
Offre de thèse (réf. RNA-2026)

Titre : Écosystème numérique de transparence et de gouvernance des données

Résumé :

Dans la société numérique actuelle, la multiplication des échanges d'information soulève de nombreux défis pour les différents acteurs amenés à intervenir à l'échelle locale, à l'instar des fournisseurs privés de services numériques ou des collectivités territoriales. Ces défis qui interviennent à tous les niveaux d'un traitement de données, depuis son partage jusqu'à son stockage, appellent en particulier à créer de nouvelles règles d'usage, de transparence, de qualité ou encore de traçabilité des données échangées. Communément partagées, ces nouvelles règles sont en effet susceptibles de favoriser la confiance entre les différents types d'acteurs à l'égard des données échangées et encouragent un partage équitable des ressources numériques, notamment entre les acteurs publics et privés.

De précédents travaux (thèse soutenue en septembre 2024) nous ont permis de valider notre approche du partage de l'information, tant du point du modèle théorique (outils du web sémantique : ontologies, raisonnement, moteurs d'inférence, etc.) que du point de vue de la faisabilité technique (architecture décentralisée où le fournisseur de la donnée et le consommateur ne sont pas connectés, encapsulation des règles de partage dans une licence qui « suit » la donnée, quelques soient les plateformes intermédiaires de distribution, intégration dans des infrastructures existantes de l'Internet des objets ou « IoT »).

L'objectif du présent projet, au travers du financement d'une allocation doctorale, est d'enrichir notre architecture de partage de l'information selon 3 axes complémentaires. Le premier consistera à augmenter l'expressivité du langage de règles en ajoutant de nouveaux concepts à la définition des licences (ex : obligations). Le second est de formaliser la notion d'explication des décisions. Les règles peuvent en effet aboutir à des décisions contradictoires. Il est donc important de pouvoir expliquer les inférences calculées (par les raisonneurs) et les arbitrages réalisés, les critères utilisés, etc. Ce point peut être lié à la notion de « preuve de conformité ». Le troisième axe de travail concerne la modélisation des « sanctions » en cas de violation d'une règle. Ce point pourrait d'ailleurs être extrapolé à des « bonus » si, à l'inverse, la règle est bien respectée.

Laboratoire et environnement scientifique

Cette thèse s'inscrit dans la continuité de travaux déjà engagés sur une architecture décentralisée de partage de l'information fondée sur des ontologies, des licences formelles qui « suivent » la donnée et des mécanismes de raisonnement applicables à des environnements distribués, notamment IoT. Elle mobilise les compétences du LIUPPA et de la nouvelle chaire CyberSécurité en gouvernance des données, Web sémantique, contrôle d'usage, collecte et analyse de données, ainsi qu'en expérimentation avec des acteurs territoriaux publics et privés.

Contexte

Avec le développement d'Internet et des réseaux de communication, la circulation des données a acquis une place primordiale au sein de la société, en particulier au sein des villes. À travers la mise en place d'outils de captation des données et d'actionneurs, à l'instar d'objets connectés ou « IoT », il est désormais possible de collecter une quantité et une diversité croissante de données ainsi que d'agir sur l'environnement en temps réel, à l'exemple de capteurs permettant d'ajuster la température dans des

bâtiments publics. In fine, la captation et la diffusion de données au sein des villes donne lieu à la création de nouveaux services dans des domaines aussi divers que les transports, l'énergie, la gestion des déchets, la gestion de l'eau ou les flux de circulation. Toutefois, l'absence de contrôle sur l'usage et la fiabilité de ces données tend à constituer un frein au partage de données, tant pour le « producteur » de données que pour le « consommateur ».

