Rapport de Stage M2 – TI

Valentin MOUCHE

Master 2 – Technologies de l'Internet

Université de Pau et des Pays de l'Adour

Encadrants :M. Munier, LIUPPA, France ; N. Laamech, UM6P, Maroc

Lieu du stage : IUT de Mont-de-Marsan

Juillet 2025

Remerciements

Je tiens à exprimer ma sincère gratitude à toute l'équipe du site de l'IUT de Mont-de-Marsan pour leur accueil chaleureux et leur soutien tout au long de ce stage.

Je souhaite remercier tout particulièrement M. Manuel Munier, qui m'a donné cette opportunité d'effectuer un stage en recherche, me permettant ainsi de découvrir un environnement scientifique enrichissant et de développer de nouvelles compétences.

Je remercie également mes encadrants académiques, Manuel Munier et Nouha Laamech, pour leur suivi attentif, leurs conseils avisés et leurs remarques constructives qui ont grandement contribué à l'amélioration de ce travail.

Enfin, je remercie chaleureusement mes collègues ainsi que toutes les personnes qui ont partagé leurs connaissances et m'ont accompagné avec bienveillance dans cette expérience enrichissante.

Merci à tous pour votre précieux soutien.

Résumé

Ce rapport présente les travaux réalisés dans le cadre de mon stage de fin d'études de Master 2 Technologies de l'Internet, effectué au sein du laboratoire LIUPPA, à Mont-de-Marsan, sur une durée de six mois.

L'objectif principal du stage était de concevoir une ontologie générale de la notion de métadonnée, capable de s'adapter à divers domaines d'application tout en assurant scalabilité et interopérabilité. La problématique repose sur le constat que, malgré leur importance croissante, les métadonnées manquent d'une définition commune et généralisable, en raison de la diversité des usages et des contextes dans lesquels elles apparaissent.

La première phase du stage a consisté en une étude approfondie de l'état de l'art, à la fois sur les métadonnées et sur les ontologies, afin d'identifier les concepts-clés et les bonnes pratiques de modélisation sémantique. Cette recherche a conduit à la conception de plusieurs schémas conceptuels, sous forme de diagrammes simplifiés, permettant de structurer progressivement une représentation unifiée des métadonnées.

Une fois un modèle satisfaisant élaboré, celui-ci a été implémenté sous forme d'ontologie à l'aide de l'outil Protégé, en respectant les standards RDF et OWL. Cette ontologie a ensuite été intégrée dans l'environnement IDSM-O, une ontologie préexistante développée au LIUPPA, dans le but de tester sa compatibilité et sa pertinence. Les tests ont montré une bonne intégration, validant ainsi à la fois la structure et la cohérence du modèle proposé.

Enfin, ce travail a été formalisé sous la forme d'un article scientifique, visant à documenter l'approche méthodologique, les choix de modélisation, ainsi que les perspectives d'évolution du modèle.

Table des matières

Remerciements **1**

Résumé **2**

1 Introduction **4**

2 Présentation de l'entreprise **5**

3 Contexte et problématique **6**

4 Méthodologie / Travaux réalisés **8**

4.1 Étude de l'état de l'art 8

4.2 Conception des modèles 8

4.3 Implémentation avec Protégé 11

4.4 Intégration à une ontologie existante 13

5 Rédaction d'un article scientifique **14**

5.1 Choix de la conférence SECAI 14

5.2 Structure de l'article 15

5.3 Changement de conférence 15

5.4 Soumission et perspectives 16

6 Conclusion **17**

Chapitre 1

Introduction

Dans un contexte marqué par une explosion des données numériques et par la montée en puissance des technologies du Web sémantique, la question de la structuration et de la description des ressources numériques devient centrale. L'information brute, aussi abondante soit-elle, ne devient exploitable qu'à partir du moment où elle est accompagnée de métadonnées claires, cohérentes et interprétables par les machines comme par les humains.

Les métadonnées, en tant que données décrivant d'autres données, jouent un rôle fondamental dans l'organisation, la recherche, l'indexation et l'interopérabilité des contenus numériques. Elles sont omniprésentes dans des domaines aussi variés que les bibliothèques numériques, les systèmes d'information distribués, l'Internet des objets ou encore les plateformes de données ouvertes. Toutefois, malgré leur importance croissante, la notion de métadonnée reste marquée par une grande hétérogénéité de définitions, de formats et de vocabulaires. Cette diversité freine la standardisation et limite l'interopérabilité entre systèmes.

