

ACADÉMIE DE MONTPELLIER
UNIVERSITÉ MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC

MÉMOIRE DE STAGE DE MASTER 2 INFORMATIQUE
effectué à la Maison de la télédétection de Montpellier

Conception d'un modèle abstrait de métadonnées pour l'interconnexion de jeux de données géoréférencées

Auteur :
Hatim CHAHDI

Encadré par :
J.C. DESCONNETS
I. MOUGENOT

28 Juin 2013

Résumé

Les images satellites sont devenues des données incontournables et leurs différents traitements apportent des éléments de réponse tangibles dans des problématiques environnementales et sociétales. Ces données sont cependant complexes à gérer, pérenniser et analyser. Une première difficulté porte sur la capacité à découvrir et à localiser des ressources, de type image satellite, susceptibles d'apporter des éléments de savoir dans un contexte et une discipline spécifique (géographie, écologie, santé, urbanisme, ...). Une seconde difficulté porte sur la capacité à intégrer de l'information provenant de plusieurs images acquises à différents intervalles, dans différents espaces et à des résolutions différentes. Dans ce travail, nous nous plaçons dans une optique de découverte, exploitation et d'intégration de données, possiblement distribuées et hétérogènes. Pour ce faire, nous nous sommes largement appuyés sur les approches basées sur les métadonnées et leurs standards afférents. L'infrastructure du web sémantique est également mise à profit afin de dégager une approche hybride, modulaire et générique, visant dans un premier temps de l'aide à la découverte d'images satellites sur le web. La même approche pourra être ré-exploitée, dans un second temps, dans une visée d'intégration de données géoréférencées.

Table des matières

Table des matières	2
Introduction et Motivations	4
I Etat de l'art	6
1 Métadonnées et annotations dans le contexte du web sémantique	6
1.1 Introduction	6
1.2 Des langages pour décrire les ressources du web	7
1.2.1 RDF	7
1.2.2 RDFS	8
1.2.3 OWL	9
1.3 Rôles des métadonnées	10
2 Interconnexion de données hétérogènes	11
2.1 Hétérogénéité entre métadonnées	13
2.2 Interopérabilité des métadonnées	14
3 Au delà des standards de métadonnées	16
3.1 Le standard Dublin Core	16
3.2 Profil d'application	17
3.2.1 Principes généraux	17
3.2.2 Bonnes pratiques autour de la construction	19
3.2.3 Profil d'application Dublin Core (DCAP)	20
3.3 Références croisées, crosswalks ou mapping de métadonnées	22
II Stage	25
4 Démarche adoptée	25
5 Mise en œuvre de l'approche	30
5.1 Contexte applicatif	30
5.2 Exigences fonctionnelles	31
5.3 Modèle du domaine	31
5.4 Choix et étude des standards	32
5.4.1 Dublin Core	32
5.4.2 Le standard ISO 19115	33
5.5 construction du modèle pivot	41
5.6 Opérationnalisation des modèles	43
5.6.1 Dublin Core	43
5.6.2 ISO 19115 : UML to OWL	44
5.6.3 Modèle pivot	48
5.7 Crosswalks	48

5.7.1	Identification des crosswalks :	48
5.7.2	Opérationnalisation des crosswalks :	49
5.8	Transformation des métadonnées	50
6	Démonstrateur et validation de l'approche	51
6.1	Outils techniques	52
6.2	Architecture du démonstrateur	53
6.3	Tests et Résultats	54
	Conclusion et perspectives	58
	Références	59

Introduction et Motivations

De nombreux jeux de données géoréférencées sont aujourd'hui localisables, voire rendus disponibles sur le Web, et cela dans des domaines variés tels que les sciences de l'environnement et du vivant. Cependant, ces jeux de données demeurent souvent décrits au travers de différents standards de métadonnées (Darwin Core, Dublin Core, EML, ISO 19115, CSDGM, Sensor ML, ...) pour les besoins propres de chacune des communautés qui les a produites ; ce qui les rend difficilement exploitables au travers d'outils communs de recherche d'information.

Ce constat nous amène à poser une réflexion en terme d'existant autour du rôle facilitateur joué par les métadonnées dans un contexte d'exploitation de gros volumes de données et autour des démarches menant à l'interconnexion de jeux de données géoréférencées. De nombreux travaux sont conduits depuis plusieurs années afin de rendre interopérables différents standards de métadonnées et ainsi permettre la conversion de lots de métadonnées d'un standard à un autre et d'en assurer un échange et une exploitation fédérée efficace.

Dans ce contexte, l'objectif de ces travaux est de proposer une démarche originale visant à faciliter la découverte et la localisation de données spatiales, et plus particulièrement d'images acquises par télédétection dans un contexte où la description des données s'appuient sur des schémas de métadonnées hétérogènes. Aussi, nos travaux s'appuient sur la notion de métadonnées comme support à l'interopérabilité dans les processus de découverte et de localisation d'une ressource. Nous avons choisi d'aborder cette problématique en nous appuyant sur les cadres conceptuels et opérationnels disponibles pour les métadonnées ainsi que sur les langages du web sémantique.

Après avoir exposé les notions de métadonnée au travers de ses finalités dans le contexte du web sémantique, présenté les différentes approches envisageables pour assurer un accès uniforme à des ressources hétérogènes et distribuées, nous explorons le cadre conceptuel apporté par l'initiative Dublin Core[PN06, DHSW02], notamment celui portant sur la mise en œuvre de profil d'application. Enfin, l'état de l'art aborde les différentes techniques visant à rapprocher les métadonnées issues de schémas hétérogènes.

Après avoir motivé et présenté notre démarche, nous détaillons sa mise en œuvre dans le cadre du projet Equipex GEOSUD. Il vise à développer une infrastructure nationale pour l'accès libre à l'imagerie satellitaire pour la communauté scientifique et les acteurs publics.

Conformément à la démarche conçue, un profil d'application répondant aux besoins de découverte des images satellitaires est proposé. Ce profil est enrichi par un ensemble de règles de correspondance, ou crosswalks, permettant de rapprocher les éléments des différents schémas de métadonnées ciblés. Enfin, un démonstrateur est présenté et testé. Implémenté sur la base du *framework* sémantique apache Jena, nous analysons deux approches de crosswalks, l'une opérant au niveau des schémas de métadonnées, l'autre au niveau des instances de métadonnées. Nous concluons ce mémoire en exposant l'originalité de l'approche visant à l'interconnexion de données spatiales et en fournissant ses limites. Les perspectives sont nombreuses, les données pouvant être renseignées à partir d'images satellites sont multiples et une suite di-

recte de travail pourrait être de permettre une consultation fine du contenu en matière d'entités géographiques des images au travers des métadonnées.

Première partie

Etat de l'art

1. Métadonnées et annotations dans le contexte du web sémantique

1.1 Introduction

Le web met à la disposition des internautes, d'innombrables ressources qu'il s'agit de localiser, d'organiser et d'exploiter de manière adaptée. La réflexion qui a conduit au web sémantique, encore appelé web de données, a été d'envisager le web comme une gigantesque base de données, mettant les notions de métadonnées et de schéma de métadonnées au cœur de ces propositions. Les métadonnées sont définies comme des données sur des données et permettent d'organiser et de structurer les ressources disponibles afin de faciliter leur indexation pour la recherche, et optimiser leur exploitation par des agents logiciels [PG04]. Le terme ressource est pris dans son acception la plus générale, comme étant toute entité concrète ou abstraite, susceptible d'être identifiée, nommée, et manipulée à travers de multiples représentations. Dans cette vision utilitaire, les métadonnées sont souvent assimilées à des annotations sémantiques qui viennent enrichir les données présentes au sein de ressources, par exemple textuelles, disponibles sur le web. Différents standards de métadonnées, généralistes ou dédiés à une discipline comme l'écologie, l'hydrologie ou la géographie, structurent ces métadonnées. Dans les domaines des sciences de l'environnement et du géospatial, qui nous intéressent au premier chef, nous pouvons citer les standards de métadonnées Darwin Core (biodiversité), EML (écologie) ou ISO 19115 (géographie). Le standard de métadonnées Dublin Core [NJNP06] a une visée généraliste. Il fournit un socle commun d'éléments comme les éléments **title**, **creator**, ou **coverage** qui vont renseigner sur l'existence de ressources et en faciliter l'exploitation sur le long terme.

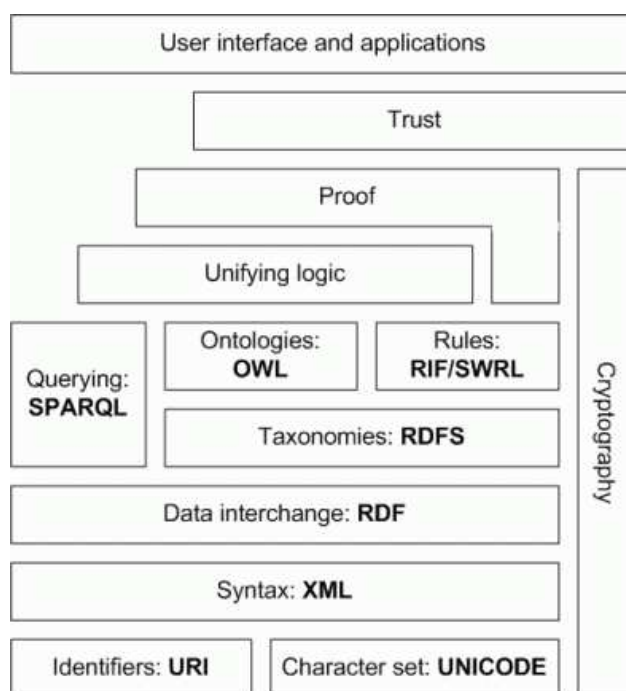


FIGURE 1 – Semantic web stack - Architecture en couches du web sémantique [SBLH06]

1.2 Des langages pour décrire les ressources du web

Afin de rendre ces ressources exploitables dans le cadre du web sémantique, le W3C¹ propose une panoplie de langages. Ces langages ont pour caractéristique d'être tous exprimables et échangeables en XML. Ils peuvent être vus comme des briques indépendantes, ou chaque langage vient répondre à des besoins d'expressivité spécifiques et peut satisfaire ses propres utilisations. Cependant, ces langages entrent dans le cadre d'une vision globale, chaque langage sert alors de support au langage du niveau supérieur, conduisant à une architecture sous forme d'empilement de couches, appelée "Semantic web Stack" (figure 1). Pour un maximum d'efficacité, les couches doivent répondre à différentes préconisations. Actuellement, seuls les niveaux du bas de la pile s'étageant jusqu'au langage OWL sont standardisés. Les niveaux supérieurs ne le sont pas encore et contiennent juste des idées et des principes qui ont fait l'objet d'implémentation partielle. Nous nous limitons à la présentation des langages RDF, RDFS et OWL qui servent de socle à l'expression de métadonnées et à leur structuration au travers de schémas dédiés. Ce n'est cependant pas leur seule fonction.

1.2.1 RDF

Le langage RDF² [MM04], à voir plutôt comme un modèle de données, offre une manière simple d'enrichir une ressource par des métadonnées, sans obliger à

1. World Wide Web Consortium : www.w3.org

2. Resource Description Framework

modifier le contenu de la ressource ni même à accéder véritablement à la ressource. RDF s'appuie sur la notion de triplet ou déclaration {Sujet Propriété Objet} qui va permettre de décrire une ressource (le sujet de la déclaration) au travers d'un couple propriété/valeur (respectivement l'élément de métadonnée et l'objet qui en est la valeur). Un schéma RDF est une collection de triplets et va constituer un graphe. Différents formats sont disponibles pour manipuler un graphe RDF : XML (RDF/XML), N3, Turtle ou encore JSON (JavaScript Object Notation).

Nous donnons un exemple de triplet : un individu, vu comme une ressource, et identifié par l'URI (Uniform Resource Identifier) `exple:individu_1` (définie par la concaténation de l'espace de noms et du nom qualifié) est décrit comme étant de type occurrence du schéma de méta-données Darwin Core. Le vocabulaire Darwin Core[WBG⁺12] facilite le partage d'informations relatives aux espèces vivantes. Ainsi les specimens identifiés et observés sur le terrain voire collectés au sein de laboratoires de recherche vont pouvoir être renseignés au sujet de différentes considérations : occurrence spatio-temporelle, traits phénotypiques, ...

Nous allons enrichir au fur et à mesure la description de la ressource `individu_1` qui nous a été inspirée par le figuier situé dans la cour du bâtiment IRD de Lavalette, et qui nous permet de mettre en oeuvre différents schémas de métadonnées provenant des sciences de l'environnement.

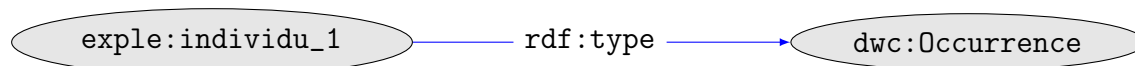


FIGURE 2 – Un exemple de déclaration RDF

Le spécimen de figuier est décrit ci-dessous, au travers de la syntaxe N3. L'espèce végétale concernée (*Ficus carica*) est référencée et la localisation géographique du figuier est notée au travers de la désignation d'un nœud anonyme (`_:blank1`).

```

<http://www.exemple.net/instance#individu_1>
  a dwc:Occurrence ;
  dcterms:created "2013-2-19" ;
  dcterms:spatial _:blank1 ;
  dcterms:subject "figuier";
  dwc:taxonConceptID
  <http://www.ubio.org/authority/metadata.php?lsid=urn:lsid:ubio.org:namebank:448169> ;
  
```

1.2.2 RDFS

Le langage RDFS (Resource Description Framework Schema)[BG04] est un vocabulaire RDF, qui enrichit le langage RDF au travers de différents éléments de modélisation. RDFS introduit notamment la notion de classe, et permet aussi d'organiser les classes et les propriétés au travers de hiérarchies. Dans notre exemple de specimen de figuier, `individu_1` est enrichi par une couverture spatiale au travers de la propriété `dcterms:spatial`. `dcterms:spatial` étend la propriété `dcterms:coverage` de la manière suivante : `dcterms:spatial rdfs:subPropertyOf dcterms:coverage`. Il est donc rendu possible de s'intéresser, de manière générale, à la couverture spatiale et/ou temporelle d'un individu.

```

<http://www.exemple.net/instance#individu_1>
  a dwc:Occurrence ;
  dcterms:created "2013-2-19" ;
  dcterms:spatial _:blank1 ;
  dwc:taxonConceptID
<http://www.ubio.org/authority/metadata.php?lsid=urn:lsid:ubio.org:namebank:448169> ;

_:blank1
  a dcterms:Location ;
  geo:lon "3.8773" ;
  geo:lat "43.61092" ;
  dwc:locality "Montpellier" ;
  dwc:coordinateUncertaintyInMeters "10000" .

```

1.2.3 OWL

OWL³ [DS04] est un langage de représentation des connaissances, qui vient s'adosser aux langages RDF et RDFS. Il introduit de nouveaux constructeurs qui viennent pallier les manques de RDF et de RDFS, dans la perspective de **la construction d'ontologies**. OWL emprunte de nombreux éléments de modélisation aux logiques de description. Il est ainsi possible de définir de nouvelles classes en appliquant des opérations d'union ou d'intersection sur des classes déjà définies, ou encore par restriction de rôles au moyen des quantificateurs universel et existentiel. OWL offre selon le sous-langage considéré (OWL Lite, OWL DL ou OWL Full [DS04]) différents compromis entre l'expressivité du modèle et la complexité et la décidabilité des mécanismes de raisonnement. En ce sens, OWL est tout indiqué pour structurer les éléments descriptifs (classes et propriétés) d'un schéma de métadonnées de façon formelle, et peut donc être utilisé pour traduire un consensus explicite sur la manière de décrire des métadonnées. Afin de mieux comprendre comment le langage OWL vient s'articuler au dessus de RDF et RDFS, et ainsi aborder les constructeurs qui seront rendus disponibles pour la spécification des éléments descriptifs des schémas de métadonnées. Nous présentons, dans un diagramme de classe UML, une portion du méta-modèle de OWL (figure 2).

3. Web Ontology Langage

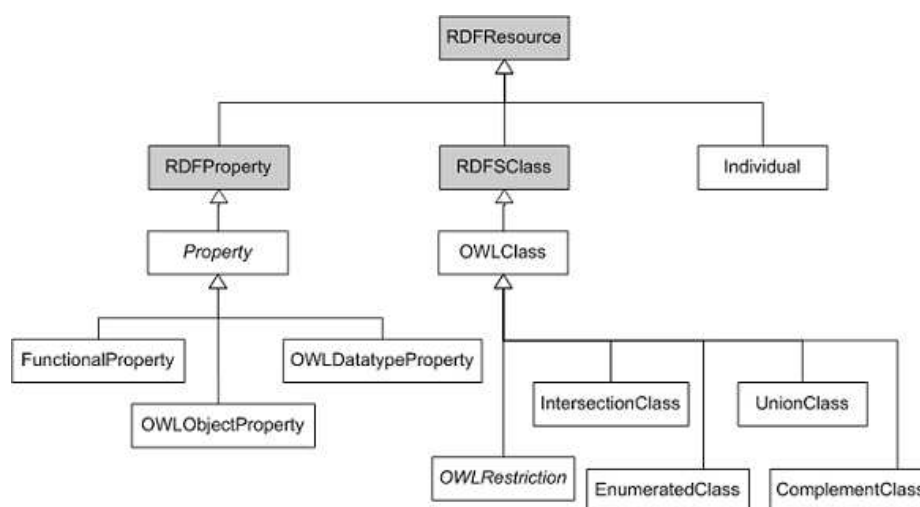


FIGURE 3 – Portion du Métamodèle OWL

1.3 Rôles des métadonnées

Les langages décrits précédemment viennent véhiculer une vision globale du web sémantique. Ils fournissent les moyens nécessaires pour décrire toute ressource disponible, ou tout au moins référençable, au travers du web, et ce de manière normalisée. Disposer de langages communs pour décrire les ressources de manière structurée, est un premier pas vers le partage de ces descriptions et conduit à ce que l'on peut qualifier d'interopérabilité⁴ syntaxique. Nous avons vu également que ces langages n'étaient pas contraints au seul périmètre des métadonnées mais pouvaient également outiller la construction de vocabulaires contrôlés voire d'ontologies pour le partage et le raisonnement sur de la connaissance d'un domaine en particulier. Nous revenons plus en avant sur la notion de métadonnées dans le contexte du web sémantique, pour en préciser les grandes fonctionnalités. Ces fonctionnalités touchent à une exploitation rationnelle de multiples ressources pouvant être mises en commun par l'intermédiaire du web. Ces différents rôles sont centrés sur la ressource. Il va ainsi s'agir pour un ensemble de ressources donné :

- d'en aider à la découverte et à la localisation,
- d'en permettre la consultation des contenus,
- d'en proposer l'organisation et le partage,
- d'en faciliter l'archivage et la pérennisation
- d'en assurer la visualisation et les traitements ultérieurs
- d'en garantir la propriété intellectuelle
- ...

