3. Liaison vers des jeux de données externes

Un second enrichissement des données initiales que permet leur représentation sous la forme d'un graphe RDF est la possibilité de les lier avec d'autres jeux de données externes publiés en respectant les principes des données liées et ouvertes. Pour démontrer le potentiel d'un tel enrichissement nous avons liés nos données représentées à l'aide de notre ontologie de domaine avec des données issues de DBpedia³ (Lehmann *et al.*, 2015), et nos données concernant les régions d'étude avec des données issues de GeoNames⁴.

2.3.1. Liaison avec Geonames

Le gazetteer GeoNames propose une base de données géographique libre et gratuite, accessible par Internet sous une licence Creative Commons, contenant plus de 10 millions de toponymes. Outre les coordonnées géographiques de chaque lieu répertorié, GeoNames propose des données comme l'altitude, la population, la subdivision administrative, le code postal et des liens vers les éventuelles pages Wikipedia correspondantes dans plusieurs langues. Outre un accès via des services web, ou à une copie brute « dump » de la base de données, GeoNames publie ses

³ <http://wiki.dbpedia.org>

⁴ <http://www.geonames.org>

données en RDF. Ainsi, chaque entité géographique (*feature*) de GeoNames est représentée comme une ressource web identifiée par une URI stable qui par négociation de contenu donne accès soit à la page HTML montrant l'entité sur une carte, soit à une description RDF de l'entité basée sur un vocabulaire défini par une ontologie OWL⁵ (Figure 6).

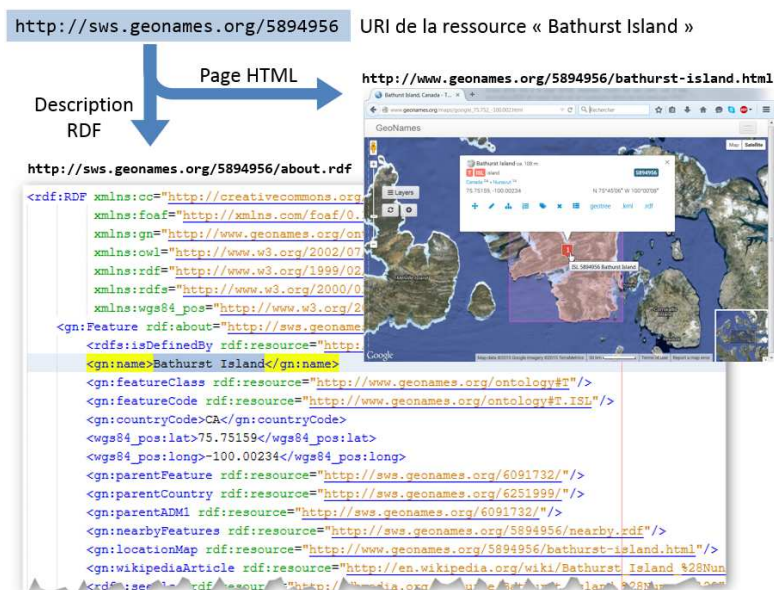


Figure 6 : Négociation de contenu avec GeoNames. Selon la valeur de l'entête accept d'une requête HTTP demandant la ressource le serveur enverra soit une page HTML, soit une description RDF de cette ressource.

Pour chacune des régions dans lesquelles se trouvent les sites, nous avons recherché dans GeoNames une entrée correspondant au nom utilisé en filtrant les résultats sur la base du pays et de la géolocalisation. Cette région du grand nord canadien étant relativement peu couverte, nous avons parfois été conduits à ajouter de nouvelles entrées dans GeoNames (ce fût le cas par exemple de l'île de Bathurst). Un lien `owl:sameAs` permet alors d'associer à une région sa représentation dans GeoNames (Figure 7).