C'est dans cette optique que s'inscrit le stage présenté dans ce rapport, réalisé au sein du laboratoire LIUPPA à l'IUT de Mont-de-Marsan. L'objectif principal était de concevoir une ontologie générique de la notion de métadonnée, c'est-à-dire un modèle formel et structuré permettant de représenter de manière unifiée les éléments constitutifs des métadonnées, tout en conservant une flexibilité suffisante pour s'adapter à différents contextes d'usage.

Ce rapport présente donc l'ensemble des travaux réalisés au cours de ce stage de six mois. Il débute par une présentation de la structure d'accueil, suivie d'un exposé du contexte et des enjeux liés à la problématique abordée. Ensuite, il détaille la méthodologie adoptée, depuis l'étude de l'état de l'art jusqu'à l'implémentation concrète du modèle à l'aide d'outils dédiés. Les résultats obtenus sont ensuite analysés, en mettant en lumière les apports, les limites et les perspectives offertes par le modèle développé.

Chapitre 2

Présentation de l'entreprise

Le stage s'est déroulé au sein du Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour (LIUPPA), sur le site de Mont-de-Marsan. Le LIUPPA est une unité de recherche en informatique rattachée à l'Université de Pau et des Pays de l'Adour (UPPA). Il fait partie de l'école doctorale des Sciences Exactes et de leurs Applications et est reconnu pour ses travaux dans divers domaines liés à l'informatique appliquée.

Les axes de recherche du laboratoire couvrent un large spectre, parmi lesquels : l'ingénierie des applications web, les systèmes d'information, les environnements intelligents, la modélisation des données ou encore les interactions homme-machine. Le laboratoire collabore avec des partenaires académiques et industriels, dans le cadre de projets régionaux, nationaux et européens.

Le site de Mont-de-Marsan constitue l'une des antennes du laboratoire, avec une activité axée sur le développement d'applications web et la gestion de systèmes d'information. L'équipe locale est composée d'enseignants-chercheurs, d'ingénieurs, de doctorants, et d'étudiants de master.

Durant mon stage, j'ai été encadré par Manuel Munier, enseignant-chercheur au LIUPPA, qui a assuré le rôle de tuteur scientifique, ainsi que par ma tutrice de stage Nouha Laamech, qui m'a accompagnée sur les aspects techniques et organisationnels du projet. Leur encadrement a été essentiel pour le bon déroulement de mon travail, en me fournissant à la fois un soutien méthodologique et une vision claire des objectifs à atteindre.

Chapitre 3

Contexte et problématique

Dans un monde de plus en plus numérique, où les systèmes d'information sont interconnectés et les données omniprésentes, la capacité à organiser, décrire et exploiter l'information devient essentielle. C'est dans ce contexte que les métadonnées jouent un rôle central.

Une métadonnée est, littéralement, une "donnée sur une donnée". Il s'agit d'une information qui décrit une autre ressource numérique ou physique, permettant d'en faciliter l'identification, la compréhension, la gestion ou l'échange. Par exemple, dans une bibliothèque numérique, le titre, l'auteur et la date de publication d'un livre sont des métadonnées qui décrivent le contenu du document. Dans le contexte du Web ou des systèmes d'information, les métadonnées peuvent également inclure des informations techniques (format, taille, type), des données de contexte (provenance, date de création, droits d'accès) ou des annotations sémantiques.

La montée en puissance du Web sémantique, des architectures orientées données et de l'interopérabilité entre systèmes a renforcé le besoin de structurer et de représenter ces métadonnées de façon précise, fiable et interprétable par les machines.

Cependant, la notion même de métadonnée reste floue et contextuelle. En effet, selon les domaines (archives, bibliothèques, systèmes distribués, IoT, etc.), les métadonnées prennent des formes et des structures variées, souvent hétérogènes et parfois incompatibles. Cette absence de définition unifiée constitue un frein à leur exploitation dans des environnements intelligents, évolutifs et interopérables.

