Offre de bourse de thèse pour le recrutement d'un doctorant sur le sujet :
"Négociation avancée de contrat d'accès aux données dans les espaces de données"
FAYOL/ISI

PRESENTATION DE MINES SAINT-ETIENNE

« École d'ingénieur.e.s responsable, moteur d'innovations à impact sociétal » traduit l'engagement de nos enseignants-chercheurs et de notre personnel administratif et technique pour relever les défis des grandes transitions du XXIe siècle. Forts d'une histoire de plus de 200 ans, de l'excellence de nos personnels et de nos étudiant.e.s, nous assurons des missions de formation, de recherche, d'innovation, de transfert vers l'industrie et de culture scientifique, technique et industrielle. Avec 2 500 élèves, 500 personnels, et un budget de 50 M€, nous rayonnons sur 3 campus dédiés à l'industrie des futurs, à la santé et au bien-être et à la souveraineté numérique et microélectronique, situés dans 3 métropoles majeures : Saint-Etienne, Lyon et Aix-Marseille-Provence. Classée par le magazine l'Etudiant dans le TOP 10 national et présente dans les classements internationaux, Mines Saint-Etienne est membre du réseau T.I.M.E. des meilleures « Technological Universities » mondiales et, par son appartenance à l'Institut Mines-Telecom, membre de l'Université Européenne EULIST.
Rejoindre aujourd'hui Mines Saint-Etienne, c'est faire le choix de contribuer à son impact sur la société et sur les défis des grandes transitions du XXIe siècle, dans un environnement stimulant, international et à taille humaine, au service de nos étudiants, enseignants, chercheurs, partenaires industriels et universitaires. C'est aussi le choix de rejoindre une équipe qui vous fera progresser dans vos compétences et vous donnera envie de continuer avec nous.

UNE ECOLE DE L'INSTITUT MINES-TELECOM :

L'Institut Mines-Télécom est le 1er groupe public de Grandes Écoles d'ingénieur.e.s et de management de France. Constitué de huit Grandes Écoles publiques et de deux écoles filiales, l'Institut Mines-Télécom anime et développe un riche écosystème d'écoles partenaires, de partenaires économiques, académiques et institutionnels, acteurs de la formation, de la recherche et du développement économique.
La stratégie 2023-2027 de Mines Saint-Etienne s'inscrit dans celle de l'Institut Mines Telecom. Elle a pour ambition d'accompagner les transitions écologique, numérique, et générationnelle et d'en former les acteurs et de soutenir la souveraineté nationale et européenne en microélectronique et numérique.
Pour mettre en œuvre cette stratégie, elle recrute une ou un doctorant en informatique

ENVIRONNEMENT DU POSTE ET DESCRIPTION DES MISSIONS :

1. Intitulé et domaines de la thèse

En français : **Négociation avancée de contrat d'accès aux données dans les espaces de données**

En anglais : **Advanced data access contract negotiation in data spaces**

- Domaine disciplinaire de rattachement et spécialité : Informatique, Gestion des données
- Mots clés : web sémantique, représentation des connaissances, espaces de données, open digital rights language, profile de données

2. Structure(s) et/ou laboratoire(s) d'accueil

Mines Saint-Étienne, Institut Henri Fayol, 29 rue Pierre et Dominique Ponchardier, 42000 Saint-Étienne

3. Inscription en thèse

- Établissement d'inscription en thèse : Ecole nationale supérieure des mines de Saint-Etienne

- École doctorale de rattachement : ED 488 : SCIENCES INGÉNIERIE SANTÉ (SIS)
- Laboratoire d'accueil : LIMOS UMR 6158

4. Encadrement

Antoine Zimmermann, Professeur IMT à Mines Saint-Étienne,
Maxime Lefrançois, Maître Assistant IMT à Mines Saint-Étienne.

