

# Données géohistoriques ouvertes et liées : les transformations territoriales de Montréal sous la loupe des graphes de connaissances

David Valentine, doctorat en sciences de l'information  
david.valentine@umontreal.ca

## Résumé

Ce projet de recherche porte sur la représentation sémantique des informations géohistoriques extraites de cartes historiques. Il vise à décrire les transformations territoriales de Montréal à l'aide des outils de modélisation du Web sémantique pour ouvrir l'écosystème montréalais des informations historiques au monde des données liées.

## Problématique

L'enrichissement sémantique des données géohistoriques sur Montréal demeure à ce jour passablement inexploité, tandis que la représentation du contenu des cartes historiques de la ville ne bénéficie d'aucune formalisation opérationnelle qui permettrait d'incorporer ce type d'information dans un réseau de données liées. Il existe pourtant une quantité importante de documents cartographiques produits à travers les époques qui représentent divers aspects du territoire montréalais.

Actuellement, l'écosystème des données géohistoriques à Montréal forme un riche ensemble de systèmes Web (systèmes d'information géographique et entrepôts de données) qui font usage de plusieurs formats de données. Les projets porteurs dans ce contexte (Olson *et al.*, 2015 ; Robichaud, 2019) n'utilisent, à notre connaissance, aucun format de données liées ou de mécanismes d'enrichissement sémantique actuellement fonctionnel. De tels formats posent plusieurs défis en lien avec l'intégration des données liées dans des systèmes d'information géographique (Chiang *et al.*, 2020 ; Mai *et al.*, 2019), qui sont des applications couramment utilisées pour la visualisation de données géohistoriques.

De son côté, la sémantisation permet d'ouvrir de nouvelles perspectives de recherche en contribuant théoriquement et techniquement à la construction des connaissances en histoire (Beretta, 2022 ; Robichaud et Lampron, 2022). Alors que les informations géographiques à elles seules ne suffisent pas à générer du sens historique, d'autres informations doivent accompagner les formes et les toponymes (Grossner *et al.*, 2022). Dans cette perspective, ce projet entend mobiliser la notion de transformation territoriale sous l'angle des dynamiques historiques sous-jacentes à l'évolution du territoire. Cette notion sera utilisée comme socle épistémologique de la construction d'un modèle définissant les relations entre temps, espace et documents.

## But de la recherche

Le but de ce projet de recherche consiste à **évaluer** l'apport potentiel des modèles ontologiques du Web sémantique pour la représentation des informations géohistoriques de documents d'archives cartographiques, dans le contexte de l'écosystème Web des données géohistoriques propre à Montréal. Il s'agit d'**expliquer** les possibilités et les limites de ces technologies sous l'angle des dynamiques des transformations historiques du territoire de la ville, tel que celles-ci peuvent être représentées à travers une approche diachronique des cartes historiques.

## Objectifs spécifiques

1. **Formaliser** les transformations du territoire montréalais en constituant un jeu de données ouvertes et liées basé sur des graphes de connaissances spatiotemporelles.
2. **Comparer** le jeu de données constitué avec les structures en usage sur le terrain afin d'évaluer les possibilités et les pistes de réconciliation et d'intégration des données.
3. **Formuler** des recommandations sur la direction que pourrait prendre le développement des jeux de données ouvertes et liées pour l'intégration des informations géohistoriques.

## Une approche diachronique structurée par le modèle RDF

Le cœur opérationnel du projet réside dans l'approche envisagée pour l'analyse des documents cartographiques. Cette analyse vise à extraire les informations pertinentes de ces documents et à créer un jeu de données grâce aux outils de modélisation du Web sémantique, dont le cadre de description RDF constitue la pierre d'assise (Cyganiak *et al.*, 2014). Basée sur une approche diachronique, l'analyse voudra mettre en relation la dimension temporelle du corpus avec les dimensions spatiale et documentaire. Le critère fonctionnel général à rencontrer en lien ce jeu de données est de permettre une représentation des transformations territoriales ayant mené aux grandes lignes de la morphologie urbaine de Montréal.

## Méthodologie

- **Devis explicatif** : cette recherche est portée par un devis explicatif pour comprendre, décrire et expliquer le fonctionnement de représentations sémantiques, de l'analyse des sources à la publication de données.
- **Échantillonnage** :
  - ▶ **Constitution d'un corpus** de cartes historiques basée sur un échantillonnage théorique : période historique **1840-1980**, des chemins de fer aux décennies de la rénovation urbaine (Teaford, 1990).
- **Collecte de données** :
  - ▶ **Extraction des informations** : approche diachronique des documents soutenue par une grille d'observation structurée et un protocole de documentation.
  - ▶ **Observation** de l'ensemble des systèmes Web de l'environnement des données géohistoriques à Montréal, incluant des systèmes d'information géographique et des entrepôts de données.
- **Modélisation** : processus de modélisation sémantique basé sur les outils du Web sémantique et structuré par le protocole *Simplified agile methodology for ontology development* (SAMOD) (Peroni, 2017).
- **Analyse des données** : visualisations basées sur des graphes de connaissances, sur le protocole SPARQL et sur l'intégration des données dans des systèmes d'information géographique pour le Web.