C'est dans ce cadre qu'a émergé le sujet de ce stage. Lors de ses travaux de doctorat, ma tutrice N. Laamech a été confrontée à la notion de métadonnée dans un contexte sémantique. Bien que cette problématique ait suscité un réel intérêt, elle n'a pu être approfondie, faute de temps et de cadrage spécifique dans le cadre de sa thèse. Le stage a ainsi été l'occasion de reprendre cette réflexion initiale, de l'enrichir, et de la transformer en un travail de formalisation structuré et ciblé.

Même si le besoin initial émanait d'un contexte précis, notre volonté a été de dépasser cette problématique locale pour proposer une définition générique, interopérable

et scalable de la notion de métadonnée. L'idée n'était pas seulement de répondre à un cas d'usage particulier, mais de concevoir un modèle sémantique réutilisable, capable de s'adapter à différents domaines d'application.

L'objectif principal du stage a donc été de proposer une ontologie générique de la notion de métadonnée, c'est-à-dire un modèle formel permettant de décrire, structurer et relier des métadonnées de manière unifiée, tout en conservant une certaine souplesse d'usage. Le principal défi a été de trouver un équilibre entre expressivité, généricité et scalabilité, afin de proposer une solution à la fois théoriquement rigoureuse et exploitable en pratique.

Chapitre 4

Méthodologie / Travaux réalisés

4.1 Étude de l'état de l'art

Le stage a débuté par une phase de recherche documentaire portant sur deux axes principaux :

— La définition, les types et les usages des métadonnées dans différents contextes.

— Les principes fondamentaux des ontologies, notamment dans le cadre du Web sémantique (OWL, RDF, SWRL).

Cette étude a permis d'identifier les concepts communs aux différentes approches de métadonnées, tels que les notions de *ressource*, *propriété*, *contexte*, *format*, *source*, etc., ainsi que les bonnes pratiques en modélisation sémantique.

Au cours de cette étape, j'ai consulté de nombreux articles scientifiques, publications académiques et spécifications techniques, ce qui m'a permis d'approfondir mes connaissances théoriques sur les métadonnées et les ontologies. Cette lecture critique m'a aidé à distinguer les approches existantes, à en percevoir les limites et à en extraire les éléments réutilisables. Progressivement, j'ai commencé à organiser ces idées sous forme de notes, de cartes conceptuelles et de premiers schémas, en vue de structurer une réflexion claire pour la suite du travail.

4.2 Conception des modèles

Sur la base de cette analyse, plusieurs schémas conceptuels ont été élaborés à l'aide de diagrammes de type « classes simplifiées », afin de structurer progressivement une ontologie de la métadonnée. Ces diagrammes ont permis de visualiser les entités centrales (ex. : `Metadata`, `Resource`, `Agent`, `Format`, `Date`, etc.) ainsi que leurs relations sémantiques (`describes`, `hasAuthor`, `hasFormat`, etc.).

Un travail itératif de discussion avec les encadrants a conduit à affiner et valider une version stable du modèle conceptuel. Chaque réunion a permis d'identifier des limites, de

proposer des améliorations ou de généraliser certaines entités.

Afin d'illustrer cette évolution, deux versions successives du modèle sont présentées ci-dessous.

