

Soutenance de Master 2 Informatique
Conception d'un modèle abstrait de métadonnées pour
l'interconnexion de jeux de données géoréférencées

Encadré par : I. MOUGENOT - J.-C. DESCONNETS

Hatim CHAHDI

Université Montpellier 2 - UMR Espace-DEV, IRD

9 Juillet, 2013



- **Introduction : Contexte et Objectif**
- **Le web sémantique**
- **Hétérogénéité et interopérabilité des métadonnées**
- **Démarche adoptée**
 - Profils d'application
 - Etapes de la démarche
- **Application de la démarche**
- **Conclusion et Perspectives**

Equipex GeoSud

Développement d'infrastructure nationale facilitant l'accès libre à l'imagerie satellitaire pour la communauté scientifique et les acteurs publics

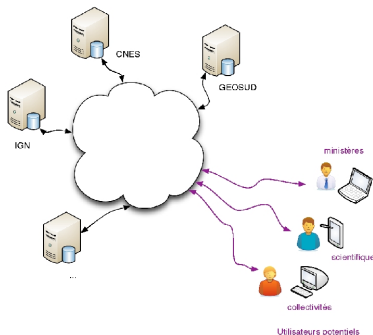


Figure: Fournisseurs de données spatiales **GEOSUD**

Approche méthodologique pour la découverte et localisation *pertinente* des images spatiales

Offrir la possibilité de localiser des images spatiales décrites par différents standards de métadonnées

- Métadonnées hétérogènes
- Normes différentes, plus ou moins rigides et non adaptées

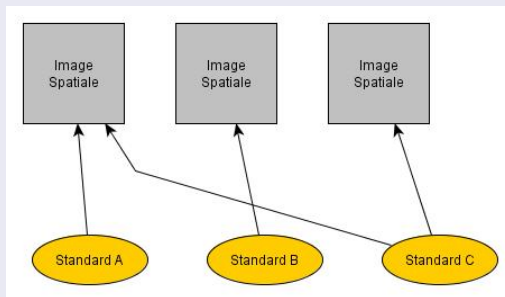


Figure: Images spatiales décrites au travers différents standards

Idée de départ et Web sémantique

Existant

Recherche sur les éléments communs entre les différents standards ou sur un seul standard spécifique

Idée de départ

Exploiter les technologies du **web sémantique**

Définitions : Web sémantique

- Nouvelle vision du web visant l'organisation des ressources au travers de métadonnées pour faciliter leur exploitation automatique par des agents logiciels
- *Framework* technologique offrant une panoplie de langages de description tous exprimables et échangeables en XML

Langages de description des métadonnées

Resource Description Framework

Initiative W3C pour décrire des ressources au travers des métadonnées.
RDF s'appuie sur la notion de triplet Sujet, Prédicat, Objet

Resource Description Framework Schema

vocabulaire RDF introduisant des éléments de modélisation simples
Concepts de classe et de propriété. Primitives d'organisation hiérarchique
des éléments. Introduction de contraintes élémentaires : `rdfs:domain`,
`rdfs:range` ...

Web Ontology Language

Langage de représentation des connaissances
Enrichit RDFS et emprunte les éléments de logiques de description
Perspective : langage pour la construction des schémas de métadonnées

Hétérogénéité des métadonnées : un frein à l'interopérabilité

L'hétérogénéité des métadonnées intervient à tous les niveaux

- Hétérogénéité structurelle : Langages de représentation, schémas différents...
- Hétérogénéité sémantique : Expressivité des langages de descriptions, différences terminologiques...

Définition

Ensemble d'éléments de métadonnées pouvant provenir de différents standards répondant à un besoin applicatif précis

Interopérabilité

S'adosser sur des standards pour favoriser l'interopérabilité

- **Etude du domaine**
- **Analyse des standards**
- **Construction et raffinement du modèle pivot**
- **Crosswalking** (liens sémantiques)

1. Étude du domaine

- Entités ciblées par les métadonnées
- Acteurs du domaine
- Relations et interactions entre les acteurs et/ou entités du domaine

2. Analyse des standards

- Prise de connaissance des standards
- Établir des tableaux de référence

3. Construction du modèle pivot

- Partir des entités présentes dans le modèle du domaine
- Sélection des éléments pertinents à partir des tableaux des standards établis

