

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **21 janvier 2025**

Nom de famille et prénom de l'auteur. e : **Madame Sarra OUELHADJ**

Titre de la thèse : Définition d'un processus durable pour l'enrichissement sémantique des open data territoriales

Résumé



Avec l'essor des technologies de l'information et le mouvement Open Data, l'accessibilité aux données publiques a fortement progressé, encouragée par des initiatives visant à accroître la transparence et l'innovation. Cependant, malgré ces avancées, les défis d'interopérabilité des données demeurent, notamment en raison de l'hétérogénéité des données. La Métropole de Lyon, engagée dans une politique d'ouverture des données, a reconnu ces enjeux, mais les obstacles liés à l'hétérogénéité sémantique des données limitent la capacité à intégrer et exploiter pleinement les données territoriales disponibles. Face aux défis de l'hétérogénéité et de l'interopérabilité des données, cette thèse CIFRE propose d'explorer des solutions pour l'enrichissement sémantique des données territoriales, afin d'améliorer leur interopérabilité et intégration. L'objectif est de rendre ces données compréhensibles et exploitables, aussi bien par des utilisateurs humains que par des machines. Dans cette perspective, les travaux s'appuient sur les principes du Web Sémantique, et introduisent des méthodes pour accompagner les producteurs de données, non-experts en Web Sémantique, dans la construction de Graphes de Connaissances. Les contributions de cette thèse incluent la conception d'un Modèle Sémantique, nommé UML annoté, destiné à l'enrichissement sémantique des données. Ce modèle sert de base à deux méthodes proposées : la méthode GES et la boîte à outils sETL. Celles-ci facilitent l'enrichissement sémantique des données en les convertissant en RDF, et en intégrant des vocabulaires partagés, sans nécessiter une expertise technique approfondie en Web Sémantique. Plus particulièrement, la méthode GES engage les producteurs de données dans la modélisation sémantique de leurs données, la proposition d'une documentation embarquée avec les données, et l'enrichissement des données avec des référentiels externes. La boîte à outils sETL exploite les schémas de données existants, fruit d'une élicitation partielle de la sémantique des données déjà conduite par les producteurs de données de la Métropole de Lyon. Ces méthodes visent à renforcer la gouvernance des données ouvertes et à faciliter leur intégration dans des initiatives stratégiques et innovantes pour les collectivités territoriales.

Mots-clés : Enrichissement sémantique, Graphe de connaissances, UML annoté, ETL, Schéma de données, Données ouvertes,