En se concentrant sur les métadonnées, plutôt que sur les ressources, les métadonnées apparaissent aussi comme un moyen d'enrichir voire de créer, par composition, un contenu informationnel, puisque les métadonnées sont aussi des données et viennent donc abonder le socle des ressources. De plus, les métadonnées sont souvent obtenues lors de processus de co-construction impliquant la participation de nombreux

4. Nous définissons ultérieurement l'interopérabilité entre systèmes

usagers. Elles sont le reflet d'un savoir et d'une capitalisation de connaissances collectives.

Pour illustrer notre propos, nous nous intéressons au rôle de recherche d'information et de consultation du contenu des ressources. Les métadonnées permettent d'indexer et de classer le contenu des ressources. Un outil de recherche qui va se baser sur des métadonnées structurées et normalisées, ayant servi à l'indexation des ressources, va permettre de retourner des résultats exhaustifs, fiables et précis.

Nous supposons que nous recherchons les rivières se situant dans le périmètre de la ville de Montpellier. L'utilisation d'un moteur de recherche classique, avec la requête "rivières de Montpellier" retourne des résultats non pertinents, et ce dès les premières pages de résultats. Ces résultats pointent, par exemple, sur des sites proposant des vacances ou des entreprises de recrutement dans Montpellier, sur des pages rédigées dans différentes langues. Seules quelques ressources sont vraiment associées aux rivières de Montpellier. Le recours à un outil de recherche se basant sur les métadonnées, permettrait de retourner des résultats plus satisfaisants au regard de la requête. Différents schémas de métadonnées peuvent être concernés : des spécifications généralistes comme le vocabulaire Dublin Core, qui vont permettre de renseigner le format de la ressource, le type bibliographique (page web, article de journal, de conférence, . . .), et/ou la langue ayant servi à la rédaction des ressources ; ou bien des standards de métadonnées dédiés comme Darwin Core, ISO19115 ou EML qui sont associés à une discipline en particulier : biodiversité, géographie ou écologie, . . .

Yannick Prié et Serge Garlatti[PG04] font état de l'utilisation d'un schéma de métadonnées appelé LOM+ pour composer dynamiquement des enseignements structurés au travers de ce schéma de métadonnées. En ce qui concerne la géographie, le standard de métadonnées ISO 19115 est largement exploité pour la description de ressources géoréférencées et va permettre de répondre aux besoins spécifiques d'aide à la découverte et de localisation de ressources. Ces derniers s'avèrent être des besoins importants notamment dans le cadre des sciences de l'environnement.

2. Interconnexion de données hétérogènes

Un rôle essentiel assuré par les métadonnées porte sur le partage des ressources. Il nécessite de manière sous-jacente de se concentrer sur l'interopérabilité des systèmes qui disposent de ces ressources. Nous reprenons à notre compte, la définition ISO de l'interopérabilité, qui peut être traduite par : "la capacité pour différents systèmes possiblement hétérogènes (en matière de système d'exploitation, plateforme logicielle, modèle de données et interfaces d'accès) d'échanger et de partager des données". Nous allons présenter comment les standards de métadonnées peuvent être utilisées afin de garantir l'interopérabilité des systèmes, en nous appuyant sur les travaux de B.Hashlhofer et W.Klas [HK10] qui proposent une synthèse détaillée des approches existantes. Nous avons choisi d'adopter leur vision de l'interopérabilité et de présenter leurs avancées en détail. Leur synthèse nous a semblé originale et exhaustive par rapport à l'état des travaux actuels.

Comme nous l'avons mentionné au début de cette étude, les sciences de l'envi-

ronnement et du vivant disposent d'un nombre important de ressources, décrites conformément à différents standards de métadonnées. Ces standards ont été créés pour répondre aux besoins de chaque communauté. Il paraît essentiel aujourd'hui de pallier cette hétérogénéité, pour pouvoir disposer d'accès transparents à toutes ses ressources, sans avoir à se soucier de leur localisation réelle.

Avant de présenter les techniques d'interopérabilité, Nous allons d'abord aborder les différents problèmes d'hétérogénéité pouvant affecter l'exploitation de métadonnées lors du partage de données. Ces problèmes sont analysés à différents niveaux d'abstraction et en considérant différents blocs de construction associés aux métadonnées. La figure 4 donne une illustration des différents niveaux distingués [HK10], cette classification des niveaux d'abstraction est inspirée du standard MOF (Meta Object Facilities) de l'OMG⁵. Toutefois ces niveaux ne prennent en considération que les métadonnées et non pas les données.

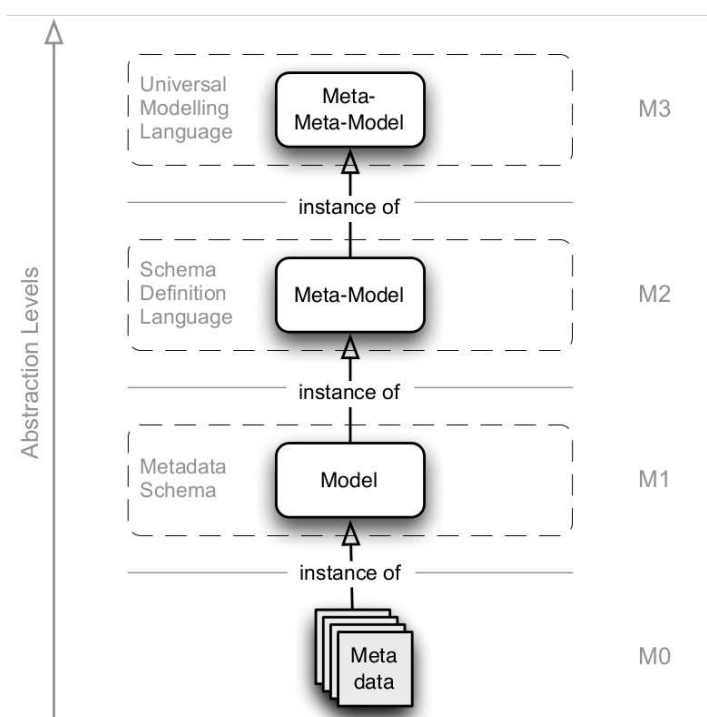


FIGURE 4 – Les différents blocs de construction des métadonnées suivant la vision MOF [HK10]

Le niveau le plus haut de la figure correspond aux **langages de définition des schémas** (M2) : C'est à ce niveau que sont fixées les constructions syntaxiques des métadonnées et leurs définitions sémantiques. Il représente, en d'autres termes, le langage au travers duquel les schémas de métadonnées seront définis. La notation UML, ou le langage OWL se situent à ce niveau. En apparence, UML et OWL ne partagent pas la même vision, et ne sont pas utilisés pour répondre aux mêmes besoins, mais ils peuvent tout les deux être choisis comme langage de représentation de

5. Object Management Group : www.omg.org/

schéma de métadonnées. Et c'est dans cette optique qu'ils se situent tout les deux à ce même niveau d'abstraction.

Le niveau intermédiaire correspond au **schémas de métadonnées**. Ils sont constitués de collections d'éléments dotés d'une sémantique précise et organisant les métadonnées. C'est à ce niveau que porte le consensus sur le vocabulaire pour décrire les métadonnées. Ainsi, la définition des éléments des standards de métadonnées, à l'exemple du Dublin Core ou du Darwin Core sont retrouvés à ce niveau.

Le niveau le plus bas correspond aux **instances de métadonnées**. Les métadonnées et leurs valeurs décrivant les ressources, se retrouvent concrètement à ce niveau.

L'interopérabilité assurée par les métadonnées, est définie dans l'article, comme leur capacité à s'affranchir des frontières entre différents systèmes d'information, de manière à exploiter toute ressource décrite sans entrave liée au contexte ou au système qui l'a produite.

2.1 Hétérogénéité entre métadonnées

Les métadonnées sont donc envisagées sous l'angle des trois niveaux d'abstraction, et nous allons nous intéresser maintenant aux diverses sources possibles d'hétérogénéité et à leur projection sur ces différents niveaux. Cette hétérogénéité peut survenir à n'importe quel niveaux et se répercuter par la suite sur les autres niveaux puisqu'il y a une correspondance entre niveaux. Nous rappelons que dans la vision du MOF (figure 4), une relation d'instanciation se noue entre les différents niveaux, et de ce fait, les métadonnées sont l'instance de leur schéma de métadonnées, qui est lui même une instance du schéma des langages de définition.

Les auteurs distinguent deux grands types d'hétérogénéité : structurelle et sémantique.

L'hétérogénéité structurelle concerne l'hétérogénéité liée à des choix différents dans la manière de représenter les métadonnées, ou de définir les éléments de métadonnées. Cette hétérogénéité se trouve au niveau des langages de définition et des schémas de métadonnées. Elle englobe les conflits de définition, d'identification, de contrainte d'intégrité et d'abstraction sur les éléments, suivant le langage de définition choisi. Par exemple, un conflit peut apparaître quand un choix différent est fait dans la manière d'identifier les éléments : l'identification est assurée par le nom de l'élément uniquement ou bien par un identifiant qualificatif complet (espace de nom plus le nom de l'élément).

L'hétérogénéité sémantique est liée aux différences sémantiques au niveau des schémas des langages de définition et des schémas de métadonnées, ainsi qu'aux différences de représentation au niveau des instances de métadonnées. Cette hétérogénéité peut être due à des problèmes d'expressivité dans les langages de description utilisés ou survenir à la suite de l'emploi de différences terminologiques. Par exemple OWL dispose de constructeurs pour exprimer la disjonction entre deux classes, tandis que cette possibilité n'existe pas dans d'autres langages. Ce type d'hétérogénéité peut apparaître aussi au niveau des schémas de métadonnées. Par exemple, le standard Dublin Core a une visée généraliste, il n'est pas doté d'un vocabulaire pour exprimer explicitement des métadonnées spécialisées de l'environnement, comme le ferait le standard ISO 19115 par exemple. Au niveau des instances de métadonnées, l'utilisa-

tion de systèmes de mesures différents ou différentes représentations pour la même valeur entraîne aussi des hétérogénéités sémantiques.

Ces différents types d'hétérogénéité sont illustrés dans la figure 5 . Cette illustration remplace les sources d'hétérogénéité que nous venons d'aborder, sur les différents niveaux d'abstraction relatifs aux métadonnées.

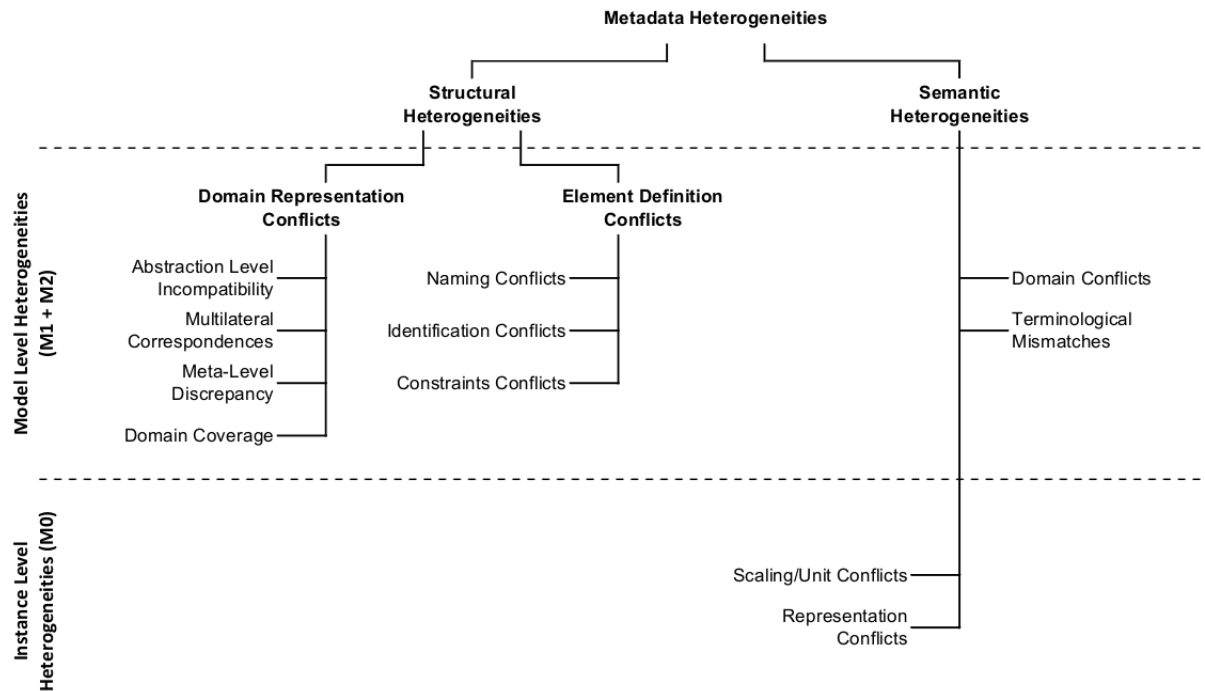


FIGURE 5 – Les différents types d'hétérogénéité des métadonnées par rapport aux niveaux d'abstraction [HK10]

2.2 Interopérabilité des métadonnées

Différentes techniques ont été développées pour mettre en pratique l'interopérabilité à partir des métadonnées. Dans cette partie, nous présentons les plus répandues. Cela nous permet de nous positionner sur les choix possibles en terme de techniques applicables dans le contexte des sciences de l'environnement et la vie.

Nous commençons par **la standardisation**, appelée aussi **la normalisation** . Cette technique se fait graduellement sur les différents niveaux d'abstraction (Figure 4), respectivement le niveau des schémas de langages de définition, le niveau des schémas des métadonnées et le niveau d'instance des métadonnées. Elle consiste à adopter un schéma de langage de définition unique pour la description des métadonnées, poursuivi par un accord sur le schéma de métadonnées à utiliser (vocabulaire contrôlé). Elle peut aller au niveau des instances si un consensus sur les valeurs peut être établi (en définissant des listes ou des intervalles de valeurs possibles).

Une autre technique d'interopérabilité consiste à adopter un méta-modèle : **Meta-model agreement** [HK10]. Cette technique est adaptée pour les institutions qui

adoptent déjà leurs propres standards et schémas de métadonnées. Il est en effet difficile d'appliquer un standard unique sur toutes les métadonnées. Dans les sciences de l'environnement et du vivant par exemple, un nombre important de ressources sont décrites avec des métadonnées s'appuyant déjà sur des standards. Il existe différentes approches pour mettre en œuvre cette technique. En adoptant par exemple l'approche du MOF [Gro06], ou en définissant un méta-modèle abstrait référencé sur le niveau des schémas de langages de définition. On peut aussi aller vers un modèle **conceptuel global abstrait**. Cette technique reste relativement assez difficile à mettre en œuvre. Parce qu'il faut prendre en considération tous les standards employés pour décrire les métadonnées à rendre interopérable. Mais elle permet théoriquement de pallier efficacement l'hétérogénéité et fournir un moyen efficace d'interopérabilité. Nous allons dans la partie suivante présenter le standard Dublin Core et son utilisation pour la mise en œuvre de profils d'application. C'est une approche qui s'inspire beaucoup de cette technique.

Une autre technique pour l'interopérabilité est le **Mapping de Métadonnées** (ou *crosswalk*). Cette technique englobe une sous-technique de *mapping* de schémas et une autre pour la transformation d'instances. Le *mapping* de métadonnées pourrait être découpé en quatre étapes majeures (figure 6) : La première consiste à trouver les relations sémantiques et structurelles entre les éléments des deux schémas. La deuxième étape est la phase de formalisation des relations trouvées. La troisième permet d'établir ces relations. La dernière consiste à la maintenance possiblement évolutive de ces relations sur le long terme.