4. Raffinement du modèle

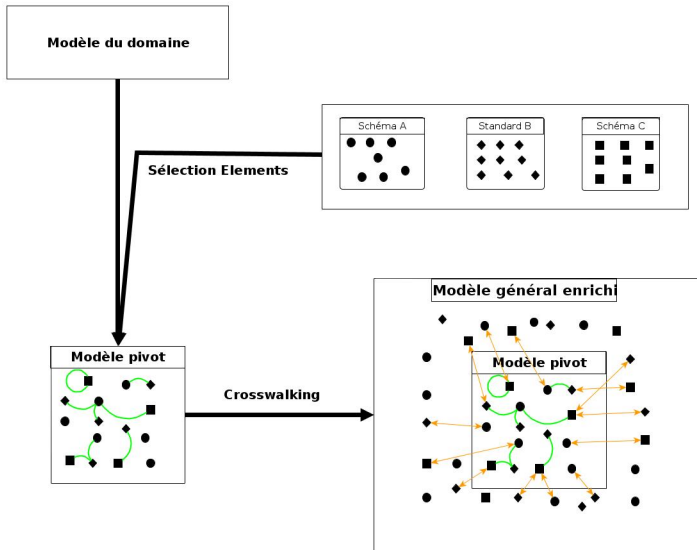
Modifier ou restreindre la signification d'un élément, sans le changer fondamentalement

5. *Crosswalking*

Définition : Établir des liens sémantiquement équivalents ou similaires entre deux schémas de métadonnées différents

Objectif : Exploiter des instances de métadonnées définis au travers de schémas de métadonnées

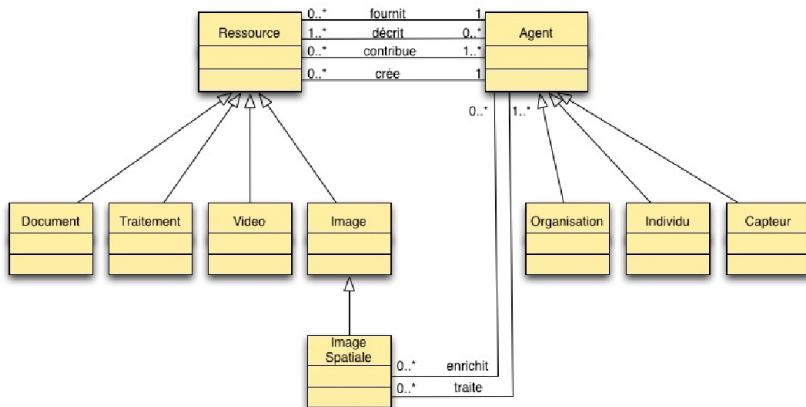
Démarche



Deux variantes :

- Au niveau du schéma de métadonnées
- Au niveau des instances de métadonnées

Étude du domaine : Imagerie Spatiale



1. Dublin Core

- Générique, standard de découverte
- Quinze éléments pour la description de ressources de type quelconque : contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type
- Implémenté officiellement dans les langages du web sémantique

1. Dublin Core

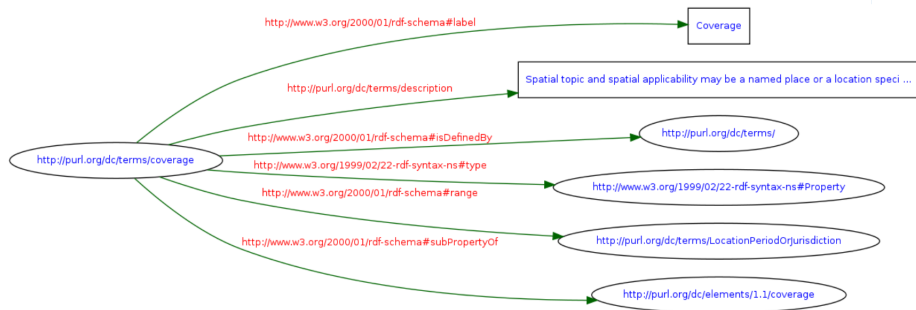


Figure: Représentation graphique de la définition d'un élément DC en RDF

2. ISO 19115

- Standard spécialisé
- Norme Internationale Pour l'information géographique
- Organisé en douze rubriques dont deux obligatoires