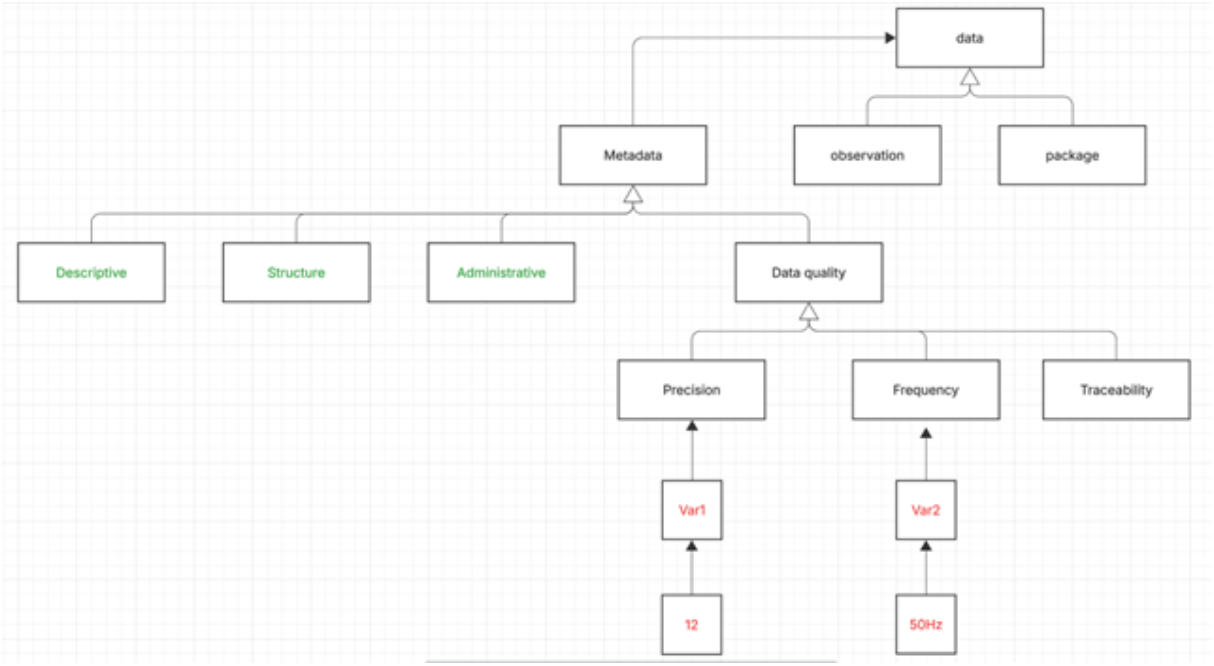

FIGURE 4.1 – Première version du schéma conceptuel, version exploratoire

La première version (voir Figure 4.1) a été conçue comme un point de départ pour discuter de la structuration des métadonnées. Elle introduit des entités de base telles que `Metadata` et `Resource`, avec des relations simples. À ce stade, la réflexion portait principalement sur l'identification des concepts fondamentaux.