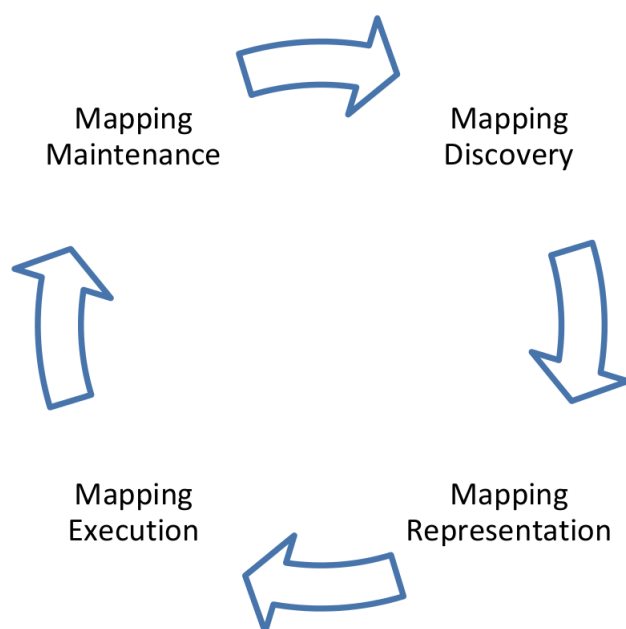


FIGURE 6 – Les quatre étapes majeures de la technique de *mapping* de métadonnées[HK10]

En présentant toutes ces techniques, nous démontrons qu'il existe plusieurs solutions pour aborder l'interopérabilité. Mais néanmoins, le choix d'utilisation d'une de

ces techniques doit être fait en étudiant le contexte d'application ainsi que d'autres facteurs. B.Haslhofer et W. Klas ne retiennent que les techniques **de standardisation** et **de mapping des métadonnées**. D'après les comparaisons qu'ils ont effectué, seules ces deux techniques couvrent des champs d'hétérogénéité assez larges. Elles ont aussi l'avantage de pouvoir être utilisées efficacement avec d'autres techniques d'interopérabilité sur le niveau d'instance de métadonnées, et cela afin de couvrir toutes les formes d'hétérogénéité (figure 5). Un autre paramètre important intervient dans le choix de la technique d'interopérabilité à utiliser, c'est le coût de la mise en œuvre. C'est un élément clé et à cet effet, Hashlhofer et W. Klas [HK10] présentent une analyse des facteurs de coût entre la standardisation et le "mapping" des métadonnées afin d'aider les experts à faire leurs choix.

3. Au delà des standards de métadonnées

Nous allons nous centrer dans cette section, sur une variante de ce que Haslhofer et Klas [HK10] désignent par technique du *meta-model agreement*. Nous pensons en effet que la méta-modélisation est une piste à explorer. L'initiative Dublin Core⁶ propose dans ce contexte le modèle abstrait du Dublin Core **DCAM** qui offre une représentation générique de toute ressource décrite par un ensemble de métadonnées. L'idée est de compléter les standards de métadonnées par différentes lignes directrices facilitant la mise en place d'applications tirant parti de métadonnées. Nous revenons dans un premier temps sur le standard Dublin Core.

3.1 Le standard Dublin Core

Avec l'émergence du web sémantique, différents standards de description ont pris une place prépondérante et constituent aujourd'hui des références incontournables pour la production de métadonnées. Ces standards peuvent être distingués en deux types : des **standards spécialisés** pour des domaines de connaissances bien précis, à l'exemple du standard ISO 19115 principalement dédié aux métadonnées géographiques ; et des standards généralistes, appelés **standards de découverte**, visant la description des ressources de manière générale. Ces derniers standards sont souvent constitués d'un ensemble restreint d'éléments. Parmi les plus utilisés, nous retrouvons Dublin Core, c'est un standard générique, visant à décrire des ressources de type quelconque. Dublin Core s'appuie sur quinze éléments de description officiels pour décrire les ressources (tableau 1).

6. DCMI Dublin Core Metadata Initiative

Élément Dublin Core	Description de l'élément
Title	Titre principal de la ressource
Creator	Personne morale ou physique à l'origine de la création de la ressource
Subjet	Sujet principal de la ressource ,Mots-clefs ou principales idées
Description	Description du contenu de la ressource décrite
Publisher	Nom de la personne morale ou physique à l'origine de la publication de la ressource
Contributor	Nom d'une personne physique ou morale qui contribue ou a contribué à l'élaboration de la ressource. Chaque contributeur fait l'objet d'un élément Contributor séparé
Date	Date d'un évènement dans le cycle de vie de la ressource
Type	Genre du contenu de la ressource
Format	Type MIME, ou format physique de la ressource
Identifier	Identificateur non ambigu, référençant la ressource dans un contexte donné
Source	Ressource dont dérive la ressource décrite, que ce soit en totalité ou en partie de la ressource en question
Langage	Langue du contenu de la ressource
Relation	Lien avec d'autres ressources.
Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle
Rights	Droits de propriété intellectuelle au sujet ou sur la ressource

TABLE 1 – Dublin Core Metadata Element Set

3.2 Profil d'application

3.2.1 Principes généraux

Un profil d'application [HP00, NMJE09], est attendu de faciliter les activités de découverte, de caractérisation et de consultation de ressources potentiellement distribuées et hétérogènes, mais appartenant toutes au même champ disciplinaire (télédétection, géographie, biologie, médecine, ...). Dans cette perspective, un profil d'application est constitué d'un ensemble d'éléments de métadonnées pouvant provenir de différents standards de métadonnées, et de lignes de conduite clarifiant les manières d'exploiter le profil, et explicitant les rôles joués par ce dernier dans un contexte applicatif précis. En filigrane, un profil d'application peut contribuer de manière significative à l'interopérabilité de sources de données, tout en s'adossant sur différents standards de métadonnées. Toute la subtilité pour un profil d'ap-

plication réside dans la capacité à répondre aux attentes fonctionnelles spécifiques d'une communauté tout en ne négligeant pas les besoins inter-communautaires en terme d'usage intégré sur des données et des services. Un principe clé est alors de n'exploiter que des standards de métadonnées pré-existants pour construire un profil d'application. Si les concepteurs du profil d'application dérogent à cette règle, ils sont alors tenus de définir et de maintenir de manière publique et sur le long terme, un nouveau standard couvrant les éléments de métadonnées nouvellement introduits. Les volontés autour des profils d'application sont donc de publier jusqu'à la manière de mettre en œuvre les démarches d'enrichissement et d'exploitation de ressources au travers de métadonnées. Les profils d'application se distinguent donc par les caractéristiques suivantes [HP00] :

- Les profils d'application peuvent être constitués à partir d'un ou plusieurs schémas de métadonnées.
- Tous les éléments de métadonnées appartiennent à des schémas existants. L'introduction d'un nouvel élément de métadonnée entraîne obligatoirement la création et la maintenance ouverte et sur le long terme, d'un nouveau schéma de métadonnées.
- Les profils d'application peuvent contraindre les valeurs pouvant être prises par les éléments de métadonnées.
- Les profils d'application peuvent raffiner⁷ les éléments des standards utilisés.

Des plate-formes (ou "frameworks") sont proposées dans différentes initiatives communautaires à l'exemple du LOM (Learning Object Metadata et schémas de métadonnées pédagogiques), du Dublin Core, ou du MODS (Metadata Object Description Schema) pour faciliter les activités de construction de profils. Ainsi, pour ce qui concerne l'initiative autour du Dublin Core, le framework nommé Singapore et le profil d'application Dublin Core (Dublin Core Application Profile ou DCAP) viennent en appui de toute démarche de définition de profil. Des préconisations méthodologiques [NMJE09] sont également rendues disponibles, et la notion de profil d'ensemble de descriptions (Description Set Profile ou DSP) fournit un cadre prescriptif à la construction d'un profil d'application et s'appuie à ce titre sur le modèle de données abstrait du Dublin Core (DCAM). Dans ce cadre, un profil d'application va traduire différentes contraintes syntaxiques portant sur les ressources qui vont être décrites au travers d'enregistrements de métadonnées constitués de propriétés (éléments de schémas de métadonnées ou éléments de données notés aussi DE pour Data Element) éligibles dans le contexte du profil d'application, et de valeurs pour ces propriétés qui respectent les règles formulées pour ce qui relève des vocabulaires de valeurs référencés (vocabulary encoding schemes) et de formats d'encodage prédéfinis (syntax encoding schemes). Un profil d'application œuvre donc au niveau syntaxique plutôt que sémantique (interopérabilité structurelle entre profils d'application 7).

7. Un raffinement restreint la signification d'un élément, mais sans la changer fondamentalement.

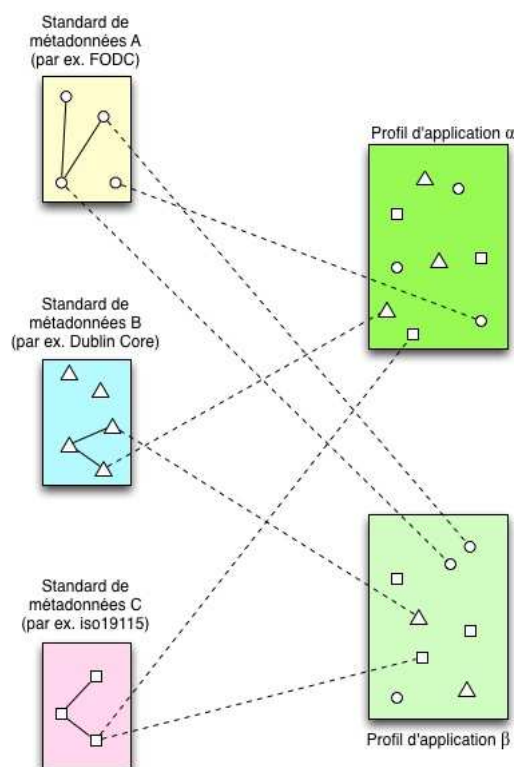


FIGURE 7 – Profil d'application

3.2.2 Bonnes pratiques autour de la construction

Un profil d'application se définit au travers de différentes activités de structuration et de documentation qui sont soit rendues obligatoires, soit laissées optionnelles à l'appréciation du concepteur du profil. Les activités obligatoires comprennent de manière successive :

- les exigences fonctionnelles qui précisent le périmètre d'utilisation du profil d'application,
- le modèle de domaine qui organise les entités prises en charge par le profil d'application. Dans le même esprit que pour les exigences fonctionnelles, il s'agit de définir la portée du profil d'application au travers d'un modèle. Très fréquemment, un modèle conceptuel est ainsi construit sous la forme d'un diagramme de classes UML. Les classes représentent les ressources qui vont pouvoir être décrites par les éléments de métadonnées présents dans les standards. Dans notre contexte, ces classes vont pouvoir être, par exemple, la classe **Image** ou encore la classe **Capteur**
- le profil d'ensemble de descriptions (DSP ou Description Set Profile) qui permet de définir les éléments de métadonnées qui sont retenus pour décrire les entités d'intérêt dans le périmètre restreint du profil d'application.

Des activités de documentation, comme par exemple la rédaction d'un guide de mise en application du profil, peuvent venir compléter de manière optionnelle la spécification du profil d'application. De même, il peut être utile de fournir certaines préconisations concernant les syntaxes d'encodage (en particulier XML et RDF)

adaptées dans le contexte du profil.

3.2.3 Profil d'application Dublin Core (DCAP)

L'initiative Dublin Core met à disposition différentes spécifications pour accompagner la construction de profils d'application. A l'actif du Dublin Core, la plateforme de Singapore 8 propose une organisation en trois couches des différentes briques (pour l'essentiel langages de description, modèles et vocabulaires contrôlés) qui vont concourir à la construction du profil. Au sein de cette architecture, nous ne détaillons que le modèle abstrait du Dublin Core (Dublin Core Abstract Model ou DCAM) qui apparaît comme très pertinent dans le contexte d'une exploitation commune de différents standards de métadonnées.

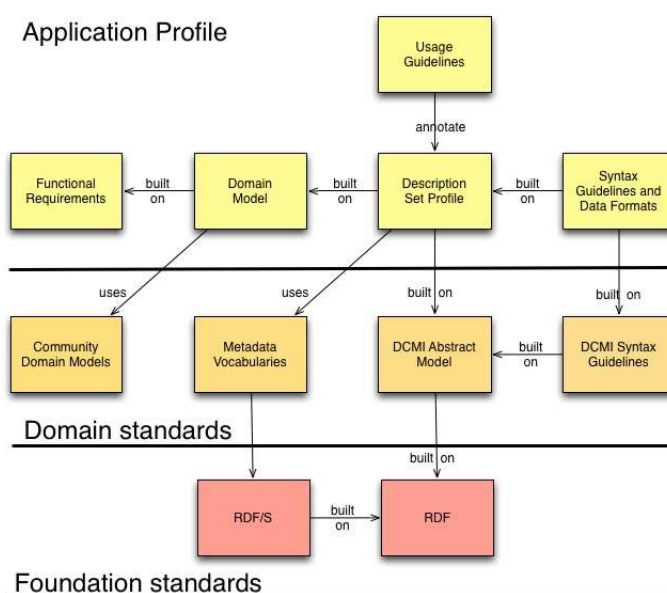


FIGURE 8 – Architecture de la plateforme de singapore [extrait de la documentation]

Le modèle abstrait du Dublin Core [PN06] établit une différence nette entre la ressource décrite et les métadonnées venant décrire cette ressource et intègre, à cet effet, deux modèles nommés "DCMI⁸ resource model" et "DCMI description model". La documentation mise en ligne par le DCMI⁹ fait état de différents diagrammes de classes UML qui facilitent la compréhension du modèle générique, qui en réalité comprend plusieurs modèles qui se complètent. Un diagramme de classes est également rendu disponible pour la typologie de tous les termes gérés au travers de l'initiative "DCMI", incluant classes, propriétés, schémas de vocabulaire et de syntaxe. L'ensemble des spécifications sont par ailleurs disponibles au format rdf/xml. Les attentes du DCMI sont claires : il ne s'agit pas seulement de fournir un schéma de données généralistes mais aussi de fournir des facilités pour l'exploitation de métadonnées dans tout contexte d'application. Ainsi tous les besoins sont

8. DCMI : Dublin Core Metadata Initiative

9. <http://dublincore.org/documents/abstract-model/>

pris en considération : schéma de métadonnées, schéma de métadonnées qualifié pour aller plus loin dans la structuration des éléments de métadonnées mais aussi dans le contrôle des valeurs prises par ces éléments au travers de vocabulaires contrôlés appropriés et de formats d'encodage adéquats, exploitation conjointe de différents schémas de métadonnées, Nous proposons, dans la Figure 9, le diagramme de classes UML décrivant une ressource ("DCMI resource model"), qui exprime, de façon très générale, l'enrichissement d'une ressource au travers de collections de métadonnées. Un tel cadre conceptuel peut servir de socle à l'établissement de références croisées entre métadonnées provenant de différents schémas (mapping de métadonnées ou "crosswalks") ou bien à la définition de profils d'application (approche "mix and match") qui consistent à faire le choix d'éléments de métadonnées pertinents provenant de différents schémas de métadonnées.

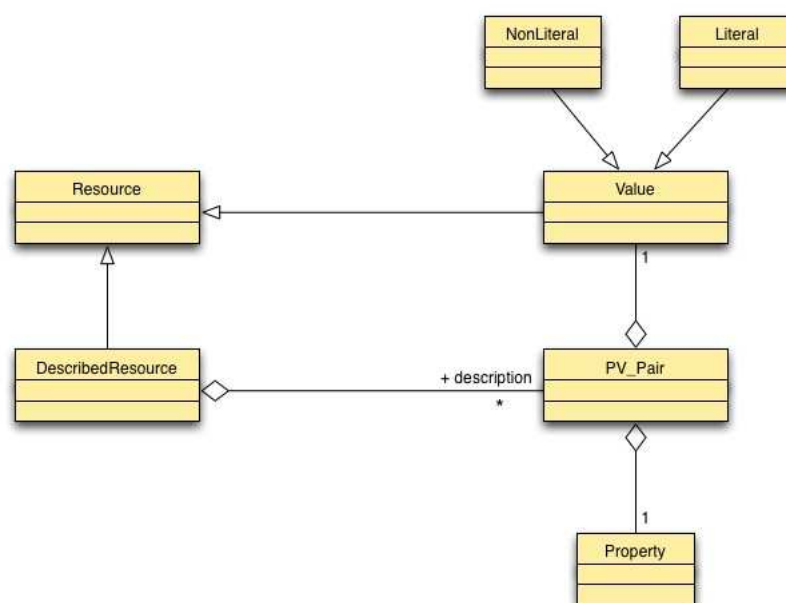


FIGURE 9 – modèle abstrait de ressources (DCAM)

Le modèle DCMI spécifie les règles suivantes :

- Chaque ressource est décrite en utilisant une ou plusieurs paires propriété-valeur.
- Chaque paire propriété-valeur est composée d'une et une seule propriété et une et une seule valeur.
- Chaque valeur est une ressource : physique, entité numérique ou conceptuelle ou littérale qui est associée à une propriété lorsqu'un couple propriété-valeur est mis à contribution pour décrire une ressource. Par conséquent, chaque valeur est soit une valeur littérale ou une valeur non littérale.
- Une valeur non littérale est une valeur qui est une entité physique, numérique ou conceptuelle.
- Un littéral est une entité qui utilise une chaîne Unicode en tant que forme lexicale, avec une étiquette de langue facultative ou type de données, pour désigner une ressource.