2. ISO 19115 : Extraction du cœur du standard pour les besoins de découverte

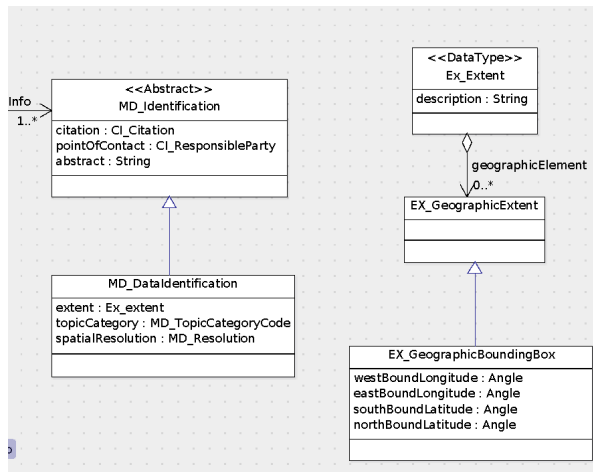


Figure: Extrait du cœur du standard ISO19115

2. ISO 19115: Non disponible en OWL

Étape intermédiaire : transformation du diagramme de classes UML en OWL

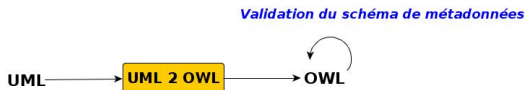


Figure: transformation de UML en OWL

Résultats

- Établissement des règles de transformation
- Application des règles sur le cœur du standard
- Validation du schéma de métadonnées OWL
- 104 Entités : 19 classes OWL, 29 DatatypeProperty, 21 ObjectProperty, 35 Restriction...

3. Modèle pivot

- disponibilité des schémas de métadonnées en langages du web sémantique
- sélection des éléments pertinents
- construction du schéma de métadonnées du modèle pivot
- raffinement des éléments du modèle pivot

Modèle pivot : Similaire à un profil d'application, c'est le schéma de métadonnées de base constitué d'éléments minimum.

4. Identification des crosswalks : Utilisation des tableaux du standard ISO 19115 et du modèle pivot

Modèle pivot	ISO 19115	Remarque
dcterms:title	CI_Citation.title	Propriétés équivalentes
dcterms:subject	MD_DataIdentification .topicCategory	subject plus général que topic Category
dcterms:subject	MD_Metadata .Key- words	Propriétés équivalentes
dcterms:Agent	CI_ResponsibleParty	Agent plus général que Responsible Party

Table: Aperçu des *crosswalks*

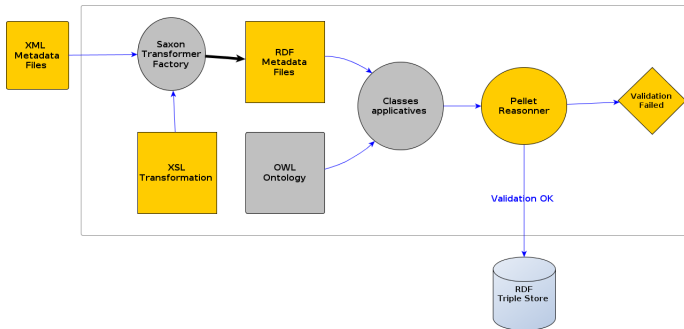
Résultat : Identification des liens sémantiquement équivalents ou similaires

4.1 Implémentation des crosswalks : Utilisation des primitives OWL

- *owl:equivalentProperty* pour les propriétés équivalentes
- *owl:equivalentClass* pour les classes équivalentes
- *rdfs:subPropertyOf* et *rdfs:subClassOf* pour la hiérarchisation

Résultat : Obtention d'un modèle enrichi sémantiquement par les *crosswalks* capable de reconnaître les éléments des autres standards liés

XML to RDF : Développement du transformateur



Résultat : Disponibilité des instances de métadonnées en RDF
500 fiches XML de métadonnées GEOSUD transformées

Transformation des métadonnées : niveau Schéma de métadonnées

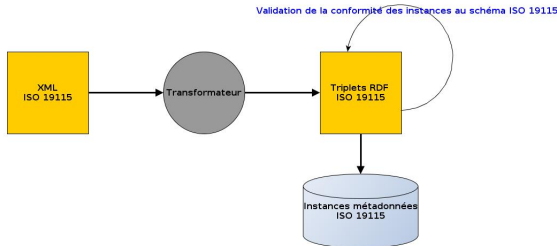


Figure: Transformation des métadonnées XML en RDF

Résultats : Les métadonnées changent de format mais restent conformes à leur standard origine