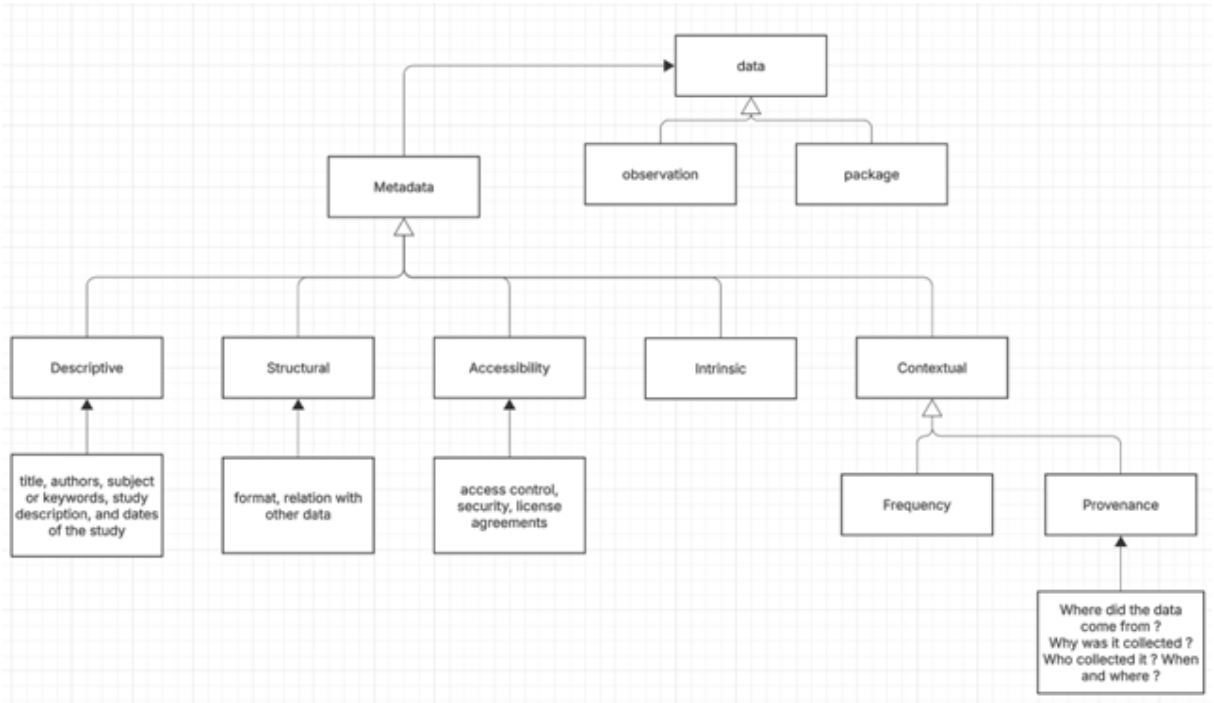

FIGURE 4.2 – Deuxième version du schéma conceptuel, version stabilisée

La deuxième version (voir Figure 4.2) est le fruit d'un processus d'affinement. Elle intègre des entités plus précises ainsi que des relations sémantiques enrichies, telles que `hasAuthor` ou `hasFormat`, en vue de permettre une description plus fine et générique des métadonnées.

Enfin, il est important de noter qu'au cours du processus de modélisation, la question de la qualité des données a été initialement envisagée comme un axe à part entière. Nous pensions dans un premier temps qu'il serait pertinent d'intégrer explicitement des concepts liés à la qualité (ex. : exactitude, complétude, fraîcheur, etc.) au sein de l'ontologie.

Cependant, au fil des réflexions et des discussions, il est apparu que les métadonnées, en elles-mêmes, contribuent à assurer la qualité des données. Par leur capacité à contextualiser, structurer et décrire les ressources, les métadonnées permettent d'évaluer indirectement la qualité d'un jeu de données, sans qu'il soit nécessaire de représenter la qualité comme une entité ou un attribut distinct.

Cette décision d'écarter la notion explicite de qualité dans le modèle final illustre un choix d'abstraction raisonné : plutôt que de dupliquer des mécanismes de contrôle ou d'évaluation, nous avons privilégié une modélisation centrée sur la sémantique, laissant à l'utilisateur ou à d'autres couches applicatives la possibilité d'exploiter les métadonnées à des fins d'évaluation de la qualité.

En résumé, cette phase de modélisation a permis de poser les bases d'une ontologie claire et évolutive, en affinant progressivement la structure autour des métadonnées et en recentrant le modèle sur l'essentiel.

4.3 Implémentation avec Protégé

Avant de passer à l'implémentation, il est utile de rappeler ce qu'est une ontologie dans le contexte du Web sémantique. Il s'agit d'une représentation formelle de la connaissance d'un domaine, composée de concepts (ou classes) et des relations qui les unissent. Une ontologie permet de structurer les données de manière explicite, afin qu'elles soient compréhensibles par les humains mais aussi exploitables automatiquement par les machines. Elle facilite ainsi l'interopérabilité des systèmes et permet l'inférence de nouvelles connaissances à partir des données existantes.

Une fois le modèle conceptuel validé, il a été implémenté sous forme d'ontologie à l'aide de l'outil *Protégé*[1]. Développé par l'université de Stanford, Protégé est un environnement libre et largement utilisé pour concevoir, éditer et visualiser des ontologies fondées sur les standards RDF (Resource Description Framework) et OWL (Web Ontology Language).

L'ontologie a été construite à partir des classes principales identifiées dans le modèle conceptuel (voir Figure 4.2), telles que `Data`, `Metadata`, `Descriptive`, `Structural`, etc. Ces classes ont été organisées selon un mécanisme d'héritage conforme à la structure hiérarchique du schéma précédent.

Les relations entre ces classes ont été modélisées à l'aide de deux types de propriétés :

— **Object properties** : elles relient deux instances de classes. Par exemple, la propriété `hasMetadata` permet d'associer une instance de `Data` à une ou plusieurs instances de `Metadata`.

— **Data properties** : elles relient une instance de classe à une valeur littérale (texte, nombre, date, etc.). Dans notre cas, la propriété `value` a été définie de manière non typée, afin de laisser à l'utilisateur le choix de spécifier le type de valeur lors de l'instanciation des métadonnées. De plus, on a une autre propriété `label` qui nous permet d'étiquetter les métadonnées.

Des restrictions sémantiques ont également été introduites, telles que les domaines et portées des propriétés, ou encore des contraintes de cardinalité, afin de garantir la cohérence logique de l'ontologie. Il est également possible d'imposer un typage spécifique pour certaines classes : par exemple, pour chaque instance de la classe `Title` (sous-classe de `Descriptive`, elle-même sous-classe de `Metadata`), on peut restreindre la propriété `value` à n'accepter que des littéraux de type `xsd:string`. Cependant, il est important de noter qu'une métadonnée n'est pas statique, elle peut évoluer au cours de sa durée de vie, par exemple pour les métadonnées de provenance qui peuvent être mise à jour à chaque fois qu'un utilisateur rajoute des informations ou modifie la métadonnée initiale.