2.3.2. Liaison avec DBpedia

Nous avons effectué la liaison avec DBpedia en cherchant pour chacune des classes de notre ontologie de domaine une ressource correspondante dans DBpedia. Pour effectuer cet alignement, nous avons écrit un programme comparant le label des concepts issus de l'ontologie du CUASHI avec les labels des ressources

⁵ http://www.geonames.org/ontology/ontology_v3.1.rdf

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

DBpedia (plus de 11 000 000 de ressources). Une fois les labels normalisés (casse minuscule, suppression des espaces et caractères non alphabétiques) une recherche de labels identiques est effectuée. Ainsi plus de 2000 des 4000 concepts définis ont pu être automatiquement retrouvés dans DBpedia et un lien `rdfs:seeAlso` a été ajouté entre la classe de notre ontologie et la ressource DBpedia correspondante (Figure 7). Pour les concepts pour lesquels cette étape a échoué, une comparaison basée sur une distance de Levenshtein (Levenshtein 1966) a été effectuée et une liste des meilleurs candidats potentiels a été produite, nécessitant ensuite une étape manuelle pour éventuellement effectuer une liaison.

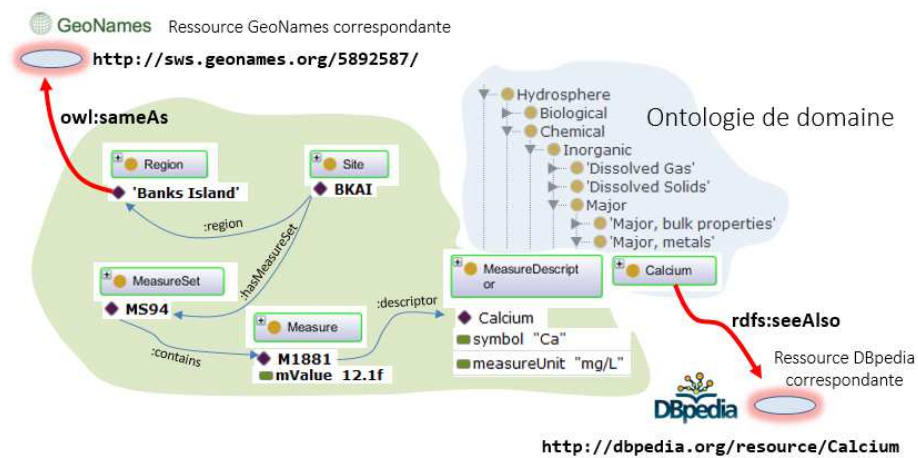


Figure 7 : Liaison des données vers GeoNames et DBpedia.

3. Architecture et implémentation du système

Pour explorer et analyser les données d'observation ainsi représentées, nous proposons une interface web cartographique qui permet de requêter les données au travers de SPARQL et son extension spatiale GeoSPARQL. Pour cela les données brutes issues des fichiers excel produits lors des campagnes de prélèvement ont été converties en RDF et stockées dans une base de données spécialement conçue pour le stockage et la récupération de données RDF appelé « triplestore ».

3.1. Stockage des données RDF

Bien que publié en 2012, peu de triplestores supportent actuellement le standard GeoSPARQL (Athanasίου *et al.*, 2013): Strabon (Kyzirakos *et al.*, 2012), uSeekM⁶ et

⁶ <https://dev.opensahara.com/projects/useekm/>

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

Parliament.⁷ Notre choix s'est porté sur Parliament, qui offre un relativement bon compromis entre facilité d'installation et d'utilisation, un support de GeoSPARQL et des performances acceptables (bien que dans l'état actuel loin des performances offertes par les bases de données spatiales (Patroumpas *et al.*, 2014)).

Sur le serveur Parliament, un graphe RDF est créé dans lequel sont chargés :

- les ontologies utilisées : l'ontologie de domaine (préfixe cuahsionto:), l'ontologie d'application (préfixe ccionto:);
- les données d'observation issues des fichiers excel et transformées en RDF en utilisant les vocabulaires définis par les ontologies précédentes et le vocabulaire GeoSPARQL pour leur dimension spatiale.

Lors du chargement, Parliament effectue automatiquement un certain nombre d'inférences : inférences RDFS plus un certain nombre d'inférences OWL (classes et propriétés, équivalentes, propriétés inverses, symétriques, fonctionnelles, inverses fonctionnelle et transitives). Une fois chargées, ces données peuvent être interrogées via des requêtes GeoSPARQL transmises (via http) au point d'accès du serveur Parliament (serveur Jetty + Joseki).

⁷ <http://parliament.semwebcentral.org>

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEomatics conference, SAGEO 2015.