5. Contexte et objectifs

La stratégie européenne en matière de données vise à renforcer la souveraineté et la compétitivité de l'UE en facilitant l'échange et l'exploitation des données au sein d'espaces de données (Curry et al., 2022). Ces espaces consistent en des infrastructures distribuées définies par un cadre de gouvernance, qui permettent des transactions de données fiables entre les participants tout en prenant en charge la confiance et la souveraineté des données (Data Spaces Support Centre, 2023). L'enjeu est ainsi de créer un écosystème mondial où les flux de données sont aussi fluides et contrôlés que les chaînes d'approvisionnement traditionnelles, impliquant des défis en matière de partage fiable, de contrôle d'utilisation et de modèles commerciaux (Bacco et al., 2024).

Les travaux de Theissen-Lipp (2024) mettent en lumière les fondations sémantiques des espaces de données. Différentes frameworks émergent – Gaia-X, IDS, EOSC, Solid Data Spaces – tous sont basés sur les standards du web, et emploient les technologies du web sémantique pour faciliter la gestion des données dans le respect des principes FAIR (Hauff et al., 2024; Meckler et al., 2023). Par exemple le vocabulaire IDS-IM (Bader et al., 2020) reprend et étend DCAT pour la description du catalogue des données (Albertoni et al., 2024), DQV pour leur qualité (Debattista et al., 2016), ou Open Digital Rights Language (ODRL) pour les conditions d'accès aux données (Iannella and Villata, 2018). La structure des données pourrait être décrite plus finement avec des formalismes comme JSON-Schema (Schema, 2022) ou SHACL (Knublauch and Kontokostas, 2017), potentiellement générées automatiquement via des frameworks comme le Semantic Data Link qui emploie LinkML (Stäbler et al., 2024).

Dans un espace de données, un contrat d'accès est négocié entre un fournisseur de données, et un consommateur de données. Une direction de recherche inexplorée à ce jour consiste à imaginer la variété du type de négociation possible, en particulier, des processus flexibles de négociation des données où le format ou la sémantique de la donnée elle-même peut être négociée. On peut imaginer par exemple qu'un client n'ait besoin que d'un sous-ensemble d'un jeu de données offert par un fournisseur. Il pourrait offrir un tarif moins élevé sur le principe qu'il n'accède pas à toute la donnée, cependant le fournisseur devra effectuer un traitement supplémentaire sur la donnée, qui pourrait justifier un tarif plus élevé.

Le concept de négociation de contenu sur le web décrit des mécanismes qui permettent de servir différentes versions d'un document à la même URI (Svensson et al., 2024). Dans sa thèse, Taghzouti (2024) a défendu l'intérêt d'une négociation de contenu fine et souple en tenant compte du fait que la négociation de contenu n'est pas un processus monolithique, mais plutôt composé de plusieurs étapes. Appliqué aux contrats d'échange de données dans les espaces de données, la négociation de contenu peut servir de point de départ pour certains mécanismes avancés qui permettent d'adapter les échanges en fonction des spécificités des parties prenantes, de leurs préférences et de leurs capacités technologiques. En particulier, on s'intéressera aux approches de négociation de contenu par profil (Taghzouti, 2023).

L'objectif de cette thèse est de proposer une formalisation et un cadre méthodologique pour des mécanismes avancés de négociation de contrat d'accès aux données au sein d'un espace de données. L'hypothèse centrale est que l'enrichissement des descriptions des jeux de données et des offres et requêtes de contrat permettrait d'explorer de nouveaux scénarios de négociation, adaptés aux besoins spécifiques des acteurs impliqués. L'ambition est que les propositions s'appuient sur les standards ouverts du Web, afin qu'elles puissent être intégrées aux cadres émergents des espaces de données (Gaia-X, IDS, EOSC, Solid Data Spaces).