3.3 Références croisées, crosswalks ou mapping de métadonnées

Les éléments présents dans différents standards de métadonnées peuvent être rapprochés au travers de liens désignés par "crosswalks", qui peuvent, par exemple dans les cas les plus simples, traduire une équivalence entre deux éléments et/ou instances de métadonnées définis dans deux schémas de métadonnées indépendants l'un de l'autre. A titre d'illustration, l'élément nommé **Originator** dans le standard FGDC, correspond à l'élément nommé **Author** dans le standard Marc21 et **Creator** dans le standard Dublin Core. Les objectifs d'une telle mise en correspondance est de pouvoir interroger de manière unifiée des jeux de métadonnées qui à l'origine n'exploitent pas les mêmes standards, ou encore d'automatiser la conversion d'un jeu de métadonnées d'un standard à l'autre. Les principes suivis sont semblables aux principes appliqués dans le contexte de l'alignement d'ontologies [ES07]

Différentes définitions des crosswalks sont données dans la littérature, et nous retenons la définition proposée dans les travaux de [Su02] : définir des crosswalks entre deux schémas de métadonnées notés A et B, revient à établir des liens, entre les concepts et propriétés du schéma A sémantiquement équivalents ou similaires dans le schéma B, et inversement. Ces liens, qui jouent le rôle de charnières sémantiques, sont souvent binaires et impliquent alors une entité de chaque modèle de métadonnées. Ils peuvent toutefois être aussi n-aires et ainsi mettre en relation plusieurs éléments de chacun des deux schémas de métadonnées [RWB11]. De même ces liens peuvent traduire non seulement des relations d'équivalence, mais aussi d'inclusion ou de plus grande spécificité/généralité. Par exemple, l'élément **Place keyword** du standard FGDC est associé à l'élément **Coverage** du standard Dublin Core par un lien que l'on pourrait désigner par **hasBroader** (a pour moins spécifique).

La construction d'un système de références croisées demeure un problème ouvert et nécessite d'introduire des mécanismes suffisamment flexibles pour prendre en charge les différents conflits de mise en concordance qui peuvent survenir. Dans notre cadre d'étude, et dans cette perspective, nous nous intéressons aux technologies du web sémantique, et en particulier, au langage OWL qui va nous permettre de définir les propriétés objet adaptées à de l'alignement de schémas de métadonnées. OWL introduit en effet des primitives de modélisation qui vont concourir à l'établissement des crosswalks. Nous pouvons citer, de manière non exhaustive, certaines propriétés objet : `owl:equivalentProperty`, `owl:sameAs`, `rdfs:subClassOf` ou encore `owl:unionOf` qui vont permettre de poser des comparaisons entre instances, classes et propriétés. Pour éclairer les problèmes posés, nous introduisons deux schémas de métadonnées, définis en OWL pour l'exemple, qui portent tout deux sur de la description d'images. Les métadonnées sont déclarées sous la forme de propriétés OWL de type `DatatypeProperty` et `ObjectProperty` et révèlent un niveau d'enrichissement élémentaire portant sur l'auteur, la date ou encore le lieu de la prise de vue.

```
<rdf:RDF
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <owl:Ontology rdf:about="http://www.exemple.net/ontologies/schema-A"/>

  <owl:Class rdf:about="http://www.exemple.net/schema-A#Auteur"/>
```

```

<owl:Class rdf:about="http://www.exemple.net/schema-A#Image"/>

<owl:ObjectProperty rdf:about="http://www.exemple.net/schema-A#A_Auteur">
  <rdfs:range rdf:resource="http://www.exemple.net/schema-A#Auteur"/>
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Image"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#date"/>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#nom">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Auteur"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#pays">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Image"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#prenom">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Auteur"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#sujet">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Image"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#taille">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Image"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-A#ville">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-A#Image"/>
</owl:DatatypeProperty>

</rdf:RDF>

<rdf:RDF
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <owl:Ontology rdf:about="http://www.exemple.net/ontologies/schema-B"/>

  <owl:Class rdf:about="http://www.exemple.net/schema-B#Photo"/>

  <owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-B#adresse">
    <rdfs:domain rdf:resource="http://www.exemple.net/schema-B#Photo"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-B#date">
    <rdfs:domain rdf:resource="http://www.exemple.net/schema-B#Photo"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-B#heure">
    <rdfs:domain rdf:resource="http://www.exemple.net/schema-B#Photo"/>
  </owl:DatatypeProperty>

```



```

<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-B#photographe">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-B#Photo"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="http://www.exemple.net/schema-B#sujet">
  <rdfs:domain rdf:resource="http://www.exemple.net/schema-B#Photo"/>
</owl:DatatypeProperty>

</rdf:RDF>

```

Nous pouvons d'ores et déjà identifier des conflits de nommage (classe Image dans le schéma A et classe Photo dans le schéma B), des conflits structurels (photographe qui est une DatatypeProperty dans le schéma A et A_Auteur qui est une ObjectProperty dans le schéma B) qu'il s'agit de prendre en charge lors de la définition des crosswalks. De même des éléments de métadonnées se retrouvent sans correspondance dans le schéma A comme dans le schéma B.

Deuxième partie

Stage

4. Démarche adoptée

Les standards de métadonnées vont faciliter la description et l'enrichissement des données. ils sont cependant dédiés à une problématique en particulier et évoluent à intervalles réguliers. De nouveaux standards viennent se surajouter et il se pourrait aussi que certains standards ne soient plus maintenus et deviennent obsolètes.

Nous nous focaliserons dans notre approche sur la manière d'exploiter en commun différents standards de métadonnées afin de décrire au mieux des ressources géoréférencées. Ces ressources englobent en particulier des images satellites dont l'acquisition se fait aujourd'hui de manière continue et qui viennent alimenter la production d'indicateurs clés apportant des éléments de réponse face à des enjeux environnementaux (érosion de la biodiversité, gestion de risques épidémiologiques, déforestation de l'Amazonie, fragilisation des agrosystèmes oasiens ...). L'idée est de rendre les jeux de métadonnées venant enrichir ces ressources interchangeables (en d'autres termes permettre la conversion d'un standard de métadonnée à un autre lorsque cela est possible) et également de pouvoir poser des interprétations sur la base de ces jeux de descriptions. Dans ce contexte, nous nous appuyons sur les notions de profils d'application [CB08] [HP00] et des crosswalks [Su02] [ES07] qui offrent des solutions à une manipulation concertée de différents standards.

Nous avons expliqué, dans les deux sections précédentes, comment les approches des profils d'application et des crosswalks proposent d'opérer l'interopérabilité. Nous avons choisi de nous intéresser à ces deux approches, pour ce qu'elles offrent en terme d'efficacité et de potentialité d'exploitation des métadonnées. Dans cette partie, nous allons détailler les motivations qui nous ont poussés à proposer une approche hybride, tirant parti à la fois des profils d'application et des crosswalks. L'approche nous a permis de dégager un profil d'application que l'on pourrait qualifier de sémantique puisque enrichi de liens sémantiques. La démarche est ensuite validée par une mise en application.

Motivations de l'approche Les profils d'application offrent la possibilité d'utiliser des éléments de différents standards dans un contexte spécifique, comme par exemple la consultation de ressources pédagogiques distribuées, ou encore l'aide à la localisation d'images satellites, et de rendre efficace l'utilisation de ces éléments de métadonnées, à cet effet. Elle peut répondre à ce titre à des besoins spécifiques et bien ciblés. Mais elle n'est adaptée qu'à des cas d'utilisation et des schémas de données prédéfinis, et ne prévoit pas d'ouverture sur d'autres schémas après construction du profil d'application.

D'un autre côté, l'approche des *crosswalks* offre une solution efficace pour rendre différents schémas de métadonnées, interopérables. Elle nécessite néanmoins la disponibilité d'un schéma préalable, qui servira comme modèle d'appui ou modèle pivot, pour la définition de crosswalks et peut être délicate à mettre en œuvre sur des

schémas structurellement et sémantiquement très différents.

L'idée est donc de favoriser l'ouverture des standards de métadonnées exploités dans le profil d'application construit à différentes conversions qui vont favoriser un accès et une mutualisation efficace des ressources géoréférencées. Cette ouverture des standards est assurée par les *crosswalks* et une étape importante de notre approche consiste à construire le modèle pivot, à partir des éléments de métadonnées choisis, sur lequel viendront s'articuler les *crosswalks*. La construction de ce modèle s'effectuera suivant la méthodologie de construction de profil d'application, proposée par la communauté Dublin Core [CB08]. Une fois le modèle pivot défini, des *crosswalks* seront dégagés pour lier les différents éléments du modèle aux standards et schémas de métadonnées à rendre interopérables.

Étude du domaine La première étape de l'approche est d'établir un modèle du domaine. Ce modèle de domaine doit décrire les entités - acteurs et/ou ressources du domaine - que les métadonnées doivent cibler, ainsi que les relations et interactions qui existent entre ces entités. Le modèle du domaine peut prendre le format d'un simple modèle, ou bien rentrer dans les détails les plus précis du domaine concerné. En effet, le degré de complexité dépend des besoins fonctionnels du profil souhaité en termes de précision et/ou efficacité. C'est aux experts chargés de la modélisation de décider du degré de cette complexité.

La définition du modèle du domaine reste une étape importante, elle permet d'acquérir une vision claire sur les besoins réels sans s'attacher en avance à un standard ou schéma préétabli.

Définition des termes de métadonnées La sélection des termes de métadonnées est fortement dépendante du modèle du domaine, elle consiste à choisir l'ensemble des éléments de métadonnées qui joueront le rôle de descripteurs des entités et des relations exprimées dans le modèle du domaine. Ces termes serviront de corpus pour la conception du schéma de métadonnées. Ce schéma sera appelé **modèle pivot**. Afin de mieux cerner les termes de métadonnées à exploiter, il est conseillé l'utilisation d'une définition textuelle de ces termes couplée avec le type de valeurs que ces termes pourraient prendre. Afin de dégager ces termes, nous avons entrepris de constituer un questionnaire dont les réponses aux questions facilitent le choix des éléments de métadonnées. Des exemples de questions extraites du questionnaire sont listés ici :

- Quelles propriétés se révèlent les plus pertinentes pour décrire les entités du modèle ?
- Quelles propriétés servent à décrire les relations qui existent entre les différents acteurs et ressources ?
- Faut-il typer les valeurs des propriétés ou les laisser sans restrictions ?
- Faut-il utiliser un vocabulaire contrôlé pour les valeurs des propriétés ? Existe-t-il des listes pré-définies par des organismes internationaux traitants le même type de valeurs (liste langues, pays, ...) ?

A partir des réponses obtenues, les éléments de métadonnées, qui couvrent les différents besoins de description des ressources du modèle, sont alors dégagés.

Dans l'essentiel des contextes d'étude, les éléments de métadonnées sont à puiser

dans des standards de métadonnées existants et l'étude de ces standards est alors rendue nécessaire.

Analyse des standards Une étude approfondie des standards de métadonnées permet d'exploiter à la fois les éléments et la structuration des éléments de manière pertinente. Nous proposons, à la suite de cette étude, de capitaliser la connaissance acquise au travers de tableaux descriptifs de chaque standard. Chaque tableau doit comporter le nom de l'élément, sa position hiérarchique, sa description et le format du contenu (DataType). Ces tableaux serviront de référence à consulter, simplifiant ensuite la construction du modèle pivot.

Profil d'application : Construction d'un modèle pivot A partir des standards étudiés et des termes de métadonnées établis, il s'agit de dégager un profil d'application comportant les éléments les plus pertinents pour la description des ressources. Cela se passe par la sélection des éléments, qui répondront le mieux aux termes de métadonnées établis pour la description du domaine étudié, à partir des tableaux des standards. Ces éléments ne doivent comporter aucune redondance, par exemple deux termes de deux standards différents ayant une grande similarité ne doivent pas être utilisés en simultané. Une fois ces éléments obtenus, ils vont être liés sémantiquement pour construire le schéma de métadonnées de notre domaine d'étude. Ce schéma constituera le modèle pivot.

Après la construction du modèle avec les éléments des différents schémas, des raffinements peuvent être effectués pour mieux répondre aux exigences fonctionnelles du modèle du domaine. Ces raffinements peuvent prendre plusieurs formes, ils peuvent porter par exemple sur la modification des définitions des éléments, apporter plus de précisions sur un élément. Ils peuvent aussi porter sur les valeurs de ces éléments, utilisation d'un vocabulaire contrôlé à la place d'une valeur libre, ou bien assouplir les contraintes sur le type de valeur d'un élément. Ces raffinements servent à adapter les éléments utilisés des autres schémas à leur utilisation dans le modèle pivot, pour apporter une cohérence général dans ce modèle.

Crosswalks A ce stade, nous disposons d'un modèle pivot qui servira comme schéma unificateur de métadonnées pour la description des ressources ciblées. Dans le but de renforcer les potentialités intégratrices de notre modèle, nous allons établir des *crosswalks* entre le modèle pivot et les autres standards étudiés. Nous allons nous servir des tableaux qui résultent de l'étude des standards et du modèle pivot pour identifier les *crosswalks* possibles.

Ensuite, ces *crosswalks* seront récapitulés sous la forme d'un tableau, qui va préciser les éléments impliqués dans chaque *crosswalk* et les transformations nécessaires pour le passage des standards au modèle pivot. Après cette étape, nous allons obtenir un schéma de métadonnées sémantiquement enrichi, couvrant à la fois le domaine ciblé au travers le modèle pivot, et les autres standards et schémas au travers des *crosswalks*. Nous appellerons ce schéma le **modèle pivot enrichi**.

Opérationnalisation du modèle : Après la spécification du modèle pivot et des *crosswalks*, l'opérationnalisation du modèle est la phase qui consiste à traduire ces

modèles dans le langage RDF. Nous utiliserons les technologies du web sémantique présentées précédemment.

Schéma récapitulatif : Nous résumons l'approche hybride avec le schéma présenté en figure 10, qui illustre l'approche décrite ci-dessus et récapitule les différentes étapes détaillées.

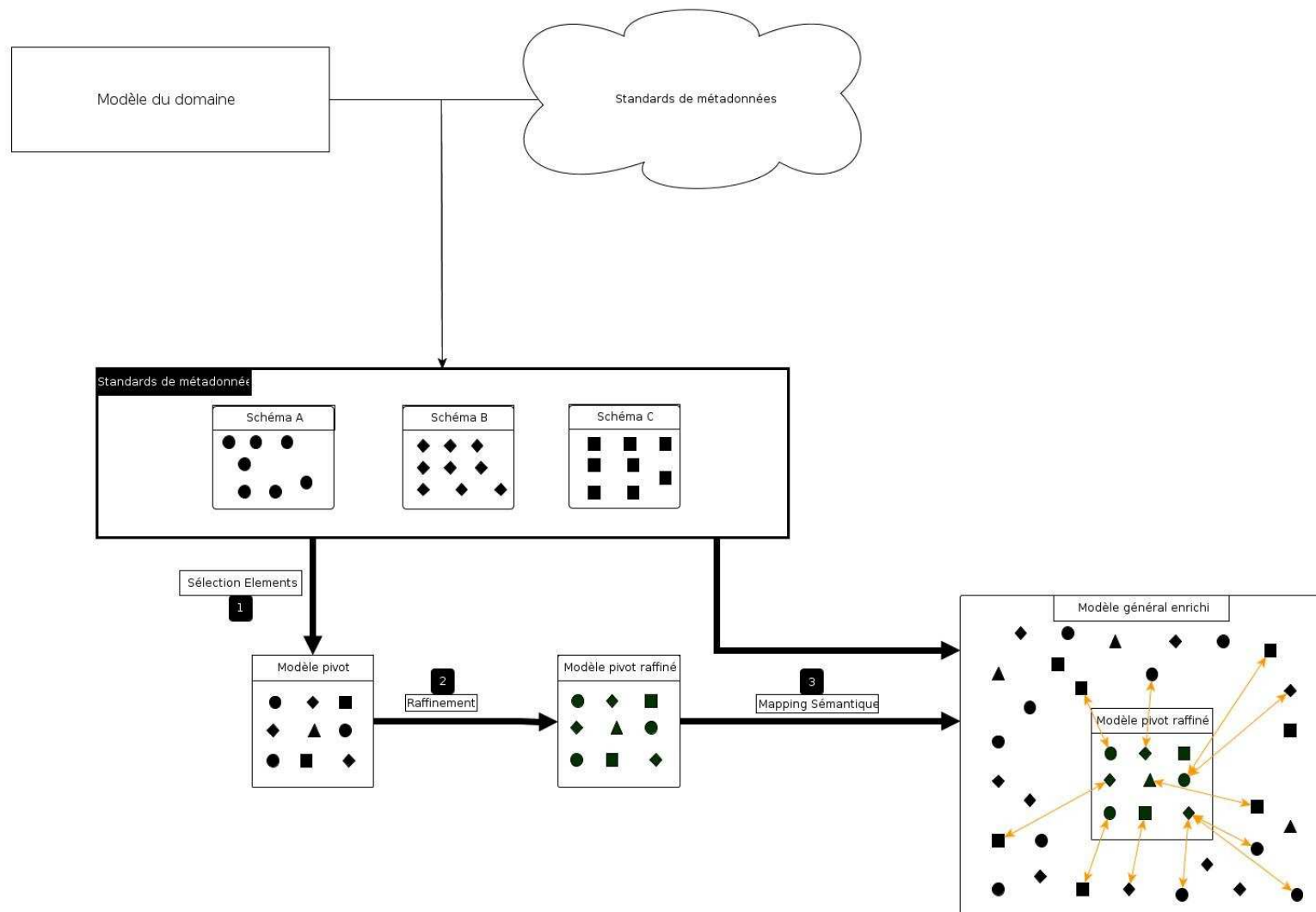


FIGURE 10 – Schéma illustratif des étapes de l'approche hybride

5. Mise en œuvre de l'approche

Après avoir détaillé les différentes étapes de l'approche que nous proposons. Cette partie va décrire le contexte d'application et détailler sa mise en œuvre.

Présentation projet EQUIPEX GEOSUD : Le stage se déroule dans le cadre du projet GEOSUD. C'est un projet ANR Equipex¹⁰, bénéficiant du Programme Investissements d'Avenir. Il vise à développer une infrastructure nationale pour l'accès libre à l'imagerie satellitaire pour la communauté scientifique et les acteurs publics. Il permet, entre autres, d'assurer pendant cinq années l'acquisition et la mise à disposition de couvertures satellitaires, à haute résolution, annuelles du territoire français. Il a aussi pour objectif la mise en réseau de la communauté scientifique, des acteurs publics travaillant sur la gestion des milieux et des ressources du territoire français. Ce projet regroupe quatorze partenaires institutionnels des différentes communautés, ainsi que des acteurs privés du domaine TIC et Environnement.