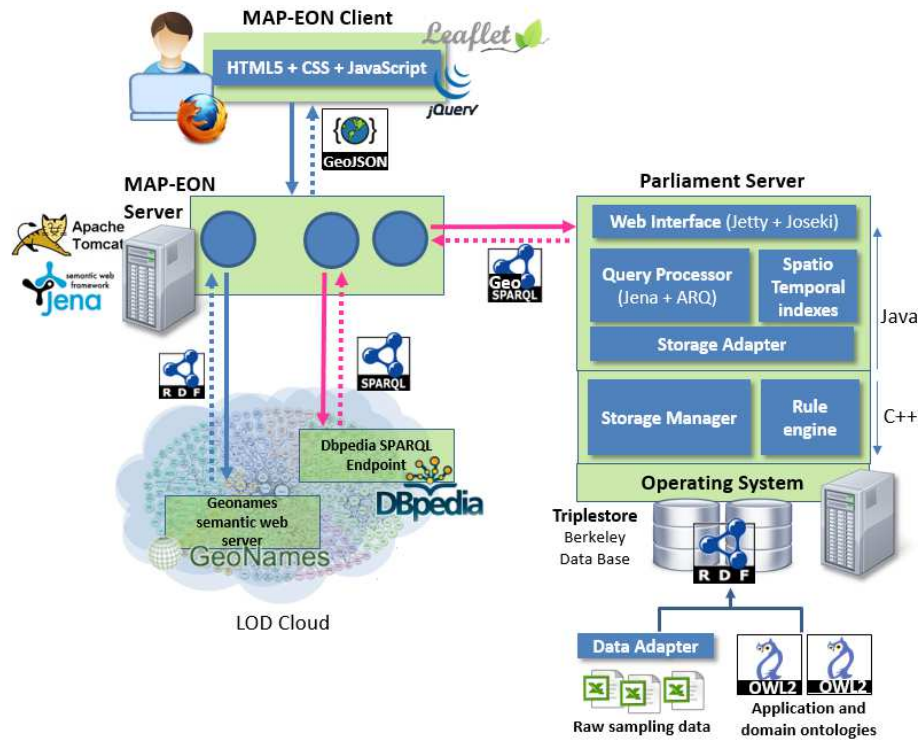


Figure 8 : Architecture du système MAP-EON

3.2. L'application web MAP-EON

L'application MAP-EON se présente sous la forme d'une application web. L'interface utilisateur s'exécute dans un navigateur web sur le poste client. Réalisée avec les technologies HTML5, CSS3 et JavaScript elle s'appuie sur les bibliothèques Leaflet⁸ pour l'interface cartographique, et jQuery⁹, jQueryUI¹⁰ pour la gestion de l'interaction et les communications AJAX avec le serveur.

Le serveur MAP-EON a quant à lui été implémenté en utilisant les technologies web JEE. Un certain nombre de servlets, s'exécutant dans un conteneur Apache Tomcat, se chargent de traiter les requêtes issues du client. Comme, nous le verrons par la suite, ces traitements se traduisent pour certains en l'envoi de requêtes SPARQL vers des points d'accès SPARQL (le serveur Parliament, les serveurs

⁸ <http://leafletjs.com>

⁹ <https://jquery.com/>

¹⁰ <https://jqueryui.com/>

DBpedia) ou à récupérer des données RDF (serveur GeoNames). Pour effectuer ces traitements RDF et SPARQL nous nous appuyons sur le framework Java open source Apache Jena¹¹. Les données retournées par le serveur au client se font en utilisant les formats d'échange JSON et GeoJSON (Buttler *et al.*, 2008) pour les informations géographiques ayant vocation être affichées sur l'interface cartographique.

L'interface de l'application MAP-EON permet à l'utilisateur de visualiser les différents sites de prélèvement sur une carte interactive. La sélection d'un site sur la carte permet d'accéder aux différentes observations effectuées sur ce site. Dans la fenêtre de dialogue qui apparaît l'utilisateur dispose de trois onglets (Figure 9) qui lui permettent :

- de filtrer les mesures à afficher. Le filtrage est réalisé grâce à la hiérarchie des concepts de l'ontologie de domaine (les inférences RDFS effectuées au chargement des données ont permis de rajouter automatiquement aux descripteurs de mesures un lien `rdfs:subClassOf` vers chacune des super-classes de la classe de l'ontologie `cuahsi`. Ne sont affichées que les observations dont le descripteur de mesure est un sous concept (une sous classe) des concepts sélectionnés.
- de visualiser toutes les observations pour une date donnée,
- de visualiser toutes les observations pour une mesure donnée (série temporelle).

