



Import des données massives, quel impact sur les données et la communauté d'OpenStreetMap ?

Stage de recherche proposé à l'IGN, Laboratoire LaSTIG, Equipe MEIG

Contexte

OpenStreetMap (OSM) est de loin l'initiative collaborative connaissant le plus de succès parmi les initiatives collaboratives qui ont vu le jour depuis une quinzaine d'années. Sa force réside dans son organisation horizontale, la souplesse du modèle et la motivation des contributeurs pour éditer et améliorer la qualité de l'information géographique. Sa notoriété est due également à la communauté internationale de chercheurs qui a fait de l'écosystème défini autour d'OSM un objet de recherche pour étudier les aspects sociaux (*e.g.* motivation, pratiques des contributeurs), de qualité (*e.g.* comparer les données OSM avec des données institutionnelles) et d'usages (*e.g.* utiliser OSM pour répondre à des besoins sociétaux ou industriel).

Avec la multitude de données géographiques ouvertes disponibles, l'import massif des données devient une pratique courante dans OpenStreetMap (*e.g.* imports de stations de métro de la RATP, de Corinne Land Cover, ou de données cadastrales). Le risque est bien évidemment qu'OSM devient un agrégateur de données et que le contributeur soit démotivé pour participer. De plus, OSM est considérée comme une initiative volontaire et les données issues de la plateforme OSM sont nommées dans la littérature scientifique des données volontaires. Or, étant donné cette pratique la question sur l'aspect volontaire de données se pose.

Sujet

Dans ce contexte, l'objectif de ce stage est d'étudier la pratique des imports massifs et l'évolution des données issues de cette pratique.

Dans un premier temps il s'agit d'identifier les types de données qui ont été importés massivement dans OSM en France et d'analyser la manière dont ces imports sont effectués. Dans un deuxième, il s'agit d'analyser l'évolution des données issues des imports massifs et d'identifier le type d'évolution (*e.g.* changement, amélioration de la qualité, cohérence avec les données OSM). L'étudiant proposera une typologie des évolutions et des méthodes qui permettront leur identification. Enfin, une réflexion sera menée concernant l'impact de l'import massif sur la motivation des contributeurs OSM et sur le concept volontaire attribués historiquement aux données OSM. Pour répondre à cette question, l'étudiant proposera une méthode qui permettra de modéliser et d'identifier des profils de contributeurs dont la motivation est de modifier les données issues des imports massifs, ou au contraire ceux dont la motivation a diminué à cause des imports massifs.

Les tâches suivantes ont été identifiées :

- Familiarisation avec les données de la plateforme OSM, le wiki, OSM et le forum.
- Identifier les sources importées dans la plateforme OSM
- Proposer et développer une méthode permettant d'analyser l'évolution des données massivement importées dans OSM
- Définir une typologie des évolutions

- Proposer une typologie des contributeurs en fonction de leur participation aux modifications des données issues des imports massifs
- Rédaction du rapport

Profil recherché

Master 2 en sciences de l'information géographique (SIG) ou en informatique avec un gout pour l'information géographique. Des compétences en programmation (java, R ou python) et en base de données géographiques sont nécessaires. Des connaissances sur la modélisation des connaissances est un plus.

Durée, rémunération et accueil

Le stage d'une durée de 5 mois sera réalisé au sein de l'équipe MEIG du Laboratoire LaSTIG de l'IGN (73 Avenue de Paris, Saint-Mandé (94) – Métro 1).

Le début de stage est prévu à partir de février 2021. Le stage est gratifié selon la législation française.

Contact et encadrement :

Si vous êtes intéressé(e), veuillez envoyer votre candidature (CV, lettre de motivation, relevés de notes –si disponibles) à : Ana-Maria Raimond (ana-maria.raimond@ign.fr) et Arnaud le Guilcher (Arnaud.Le-Guilcher@ign.fr).