Problématique scientifique

1) Les enjeux liés aux échanges de données

De ces échanges de données (et pas uniquement les données à caractère personnel) subvient alors un enjeu majeur: celui de la maîtrise et de la protection des données circulant entre les différents acteurs. Il ne s'agit pas de s'intéresser à la sécurité technique des différents réseaux mis en place, tels que la 4G, la 5G, ou des protocoles de communications d'objets connectés comme LoraWAN, mais d'envisager une question plus large, relevant de la gouvernance des échanges¹. L'enjeu relève davantage de l'accès aux données par le citoyen, de leur diffusion, de leur traçabilité ou des garanties d'intégrité apportées, dans un contexte où les échanges de données deviennent aujourd'hui, et de plus en plus, des éléments essentiels au fonctionnement des villes, en France comme à travers le monde.

Aujourd'hui, un ensemble de textes légaux et réglementaires visent déjà à orienter les usages des données, par exemple en encourageant leur diffusion ou en limitant l'utilisation de certains types de données. En France notamment, la Loi pour une République Numérique du 7 octobre 2016 (LRN) a pour but d'inciter les administrations à ouvrir et à partager l'ensemble des données publiques qu'elles exploitent afin de faciliter la réutilisation de celles-ci par les entreprises et les citoyens. Dans le cadre des données à caractère personnel, citons également le Règlement Général européen sur la Protection des Données (RGPD) en vigueur depuis le 25 mai 2018. Celui-ci a pour but d'encadrer l'utilisation des données personnelles afin que cette utilisation ne porte pas atteinte aux droits et libertés des individus, en particulier leur droit à la vie privée.

Outre ces deux textes juridiques importants, la Commission européenne, le Parlement et le Conseil se sont accordés fin 2023 sur une proposition de règlement européen sur l'intelligence artificielle (ou « AI Act ») entré en vigueur en août 2024. Ce texte vise à établir un cadre pour l'utilisation des données dans les modèles d'IA en fonction de risques identifiés.

Depuis septembre 2023 également, le Règlement européen sur la gouvernance européenne des données (ou « Data Governance Act »)² apporte de nouvelles règles en matière de partage de données au sein du marché intérieur. Ce dernier a pour objectif de favoriser le libre partage des données à travers l'Union en encourageant l'instauration d'espaces communs de partage de données et la mise en place d'intermédiaires de données, notamment dans les domaines de la santé, de la mobilité, de l'énergie et de l'agriculture. Le but est ainsi de développer l'accès, la portabilité et l'interopérabilité des données au sein de l'Union, en conciliant le secret des affaires, le respect de la confidentialité des données personnelles et la libre réutilisation des données.

S'il apporte certaines avancées en matière de partage et de mise à disposition des données, notamment à travers les intermédiaires de données, le DGA demeure cependant lacunaire sur les formes concrètes que sont susceptibles de prendre les partages de données entre les acteurs publics et privés au sein des États membres. Ainsi, il tend à manquer, sur le terrain, des règles de gestion, mais aussi d'organisation des responsabilités, qui articulent de manière adéquate l'ouverture, la protection et le partage des données.

1 La gouvernance peut être définie comme : « les formes de pilotage, de coordination et de direction des individus, des groupes, des secteurs, des territoires, et de la société, au-delà des organes classiques du gouvernement. [...] avec trois points centraux: l'idée de donner une direction à la société, de mobiliser une coalition et enfin d'exercer une contrainte, soit trois dimensions essentielles du politique ». V. Patrick Le GALÈS, « Gouvernance », Dictionnaire des politiques publiques, Laurie BOUSSAGUET, Sophie JACQUOT et Pauline RAVINET, 5e édition., Paris: Presses de Sciences Po, 2018, p. 299–308

2 Règlement (UE) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724

2) Une démarche « sécurité de l'information »

Pour mettre en pratique une telle gestion des responsabilités, il sera nécessaire de disposer d'un « outillage technologique » sur lequel nous pourrions nous appuyer. Il s'agit d'une vision plus technique du sujet présentée en termes de politique de sécurité (les règles de partage : contrats, règles d'usage, etc.), de métadonnées (traçabilité, etc.).