Il se double d'un enjeu de positionnement national pour les dix à quinze prochaines années, dans le cadre de GMES¹¹, de GEO¹², de *l'Espace au service du Développement Durable et de la Société*, et de l'arrivée de nouveaux acteurs (pays émergents tels que Brésil, Inde, Chine ...) dans le domaine du spatial.

Ce projet est motivé par le constat de sous utilisation des données spatiales issues de la télédétection dans les deux communautés visées. La difficulté d'accès aux données satellitaires à haute et très haute résolution, peut s'expliquer par plusieurs facteurs concomitants : le coût et le type de licence associé à l'utilisation de l'image, la méconnaissance de l'offre, le déficit de formation aux techniques d'exploitation de ces données, le manque de capitalisation et d'animation autour des méthodes de traitements et d'intégration de ces données avec d'autres sources de données (observations in situ, données cartographiques de référence, ...).

5.1 Contexte applicatif

Notre travail se concentre sur la problématique d'accès aux données satellitaires, plus précisément à la découverte d'images spatiales. La recherche d'information géo-spatialisée reste à ce jour une tâche difficile, et présente des difficultés de mise en œuvre. Nous allons exploiter le framework technologique offert par le web sémantique, combiné avec la démarche que nous proposons, pour apporter une solution efficace à l'aide à la découverte d'images spatiales.

La construction d'un modèle de données prenant en charge des aspects spatio-temporels et environnementaux est complexe et, en particulier, nécessite une connaissance approfondie de tout les standards des disciplines afférentes. Par exemple, l'OGC (Open Geospatial Consortium) rend disponible des standards de métadonnées très fouillés à l'exemple du standard ISO 19115. Nous avons donc choisi pour une première application et validation de notre approche de partir sur les éléments essentiels des différents standards, appelés cœurs du standard. Ces éléments sont suffisants

10. Équipements d'excellence

11. Global Monitoring for Environment and Security

12. Group on Earth Observations

pour une application élémentaire de découverte des ressources. Il faut noter aussi que le modèle de données obtenu peut toujours être étendu par la suite pour répondre à d'autres rôles que celui de la découverte.

Nous détaillons les différentes activités qui ont conduit à la construction d'un profil d'application sémantique. A cet effet, nous reprenons à notre compte les activités rendues obligatoires pour toute construction d'un profil d'application et nous y ajoutons une activité de définition de *crosswalks*.

5.2 Exigences fonctionnelles

Le profil d'application en cours de construction est attendu de faciliter le processus de découverte et de localisation de ressources géoréférencées. Nous nous intéressons tout particulièrement aux images spatiales qui concentrent toute l'attention de l'équipex GEOSUD en matière d'acquisition, de gestion et de valorisation de ces images. L'information potentiellement contenue dans ces images est multiple et complexe. L'idée est dans un premier temps, de porter une première réflexion sur la manière de faciliter les processus d'aide à la découverte et à la localisation d'images d'intérêt pour une communauté de pratiques (par exemple, urbanistes, géographes ou écologues du paysage). Dans un second temps, un deuxième objectif pourra être de faciliter les processus permettant la consultation portant sur le contenu des images sur la base des métadonnées complétant ces images.

5.3 Modèle du domaine

Le modèle décrit les entités qui jouent un rôle central dans notre contexte de travail. Cette modélisation est le fruit d'une réflexion portée sur les besoins généraux du domaine et les tâches spécifiques auxquelles doit répondre l'application. Ces entités sont, d'une part les ressources qui vont être rendues localisables et consultables au travers de l'application, et d'autre part les agents qui vont soit créer la ressource, soit contribuer au traitement/enrichissement de la ressource. Une **Ressource** est spécialisée en **Document**, **Traitement**, **Vidéo** et **Image**. La classe **Image** est spécialisée en **Image Spatiale** qui est préférentiellement au centre de nos problématique. Dans les problématiques liées au géospatial et à la télédétection, les traitements, appliqués notamment sur les images, revêtent énormément d'importance. Nous les représentons donc au travers d'une classe à part entière, de manière à pouvoir renseigner également les usagers sur les traitements qu'ils seraient susceptibles de lancer sur les images. La classe **Agent** généralise les acteurs humains ou électroniques (**Organisation**, **Individu**, **Capteur**) qui vont être à l'origine des ressources et de leurs descriptions.

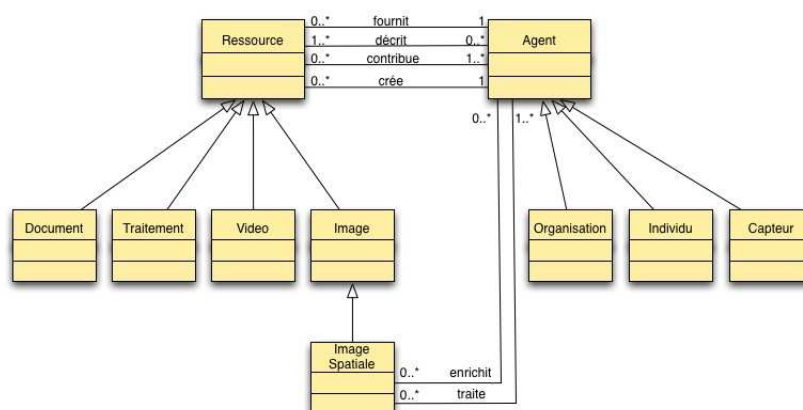


FIGURE 11 – Diagramme de classes représentant le modèle du domaine

5.4 Choix et étude des standards

Il s’agit maintenant de nous doter d’un profil d’ensemble de descriptions qui va prendre la forme, dans notre travail, d’un modèle pivot. Il nous a semblé pertinent de nommer ce modèle, *modèle pivot*, puisque c’est sur ce modèle que viendront s’articuler les charnières sémantiques “*crosswalks*” qui vont rendre l’ensemble du système interopérable.

Nous avons entamé une étude des standards qui sont en adéquation avec notre domaine. Cette étape est très importante. Elle permet d’acquérir une connaissance essentielle à l’élaboration du profil ou modèle pivot et de faire une utilisation appropriée des termes des standards. La bonne connaissance des standards va aussi nous servir ultérieurement lors de la spécification des *crosswalks*.

L’étude détaillée des standards ne porte ici que sur deux standards : le standard Dublin Core qui joue le rôle du standard généraliste permettant de décrire tout type de ressource, et le standard ISO 19115 qui est dédié à la description de l’information géographique. Nous avons exploré d’autres standards de métadonnées à l’exemple du standard FGDC toujours dans le domaine de l’information géographique, et du standard Darwin Core pour ce qui relève de l’écologie. Cependant, l’étude d’un standard se révèle très chronophage et nous avons volontairement limité cette activité aux deux standards les plus à même de servir notre approche. L’analyse des standards Dublin Core et ISO 19115 nous a permis de dégager des tableaux synthétiques dans lesquels sont listés et définis les éléments de métadonnées retenus pour la construction du modèle pivot.

5.4.1 Dublin Core

Nous avons commencé par étudier le standard Dublin Core. Il est parmi les standards les plus utilisés pour la description des ressources et a pour objectif premier de fournir des éléments permettant de décrire **génériquement** tout type de ressource, offrant ainsi aux communautés des moyens efficaces pour l’aide à la découverte et à la consultation de ressources au travers de systèmes généralistes. Dublin Core dispose à cet effet de quinze éléments de description, qui constituent le cœur du standard.

Ces éléments, ainsi que le modèle abstrait DCAM, ont été présentés dans la section 3.2 de l'état de l'art.

Espaces de noms : L'initiative Dublin Core maintient plusieurs vocabulaires, chacun venant servir différentes intentions. Chaque vocabulaire est caractérisé par un espace de noms qui lui est spécifique. Nous revenons rapidement sur la notion d'espace de noms qui offre un moyen d'assigner des identifiants uniques aux éléments appartenant aux standards de métadonnées concernés. Ces identifiants étant composés généralement de l'espace de noms combiné avec le nom du terme¹³. Dans le périmètre du Dublin Core, il va donc être possible de manipuler simultanément les vocabulaires et leurs éléments avec l'assurance de pouvoir identifier à tout moment, chacun de ces éléments.

Les termes assignés et gérés par le DCMI sont regroupés en une collection connue sous le nom de DCMI namespaces. Le tableau suivant décrit l'utilisation de ces espaces de noms.

Abréviation de l'espace de noms	Espace de noms	Description de l'espace de noms
dcterms	http://purl.org/dc/terms/	Toutes les propriétés DCMI, les classes et les schémas d'encodages (Sauf celles indiquées au dessous)
dcmitype	http://purl.org/dc/dcmitype/	Les classes du "DCMI Type Vocabulary"
dcam	http://purl.org/dc/dcam/	Les termes utilisés dans le modèle abstrait DCMI
dc	http://purl.org/dc/elements/1.1/	Dublin Core "Metadata Element Set", Version 1.1 (les 15 éléments originaux)

TABLE 2 – Espaces de noms du standard Dublin Core

Politique des espaces de noms Dublin Core : Les vocabulaires *dcterms* et *dc* laissent apparaître une évolution naturelle du schéma d'éléments de métadonnées vers plus de sophistication (le vocabulaire *dc* est le schéma d'origine, il est conservé pour des raisons de compatibilité ascendante). Les vocabulaires *dcterms* permet de définir des qualificatifs sur les éléments de métadonnées, voire de préciser le contenu attendu par ces éléments. Le vocabulaire *dcam* et *dcmitype* sont en étroite relation avec les profils d'applications Dublin Core **DCAP**, et l'ouverture du standard Dublin Core vers d'autres standards de métadonnées plus spécifiques à une thématique en particulier.

5.4.2 Le standard ISO 19115

Notre contexte d'application est centré sur la découverte d'images spatiales. Le standard ISO 19115 [ISO03] a le statut de norme internationale. Il est une référence

13. Terme : Propriété, Classe ...

pour l'information géographique dans le domaine des métadonnées. Nous avons donc entamé une étude de ce standard.

Il présente une synthèse exhaustive d'éléments, visant à unifier les modes de description des ressources géographiques. Il comporte 12 rubriques principales, dont 9 communes à toutes les ressources, que nous allons présenter ci-dessous :

Metadata : Information pour l'identification des principales caractéristiques de la métadonnée. Cette rubrique regroupe des éléments de métadonnées principaux pour la gestion des différentes fiches de métadonnées (date de création, standard et langage utilisé. . .)

Identification : Information pour l'identification des principales caractéristiques de la ressource. Ce sont des éléments de métadonnées principaux pour décrire la ressource (comme le Titre, la caractérisation spatiale et thématique. . .).

Seules ces deux premières rubriques sont obligatoires, les suivantes sont optionnelles.

Constraints : Information sur les contraintes d'accès à la ressource. Ces éléments de métadonnées permettent de savoir si une ressource est en accès libre ou restreint à certains utilisateurs et organismes, cette rubrique peut également être appliquée pour décrire les contraintes d'accès associées à la fiche de métadonnée elle-même (existence d'un lien d'agrégation entre MD_Metadata et cette rubrique).

Data quality : Information sur la qualité de la ressource. Dans cette rubrique, il s'agit de rendre compte du contexte dans lequel la ressource a été produite, en particulier de sa généalogie : la méthodologie employée, la qualité du résultat (en particulier d'un point de vue spatial - la précision d'une carte par exemple - ou temporel. . .), d'attribuer une note/un indicateur de qualité. Ceci permet de savoir si une ressource est compatible avec les attentes d'un utilisateur pour une application donnée (traitements envisagés après acquisition. . .)

Maintenance information : Information sur la maintenance et la mise à jour des données/métadonnées. Cette rubrique informe sur la fréquence de mise à jour de la ressource (régulière ou non), son état (ressource devenue obsolète) et les parties de la ressource concernées par d'éventuelles mises à jour.

Distribution : Information sur les modalités de distribution de la ressource. Ces éléments permettent de savoir qui contacter et où se procurer la ressource (le nom, adresse, téléphone. . . du distributeur) mais aussi sur quel support la ressource est disponible (numérique, papier, le format sur lequel elle est enregistrée ou disponible. . .), par quel moyen elle est accessible (en ligne, courrier . . .), éventuellement à quel prix . . .

Metadata extension information : Information pour l'ajout d'éléments de métadonnées. Dans le cas où les éléments proposés par l'ISO ne suffisent pas à décrire les caractéristiques de certaines ressources, cette rubrique permet d'étendre le modèle de métadonnées en proposant de nouveaux éléments de métadonnées. Pour cela, il faut renseigner les propriétés des éléments créés. La manipulation de cette rubrique présente moins d'intérêt pour l'utilisateur que pour l'administrateur. Par contre, elle est capitale pour l'adaptation de la norme à un contexte donné.

Portrayal Catalogue Reference : Cette classe contient un unique attribut dont l'objectif est de décrire le catalogue de règles de (re)présentations utilisées dans la ressource.

Application Schema Information : Schéma conceptuel de la ressource. Ce pa-

quet contient la description du schéma conceptuel et du logiciel utilisés pour générer la ressource décrite.

Remarque : En général, ces deux dernières rubriques disparaissent des profils car elles sont difficiles à renseigner et peu claires pour les utilisateurs. De même une partie de la rubrique qualité n'est exploitable que pour de l'information géographique au sens strict.

Trois rubriques servent uniquement à la description des ressources de type information géographique :

Spatial representation : Information sur le mode de représentation spatiale de la ressource. Selon la nature de la ressource décrite, les éléments de métadonnées de cette rubrique permettent la description détaillée des représentations vecteur ou raster à travers des attributs tels que la résolution, les caractéristiques du géoréférencement, la nature des objets géométriques présents. . .

Reference system : Information sur le système de référence. On trouve dans cette section les informations relatives à la description des systèmes de référence spatial (géographique et vertical) et temporel utilisés, le système géodésique, ellipsoïde. . .

Content information : Information sur le contenu de la ressource. Ces éléments portent sur des caractéristiques plus techniques, en particulier pour le cas d'images (aérienne, satellite. . .) pour lesquelles la compréhension fait appel à des notions plus fines de télédétection (caractéristiques des capteurs, longueurs d'ondes, catalogue des objets, description du contenu des données raster . . .).

A ces 12 rubriques s'ajoutent des classes DataTypes communes à de nombreuses rubriques, ces DataTypes servent à mettre en place une sorte de vocabulaire contrôlé en définissant de manière précise la forme et le type des données attendus.

Extraction du cœur du standard : Pour notre approche, nous avons décidé de nous intéresser uniquement au cœur du standard ISO 19115, qui est suffisant pour répondre à nos besoins de découverte et de localisation. Ce cœur, défini par le standard, propose les éléments de description portant principalement sur l'identification de la ressource et d'autres des éléments fournissant, à minima, des informations sur sa généalogie, sa distribution, son empreinte spatiale. Nous présentons en figure 12, le diagramme de classes représentant ces éléments.

En utilisant le diagramme précédent et la norme ISO 19115. Nous avons entamé une étude du standard qui a nous a permis d'élaborer le tableau suivant. Ce tableau représente une synthèse du standard et va nous servir comme référence.