6. Défis scientifiques

La négociation avancée de contrats d'accès aux données dans les espaces de données soulève plusieurs défis scientifiques, à la croisée de l'interopérabilité des données, de la modélisation sémantique et des mécanismes de gouvernance des échanges. Les travaux pourront être structurés autour des deux axes suivants :

A1. **Modélisation sémantique des jeux de données, des offres et des requêtes 'accès.** La flexibilité dans la négociation de contrat d'accès aux données nécessite une formalisation précise des métadonnées décrivant les jeux de données, y compris des informations concernant leur structure (ex. JSON Schema) et sémantique (ex. vocabulaire, shape SHACL). Ces informations devraient pouvoir être utilisées dans des descriptions des offres et requêtes d'accès exprimées en ODRL (ex. pour définir que les données doivent satisfaire une contrainte de structure ou de sémantique).

A2. **Mécanismes avancés de négociation de contrat.** Pour qu'un accord puisse être satisfaisant pour les deux parties, des mécanismes de calcul dynamique de compromis doivent être définis sur la base des offres et des requêtes de contrat. Des pistes de contributions résident dans l'adaptation dynamique de la structure et de la sémantique des jeux de données, en parallèle de l'adaptation des permissions, interdictions, et devoirs associés (y compris le mode de calcul de la compensation financière).

7. Caractère innovant

Cette thèse propose une approche novatrice en explorant des mécanismes avancés de négociation des jeux de données dans les espaces de données, où un compromis peut résulter d'une séquence d'échanges d'offres et de contre-offres. A notre connaissance, aucun travail passé ou en cours ne considère que la structure et la sémantique des données elles-mêmes peut être l'objet de négociations, et être adaptée en fonction des préférences et capacités des parties prenantes. En mobilisant les technologies du web sémantique et les mécanismes d'échange sur le web, cette recherche ouvre la voie à de nouveaux cas d'usage de partage de la donnée, tout en garantissant interopérabilité et souveraineté des acteurs, et favorisant ainsi la création de valeur dans les espaces de données.

8. Résultats attendus et valorisation

Les travaux de cette thèse devraient aboutir à :

- Une formalisation de la négociation de contrat d'accès aux données, intégrant différents types d'information concernant les données, y compris leur structure et sémantique, ainsi que les permissions, interdictions, et devoirs associés
- Un cadre méthodologique et un prototype de mécanisme avancé de négociation de contrat, exploitant ces informations pour calculer dynamiquement des compromis
- Des scénarios d'évaluation concrets, testant la faisabilité et l'efficacité du cadre proposé dans des environnements synthétiques ou réels d'échange de données, en lien avec des initiatives comme IDS.
- Des contributions aux standards et communautés académiques et industrielles, par la publication de résultats dans des conférences et revues en web sémantique ou graphes de connaissances distribués.
- Une valorisation technologique via la publication de code open source, ouvrant la voie à une intégration de la négociation sémantique dans les futures architectures des espaces de données.

9. Impact

Cette thèse aura un impact scientifique en avançant l'état de l'art sur la négociation de contrats d'accès aux données. Sur le plan technologique, elle apportera des pistes d'amélioration des cadres émergents des espaces de données. En outre, son impact économique résidera dans le fait de rendre possible de nouveaux modèles commerciaux basés sur une personnalisation fine des transactions, ouvrant la voie à la création de davantage de valeur dans les espaces de données. Ce faisant, cette thèse contribuera notamment aux objectifs de développement durable n°9 (9.1, 9.2, …)

10. Partenariats éventuels

Possibilité de coopération sur différents projets de recherche sur des thématiques connexes, tout en avançant de manière indépendante (pas d'implication dans les livrables). Exemple : projet LoireConnecTID, projet EU OMEGA-X avec EDF R&D. Implication dans des associations comme IDSA, AIOTI, BDVA, ou organismes de normalisation (W3C, ETSI TC SmartM2M, bientôt TC Data, AFNOR).

11. Références bibliographiques

Albertoni, R., Browning, D., Cox, S.J.D., Beltran, A.G., Perego, A., Winstanley, P., 2024. Data catalog vocabulary (DCAT) (W3C Recommendation). World Wide Web Consortium.