Si l'on aborde cette problématique sous l'angle de la sécurité de l'information et avec une démarche de gestion des risques (ex: norme ISO/IEC 27005:2022³, méthode EBIOS Risk Manager⁴), l'objectif consiste à proposer et à mettre en place des mécanismes permettant de superviser les échanges d'informations entre une multitude d'organisations (les acteurs). Ces entités sont indépendantes les unes des autres, c'est-à-dire qu'elles sont libres de mettre en place leur propre politique de sécurité (de l'information); il n'y a pas, à priori, d'autorité centrale. Elles ont en outre chacune leurs propres objectifs et critères d'évaluation, pouvant même éventuellement être concurrentes sur certains points: enjeux économiques, critères et niveaux de sécurité, contraintes réglementaires, etc.

Un certain nombre de technologies et de modèles existent déjà pour garantir des propriétés de sécurité sur les échanges de données entre différents systèmes d'information inter-connectés : cryptographie (chiffrement, signature, tatouage, etc.), journalisation, blockchain, contrôle d'accès et contrôle d'usage, etc. Elles permettent d'acheminer « correctement » une donnée d'un point A à un point B. Mais vis-à-vis de l'information véhiculée par cette donnée, qu'en est-il de la confiance envers le système qui l'a émise, de la conformité de l'usage qui en sera fait, etc. ?

3) Notre approche

Le défi est de proposer des mécanismes pour mettre en pratique une politique de gouvernance des données et une gestion des responsabilités qui en découlent. L'idée est de pouvoir vérifier automatiquement si les règles de partage qui auront été définies sont bien respectées, ou encore que ces outils soient capables de raisonner pour identifier les causes d'une éventuelle violation et d'en déduire par exemple les responsabilités des différents acteurs.

Ceci nécessite de pouvoir spécifier formellement les règles de partage (dans une « logique mathématique ») et d'utiliser ensuite des techniques d'inférence pour le raisonnement. Les « règles » sont exprimées sous la forme de prédicats logiques sur lesquels il sera possible de « raisonner », c'est-à-dire de déduire (notion d'inférence en logique) de nouvelles connaissances (des faits) sur la base des connaissances déjà exprimées. Intuitivement ces règles sont écrites sous la forme « si... alors... ». On parle de règles de contrôle d'accès (à l'information) ou de contrôle d'usage (de l'information). Exprimer ces règles dans un langage formel permet ensuite d'automatiser leur traitement au travers d'outils tels que les raisonneurs ou les moteurs d'inférence.

Cette approche est parfaitement réaliste. Dans [MEDES'22, KES'23, CRiSIS'24] nous présentons une telle architecture pour le partage des informations basée sur les ontologies. Dans une approche orientée autodétermination informationnelle, et grâce aux outils du Web sémantique, les producteurs de données peuvent ainsi exprimer les règles de contrôle d'usage qu'ils souhaitent voir appliquées, ceci sous la forme de licences (OWL pour exprimer les entités, SWRL pour les règles). Un prototype « proof of concept » a également été développé pour confirmer la faisabilité de notre approche et réaliser quelques évaluations de performances.

4) Les communs pour changer de paradigme

À la fin des années 1970 l'économiste Elinor Ostrom envisage un nouveau modèle de gestion des données par les « Communs » qui constitue un nouveau paradigme économique au sein duquel la valeur d'usage

3 ISO/IEC 27005:2022 : Sécurité de l'information, cybersécurité et protection de la vie privée – Préconisations pour la gestion des risques liés à la sécurité de l'information

4 EBIOS : Expression des Besoins et Identification des Objectifs de Sécurité (ANSSI)

prime sur la propriété, à l'instar de ce qui est observé pour les applications d'auto-partage, de « *couchsurfing* » ou de « *crowdfunding* ».

Les NTIC offrirait dans ce cadre un second souffle au modèle des Communs, sous réserve que ce partage de l'information reste « maîtrisé », tant du point de vue de l'usage de l'information par les différents acteurs que de leurs responsabilités respectives. Notre vision du partage de l'information « orientée communs numériques » est en adéquation avec la stratégie européenne pour les données qui se concrétise au travers du DGA, du « Data Act », etc. [CRiSIS'24].