## Qualité de la recherche

- **Modélisation** : systématisme du processus et cohérence du modèle garantis par les tests du protocole SAMOD.
- **Données** : les données ouvertes liées sont compatibles avec les principes FAIR pour la réutilisation des données de la recherche (Koster et Woutersen-Windhouver, 2018).
- **Audit** : sollicitation des communautés d'intérêts pour un audit qui déterminera les suites à donner au projet.

## Importance et retombée

- Description et encodage des transformations territoriales de Montréal à l'aide d'un modèle offrant de nouvelles perspectives sur le phénomène historique observé à travers le patrimoine documentaire.
- Amélioration de notre compréhension des modèles de données liés pour l'histoire montréalaise grâce à une proposition applicative d'enrichissement sémantique autour de l'écosystème des données géohistoriques.
- Mise en valeur des documents et amélioration des conditions de consultation par la création de liens formels entre le document d'archives et les éléments constitutifs des transformations territoriales.

## Références

- Beretta, F. (2022). Interopérabilité des données de la recherche et ontologies fondationnelles : un écosystème d'extensions du CIDOC CRM pour les sciences humaines et sociales. Dans N. Lasolle, O. Bruneau et J. Lieber (dir.), *Actes des journées humanités numériques et Web sémantique* (p. 2-22). <https://doi.org/10.5281/ZENODO.7014341>
- Chiang, Y.-Y., Duan, W., Leyk, S., Uhl, J. H. et Knoblock, C. A. (2020). *Using historical maps in scientific studies: applications, challenges, and best practices*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-66908-3>
- Cyганиак, R., Wood, D. et Lanthaler, M. (2014, 25 février). *RDF 1.1 Concepts and abstract syntax*. World Wide Web Consortium. <https://www.w3.org/TR/rdf11-concepts/>
- Grossner, K., Grunewald, S. et Mostern, R. (2022). Bringing places from the distant past to the present: A report on the World Historical Gazetteer. *International Journal on Digital Libraries*. <https://doi.org/10.1007/s00799-022-00341-2>
- Koster, L. et Woutersen-Windhouver, S. (2018). FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections. *The Code4Lib Journal*, 40. <https://journal.code4lib.org/articles/13427>
- Mai, G., Janowicz, K., Yan, B. et Scheider, S. (2019). Deeply integrating linked data with geographic information systems. *Transactions in GIS*, 23(3), 579-600. <https://doi.org/10.1111/tgis.12538>
- Olson, S., Sweeny, R. C. H. et Centre interuniversitaire d'études québécoises. (2015). *Montréal, l'avenir du passé*. Centre interuniversitaire d'études québécoises. <https://map.cieq.ca>
- Peroni, S. (2017). A simplified agile methodology for ontology development. Dans M. Dragoni, M. Poveda-Villalón et E. Jimenez-Ruiz (dir.), *OWL : Experiences and directions – Reasoner evaluation. 13th international workshop, OWLED 2016, and 5th international workshop, ORE 2016* (p. 55-69). Springer International Publishing. [https://doi.org/10.1007/978-3-319-54627-8\\_5](https://doi.org/10.1007/978-3-319-54627-8_5)
- Robichaud, L. (2019). Prendre le tournant spatial en histoire : le Laboratoire d'histoire et de patrimoine de Montréal et le Système de cartographie de l'histoire de Montréal (SCHEMA). Dans J. Burgess et P.-A. Linteau (dir.), *Histoire et patrimoine. Pistes de recherche et de mise en valeur* (p. 213-236). Presses de l'Université Laval. <https://www.pulaval.com/libre-acces/9782763743295/>
- Robichaud, L. et Lampron, C. (2022, 3 juin). *La description sémantique des lieux. Réflexions sur la mise en relation des références spatiales extraites d'un document d'archive* [Conférence]. Journée annuelle du Laboratoire d'histoire et de patrimoine de Montréal, Montréal. <https://lhpm.uqam.ca/production-scientifique/communications/>
- Teaford, J. C. (1990). *The rough road to renaissance: Urban revitalization in America, 1940-1985*. Johns Hopkins University Press.