Élément ISO 19115	Description de l'élément	Nom Complet : Position Hiérarchique	Format du contenu
Resource title	Titre principal de la ressource	MD_Metadata > MD_DataIdentification.citation > CI_Citation.title	Texte libre
Resource reference date	la date référence de la ressource citée	MD_Metadata > MD_Data Identification.citation > CI_Citation.date	CI_Date (date :Date; date-Type :CI_DateType Code)
Resource responsible party	L'identification et les coordonnées des personnes et organisations associés à la ressource décrite	MD_Metadata > MD_Data Identification.pointOfContact > CI_ResponsibleParty	CI_ResponsibleParty
Geographic location of the Resource	L'emprise spatiale de la Resource	MD_Metadata > MD_Data Identification.extent > EX_Extent > EX_Geographic Extent > (EX_Geographic-BoundingBox ou EX_Geographic Description ou EX_BoundingPolygon)	EX_Geographic-BoundingBox ou EX_Geographic Description ou EX_BoundingPolygon
Resource language	Langue utilisé par Resource	MD_Metadata > MD_Data Identification.language	CodeList (ISO-639-2)
Resource character set	Le nom complet de la norme de codage des caractères utilisé pour le Dataset	MD_Metadata > MD_Data Identification.characterSet	CodeList (MD_characterSet Code)
Resource topic category	La principale catégorie de la Resource	MD_Metadata > MD_Data Identification.topicCategory	Texte libre
Spatial resolution of the dataset	Facteur qui fournit une compréhension générale de la densité des données spatiales dans la ressource	MD_Metadata > MD_Data Identification.spatialResolution > MD_Resolution.equivalentScale ou MD_Resolution.distance	Texte libre

TABLE 3 – ISO 19115 Core Element Set 1/3

Élément ISO 19115	Description de l'élément	Nom Complet : Position Hiérarchique	Format du contenu
Abstract describing the resource	Résumé du contenu des ressources	MD_Metadata > MD_Data Identification. abstract	Texte libre
Distribution format	Format avec lequel les ressources sont distribuées ainsi que la version du format utilisé	MD_Metadata > MD_Distribution > MD_Format.name et MD_Format.version	Texte libre
Additional extent information for the dataset (vertical and temporal)	Extension des informations sur la ressource : cette extension peut prendre la forme d'une extension temporelle ou verticale	MD_Metadata > MD_Data Identification. extent > EX_Extent > EX_TemporalExtent or EX_VerticalExtent	Libre
Spatial representation type	représentation spatiale d'une information géographique	MD_Metadata > MD_Data Identification. spatial-Representation Type	CodeList MD_Spatial representation Type-Code
Reference system	Informations sur le système de référence	MD_Metadata > MD_ReferenceSystem A.9 .referenceSystem Identifier	RS :Identifier fig
Lineage	informations sur les événements ou les données de base utilisées dans la construction des données spécifiées par la portée ou le manque de connaissances sur la lignée	MD_Metadata > DQ_DataQuality. lineage > LI_Lineage .Statement	Texte libre
On-line resource	Les ressources disponibles en ligne pour : contacter des individus ou des organisations, ou consulter du contenu	MD_Metadata > MD_Distribution > MD_Digital TransferOption. onLine > CI_OnlineResource .Linkage	URL
Metadata file identifier	identifiant unique pour la fiche de métadonnées	MD_Metadata. fileIdentifier	Texte libre

TABLE 4 – ISO 19115 Core Element Set 2/3

Élément ISO 19115	Description de l'élément	Nom Complet : Position Hiérarchique	Format du contenu
Metadata Standard Name	Le nom du standard de métadonnées utilisé (nom du profil inclus)	MD_Metadata. metadata Stan- dardName	Texte libre
Metadata standard version	La version du standard de métadonnées utilisé	MD_Metadata. metadata Stan- dardVersion	Texte libre
Metadata language	Langue utilisée pour documenter les métadonnées	MD_Metadata. lan- guage	Libre
Metadata character set	Nom complet de la norme de codage des caractères utilisée pour le codage des métadonnées	MD_Metadata. cha- racterSet	CodeList (MD_characterSet Code)
Metadata point of contact	La partie qui peut fournir la ressource concernée	MD_Metadata. contact > CI_Responsible Party	Libre
Metadata date stamp	la date de création de la métadonnée	MD_Metadata. da- teStamp	Libre

TABLE 5 – ISO 19115 Core Element Set -3/3

5.5 construction du modèle pivot

En partant du modèle du domaine que nous avons établi, et en utilisant les tableaux produits après l'étude des standards. Nous avons sélectionné un ensemble d'éléments pour construire un modèle pivot pour la découverte d'images. Notre choix s'est porté sur les éléments qui nous ont paru les plus pertinents dans la description des éléments présents dans le modèle. Nous avons ainsi constitué le tableau suivant :

Nom Élément	Description de l'élément	Nom du standard origine	Raffinement
Title	Titre principal de la ressource	Dublin Core	Sans changement
Creator	Personne morale ou physique à l'origine de la création de la ressource	Dublin Core	gmd :CI_ ResponsibleParty
Subject	Sujet principal de la ressource ,Mots-clefs ou principales idées	Dublin Core	Texte libre
Description	Description du contenu de la ressource décrite	Dublin Core	Texte libre
Publisher	Nom de la personne morale ou physique à l'origine de la publication de la ressource	Dublin Core	gmd :CI_ ResponsibleParty
Contributor	Nom d'une personne physique ou morale qui contribue ou a contribué à l'élaboration de la ressource. Chaque contributeur fait l'objet d'un élément Contributor séparé	Dublin Core	CI_ResponsibleParty
Date	Date d'un évènement dans le cycle de vie de la ressource	Dublin Core	gmd :CI_Date

TABLE 6 – Élément du modèle pivot 1/3

Nom Élément	Description de l'élément	Nom du standard origine	Raffinement
Type	Genre du contenu de la ressource	Dublin Core	<code>rdfs :Class</code>
Format	Type MIME, ou format physique de la ressource	Dublin Core	Texte libre
Source	Ressource dont dérive la ressource décrite, que ce soit en totalité ou en partie de la ressource en question	Dublin Core	Texte libre
Language	Langue du contenu de la ressource	Dublin Core	Code Liste (ISO639-2)
Relation	Lien avec d'autres ressources	Dublin Core	URI
Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle	Dublin Core	EX_Bounding Box
Rights	Droits de propriété intellectuelle au sujet ou sur la ressource	Dublin Core	<code>dcterms :RightsStatement</code>
Identifier	Identificateur non ambigu, référençant la ressource dans un contexte donné	Dublin Core	Texte libre
Dataset character set	Le nom complet de la norme de codage des caractères utilisé pour le Dataset	ISO 19115	CodeList (MD_character SetCode)

TABLE 7 – Éléments du modèle pivot 2/3

Nom Élément	Description de l'élément	Nom du standard origine	Raffinement
Spatial resolution of the dataset	Facteur qui fournit une compréhension générale de la densité des données spatiales dans Dataset	ISO 19115	Texte libre
Reference system	Informations sur le système de référence	ISO 19115	gmd : RS_Identifier
Metadata file identifier	identifiant unique pour la fiche de métadonnées	ISO 19115	Texte libre
Metadata Standard Name	Le nom du standard de métadonnées utilisé (nom du profil inclus)	ISO 19115	Texte libre
Metadata standard version	La version du standard de métadonnées utilisé	ISO 19115	Texte libre
Metadata language	Langue utilisée pour documenter les métadonnées	ISO 19115	Code liste (ISO 639-2)
Metadata character set	Nom complet de la norme de codage des caractères utilisée pour le codage des métadonnées	ISO 19115	Texte libre

TABLE 8 – Éléments du modèle pivot 3/3

5.6 Opérationnalisation des modèles

Il nous faut maintenant travailler sur un langage de description des données unifié pour manipuler l'ensemble des éléments de métadonnées ISO19115 et Dublin Core. Nous avons naturellement choisi RDF et par extension RDFS et OWL. Ces langages vont nous permettre de représenter sémantiquement les éléments et les différents liens que nous voulons manipuler, et former ainsi nos schéma de métadonnées.

5.6.1 Dublin Core

Le standard Dublin Core dispose d'une implémentation officielle en RDF/RDFS. Chaque élément du standard est décrit au travers de différentes propriétés :

- **rdfs :label** : donne le nom local de l'élément.
- **rdfs :comment** : donne une description de l'élément, une définition par exemple.
- **rdfs :isDefinedBy** : contient l'espace de nom qui définit l'élément.
- **rdf :type** : donne le type RDF de l'élément décrit (classe, propriété ...).
- **dcterms :issued** : date de la première publication de l'élément.
- **dcterms :modified** : date de la dernière modification de l'élément.
- **rdfs :range** : représente le type des valeurs de l'élément.
- **rdfs :domain** : représente la classe d'appartenance de l'élément, quand elle existe.
- **rdfs :subPropertyOf** ou **rdfs :subClassOf** : donne la propriété ou classe mère de l'élément, elle représente souvent les propriétés de l'ancien espace de nom Dublin Core.
- **rdfs :replaces** : l'URI de l'élément remplacé par celui là, quand il existe.

La figure 13 donne un exemple de représentation graphique de la description de l'élément "title" du standard Dublin Core :

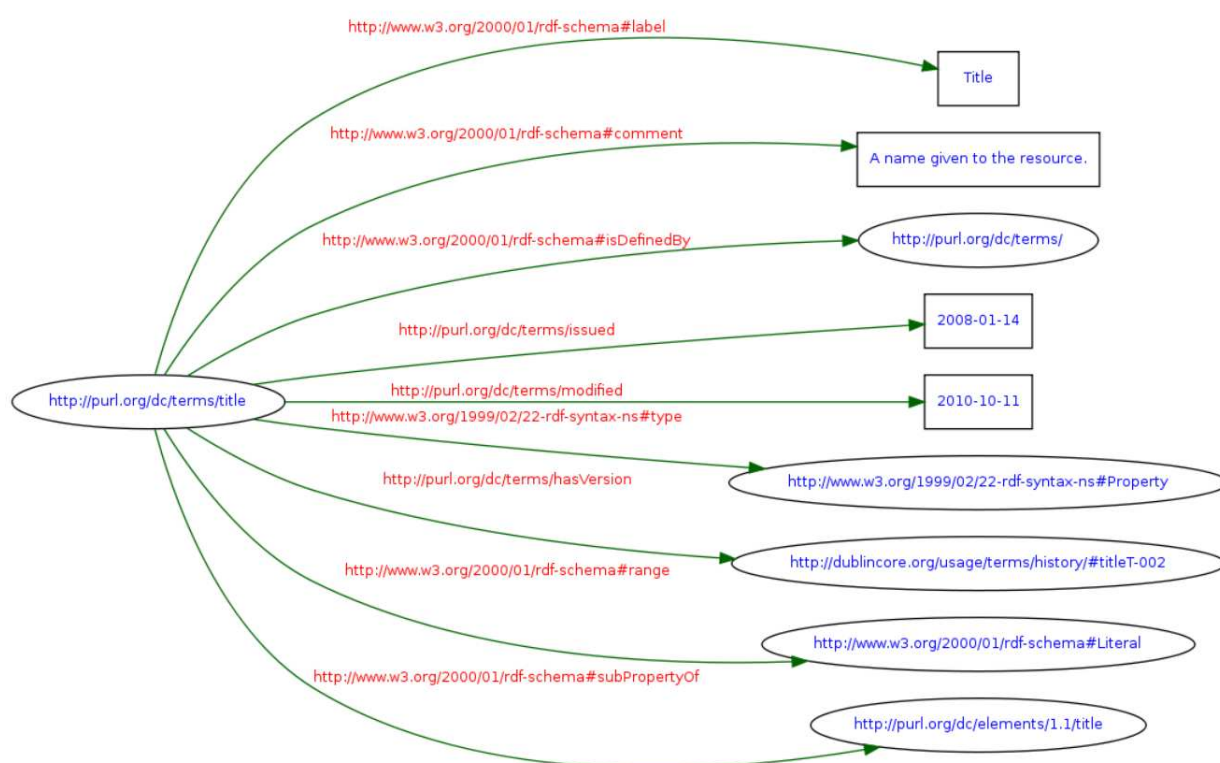


FIGURE 13 – Représentation de l'élément dcterms :title

Nous allons utiliser cette implémentation comme schéma de métadonnées Dublin Core, elle va par la suite être exploitée dans la construction du modèle pivot.

5.6.2 ISO 19115 : UML to OWL

La norme ISO19115 ne dispose pas d'implémentation dans les langages du web sémantique. Il fallait donc traduire la norme disponible sous la forme de diagrammes

de classes UML en OWL, pour obtenir le schéma de métadonnées.

Pour cela, nous avons consulté les travaux qui ont été menés sur les transformations automatiques des modèles UML à OWL, notamment ceux qui s'appuient sur la norme OMG "Ontology Definition Metamodel" [Gro09]. Nous avons cependant déduit que ces solutions n'étaient pas encore matures et nécessitaient plus de stabilité pour être exploitées. Car il n'existe actuellement aucun procédé permettant de transformer automatiquement, sans perte de contenu et d'expressivité, un diagramme de classe UML en OWL. Nous avons choisi de nous orienter vers une transformation manuelle. Ce choix à été fait pour deux raisons principales. La première est que cette transformation ne sera effectuée qu'une seule fois et que ce processus n'est pas récurrent. La deuxième est que nous considérons cette étape primordiale et qu'il faut maîtriser le cycle de transformation pour obtenir un schéma de qualité.

Les sous langages du standard OWL : Le standard OWL [DS04] offre trois sous langages d'expression croissante : **OWL Lite**, **OWL DL**, **OWL Full**.

OWL Lite offre la possibilité d'établir une hiérarchie de classification et d'exprimer les caractéristiques de contraintes simples. Il prend en charge par exemple les contraintes de cardinalité, mais ne permet comme valeurs de ces cardinalités que 0 ou 1.

OWL DL est un sous langage adossé aux logiques de description. Il offre une expressivité maximale en garantissant la décidabilité des calculs (toutes les inférences sont garantis d'être calculé). Il inclut toutes les constructions du standard OWL avec des restrictions telles que le type de séparation (une classe ne peut pas aussi être un individu ou une propriété ...).

OWL Full est un sous langage destiné aux utilisateurs voulant bénéficier d'une expressivité maximale et d'une liberté syntaxique de RDF, sans garantie de calcul (une classe peut être, à la fois, un individu ou une collection d'individus ...)

Transformation du modèle : Nous avons choisi d'utiliser OWL DL comme langage d'implémentation d'ISO19115. Il nous a semblé le plus approprié pour supporter l'interopérabilité avec les diagrammes de classes statiques UML de la norme. OWL-DL s'appuie sur les logiques de description. Il permet donc un meilleur raisonnement et apporte une expressivité suffisante sur l'utilisation des restrictions pour les cardinalités et autres. Nous avons éliminé OWL Lite pour son manque apparent d'expressivité. OWL Full, quant à lui, n'a pas été utilisé. Il offre une grande flexibilité et permet à l'utilisateur de déclarer des constructions qui ne sont pas standardisées, ce qui limite considérablement l'interopérabilité des ontologies exprimées avec ce sous langage.

Nous avons pris en compte, durant le procédé de transformation, les différences qui existent entre UML et OWL. Et nous avons essayé d'éviter toute perte de contenu ou d'expressivité. Pour établir les règles de transformations, nous avons pris des choix pour clarifier des cas particuliers.

Nommage : Le contexte est pris en compte dans les diagrammes de classes UML, de ce fait, il est possible d'utiliser le même nom dans deux contextes différents. Deux classes Image et Partie Responsable peuvent toutes les deux avoir la propriété "Identifiant". En OWL, ce n'est pas permis. Le contexte est global, et deux propriétés

ayant le même nom sont considérées identiques. Pour pallier à ce problème, nous avons décidé de combiner le nom de la classe avec celui du nom locale de la propriété pour obtenir une nomenclature sans ambiguïté.

Classe UML : La primitive owl :Class sera utilisée pour représenter les classes UML. En OWL, les attributs et les associations sont déclarés en dehors de la classe. Seules les restrictions et les spécialisations peuvent être ajoutées à la déclaration de la classe.

Classe abstraite : OWL ne dispose pas d'équivalent pour la classe abstraite UML. Plusieurs solutions sont possibles, l'utilisation de propriété d'annotation pour décrire le type de classe par exemple. Nous nous sommes inspirés d'une solution exploitée dans les base de données, et nous avons décidé que pour chaque classe abstraite, toutes les relations et les attributs de cette classe vont migrer vers les classes filles.

Attribut UML : Suivant le type des attributs, ils peuvent être transformés soit owl :DatatypeProperty ou en owl :ObjectProperty.

Pour les attributs de type simple (String, Int...), ils seront traduits en OWL en utilisant trois propriétés : owl :DatatypeProperty représentera le nom de l'attribut, rdfs :domain désignera la classe de l'attribut et rdfs :range le type de l'attribut.

Pour les attributs de type complexe, on utilisera la propriété owl :ObjectProperty pour les représentés, et rdfs :range pointera vers la classe qui constitue le type en UML.

Association UML : Les associations vont être exprimées en OWL à l'aide de owl :objectProperty, les agrégations et les compositions n'ont pas d'équivalents en OWL. Ils vont être représentés à l'aide de owl :ObjectProperty, rdfs :domain et rdfs :range.

Spécialisation : Les spécialisations / généralisations vont être exprimées avec la primitive rdfs :subClassOf.

Cardinalités et Contraintes OCL : OWL offre la possibilité de spécifier les cardinalités sur les propriétés et les classes à l'aide de owl :minCardinality, owl :maxCardinality et owl :cardinality. Ces primitives peuvent être combinées avec owl :Restriction pour spécifier aussi les contraintes OCL sur les attributs.

Énumération et Code Liste : On utilise owl :oneOf, Cette primitive permet d'énumérer un ensemble de valeurs ou caractéristiques possibles sur une classe ou propriété.

Schéma de métadonnées ISO19115 : A partir de ces règles, nous avons traduit le cœur de la norme ISO19115 en OWL. Une courte partie du schéma de métadonnées obtenu est présentée dans l'annexe B. Pour illustrer la démarche, nous allons présenter la transformation d'une classe simple de la norme ISO19115.

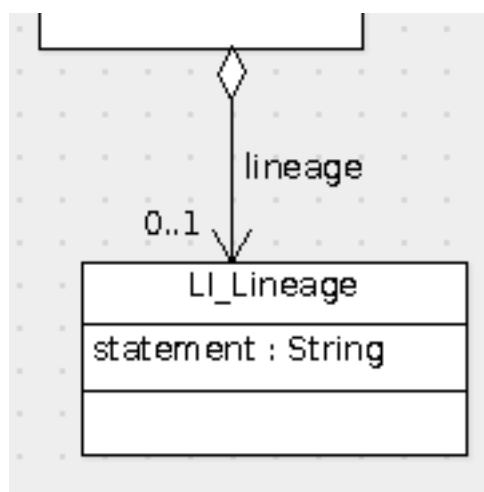


FIGURE 14 – Classe LI_Lineage de ISO19115

La classe LI_Lineage comprend un attribut de type simple, et une agrégation *li-
neage* avec la classe DQ_DataQuality. La représentation OWL correspondant à cette
classe est la suivante :

```

<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#LI_Lineage">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    LI_Lineage</rdfs:label>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.teledetection.fr/iso19115/gmd#
        LI_Lineage.statement"/>
      <owl:maxCardinality rdf:datatype="&xsd;nonNegativeInteger">1
      </owl:maxCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</rdf:Description>
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd
#LI_Lineage.statement">
  <rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    statement</rdfs:label>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
  <rdf:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#LI_Lineage"/>
  <rdf:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Description>

<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
LI_Lineage.inverseLineage">

<rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
inverse lineage</rdfs:label>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>

```



```

<owl:inverseOf rdf:resource="http://www.teledetection.fr/iso19115/gmd#
DQ_DataQuality.lineage" />
  <rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#
  LI_Lineage"/>
  <rdfs:range rdf:resource="http://www.teledetection.fr/iso19115/gmd#
  DQ_DataQuality"/>
</rdf:Description>

```

Cette démarche a été appliquée sur tout le modèle UML du cœur de ISO19115 présenté précédemment. Après l'obtention du schéma de métadonnées OWL, nous avons utilisé un raisonneur pour s'assurer de sa conformité avec la norme OWL DL.