Bacco, M., Kocian, A., Chessa, S., Crivello, A., Barsocchi, P., 2024. What are data spaces? Systematic survey and future outlook. Data in Brief 57, 110969.

Bader, S., Pullmann, J., Mader, C., Tramp, S., Quix, C., Müller, A.W., Akyürek, H., Böckmann, M., Imbusch, B.T., Lipp, J., Geisler, S., Lange, C., 2020. The International Data Spaces Information Model – An Ontology for Sovereign Exchange of Digital Content, in: The Semantic Web – ISWC 2020, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 176–192. https://doi.org/10.1007/978-3-030-62466-8_12

Curry, E., Scerri, S., Tuikka, T., 2022. Data spaces. Springer.

Data Spaces Support Centre, 2023. 1st collaborative discussion paper Why data spaces A business and user's perspective.

Debattista, J., Dekkers, M., Guéret, C., Lee, D., Mihindukulasooriya, N., Zaveri, A., 2016. Data on the Web Best Practices: Data Quality Vocabulary (W3C Working Group Note). World Wide Web Consortium.

Hauff, M., Comet, L.M., Moosmann, P., Lange, C., Chrysakis, I., Theissen-Lipp, J., 2024. FAIRness in Dataspaces: The Role of Semantics for Data Management, in: The Second International Workshop on Semantics in Dataspaces, Co-Located with the Extended Semantic Web Conference.

Iannella, R., Villata, S., 2018. ODRL Information Model 2.2 (W3C Recommendation). W3C.

Knublauch, H., Kontokostas, D., 2017. Shapes Constraint Language (SHACL) [WWW Document]. URL https://www.w3.org/TR/shacl/ (accessed 3.10.25).

Meckler, S., Dorsch, R., Henselmann, D., Harth, A., 2023. The Web and Linked Data as a Solid Foundation for Dataspaces, in: Companion Proceedings of the ACM Web Conference 2023. Presented at the WWW '23: The ACM Web Conference 2023, ACM, Austin TX USA, pp. 1440–1446. https://doi.org/10.1145/3543873.3587616

Schema, J., 2022. Json schema. JSON Schema. Online: https://json-schema. org/(abgerufen am: 22.5. 2023).

Stäbler, M., Moosmann, P., Dittmer, P., Wang, D., Köster, F., Lange, C., 2024. Semantic Data Link: Bridging Domain-Specific Needs with Universal and Interoperable Semantic Models. 10.1007/978-3-031-60626-7.

Svensson, L.G., Atkinson, R., Car, N.J., Taghzouti, Y., 2024. Content negotiation by profile (W3C Editor's Draft). World Wide Web Consortium.

Taghzouti, Y., 2024. Semantic content negotiation for knowledge exchange between heterogeneous systems (phdthesis). Université de Lyon.

Theissen-Lipp, J., 2024. Semantic foundations of dataspaces. RWTH Aachen University.

PROFIL RECHERCHÉ :

Le candidat idéal est titulaire d'un diplôme Bac + 5 ou équivalent, avec des bonnes connaissances en gestion des données, interopérabilité sémantique, et web sémantique. Proactif, autonome, organisé et rigoureux dans la gestion des projets, il possède des compétences analytiques solides pour concevoir des solutions adaptées à des problématiques complexes, un excellent esprit de synthèse, et des aptitudes en communication écrite et orale en français et en anglais. Il peut souhaiter démarrer un projet d'entreprenariat pour valoriser ses travaux de thèse, ou poursuivre dans le milieu académique.