L'ensemble a ainsi été structuré de manière hiérarchique et cohérente, fournissant une base formelle et extensible pour décrire, interconnecter et exploiter des métadonnées dans différents contextes d'usage.

1. `https://protege.stanford.edu/`

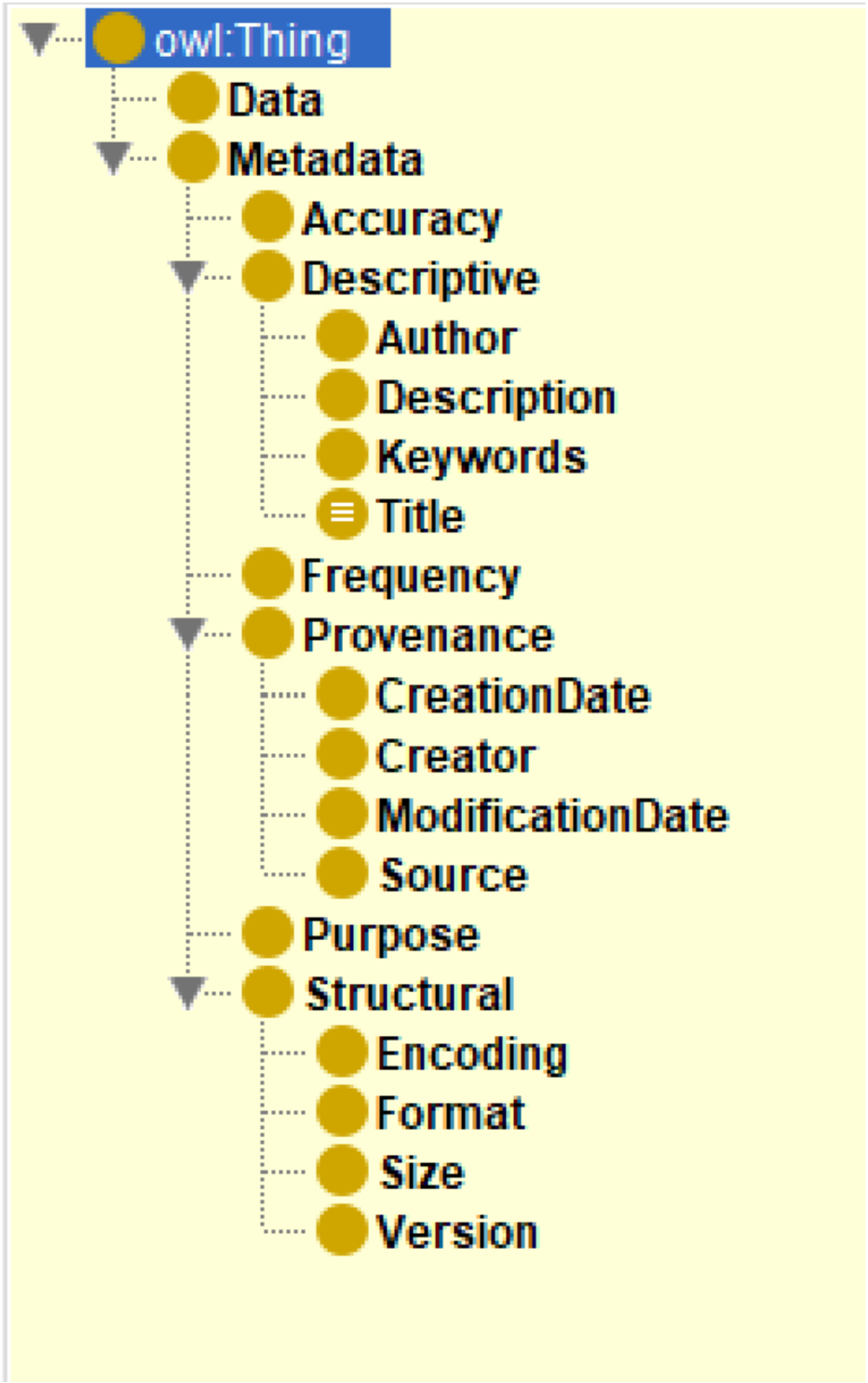

FIGURE 4.3 – Vue des classes de l'ontologie dans Protégé

4.4 Intégration à une ontologie existante

L'ontologie développée a ensuite été testée en l'intégrant à `IDSM-O`, une ontologie existante conçue par N. Laamech dans le cadre d'un projet antérieur. Cette étape avait pour objectif de valider la cohérence, la flexibilité et l'interopérabilité de notre modèle au sein d'un environnement sémantique concret, manipulant des instances de données réelles ou représentatives.

Pour cela, l'ontologie a été fusionnée avec `IDSM-O` dans Protégé, puis des instances ont été créées afin d'illustrer différents cas d'usage. Des règles logiques ont également été formulées à l'aide du langage `SWRL` (Semantic Web Rule Language), permettant d'exprimer des inférences supplémentaires ou des comportements attendus à partir des données modélisées.

L'ensemble a ensuite été validé à l'aide du moteur de raisonnement `Pellet`, qui a permis de vérifier la cohérence de l'ontologie étendue, d'exécuter les règles SWRL et de déduire automatiquement de nouvelles informations à partir des axiomes existants. Cette phase de raisonnement a confirmé que le modèle était logiquement consistant et opérationnel dans un contexte applicatif.

En complément, une seconde phase d'intégration a été réalisée avec l'ontologie `SSN` (Semantic Sensor Network), une ontologie standard proposée par le W3C pour modéliser les capteurs, observations et systèmes de mesure dans des environnements connectés. Cette intégration visait à tester la capacité de notre ontologie à s'adapter à un cadre normatif existant et très utilisé, notamment dans le domaine de l'Internet des Objets (IoT). Elle a permis de valider que les concepts de métadonnées développés pouvaient être articulés avec des modèles plus larges décrivant des observations et leurs contextes de collecte, renforçant ainsi la portée interopérable et générique du modèle proposé.

