

Titre de la thèse :

CONCEPTION ET ÉVALUATION D'UN ENTREPÔT DE DONNÉES HÉTÉROGÈNES POUR LE SUIVI DU RÔLE DE L'AGRICULTURE SUR LE TERRITOIRE DU LANGUEDOC-ROUSSILLON

Unité de recherche : TETIS - Télédétection Environnement Télédétection et Information Spatiale

Directeur de la thèse Flavie CERNESSON, Co-Directeur Maguelonne TEISSEIRE, Co-Encadrant Lucile SAUTOT

Concours GAIA - Début de la thèse le 1 octobre 2018 - Date limite de candidature 16 mai 2018

Établissement : Institut des sciences et industries du vivant et de l'environnement (AgroParisTech) - École doctorale GAIA - Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau - Spécialité STE - Sciences de la Terre et de l'Eau

Description de la problématique de recherche

Le diagnostic et le suivi en ingénierie territoriale s'appuient sur de nombreuses données hétérogènes issues de sources variées : images satellites, cartes d'artificialisation des sols, modèles numériques de terrain, données agricoles, données environnementales, documents de planification et réglementaires etc. L'hétérogénéité des données est une notion qui intègre plusieurs composantes : 1) la nature des données (tableaux de mesures qualitatives, tableaux de mesures quantitatives, tableaux de mesures mixtes, images, textes, graphes) ; 2) les échelles spatiotemporelles auxquelles peuvent être récoltées les données ; 3) les thématiques auxquelles se rattachent les données. Ces différentes composantes sont indissociables et complexifient la mise en relation de données dans les systèmes d'information. Les outils SIG classiquement utilisés pour l'exploration de données spatiales permettent ainsi de faire le lien entre une thématique et objets spatiaux au travers d'une couche de données spatiales associée à une table attributaire. Les SIG permettent d'agréger les informations sur une couche, et les opérations d'analyse sont menées pas à pas en manipulant une ou deux couches à la fois. Les entrepôts de données sont l'un des types phares de plate-formes numériques permettant le stockage, le croisement, l'agrégation et l'analyse de gros volumes de données. Cependant, la littérature scientifique montre que cette question est loin d'être réglée, que ce soit du point de vue théorique ou du point de vue opérationnel. En effet, la conception des entrepôts reste rigide en considérant les « faits » devant être quantitatifs et les « dimensions » devant être des catégories organisées en hiérarchie. « Faits » et « dimensions » dépendent donc de la nature des données. De nombreux travaux existent sur les composantes « échelles spatio-temporelles » et « thématiques ». La thèse comporte donc un premier verrou qui consiste, tout en s'appuyant sur les savoirs existant, en la proposition d'un nouveau paradigme de modélisation des données, afin de pouvoir réaliser des agrégations et des analyses sur des données de natures différentes. Un second verrou porte sur l'évaluation des performances de ce type d'outil, qui, ne peut passer qu'au travers de cas concrets impliquant des utilisateurs pouvant être non-informaticiens. Ce sont à ces verrous que les travaux proposés dans cette thèse souhaitent répondre en (1) définissant un modèle d'entrepôt de données permettant la gestion et le croisement de données hétérogènes et (2) offrant les outils d'analyse de ces nouveaux types d'information au travers d'un cas concret mettant en avant le lien agriculture/territoire en Languedoc-Roussillon.

Lien sur le sujet détaillé et comment candidater :

https://www.adum.fr/as/ed/voirproposition.pl?langue=fr&site=gaia&matricule_prop=20110