

## Stage M2 informatique ou 5A école d'ingénieur (5 ou 6 mois) : Approche sémantique pour reconstituer l'évolution de données territoriales historiques

poursuite en thèse possible

**Mots clés – compétences souhaitées :** web sémantique, humanités numériques, systèmes d'information géographiques, intégration et raisonnement spatial et temporel sur des données hétérogènes

**Problématique :** Dans le cadre des humanités numériques, les systèmes d'information permettent de représenter des données ayant des coordonnées spatiales et, le plus souvent, datées. Pour les informations géographiques, la technologie associée est généralement celle des bases de données couplées à des fonctionnalités spécifiques pour les manipuler, les comparer ou calculer. Ce type d'approches est utilisé par exemple pour représenter des découpages territoriaux liés à des activités humaines (administratifs, juridiques, fiscaux, ...). L'intégration, la comparaison de plusieurs découpages de ce type sur un même territoire, et l'étude de leur évolution dans le temps, suppose de recouper des données aux formats et contenus hétérogènes. Dans ce contexte, une approche sémantique apporte une solution à l'hétérogénéité et facilite l'interopérabilité : elle consiste à définir un vocabulaire unique permettant de décrire les contenus des sources de données, et à produire des graphes de connaissances dont les nœuds sont les entités géographiques et les entités associées, et dont les arcs expriment la nature des relations entre entités ou valeurs. Cependant, la mise en relation de découpages historiques se heurte à plusieurs problèmes : l'évolution de chacune des unités de territoire ; la différence des dates d'évolution selon les types d'unités territoriales ; l'évolution des types d'unités, de la nature des activités humaines prises en compte ou de la connaissance associée. La mise en correspondance et l'étude de ces données supposent donc des raisonnements sur le temps et l'espace, mais aussi des connaissances historiques ou territoriales.

**Objectif du stage :** Le stage a pour but de répondre à l'intégration de données territoriales historiques dans le cadre du projet OBARDI. Il s'agira d'abord d'étudier les différents schémas ou les structures des sources de données à considérer, et de reprendre l'état de l'art sur la représentation sémantique de données géographiques territoriales (ontologies existantes), pour définir un vocabulaire facilitant l'intégration de données de différentes sources. Une deuxième étape consistera à définir des patrons de génération de graphes RDF conformes à ce vocabulaire à partir des différentes sources. Enfin, une 3<sup>e</sup> étape visera la mise en relation spatiale et temporelle des éléments de territoire provenant des différentes sources à l'aide de raisonnement automatique et afin de tracer leur évolution.

**Contexte :** Le stage se déroulera dans le cadre du projet de recherche ANR ObARDI (janvier 2021–décembre 2024) de l'axe «Révolution numérique : rapports au savoir et à la culture», dont l'un des objectifs est de constituer un ensemble de bases de données structurées sur les dynamiques institutionnelles et sociales de la France de l'Ancien Régime (XVII<sup>e</sup>–XVIII<sup>e</sup> siècles).<sup>1</sup> Ce projet est coordonné par Victor Gay, enseignant-chercheur en histoire économique à l'Université Toulouse 1 Capitole, en collaboration avec les laboratoires en histoire TEMPORA, ainsi que le laboratoire en informatique de l'IRIT pour ce qui concerne le contenu du stage proposé. Ce projet prévoit un financement de thèse sur la thématique abordée dans le cadre du stage.

**Encadrement :** Nathalie Aussenac-Gilles (aussenac@irit.fr), Nathalie Hernandez (hernande@irit.fr)

**Dates :** d'avril à septembre 2021 (5 ou 6 mois)

**Lieu du stage :** IRIT - Université de Toulouse

**Rémunération :** prime de stage mensuelle 600€

**Candidature :** envoyer un CV et une lettre de motivation aux encadrantes.