5.6.3 Modèle pivot

Nous disposons à ce stade des schémas de métadonnées de Dublin Core et ISO19115 en RDF / RDFS et OWL. l'opérationnalisation du modèle pivot consiste à sélectionner les éléments des deux précédents schémas pour assurer l'implémentation du modèle en RDF.

5.7 Crosswalks

5.7.1 Identification des crosswalks :

Après l'obtention du modèle pivot, la prochaine étape est de l'enrichir avec les *crosswalks*. Comme nous l'avons expliqué, le but est d'identifier les éléments sémantiquement équivalents ou similaires. Nous avons donc mené une étude sur le modèle pivot et le standard ISO19115, qui a permis d'identifier ces *crosswalks*. Nous proposons dans le tableau suivant les résultats de cette étude. Le tableau indique les éléments impliqués et les règles de transformation ou conditions à respecter pour passer de ISO19115 au modèle pivot.

Élément du modèle pivot	Élément ISO19115	Condition / Remarque
dcterms :title	CI.Citation .title	Pas de condition
dcterms :creator	CI.ResponsibleParty .individualName	CI.responsibleParty.role = "originator"
dcterms :creator	CI.ResponsibleParty .organisationName	CI.responsibleParty.role = "originator"
dcterms :subject	MD.DataIdentification .topicCategory	subject plus général que topicCategory
dcterms :subject	MD.Metadata .Keywords	Pas de condition
dcterms :description	MD.DataIdentification .abstract	description plus général que abstract
dcterms :publisher	CI.ResponsibleParty	role = publisher
dcterms :language	MD.DataIdentification .language	Pas de condition
dcterms :date	MD.Metadata .dateStamp	Pas de condition
dcterms :type	MD.SpatialRepresentation .typeCode	type plus général que type-Code
dcterms :format	MD.format	Pas de condition
dcterms :identifiant	CI.OnlineResource .linkage	Pas de condition
dcterms :coverage	EX.BoundingBox	Pas le même type
dcterms :contributor	MD.DataIdentification.credit	Pas de condition
dcterms :Agent	CI.ResponsibleParty	Pas de condition

TABLE 9 – Éléments du modèle pivot 3/3

5.7.2 Opérationnalisation des crosswalks :

Pour implémenter les *crosswalks*, deux solutions vont être explorées. Le *crosswalking* peut être établi soit au niveau des schémas de métadonnées, soit au niveau des instances. Nous avons décidé de mettre en œuvre les deux solutions et de comparer les résultats obtenus.

Crosswalking au niveau du schéma de métadonnées : Cette première approche consiste à opérationnaliser les correspondances au niveau des schémas de métadonnées. Concrètement, cela reviendrait à enrichir le modèle pivot avec les éléments ISO19115, et d'utiliser les primitives offertes par OWL pour exprimer les *crosswalks*.

Pour illustrer l'utilisation de OWL, nous allons présenter un exemple de transformation.

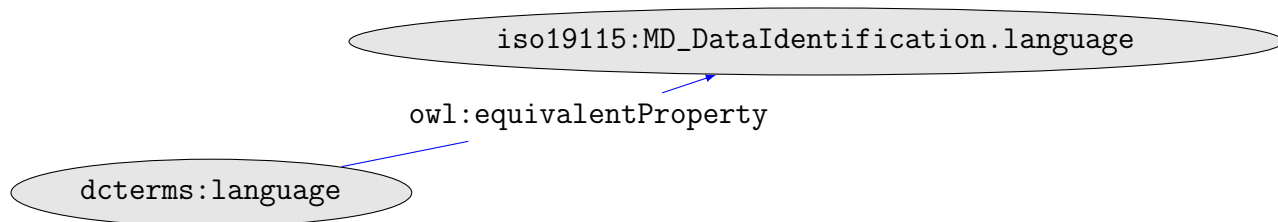


FIGURE 15 – crosswalk Dublin Core - ISO19115 (1)

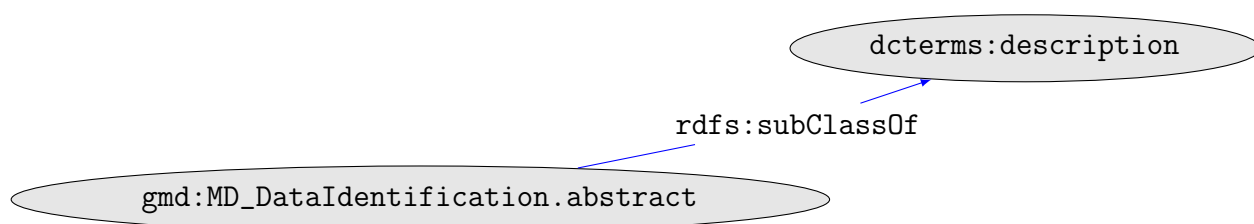


FIGURE 16 – crosswalk Dublin Core - ISO19115 (2)

Une fois ces *crosswalks* implémentés, ils enrichiront le modèle pivot et seront pris en compte lors des calculs des inférences. Ce qui va nous permettre d’exploiter à la fois les métadonnées conformes à notre modèle pivot, et les métadonnées ISO19115.

Crosswalking au niveau des instances de métadonnées : Le *crosswalking* peut aussi être établi au niveau des instances de métadonnées. Nous allons utiliser pour cet effet des feuilles de style XSLT. Ces feuilles de style vont transformer les métadonnées ISO19115 en instances de métadonnées conformes au modèle pivot, et ceci en appliquant les crosswalks établis dans le tableau précédent. Comme le montre le schéma suivant.

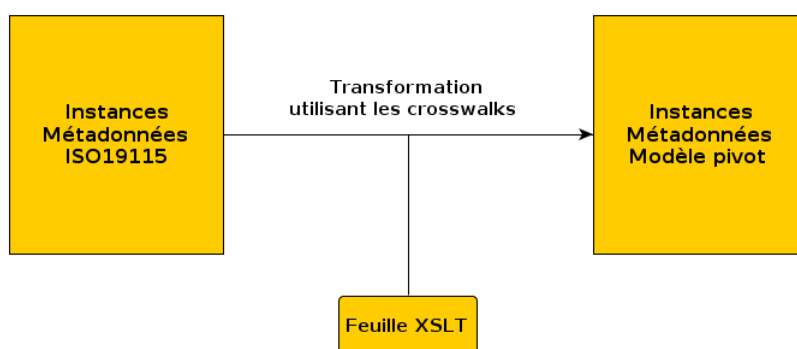


FIGURE 17 – Crosswalking au niveau des instances de métadonnées

5.8 Transformation des métadonnées

XML est le format d’échange le plus répandu dans le web. Plusieurs services web, fournissent des interfaces d’acquisition de métadonnées standardisés en XML.

C'est pour cela que nous avons décidé de nous intéresser à ce format. Dans le cadre de GEOSUD, des fiches de métadonnées en XML sont disponibles suivant les spécifications ISO19139 (implémentation XML de ISO19115) et Dulin Core. Ces données ne sont pas en RDF. Nous avons implémenté un module de transformation de données, permettant le passage des fiches XML en métadonnées RDF. Comme le montre la figure suivante, un processus d'extraction et de stockage de triplets RDF à partir de fichiers de métadonnées XML est proposé.

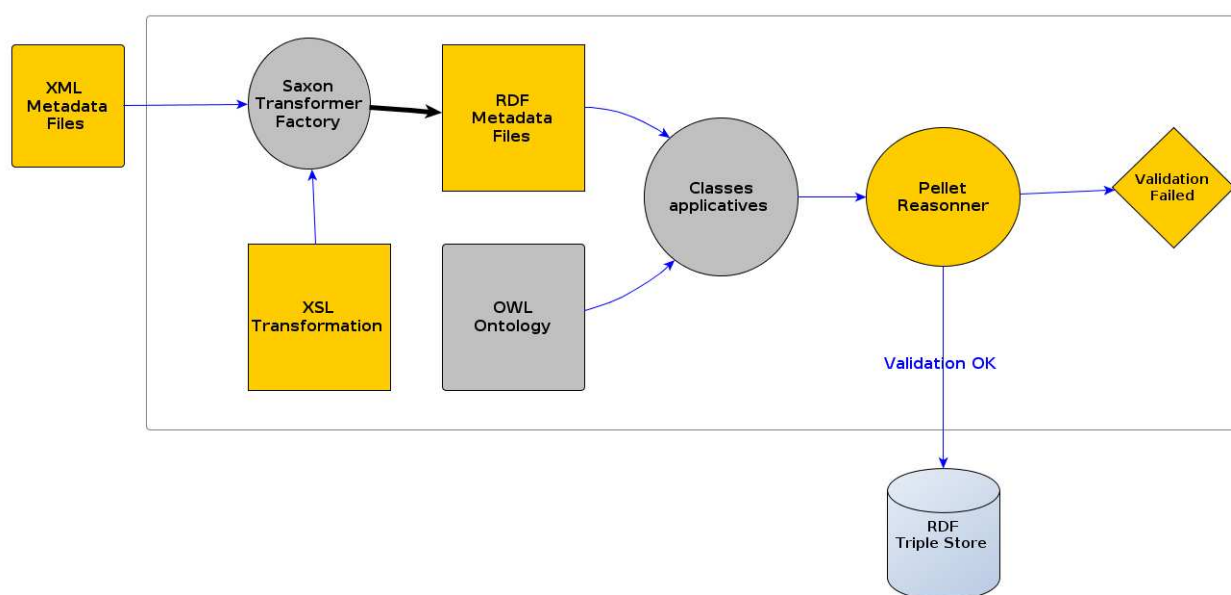


FIGURE 18 – Architecture du transformateur de données

Le transformateur utilise des outils du web sémantique combinés avec des feuilles de style XSLT. Il prend en entrée la fiche de métadonnées XML. Des feuilles de style XSL sont ensuite utilisées pour extraire les informations, et produire des triplets RDF. La qualité des métadonnées est importante, nous avons donc intégré dans notre transformateur une étape de validation des données. Cette étape vise à vérifier la conformité des instances de métadonnées, obtenues, à leurs schémas de métadonnées. Nous avons à cet effet utilisé le raisonneur OWL DL Pellet. Ce raisonneur prend en entrée les instances à vérifier et le schéma de métadonnées correspondant (le schéma OWL ISO19115 ou le modèle pivot dans notre cas). Nous utilisons ensuite ce raisonneur pour calculer les inférences, et vérifier la cohérence des instances de métadonnées avec le schéma. Les métadonnées RDF validées sont stockées dans par la suite dans des *triples store* RDF.

6. Démonstrateur et validation de l'approche

6.1 Outils techniques

Nous allons dans cette partie présenter les outils techniques et l'architecture utilisés pour développer notre démonstrateur. Nous avons utilisé **Java** comme langage d'implémentation, notre choix s'est porté sur ce langage parce qu'il dispose de bibliothèques et *frameworks* implémentées conformément aux recommandations W3C pour le web sémantique. Il est aussi le langage utilisé par mon équipe de recherche pour implémenter les solutions en rapport avec le projet GEOSUD.

Jena : framework de construction d'applications Web Sémantique Jena¹⁴ est un *framework* Java pour créer des applications Web sémantique. C'est un projet *open source* soutenu et hébergé par la fondation Apache¹⁵. Il fournit un ensemble d'outils et de bibliothèques pour faciliter la manipulation, la lecture, l'écriture, le stockage et l'interprétation des déclarations RDF.

Le *framework* Jena comprend entre autres :

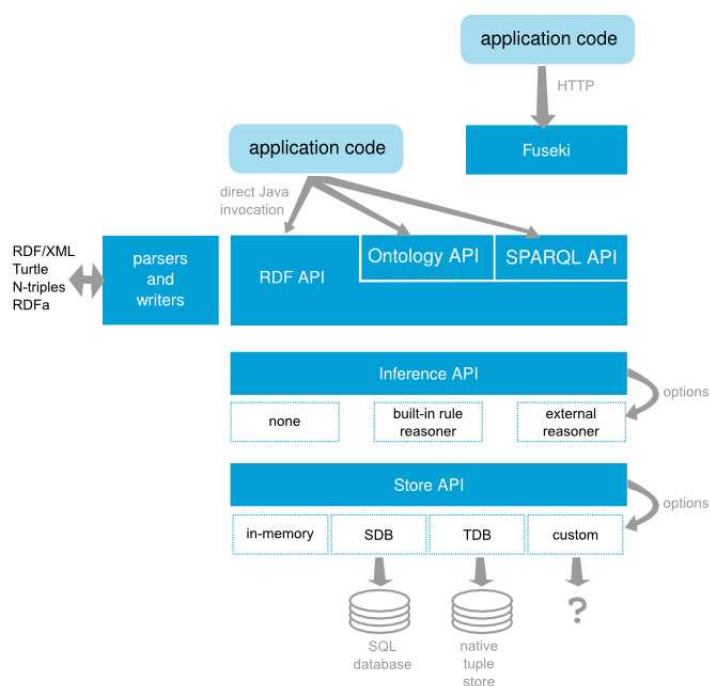
- une API pour la lecture, l'écriture et le traitement de données RDF sérialisées ,en XML, N-triples et formats Turtle,
- une API OWL pour la manipulation des ontologies,
- un moteur d'inférence à base de règles pour raisonner sur des sources de données RDF et OWL,
- des modules pour la persistance des triplets RDF : TDB , SDB.

Le schéma suivant, repris de la documentation officielle de Jena¹⁶, présente l'architecture globale du *framework* avec les différents modules disponibles.

14. Apache Jena : <http://jena.apache.org/>

15. The Apache Software Foundation : <http://www.apache.org/>

16. jena.apache.org/about.jena/architecture.html

FIGURE 19 – Architecture du *framework* Jena

Pellet : raisonneur OWL en Java Pellet est un raisonneur qui traite notamment, la recommandation W3C **OWL DL**. Il figure parmi les raisonneurs OWL les plus utilisés. Nous l'avons utilisé pour valider nos schémas de métadonnées avec les normes OWL. Le raisonneur nous permet aussi de vérifier la conformité des instances de métadonnées avec leurs schémas.

6.2 Architecture du démonstrateur

En utilisant les différents outils présentés précédemment, nous avons développé un démonstrateur fonctionnel, se basant sur les technologies du web sémantique, et répondant aux besoins de découverte d'images spatiales. Ce qui nous a permis de rendre accessible les modèles de métadonnées conçus lors de notre mise en œuvre et les métadonnées GEOSUD transformées. La figure suivante présente l'architecture de notre application.

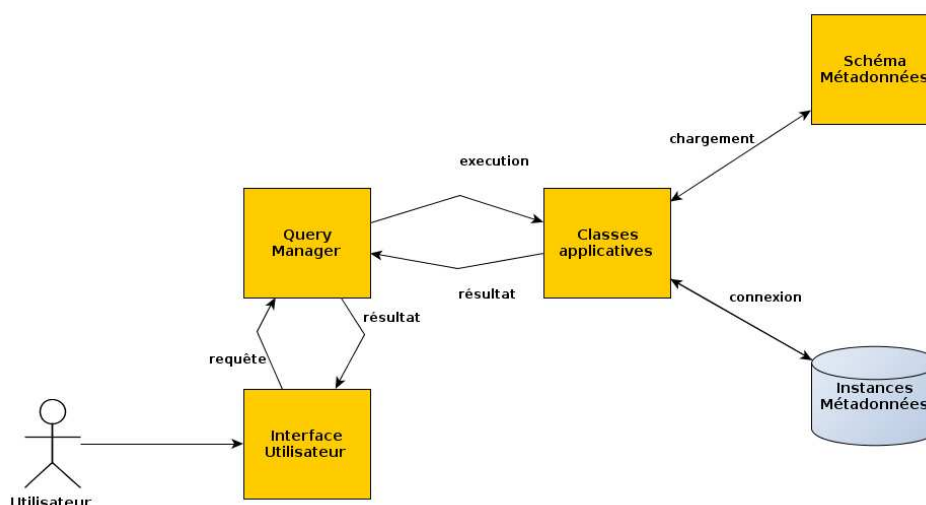


FIGURE 20 – Architecture du démonstrateur

6.3 Tests et Résultats

Pour valider notre démarche, nous avons élaboré des tests visant à vérifier différents critères :

- Tester l’efficacité du schéma de métadonnées défini pour la découverte d’images spatiales,
- Tester l’accès aux instances de métadonnées.

Les jeux de métadonnées : Nous avons stocké les jeux de métadonnées RDF, que nous avons transformé précédemment, dans deux *triple store*. Nous rappelons que nous voulons explorer les deux approches de *crosswalking* présentées. Nous disposons donc pour les tests de deux jeux de métadonnées.

Un jeu de métadonnées ayant des instances toutes conformes au modèle pivot (*crosswalks* appliqués au niveau des instances), et d’un autre jeu de métadonnées conformes à leurs schémas de métadonnées origine (ISO19115 ou Dublin Core) mais possédant comme schéma de métadonnées, pour l’interrogation un modèle pivot enrichi (modèle pivot + *crosswalks*).

Le premier jeu de métadonnées (toutes les instances transformées vers le modèle pivot) contient 171090 triplets RDF. Le deuxième jeu de métadonnées (instances de métadonnées ISO19115 et Dublin Core) contient 181168 triplets RDF.

Le langage d’interrogation SPARQL : Le W3C propose une recommandation définissant un langage dédié nommé **SPARQL**[PS⁺08]. Ce langage fonctionne à base de requêtes exprimant les motifs recherchés dans le graphe RDF pour en extraire des informations.