Jeu de métadonnées obtenu : **181.168** triplets RDF ISO et DC

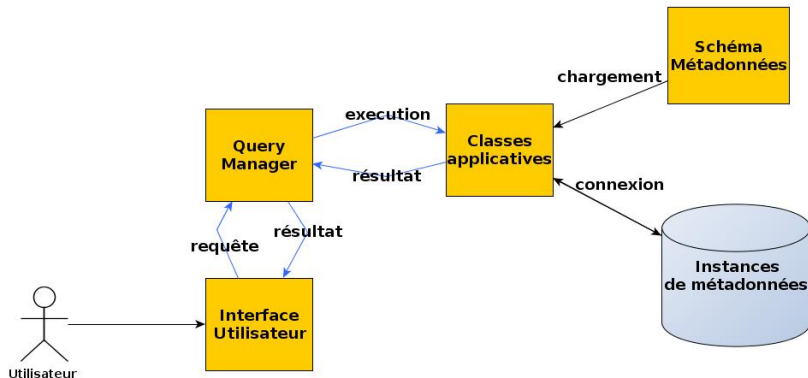


Figure: Architecture du démonstrateur

Exemple d'exécution : Schéma de métadonnées

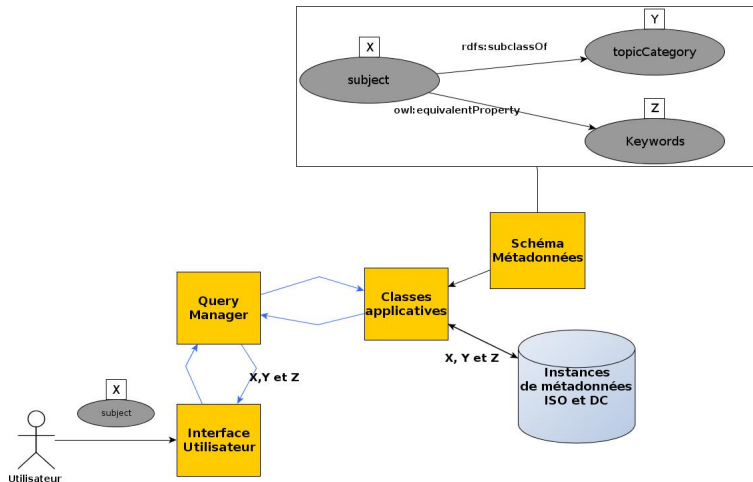


Figure: Exemple d'exécution

4.2 Implémentation des crosswalks aux niveau des instances de métadonnées

- utilisation du tableau des correspondances établies
- Implémentation de ces correspondances à l'aide de XSLT

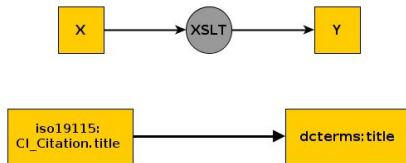


Figure: *crosswalking* des instances

Résultats :

- Instances de métadonnées en RDF Conforme au **modèle pivot**
- Jeu de métadonnées obtenu : **171.090** triplets RDF

4.2 Implémentation des crosswalks aux niveau des instances de métadonnées

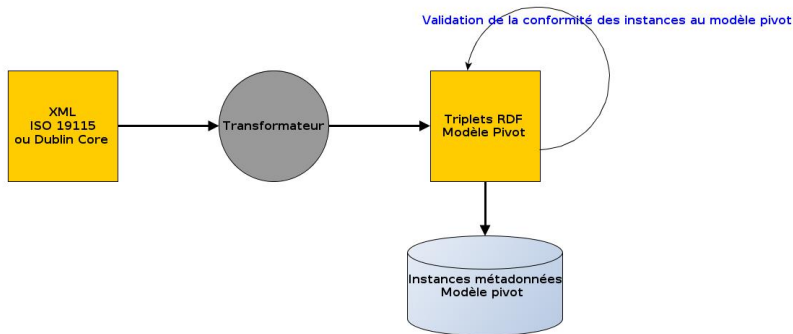


Figure: transformation des métadonnées en RDF conforme Modèle pivot

Résultat : Instances de métadonnées en RDF Conforme au **modèle pivot**
Interrogation directe du modèle pivot

Schéma de métadonnées

- Exploitation des possibilités de OWL
- Mise en œuvre moins coûteuse
- Temps d'exécution plus important ($\times 1.3$ en moyenne)

Instances de métadonnées

- Instances de métadonnées homogènes
- Mise en œuvre plus coûteuse
- Temps d'exécution moins important

Besoins de tests plus approfondis :

- Requêtes plus complexes
- Passage à l'échelle : 10.000 fiches de métadonnées dans le cadre de l'Equipex GEOSUD

Résultats et apports:

- Proposition d'une approche modulaire palliant l'hétérogénéité
- Démonstration de la faisabilité de l'approche : Application sur la découverte d'images spatiales
- Implémentation des deux variantes du *crosswalking*

Perspectives :

- Améliorer l'expressivité des requêtes
- Mener des tests de performance plus approfondis
- Application de la démarche pour le traitement d'images spatiales

Merci pour votre attention