Chapitre 5

Rédaction d'un article scientifique

Les deux derniers mois de mon stage ont été principalement consacrés à la rédaction d'un article scientifique visant à présenter les travaux réalisés. L'objectif était de soumettre cet article à une conférence académique dans le domaine des métadonnées et du web sémantique.

5.1 Choix de la conférence SECAI

Après réflexion, nous avions initialement ciblé **SECAI 2025 (Workshop on Security and Artificial Intelligence)** [1], organisée en collaboration avec ESORICS, et qui se tiendra à Toulouse les 25 et 26 septembre 2025. SECAI propose un forum d'échange entre chercheurs et industriels autour de l'intersection entre l'intelligence artificielle et la cybersécurité. Les axes d'intérêt incluent notamment :

— l'*AI for security* (IA au service de la sécurité) ;
— le *malicious use of AI* (usage malveillant de l'IA) ;
— la *security for AI* (sécurisation des systèmes IA).

Tous les articles acceptés pour SECAI 2025 sont publiés dans la collection LNCS (Springer).

Nous avions donc orienté notre article autour de l'utilisation des métadonnées dans les *pipelines d'IA*, c'est-à-dire les chaînes de traitement automatique qui, à partir de données brutes, appliquent des étapes successives (prétraitement, entraînement, validation, déploiement, etc.) pour produire un modèle ou un résultat intelligent. Les métadonnées permettent alors de tracer, qualifier ou sécuriser chaque étape de cette chaîne.

Cependant, au fil de l'écriture, nous avons constaté que notre contribution prenait davantage de sens du point de vue de la **confiance** et de la **sécurité de la donnée** plutôt que strictement du point de vue de l'IA. Cela nous a naturellement amenés à réorienter notre article vers une autre conférence : **STM 2025**.

1. `https://esorics2025.sciencesconf.org/?lang=en`

5.2 Structure de l'article

Une fois la conférence ciblée, j'ai pu débuter la rédaction de l'article en respectant les contraintes imposées par SECAI. Le format suivi était celui de la série LNCS (Lecture Notes in Computer Science), avec une limite maximale de 8 pages (ou 10 avec les annexes).