- **SELECT** permet d’énumérer les variables que l’on souhaite récupérer en sortie.
- **WHERE** spécifie le sous-graphe correspondant à la recherche.

A l’intérieur de la clause WHERE, d’autres clauses permettent de préciser la requête.

- **FILTER** permet de filtrer les résultats en fonction des valeurs des variables ; Le filtrage est applicable aux intervalles de valeurs numériques ou bien aux expressions régulières sur les chaînes de caractères.
- **UNION** exprime une alternative entre deux expressions.
- **OPTIONAL** détaille une partie de requête facultative, au sens où le nœud mentionné n'est pas obligatoirement requis.

SPARQL dispose aussi d'options pour la présentation des résultats. Il permet par exemple d'ordonner de manière croissante ou décroissante par rapport à une variable, grâce aux clauses **ORDER BY DESC** ou **ORDER BY ASC**.

Nous allons utiliser ce langage pour interroger les deux sources de données RDF (*triples store*) alimentées à l'issue de la transformation des métadonnées ISO19115 XML, et de l'application des deux approches de *crosswalking*.

Exemples de requêtes :

Restituer les adresses de téléchargement des images de très haute résolution

En télédétection, la résolution spatiale du capteur utilisé est une information très importante. Elle permet aux télédéTECTEURS de sélectionner les images spatiales exploitables pour une tâche en particulier, puisque la précision des détails discernables sur une image dépend de cette résolution. Supposons maintenant que nous souhaitons renvoyer pour les besoins d'un traitement quelconque, les images qui sont à une résolution de 5 mètres.

```
PREFIX  rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX  gmd:  <http://www.teledetection.fr/iso19115/gmd#>
PREFIX  owl: <http://www.w3.org/2002/07/owl#>
PREFIX  xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX  rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX  dcterms: <http://purl.org/dc/terms/>
PREFIX  gmi:  <http://www.teledetection.fr/iso19115/gmi#>
```

```
SELECT  ?identifiant ?link
WHERE
{
  ?s rdf:type gmi:MI_Metadata .
  ?s dcterms:identifiant ?id .
  ?s gmd:MD_Metadata.identificationInfo ?d .
  ?s gmd:MD_Metadata.distributionInfo ?dis .
  ?dis gmd:MD_Distribution.transferOptions ?tr .
  ?tr gmd:MD_DigitalTransferOptions.onLine ?on .
  ?on gmd:CI_OnlineResource.linkage ?link .
  ?d gmd:MD_DataIdentification.spatialResolution ?sp .
  ?sp gmd:MD_Resolution.distance "5.0"
}
```

Résultat :


```

identifiant : SCENE_5_041-262_04_12_13_10_59_06_1_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_041-262_04_12_13_10_59_06_1_A.zip
-----
identifiant : SCENE_5_052-258_1_05_10_16_10_50_52_1_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_052-258_1_05_10_16_10_50_52_1_A.zip
-----
identifiant : SCENE_5_037-251_9_05_10_04_11_21_20_2_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_037-251_9_05_10_04_11_21_20_2_A.zip
-----
identifiant : SCENE_5_026-252_8_04_06_12_11_36_41_2_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_026-252_8_04_06_12_11_36_41_2_A.zip
-----
identifiant : SCENE_5_046-255_05_10_15_11_09_53_1_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_046-255_05_10_15_11_09_53_1_A.zip
-----
identifiant : SCENE_5_039-259_4_05_06_22_11_23_59_1_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_039-259_4_05_06_22_11_23_59_1_A.zip
-----
identifiant : SCENE_5_052-261_05_07_05_10_33_39_1_A
lien téléchargement : ftp://geosud-sw.teledetection.fr/SCENE_5_052-261_05_07_05_10_33_39_1_A.zip

```

FIGURE 21 – Aperçu du résultat de la requête résolution = 5 mètres

Rechercher des images couvrant la commune Montpellier

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX gmd: <http://www.teledetection.fr/iso19115/gmd#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX gmi: <http://www.teledetection.fr/iso19115/gmi#>

```

```

SELECT *
WHERE
{ ?m rdf:type gmi:MI_Metadata .
  ?m gmd:MD_Metadata.metadataStandardVersion ?v .
  ?m dcterms:subject ?p
  FILTER regex(?p, "montpellier")
}

```

Nous pouvons déduire que les deux entrepôts de métadonnées répondent aux spécifications élaborées. Mais il existe néanmoins des différences notables entre les deux.

L'application des crosswalks au niveau des instances de métadonnées, transformation des triplets ISO19115 et Dublin Core en triplets conformes au modèle pivot, permet d'obtenir des entrepôts plus performant en temps d'exécution. Comme les métadonnées sont alignées sur le modèle pivot, le calcul d'inférences est moins coûteux et les requêtes moins complexes. Mais la qualité et la précision des réponses est plus faible, puisque la migration des instances de métadonnées entraîne la perte de quelques spécificités du schémas origine. par exemple le passage de l'élément `gmd:MD_DataIdentification.topicCategory` à l'élément `dcterms:subject`.

Pour le *crosswalking* au niveau des schémas de métadonnées, la qualité et la précision sont plus élevées. En effet, en conservant les instances de métadonnées dans leur

schéma origine, ils gardent leurs spécificités. Mais en contre partie, le temps du calcul des inférences au niveau du modèle est plus important, ce qui donne un temps d'exécution plus élevé.

L'expérience que nous avons menée montre que la mise en œuvre des *crosswalks* au niveau du schéma de métadonnées est moins coûteuse. Elle ne nécessite pas la mise en place d'un processus de transformation des instances de métadonnées.

Conclusion et perspectives

Dans cette partie qui conclut le mémoire, nous résumons les travaux réalisés et nous dégagons les résultats obtenus face aux objectifs fixés. Nous proposons aussi quelques perspectives à court et moyen terme de l'approche développée.

L'étude bibliographique, qui a précédé le stage, a consisté à positionner différents moyens pour rendre interopérables des sources de données en s'adossant à des standards de métadonnées. Les approches basées sur les profils d'application et les *crosswalks* ont retenu toute notre attention et nous avons alors défini une approche hybride de construction d'un profil d'application enrichi par des références croisées.

Nous avons spécifié l'approche méthodologique associée. Nous avons suivi cette approche dans le domaine des images spatiales et dégagé successivement un modèle conceptuel du domaine, un modèle pivot s'articulant sur une sélection des éléments de métadonnées des standards Dublin Core et ISO19115 et un modèle enrichi par une collection de références croisées (*crosswalks*).

L'objectif était d'exploiter le profil d'application sémantique construit pour faciliter les processus d'aide à la découverte et à la localisation d'images spatiales décrites soit au travers du standard Dublin Core, soit au travers du standard ISO19115.

Un préalable était d'unifier le format de description des instances de métadonnées venant décrire le jeu de ressources disponibles. ISO19115 ne propose pas de modèle RDF, RDFS ou OWL pour l'expression du standard. Nous avons défini un schéma de métadonnées OWL pour le cœur du standard ISO19115. Nous avons également transformé les fiches de métadonnées XML en RDF du jeu d'images que nous avons à notre disposition. au travers d'un transformateur basé sur XSLT construit à cet effet.

Nous avons construit un démonstrateur pour procéder à des premiers tests de faisabilité. Ce dernier travail est encore en phase d'expérimentation et seules quelques requêtes d'accès à des images sur la base de critères élémentaires ont été explorées. Une question clé, à notre sens, est de décider du niveau de positionnement des références croisées. Nous avons lancé une série de tests en faisant apparaître ces liens sémantiques soit au niveau du modèle pivot, soit au niveau des instances de métadonnées. ces tests ne sont pas suffisamment renseignés pour l'instant pour nous permettre de porter des conclusions et il s'agit d'une perspective à très court terme de ce travail.

Dans les perspectives à plus long terme, différents travaux peuvent s'envisager. Nous avons volontairement restreint le nombre de standards de métadonnées exploités, la richesse du modèle conceptuel du domaine ou encore la richesse sémantique des références croisées. Il s'agira donc d'enrichir l'ensemble de ces activités. De même, le démonstrateur reste pour l'instant limité; il s'agira donc de poursuivre les développements pour en démontrer toutes les capacités en l'alimentant notamment avec des jeux de métadonnées de taille réelle.

Un enjeu d'importance reste à plus long terme la consultation de contenu des images satellites au travers de notre système.

Références

- [BG04] Dan Brickley and R.V. Guha. RDF vocabulary description language 1.0 : RDF schema. W3C recommendation, W3C, February 2004. Published online on February 10th, 2004 at <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>.
- [CB08] Karen Coyle and Thomas Baker. Guidelines for dublin core application profiles. 2008.
- [DHSW02] Erik Duval, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. Metadata principles and practicalities. *D-Lib Magazine*, 8(4), 2002.
- [DS04] Mike Dean and Guus Schreiber. OWL Web Ontology Language - Reference. W3C recommendation, W3C, 10 feb 2004.
- [ES07] Jérôme Euzenat and Pavel Shvaiko. *Ontology matching*, volume 18. Springer Heidelberg, 2007.
- [Gro06] OMG : Object Management Group. Meta object facility (MOF) core specification. MOF Specification 2, OMG, Needham Heights, USA, January 2006.
- [Gro09] OMG : Object Management Group. Ontology definition metamodel (ODM). Technical Report 1.0, OMG, Needham Heights, USA, 2009.
- [HK10] Bernhard Haslhofer and Wolfgang Klas. A survey of techniques for achieving metadata interoperability. *ACM Comput. Surv.*, 42(2) :1–37, March 2010.
- [HP00] Rachel Heery and Manjula Patel. Application profiles : mixing and matching metadata schemas. *AGI - Information Management Consultants*, 10(1) :1–10, January 2000.
- [ISO03] Iso 19115 - geographic information metadata. Iso, International Organization for Standardization, Geneva, Switzerland, 2003.
- [MM04] Franck Manola and Eric Miller. RDF primer. W3C recommendation, World Wide Web Consortium, February 2004. <http://www.w3.org/TR/rdf-primer/>.
- [NJNP06] Mikael Nilsson, Pete Johnston, Ambjörn Naeve, and Andy Powell. Towards an interoperability framework for metadata standards. In *Proceedings of the 2006 international conference on Dublin Core and Metadata Applications : metadata for knowledge and learning*, DCMI '06, pages 13–23. Dublin Core Metadata Initiative, 2006.
- [NMJE09] Mikael Nilsson, Alistair J. Miles, Pete Johnston, and Fredrik Enoksson. Formalizing dublin core application profiles - description set profiles and graph constraints. In Miguel-Angel Sicilia and Miltiadis Lytras D., editors, *Metadata and Semantics*, pages 101–111. Springer US, 2009.
- [PG04] Yannick Prié and Serge Garlatti. Métadonnées et annotations dans le web sémantique. *Revue I3 Information-Interaction-Intelligence*, Numéro Hors-série Web sémantique, Edition Cépaduès, 24(6), 2004.

- [PN06] Sarah Pulis and Liddy Neville. Using the DC Abstract Model to support application profile developers. In *DC-2006 : International Conference on Dublin Core and Metadata Applications, Colima (Mexico), 3-6 October 2006*, pages 227–234, Colima, Mexico, October 2006.
- [PS⁺08] Eric Prud'Hommeaux, Andy Seaborne, et al. Sparql query language for rdf. *W3C recommendation*, 15, 2008.
- [RWB11] James Reid, Will Waites, and Ben Butchart. Going for gold—adventures in open linked metadata. 12, 2011.
- [SBLH06] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3) :96–101, May 2006.
- [Su02] Xiaomeng Su. A text categorization perspective for ontology mapping. *Norway : Department of Computer and Information Science, Norwegian University of Science and Technology*, 2002.
- [WBG⁺12] John Wiczorek, David Bloom, Robert Guralnick, Stan Blum, Markus Döring, Renato Giovanni, Tim Robertson, and David Vieglais. Darwin core : An evolving community-developed biodiversity data standard. *PLoS ONE*, 7(1) :e29715, 01 2012.

Annexe A

Élément Dublin Core	Description de l'élément	Nom Complet : Position Hiérarchique	Format du contenu
Title	Titre principal de la ressource	dcterms :title	rdfs :Literal
Creator	Personne morale ou physique à l'origine de la création de la ressource	dcterms :creator	dcterms :Agent
Subjet	Sujet principal de la ressource ,Mots-clefs ou principales idées	dcterms :subject	Texte libre
Description	Description du contenu de la ressource décrite	dcterms :description	Texte libre
Publisher	Nom de la personne morale ou physique à l'origine de la publication de la ressource	dcterms :publisher	dcterms :Agent
Date	Date d'un évènement dans le cycle de vie de la ressource	dcterms :date	rdfs :Literal
Contributor	Nom d'une personne physique ou morale qui contribue ou a contribué à l'élaboration de la ressource. Chaque contributeur fait l'objet d'un élément Contributor séparé	dcterms :contributor	dcterms :Agent
Type	Genre du contenu de la ressource	dcterms :type	rdfs :Class
Format	Type MIME, ou format physique de la ressource	dcterms :format	dcterms :MediaTypeOrExtent
Language	Langue du contenu de la ressource	dcterms :language	dcterms :LinguisticSystem

TABLE 10 – Dublin Core Metadata Element Set 1/2

Élément Dublin Core	Description de l'élément	Nom Complet : Position Hiérarchique	Format du contenu
Source	Ressource dont dérive la ressource décrite, que ce soit en totalité ou en partie de la ressource en question	dcterms :source	Texte libre
Relation	Lien avec d'autres ressources	dcterms :relation	Texte libre
Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle	dcterms :coverage	dcterms :Location-PeriodOrJurisdiction
Rights	Droits de propriété intellectuelle au sujet ou sur la ressource	dcterms :rights	dcterms :RightsStatement
Identifier	Identificateur non ambigu, référant la ressource dans un contexte donné	dcterms :identifier	rdfs :Literal

TABLE 11 – Dublin Core Metadata Element Set 2/2

Remarque : la position hiérarchique ici ne joue pas un rôle important, puisque le standard est "plat", il n'est pas organisé en paquets ou classes.

Annexe B

```

<rdf:RDF
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    xmlns:gco="http://www.teledetection.fr/iso19115/gco#"
    xmlns:gmd="http://www.teledetection.fr/iso19115/gmd#"
  <!-- Transformation du modèle UML de ISO 19115 en  OWL -->
  <!-- Classe MD_Metadata -->
  <rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#MD_Metadata">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
      MD_Metadata</rdfs:label>
    <rdfs:comment xml:lang="en">root entity which defines metadata
      about a resource or resources</rdfs:comment>
    <rdfs:comment xml:lang="fr">entité racine qui définit les métadonnées d'une
      ou plusieurs ressources</rdfs:comment>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class" />
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.teledetection.fr/iso19115/gmd#
        MD_Metadata.fileIdentifier" />
      <owl:cardinality rdf:datatype="xsd:nonNegativeInteger">1</owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.teledetection.fr/iso19115/gmd#
        MD_Metadata.contact" />
      <owl:minCardinality rdf:datatype="xsd:nonNegativeInteger">1
        </owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.teledetection.fr/iso19115/gmd#
        MD_Metadata.identificationInfo" />
      <owl:minCardinality rdf:datatype="xsd:nonNegativeInteger">1
        </owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.teledetection.fr/iso19115/gmd#
        MD_Metadata.distributionInfo" />
      <owl:maxCardinality rdf:datatype="xsd:nonNegativeInteger">1
        </owl:maxCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>

```



```

</rdf:Description>
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
MD_Metadata.fileIdentifier">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    fileIdentifier</rdfs:label>
  <rdfs:comment xml:lang="en">unique identifier of this metadata file
  </rdfs:comment>
  <rdfs:comment xml:lang="fr">identifiant unique pour la fiche de métadonnées
  </rdfs:comment>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty" />
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
  <rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Metadata"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
MD_Metadata.contact">
  <rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    contact</rdfs:label>
  <rdfs:comment xml:lang="en">party responsible for the metadata information
  </rdfs:comment>
  <rdfs:comment xml:lang="fr">l'entité responsable des métadonnées</rdfs:comment>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  <rdfs:range rdf:resource="http://www.teledetection.fr/iso19115/gmd#
    CI_ResponsableParty"/>
  <rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Metadata"/>
</rdf:Description>
<!-- Roles properties -->
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
MD_Metadata.identificationInfo">
  <rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    identificationInfo</rdfs:label>
  <rdfs:comment xml:lang="en">basic information about the resource(s)
    to which the metadata applies</rdfs:comment>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  <rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Metadata"/>
  <rdfs:range rdf:resource="http://www.teledetection.fr/iso19115/gmd#
    MD_DataIdentification"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
MD_Metadata.referenceSystemInfo">
  <rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    referenceSystemInfo</rdfs:label>
  <rdfs:comment xml:lang="en">object property between MD_Metadata
    and MD_ReferenceSystem</rdfs:comment>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  <rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Metadata"/>
  <rdfs:range rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_ReferenceSyst
</rdf:Description>
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
MD_Metadata.distributionInfo">

```

```

<rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
  distributionInfo</rdfs:label>
<rdfs:comment xml:lang="en">object property between MD_Metadata and MD_Distribution
</rdfs:comment>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
<rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Metadata"/>
<rdfs:range rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Distribution"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.teledetection.fr/iso19115/gmd#
MD_Metadata.dataQualityInfo">
  <rdfs:label xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    dataQualityInfo</rdfs:label>
<rdfs:comment xml:lang="en">provides overall assessment of quality of a resource(s)
</rdfs:comment>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
<rdfs:domain rdf:resource="http://www.teledetection.fr/iso19115/gmd#MD_Metadata"/>
<rdfs:range rdf:resource="http://www.teledetection.fr/iso19115/gmd#DQ_DataQuality"/>
</rdf:Description>

```