Grâce aux liens vers DBpedia et GeoNames que nous avons présentés au paragraphe 2.3, l'utilisateur peut accéder à des informations supplémentaires que l'application va chercher dynamiquement dans le web des données (Figure 10) :

¹¹ <http://jena.apache.org>



Figure 9 : Filtrage des observations utilisant la hiérarchie des concepts de domaine. Ne sont affichées que les observations concernant les métaux lourds et pesticides.

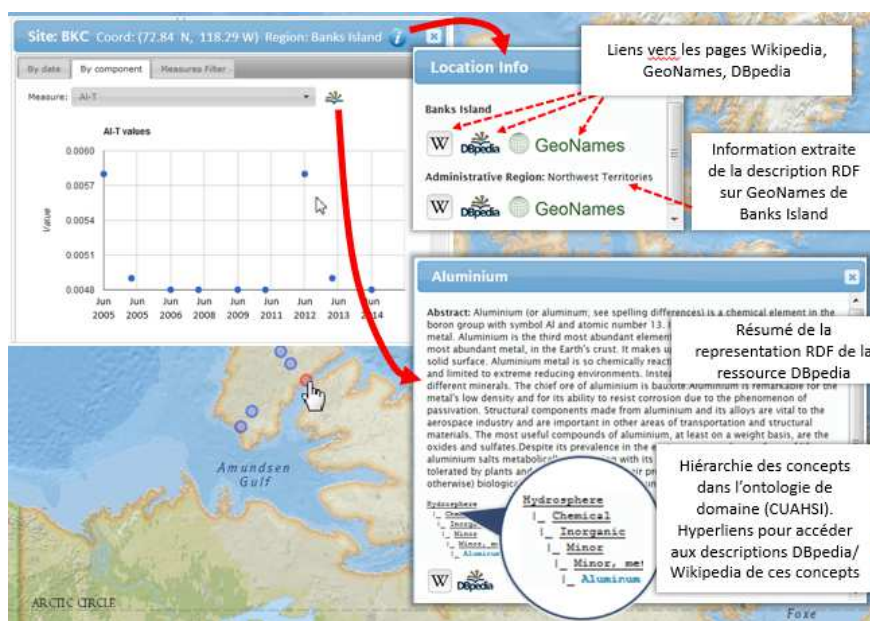


Figure 10: Recherche d'informations complémentaires dans le web des données en utilisant les liens GeoNames d'une région (ici Banks Island) ou DBpedia d'une mesure (ici l'Aluminium).

- le lien GeoNames permet de retrouver la région administrative de niveau supérieur et les pages Wikipedia décrivant une région ;
- le lien DBpedia permet pour une observation donnée d'aller chercher dans Wikipedia une description de l'élément mesuré.