Dans le contexte des échanges de données, nos travaux visent donc à mettre en place une gouvernance des données pour contrôler le partage de l'information afin de rassurer les détenteurs de données et les inciter à les mettre à disposition de la communauté. Notre problématique est ainsi orientée vers des aspects tels que la provenance, le contrôle d'usage, la qualité des données, les stratégies de partage.

Objectifs de la thèse

La thèse poursuivra trois axes principaux.

Toujours en suivant une approche holistique de la sécurité de l'information (informatique, droit, économie) avec comme priorité de replacer le citoyen au centre du dispositif, ce projet vise à enrichir notre architecture de partage des données selon 3 axes complémentaires :

- 1) Expressivité du langage de règles : il s'agit d'ajouter de nouveaux concepts pour la définition des licences, et notamment la notion d'obligation ; à savoir que si le règlement de sécurité contient simultanément des permissions, des interdictions et des obligations exprimant différents aspects de la stratégie de partage de l'information, certaines de ces règles peuvent parfois se contredirent selon le contexte.
- 2) Explication des décisions : certains outils de raisonnement du web sémantique peuvent également fournir une explication des inférences réalisées ; l'idée est ici d'exploiter cette possibilité pour fournir une explication des décisions basée sur les règles déclenchées, les arbitrages réalisés (en cas de règles contradictoires par exemple), les critères utilisés, etc. ; il serait alors possible de produire une « preuve de conformité » détaillée au consommateur qui aura interrogé la licence.
- 3) Modélisation des « sanctions » : il s'agit d'exprimer, pour chacune des règles, quelles seront les « sanctions » à appliquer en cas de violation de cette règle ; il est possible d'envisager une extension de ce point à des « bonus » si, au contraire, la règle est respectée ; ce serait une sorte de « gamification » du partage de l'information en vue de faire prendre conscience, aux différents acteurs, de la valeur des informations.

Pour mener à bien ce projet nous envisageons également d'utiliser des technologies issues du domaine de l'intelligence artificielle (IA) pour aller au-delà d'une simple représentation « syntaxique » des stratégies de partage de données :

- résolution des conflits, arbitrage des décisions : couplée aux explications des décisions, l'intégration de l'IA dans le processus décisionnel permettra de prendre de meilleures décisions avec plus de confiance
- génération des licences : en observant les partages de données entre les acteurs, grâce au Machine Learning (ML) il serait possible de profiler les usages pour construire automatiquement les règles de la licence
- création des licences : en utilisant des outils basés sur les LLM (« Large Language Models », ou grand modèles de langage) nous pourrions construire nos licences à partir de vrais contrats « papier » (apprentissage) ou des besoins exprimés en langage naturel par les acteurs (assistance) ; ce travail pourrait nécessiter une étape intermédiaire basée sur la théorie des actes du langage

Programme de travail (prévisionnel)

Un programme de recherche prévisionnel a été établi. Nous ne présentons pas ici les tâches détaillées mais uniquement les objectifs de chacune des trois années avec les livrables attendus.