POURQUOI NOUS REJOINDRE :

Mines Saint-Etienne se distingue par :

- Un environnement de travail privilégié avec un taux d'encadrement des étudiants et un taux d'environnement (fonctions support et soutien) élevé
- Des moyens expérimentaux et numériques de premier plan
- Une activité de recherche contractuelle importante (11 M€ / an de contrats en Recherche et Innovation), majoritairement avec des partenaires industriels
- 25% d'étudiants internationaux, Membre du réseau T.I.M.E. et de l'Université Européenne EULIST
- Un centre de Culture Scientifique Technique et Industrielle – La Rotonde - unique en France, qui démultiplie son impact sur la société (> 50 000 visiteurs par an)
- Classée par le Times Higher Education dans le TOP 100 Mondial et leader dans ses deux régions d'appartenance pour les Objectifs de Développement Durable « Lutte contre les changements Climatiques », « Villes et Communautés Durables », « Ressources en Eau »

- Des campus à taille humaine, en environnement urbain, accessibles en mobilités douces (vélo, marche, transports en commun …)
- Un cadre de travail agréable. Pour 83 % des répondants au baromètre social annuel de Mines Saint-Etienne, la Qualité de Vie au Travail est jugée bonne ou satisfaisante. Accessibilité en transport ou proche des axes routiers
- Des frais de transports en commun pris en charge à hauteur de 75% (sous conditions)
- Un forfait mobilité durable
- Un foyer du personnel (activités sportives, culturelles, avantages CE sur des loisirs et temps de convivialité)
- La possibilité de télétravail partiel
- 49 jours de congés et RTT

L'Institut Mines-Telecom est présent sur l'ensemble du territoire français et se caractérise par un environnement scientifique d'excellence : https://www.youtube.com/watch?v=m39m6hdNC48

INFORMATIONS COMPLÉMENTAIRES :

Conditions de recrutement :

- Nature du contrat CDD de droit public d'une durée de 36 mois
- Profil Cadre
- Prise de fonction souhaitée : 01/09/2025
- Rémunération sera fixée selon le profil du candidat, en fonction des règles définies par le cadre de gestion de l'Institut Mines Télécom
- Temps plein
- Poste basé à Saint-Étienne

Le poste est ouvert à toutes et tous avec, sur demande, des aménagements pour les candidates et candidats en situation de handicap.
Emploi ouvert aux titulaires de la fonction publique et/ou aux personnes contractuelles. Toute candidature peut faire l'objet d'une enquête administrative.

Modalités de candidatures :

Les dossiers de candidature (CV, lettre de motivation, lettre de recommandation le cas échéant, pièce d'identité) sont à déposer sur la plateforme RECRUITEE :
https://institutminestelecom.recruitee.com/o/doctorant-ou-doctorante-negociation-avancee-de-contrat-dacces-aux-donnees-dans-les-espaces-de-donnees-cdd-3-ans-2

- Date limite de candidature : 01/06/2025
- Processus de recrutement pour information (dates prévisionnelles des étapes suivantes)

Les candidats retenus après examen de leur dossier de candidature seront reçus en entretiens au fil de l'eau.

Dans le cadre de sa politique Égalité, Diversité et Inclusion, l'École des Mines de Saint Etienne est un employeur soucieux de l'équité de traitement entre les candidatures.

En savoir plus :

Pour tous renseignements sur le poste, s'adresser à :
Maxime LEFRANCOIS – Maitre Assistant et Antoine ZIMMERMANN – Professeur IMT
Mail : maxime.lefrancois@emse.fr antoine.zimmermann@emse.fr
Tel. : +33 4 77 42 66 40

Pour tous renseignements administratifs, s'adresser à :
Julie JAFFRE – Gestionnaire RH
Mail : julie.jaffre@emse.fr
Tel. : +33 (0)4 77 42 00 17

Amandine HIRONDEAU – Adjointe DRH et Gestionnaire RH
Mail: hirondeau@emse.fr
Tel. : + 33 (0)4 77 42 01 03

Liens utiles :

https://www.mines-stetienne.fr/

https://www.imt.fr/

https://www.youtube.com/watch?v=QUeuC5iQiN0

- Lien vers le site de l'entité ou toute autre ressource pertinente

La protection de vos données :
https://www.mines-stetienne.fr/wp-content/uploads/2018/12/Informations-des-candidats-sur-les-traitements-de-donn%C3%A9es-personnelles.pdf