L'article suivait une structure académique classique :

— **Abstract** : résumé succinct du travail, de ses contributions principales et des résultats obtenus.

— **Introduction** : mise en contexte, problématique, objectifs et contributions.

— **Related Work** : analyse critique des travaux existants dans le domaine des ontologies et de la gestion des métadonnées.

— **The Metadata Management Ontology** : description détaillée de l'ontologie développée durant le stage.

— **Ontology Evaluation** : méthodes d'évaluation de l'ontologie (validation implémentation dans une ontologie existante, via des règles SWRL et aussi via l'étude de métriques).

— **Conclusion** : synthèse des apports et perspectives futures.

Chaque section devait être rédigée de manière concise, rigoureuse et documentée, en s'appuyant sur des références scientifiques et des standards en vigueur.

5.3 Changement de conférence

À environ un mois de la fin de mon stage, et peu avant la date de soumission initialement prévue, nous avons réévalué la pertinence de la conférence SECAI au regard de l'évolution du positionnement scientifique de notre article. Si l'approche par les pipelines d'IA nous avait semblé pertinente au départ, l'orientation de notre travail s'est progressivement affirmée autour de la **confiance dans les métadonnées**, de leur **interopérabilité**, et de leur **structuration dans des environnements documentaires**.

Nous avons ainsi décidé de rediriger notre soumission vers la conférence **STM 2025 (Security and Trust Management)** [2], qui se tiendra à Copenhague en septembre 2025. STM s'inscrit dans les conférences affiliées à ESORICS, mais avec un angle plus large sur les technologies de confiance, la gestion des politiques d'accès, ou encore la traçabilité et l'annotation des données.

Ce changement a entraîné plusieurs ajustements importants :

— **Révision du positionnement** : le discours scientifique a été réorienté pour correspondre aux thématiques principales de STM, centrées sur la sécurité, la confiance et la gestion des données.

2. `https://nics.uma.es/stm2025/`

— **Profondeur des sections** : certaines parties, notamment la méthodologie, les contraintes d'usage et les perspectives applicatives, ont été développées davantage, profitant de la nouvelle limite de **16 pages** autorisée par STM (contre 8 pages pour SECAI).
— **Références bibliographiques** : un enrichissement ciblé de l'état de l'art a permis d'ancrer le travail dans les débats spécifiques de la communauté STM.

5.4 Soumission et perspectives

Une fois les révisions effectuées, l'article a été soumis dans les délais impartis. Au moment de la rédaction de ce rapport, la décision du comité scientifique de STM n'est pas encore connue. Une éventuelle acceptation constituerait une belle valorisation du travail accompli, et permettrait une diffusion de l'ontologie auprès de la communauté scientifique.

Ce travail de rédaction, en plus d'avoir consolidé ma compréhension du domaine, m'a permis de me familiariser avec les exigences de la publication scientifique : rigueur, synthèse, respect des formats, bibliographie normalisée, etc.

Chapitre 6

Conclusion

Ce stage a permis d'apporter une réponse partielle mais significative à un besoin fondamental dans le domaine des systèmes d'information : la structuration générique des métadonnées. L'ontologie développée constitue une base solide, réutilisable et extensible, permettant de représenter la diversité des métadonnées de manière unifiée, cohérente et formellement définie.

Au-delà de l'aspect purement technique, ce stage a été une véritable immersion dans le monde de la recherche appliquée. Il m'a permis de découvrir une nouvelle manière de travailler, fondée sur la rigueur, l'analyse critique, l'exploration bibliographique approfondie, et l'itération continue entre hypothèses, modélisations et validations. Travailler dans un environnement de recherche, au contact de chercheurs expérimentés, a grandement enrichi ma compréhension du cycle de vie d'un projet scientifique. Cette expérience a également été l'occasion d'acquérir des compétences transversales précieuses :

— synthétiser un état de l'art dans un domaine pointu ;
— concevoir une solution originale répondant à une problématique ouverte ;
— formuler et structurer une contribution scientifique selon les normes académiques ;
— rédiger un article scientifique en anglais destiné à une conférence internationale ;
— collaborer de manière autonome tout en intégrant un cadre collectif de recherche.

Enfin, ce stage m'a conforté dans mon intérêt pour les approches formelles et les technologies sémantiques, tout en me montrant le potentiel concret de ces méthodes dans des cas d'usage réels.