- **Année 1** – Fondations théoriques et socle technique
 - **Objectifs principaux**
 - Appropriation approfondie de l'état de l'art (sécurité de l'information, gouvernance, Web sémantique, politiques d'usage, AI Act, DGA, etc.).
 - Spécification formelle du langage de règles enrichi (axe 1) et cadrage des besoins pour l'explicabilité (axe 2) et les sanctions (axe 3).
 - Rédaction d'un document de cadrage (concepts, exigences, cas d'usage).
 - Mise en place du socle logiciel (ontologies, moteurs d'inférence, pipeline expérimental).
 - Définition des jeux de données et scénarios initiaux en agriculture connectée (en lien avec partenaires).
 - **Livrables principaux**
 - Rapport d'état de l'art structuré et modèle conceptuel global de l'écosystème de gouvernance.
 - Spécification formelle du langage de règles enrichi (document technique + schémas d'ontologie).
 - Prototype logiciel de base et premiers scénarios d'évaluation (données synthétiques ou partiellement réelles).
- **Année 2** – Explication, sanctions et intégration de l'IA
 - **Objectifs principaux**
 - Développer les mécanismes d'explication des décisions et de preuve de conformité (axe 2).
 - Modéliser les sanctions/bonus et articuler ce modèle avec les règles (axe 3).
 - Introduire les premiers modules IA (résolution de conflits, génération de règles à partir d'observations).
 - **Livrables principaux**
 - Module d'explication des décisions et de génération de preuves de conformité intégré à l'architecture.
 - Modèle complet sanctions/bonus et prototype de gamification du partage de données.
 - Premiers résultats d'intégration de techniques IA (résolution de conflits, suggestions de règles) avec évaluations préliminaires.
- **Année 3** – Licences « intelligentes », validation et valorisation
 - **Objectifs principaux**
 - Exploiter les LLM pour la génération/assistance de licences à partir de contrats et besoins exprimés en langage naturel.
 - Consolider et valider l'architecture globale sur des cas d'usage réalistes (agriculture connectée, communs numériques régionaux) → Définition conjointe avec les partenaires de scénarios complets (chaîne de bout en bout) : expression des politiques, génération de licences, exécution des échanges, monitoring, explication, sanctions/bonus.
 - Finaliser les contributions scientifiques et la rédaction de la thèse.
 - **Livrables principaux**
 - Pipeline complet d'édition de licences assistée par LLM, connecté à l'architecture de gouvernance.
 - Plateforme expérimentale validée sur au moins un cas d'usage réel ou pré-opérationnel.
 - Manuscrit de thèse, publications associées et livrables techniques pour le laboratoire/partenaires.

Profil recherché

- Le candidat recherché doit maîtriser la programmation Python (Java également serait un plus) ainsi que de solides bases en algorithmique et modélisation, avec un intérêt marqué pour l'IA et les systèmes complexes.
- Des compétences en logique formelle (logique modale, déontique, temporelle), IA explicable (XAI), systèmes à base de règles, modélisation de politiques d'accès (RBAC/ABAC), analyse de données et *machine learning*, ainsi que des notions de traçabilité et d'audit des systèmes sont fortement appréciées.
- Le profil attendu requiert également de la rigueur scientifique, une forte capacité de formalisation, un intérêt pour les systèmes socio-techniques et les problématiques hybrides combinant IA, gouvernance et comportements, ainsi qu'un bon niveau d'anglais scientifique.
- Les atouts recherchés incluent une expérience en recherche (stage de M2 ou publication), des compétences en IA symbolique ou XAI, un intérêt pour la gouvernance des données ou la cybersécurité, des connaissances en théorie des jeux ou systèmes multi-agents, ainsi qu'une expérience en data engineering ou en systèmes distribués.
- La thèse s'inscrit dans un environnement interdisciplinaire à l'interface de l'intelligence artificielle, des systèmes distribués et de la gouvernance des données, de l'explicabilité et de la conformité, ainsi que de l'analyse comportementale et des mécanismes incitatifs.

Localisation

- Université de Pau et des Pays de l'Adour (UPPA)
- Site de Mont de Marsan (371 rue du Ruisseau 40000 Mont de Marsan)
- Laboratoire LIUPPA (<http://liuppa.univ-pau.fr/>) – équipe ISI-Know

Contacts

- **Manuel Munier** → manuel.munier@univ-pau.fr
- Richard Chbeir → richard.chbeir@univ-pau.fr

Évaluation de la candidature

Processus de sélection

- constitution d'un jury de sélection
- sélection des candidats sur dossier de candidature
- audition des candidats et classement

Critères d'évaluation

- motivation, maturité scientifique, curiosité du candidat
- notes et classement en M1 et en M2
- appétence pour la recherche, esprit critique du candidat
- maîtrise de l'anglais

Constitution du dossier de candidature

Envoyer par email (manuel.munier@univ-pau.fr) un dossier de candidature comprenant :

- CV
- lettre de motivation
- copie du diplôme
- master (ou diplôme équivalent) : relevé détaillé des notes obtenues dans chaque matière et classement
- lettres de recommandation
- si possible, coordonnées de deux personnes références du milieu professionnel à contacter