



ANNEXE 1

Entreprises, pour proposer un sujet de thèse soutenue par le dispositif CIFRE, merci de remplir les champs suivants, et d'envoyer le document à : cifre@anrt.asso.fr

Si vous souhaitez ajouter un descriptif plus détaillé de l'offre à votre annonce, merci de le joindre accompagné de ce formulaire.

- **Nom de l'entreprise*** : Métropole de Lyon
- **Ville et code postal*** : Lyon, 69003
- **Nom du laboratoire académique partenaire (si déjà connu)** : LIRIS UMR 5205
- **Numéro de reconnaissance du laboratoire** : UMR 5205, École Doctorale InfoMaths n° 512
- **Thématique de recherche en une phrase (sans aucun caractère confidentiel) *** :

Définir un processus durable pour l'enrichissement sémantique des Open Data territoriales

- **Descriptif de la thématique de recherche (sans aucun caractère confidentiel) *** :

Au même titre que beaucoup d'autres collectivités territoriales, la Métropole de Lyon publie en Open Data plusieurs jeux de données qui sont produits soit par la Métropole de Lyon elle-même soit par ses partenaires. Comme c'est souvent le cas, ces jeux de données sont disponibles dans des formats relativement standards (e.g. CSV, JSON, GeoJSON, ...), assurant un bon niveau d'interopérabilité syntaxique. Cependant, les aspects sémantiques sont traités de façon assez superficielle. La signification des données est souvent transmise "hors bande", via de la documentation textuelle dans les métadonnées, ou bien via des communications privées entre les consommateurs de données et le support utilisateurs de la plateforme data.grandlyon.com. De plus, les consommateurs s'appuient parfois sur leurs propres interprétations, qui certainement peuvent être erronées. De ce fait, une bonne partie de la valeur potentielle des Open Data publiées est difficile à exploiter:

- la sémantique de chaque jeu de donnée doit être encodée de façon spécifique dans chaque application cliente,
- ce qui rend ces applications enclines à tomber en panne suite à des évolutions des données sous-jacentes,
- et qui empêche les consommateurs de développer aisément des applications à forte valeur ajoutée mélangeant différents jeux de données.

La notion de "Linked Data" dénote un ensemble de technologies et standards recommandés par le W3C, visant à résoudre les problèmes évoqués ci-dessus. Elle s'appuie sur un modèle de données orienté graphe (RDF) qui facilite l'intégration de données, et elle encourage l'utilisation de vocabulaires partagés munis d'une sémantique explicite. L'effort supplémentaire demandé aux producteurs de données pour se conformer aux principes des Linked Data est bénéfique à tous les consommateurs, car il rend chaque jeu de données plus simple à utiliser et à mettre en relation avec d'autres, ce qui est encore plus important.

La Métropole de Lyon compte rendre ses Open Data progressivement conformes aux principes des Linked Data. Cet objectif à long terme fournit le contexte général de cette offre de thèse.

- **Descriptif du poste*** :

Comme première mission, le/la doctorant(e) devra passer en revue les outils et vocabulaires existants, afin de repérer ceux qui sont les plus adaptés pour les Open Data publiées par la Métropole de Lyon. Potentiellement, cette tâche conduira à étendre certaines des solutions existantes.

Dans un deuxième temps, le/la doctorant(e) devrait s'assurer que le workflow de migration soit viable et maintenable dans la durée. En effet, l'intégration des aspects sémantiques est un processus continu, sans fin :

- lors de la définition du mapping entre les données existantes et les vocabulaires ciblés, des interprétations erronées (de données et de vocabulaires) peuvent facilement intervenir. Lorsqu'elles sont détectées, ces dernières nécessitent de mettre à jour le mapping.

- Les jeux de données sont mis à jour avec différentes fréquences, allant de la fréquence annuelle au temps-réel. Bien évidemment, la sémantique doit suivre chaque évolution du workflow de production/mise à jour de données, ce qui peut nécessiter également une mise à jour du mapping.

- La manière dont les données sont utilisées en pratique peut révéler des subtilités sémantiques qui n'avaient pas été prise en compte au départ, et peuvent encore nécessiter une mise à jour du mapping.

Par conséquent, le processus de migration doit être robuste vis-à-vis des changements dans les données d'origine. De plus, il doit permettre aux équipes de la Métropole de Lyon d'identifier les évolutions à apporter au processus et la meilleure façon de les implémenter.

• **Date de recrutement*** : septembre 2019

• **Adresse e-mail à laquelle le candidat doit envoyer sa candidature*** :

acerioni@grandlyon.com

pierre-antoine.champin@liris.cnrs.fr

*champs obligatoires