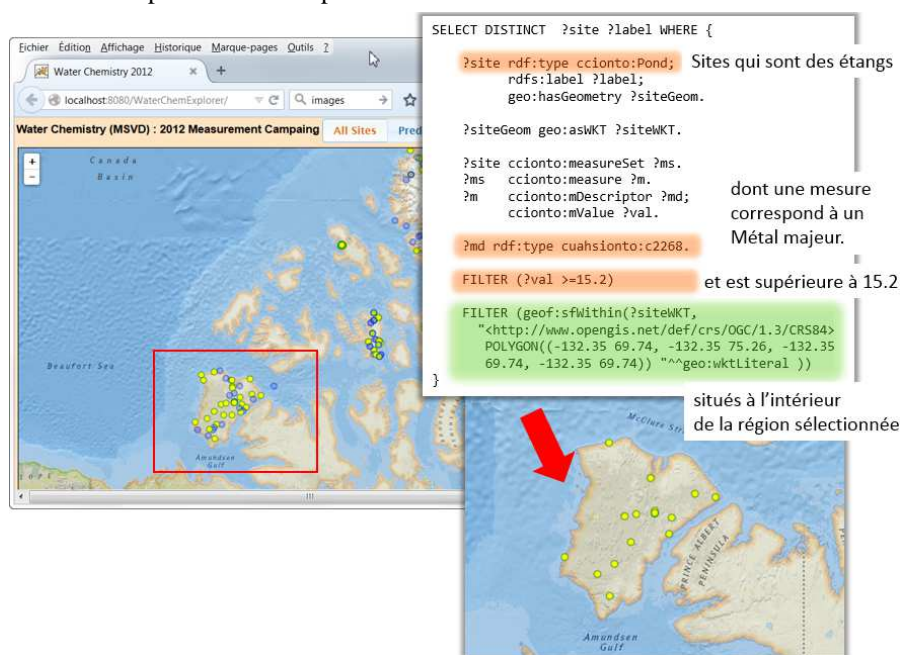


Figure 11 : Requête combinant filtre spatial et filtre « sémantique ». Seuls sont affichés les étangs (points jaunes) de la région sélectionnée et ayant une valeur de métal lourd supérieure à 15.2 mg/l.

Par ailleurs l'utilisation de GeoSPARQL permet de faire de requêtes combinant à la fois une composante spatiale et une composante sémantique. Par exemple, la requête suivante « Trouver tous les sites situés dans une zone donnée pour lesquels il existe une observation pour un métal lourd dont la valeur est supérieures à 15.2 mg/l » peut s'exprimer à l'aide d'une requête GeoSPARQL (Figure 11) qui :

- utilise les inférences de type RDFS pour ne sélectionner que les sites de type « Etang » (*Pond*) (pattern de triplet RDF `?site rdf:type ccionto:Pond`) et qui ont une observation correspondant à un métal lourd (pattern de triplet `?md rdf:type cuahsionto:c2268` où `?md` est un descripteur de mesure et `cuahsionto:c2268` l'URI du concept « Major Metal » dans l'ontologie de domaine (ontologie du CUAHSI)) ;

- utilise les capacités de traitement spatial quantitatif de GeoSPARQL en ne sélectionnant que les sites dont la géométrie est à l'intérieur zone sélectionnée: les patterns de triplets `?site geo:hasGeometry ?siteGeom. et ?siteGeom geo:asWKT ?siteWKT.` permettent de récupérer la géométrie d'un site au format WKT dans la variable `?siteWKT`; le filtre spatial `FILTER(geof:sfwithin(?siteWKT,"<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POLYGON((-132.35 69.74, -132.35 75.26, -132.35 69.74, -132.35 69.74))"^^geo:wktLiteral))` permet de ne retenir que les sites dont la géométrie est à l'intérieur de la zone sélectionnée.

Conclusion

Cet article présente l'intérêt de l'utilisation des techniques de web sémantique pour l'intégration, la publication et la visualisation de données d'observation hétérogènes dans leur contenu, leur format et leur temporalité. L'utilisation de requêtes sémantiques à l'aide d'opérateurs spatio-temporels a aussi permis de montrer toute la puissance d'un tel mécanisme dans le croisement de données brutes et la déduction de données synthétiques. Pourtant cette expérimentation a permis de mettre à jour un certain nombre d'insuffisances. Nous avons défini notre propre modèle d'observations dans notre ontologie d'application en nous appuyant sur GeoSPARQL pour la partie spatiale. Si cela présentait l'avantage de la simplicité en nous permettant de « coller » aux données brutes dont nous disposions, nous perdons en généralité et en interopérabilité. Des ontologies existent pour les observations terrestres telles OBOE (Madin *et al.*, 2007) et un travail conséquent a été mené par l'OGC pour la définition du standard Observations and Measurements (OGC, 2013). L'utilisation de ces ontologies devrait permettre une meilleure analyse sémantique de ce qu'est une observation et une mesure en définissant en particulier de façon fine les conditions d'obtention des données et spécifications des paramètres associées. Au niveau intégration, les liens vers DBpedia et Geonames ont permis de montrer l'intérêt de lier les données. Il serait intéressant d'aller un peu plus loin dans la recherche d'autres informations disponibles sur le web et qui pourraient enrichir notre application. Enfin d'un point de vue architecture, il n'existe pas encore de triplestore qui implémente complètement le standard GeoSPARQL et qui offre des performances entièrement satisfaisante. Mais l'utilisation de standards nous permet d'envisager sereinement la migration vers de nouveaux outils.

Bibliographie

- Athanasiou S., Bezati L., Giannopoulos G., Patroumpas K., Skoutas D. (2013). Market and Reaserach Overview. *GeoKnow Deliverable 2.1.1*.
- Battle R., Kolas D. (2012). GeoSPARQL: Enabling a Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web Journal*, vol. 3, n° 4, p. 355-370.
- Berners-Lee T. (2006). *Linked Data – design issues*, publié en 2006, révisé en 2009, <http://www.w3.org/DesignIssues/LinkedData.html>.

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

- Bizer C., Heath T., Berners-Lee T. (2009). Linked Data - The story so far. *International Journal on Semantic Web and Information Systems*, vol.5, n° 3, p. 1-22.
- Brickley D. Guha R. V. (2004). *RDF vocabulary description language 1.0: RDF Schema. W3C Recommendation 10 February 2004*. <http://www.w3.org/tr/rdf-schema/>.
- Butler H., Daly M., Doyle A., Gillies S., Schaub T., Schmidt C. (2008). *The GeoJSON Format Specification*, <http://geojson.org/geojson-spec.html>.
- Couch A., Hooper R., Pollak J., Martin M., Seul M. (2014), Enabling Water Science at the CUAHSI Water Data Center, *7th Int'l Congress on Env. Modelling and Software*, San Diego, California, USA.
- Cyganiak B., Wood D., Lanthaler M. (2014). *RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014*. <http://www.w3.org/tr/rdf11-concepts/>.
- Harris S., Seaborne. A. (2012). *SPARQL 1.1 Query Language. W3C Proposed Recommendation 8 November 2012*. <http://www.w3.org/tr/sparql11-query/>.
- Hitzler P., Krötzsch M., Parsia B., Patel-Schneider P.F., Rudolph. S. (2012). *OWL 2 Web Ontology Language Primer (Second Edition). W3C Recommendation 11 December 2012*. <http://www.w3.org/tr/owl2-primer/>
- Madin J., Bowers S., Schildhauer M., Krivov S., Pennington D., Villa F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, vol.2, p. 279-296.
- Lehmann J., R. Isele, Jakob M., Jentzsch A., Kontokostas D., Mendes P. N., Hellmann S., Morsey M., van Kleef P., Auer S., Bizer C. (2015). DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia, *Semantic Web Journal*, vol.6, n° 2, p. 167-195.
- Levenshtein V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, vol. 10, n° 8, p. 707-710.
- Lim D., Douglas M., Smol J. (2005). Limnology of 46 lakes and ponds on Banks Island, N.W.T, Canadian Arctic Archipelago, *Hydrobiologia*, vol. 545, n° 1, p. 11-32.
- Lim D., Smol J., Douglas M. (2008). Recent environmental changes on Banks Island (N.W.T., Canadian Arctic) quantified using fossil diatom assemblages. *Journal of Paleolimnology*, vol. 40, n° 1, p. 385-398.
- Open Geospatial Consortium (2011). *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. http://portal.opengeospatial.org/files/?artifact_id=25355
- Open Geospatial Consortium (2012). *OGC GeoSPARQL - A Geographic Query Language for RDF Data*, https://portal.opengeospatial.org/files/?artifact_id=47664
- Open Geospatial Consortium (2013). *OGC Abstract Specification, Geographic information - Observations and measurements*. https://portal.opengeospatial.org/files?artifact_id=41579
- Patrourmpas K., Giannopoulos G., Athanasiou S. (2014). Towards GeoSpatial semantic data management: strengths, weaknesses, and challenges ahead. Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas, Texas, USA.

- Parundekar R., Knoblock C.-A., Ambite J.-L. (2010). Aligning Geospatial Ontologies on the Linked Data Web, *Proceedings of the GIScience Workshop on Linked Spatiotemporal Data, 2010 (LSTD-2010)*, Zurich, Switzerland.
- Kyzirakos K., Karpathiotakis M., Koubarakis M. (2012). Strabon: A Semantic Geospatial DBMS. *11th International Semantic Web Conference (ISWC 2012)*, Boston, USA.