

Une approche d'implémentation des dictionnaires de métadonnées pour la fédération de données géographiques multisource

Landry Tongo, Georges-Edouard Kouamou, Gilbert Armand Tchudjo

▶ **To cite this version:**

Landry Tongo, Georges-Edouard Kouamou, Gilbert Armand Tchudjo. Une approche d'implémentation des dictionnaires de métadonnées pour la fédération de données géographiques multisource. Revue Africaine de Recherche en Informatique et Mathématiques Appliquées, 2014, Volume 18, 2014, pp.53-65. 10.46298/arima.1975 . hal-01300089

HAL Id: hal-01300089
https://inria.hal.science/hal-01300089

Submitted on 8 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une approche d'implémentation des dictionnaires de métadonnées pour la fédération de données géographiques multisource

Landry Tongo[1], Georges Edouard Kouamou[2], Gilbert Armand Tchudjo[3]

[1]. Institut National de Cartographie
BP 157 Yaoundé, Cameroun
ltongo@yahoo.fr
[2]. Ecole Nationale Supérieure Polytechnique
BP 8390 Yaoundé, Cameroun
georges.kouamou@uy1.uninet.cm
[3]. Faculté des Sciences, Université de Yaoundé I
BP 812 Yaoundé, Cameroun
atchudjo@gmail.com

RESUME. Les métadonnées dans le domaine des SIG sont utilisées pour décrire les sources de données existantes en vue d'en faciliter l'accès ou le partage entre différents acteurs. Le problème d'exploitation de ces métadonnées se pose lorsqu'il faut les cataloguer dans le cadre d'une plateforme de fédération des données spatiales. Nous décrivons une approche de structuration de cette composante, orientée vers les services avec une implémentation basée sur LDAP. Pour y parvenir, nous partons d'une description canonique qui unifie les principaux standards de métadonnées géographiques connus, ensuite nous définissons de nouvelles classes d'objets LDAP qui mappent les entités syntaxiques de cette description.

ABSTRACT. Spatial Metadata are used to describe the existing data sources in order to facilitate their access and sharing between different actors. The problem of exploitation of these metadata arises when they should be catalogued within the framework of a platform for spatial data federation. We describe a service oriented approach for structuring this component, with an implementation based on LDAP. To achieve this, we start from a canonical language that unifies the major known geographic metadata standards, and then we define new classes of LDAP objects which map syntactic units of canonical language.

MOTS-CLES : Métadonnées géographiques, fédération des données, Architecture Orientée Service, LDAP.

KEYWORDS: Spatial metadata, data federation, Service Oriented Architecture, LDAP.

Received, March 31, 2013
Revised, December 31, 2013
Accepted, January 7, 2014

ARIMA Journal, vol. 18 (2014), pp. 53-65.

1. Introduction

Le coût d'acquisition et de collecte des données est très élevé dans un projet de SIG. Elle est estimée environ à 80% du budget. Pour faire face à cette difficulté, les acteurs qui s'intéressent à une zone géographique commune, s'échangent les données dont ils disposent afin de mutualiser leurs efforts. Cet échange de données se fait à travers des supports de stockage, et nécessite des métadonnées descriptives qui facilitent l'accès aux données échangées [1].

Cette approche d'échange est statique et elle ne favorise pas l'instantanéité des mises à jour. Si le pourvoyeur d'une donnée effectue des modifications, l'utilisateur devra attendre l'acheminement du prochain support contenant ces modifications pour en profiter. D'où notre préoccupation de satisfaire aux besoins de construction des environnements de fédération des données géographiques qui permettrait d'automatiser le processus tout en garantissant l'accès transparent à différentes sources de données et l'instantanéité des modifications. Nous sommes confortés dans cette voie par l'évolution des technologies de l'information et de la communication. Elle est marquée par le développement de l'Internet qui favorise la distribution des systèmes d'information.

Les SIGs s'intègrent actuellement dans les systèmes d'informations interconnectés, transversaux et ouverts sur l'extérieur, appelés Infrastructure de Données Spatiales (IDS) ou géoportails. Les exemples d'IDS telles que GeOrchestra et EasySdi, sont des structures fortement centralisées en termes d'accès aux données et elles ne permettent pas l'intégration de données multisources i.e. issues des serveurs différents. Notre démarche s'inscrit dans un contexte où les données ne sont pas centralisées dans un serveur principal mais plutôt sur un ensemble de serveurs d'où l'intérêt de l'utilisation des métadonnées. La fédération doit reposer sur des plateformes logicielles qui assurent l'organisation et l'acheminement effectif des données à référence spatiale [2]. Ces plateformes de distribution nécessitent une mémoire qui répertorie les différentes sources de données mises à la disposition de la communauté. Cette mémoire peut s'organiser sous forme de dictionnaire qui est une composante essentielle et incontournable pour ce type de plateforme. La mise en œuvre d'une telle composante de mémorisation des informations sur les données à référence spatiale reste un défi pour la fédération effective des données.

Dans ce papier, nous proposons une approche d'implémentation d'un dictionnaire de métadonnées basé sur le protocole d'annuaire LDAP. Ceci en vue de tirer profit des atouts du langage de manipulation existants pour les annuaires électroniques traditionnels pour mettre en œuvre une composante de mémorisation qui servira de support pour les plateformes de distribution des données à référence spatiale. Cette

exploitation des atouts des annuaires vise à spécifier la mémorisation et accroître l'exploitation (distribution) des métadonnées.

Dans la suite, nous présenterons les motivations de notre étude basées sur la nécessité des catalogues de métadonnées dans le contexte de l'automatisation de la fédération des données géographiques. La section 3 présente un modèle de métadonnée géographique canonique ainsi que le principe de construction d'un langage canonique inhérent au modèle de dictionnaire élaboré. La section 4 décrit la structuration des concepts sous-jacents aux métadonnées géographiques et leur mise en œuvre dans LDAP. la section 5 présente le cadre de validation mis en place à travers l'implémentation d'un frontal pour tester l'utilisation des services offerts par le composant qui gère le dictionnaire de métadonnées.

2. Motivation

Les métadonnées sont un ensemble organisé d'information qui permettent de présenter et de décrire une ressource (données, personnes, machines, produits industriels, etc.). La présentation et la description étant des éléments importants dans les critères de choix, les métadonnées apparaissent dès lors importants pour la recherche et la sélection des données. Elles permettent de savoir ce qui existe et de prendre une décision dans la sélection des données adéquates par rapport à un besoin.

Une métadonnée représente l'information structurée qui décrit, explique et localise une ressource en facilitant sa recherche, son usage et sa gestion. La disponibilité de métadonnées améliore la capacité des consommateurs à trouver, à consulter et à utiliser l'information dont ils ont besoin auprès des producteurs. L'usage des métadonnées pour faciliter l'accès à des sources de données hétérogènes est accepté par les acteurs du domaine des SIG [1]. Elles facilitent aussi dans certains cas, l'échange de données à référence spatiale entre un producteur et un utilisateur de donnée. Le support contiendra au minimum la géodonnée échangée (au format du logiciel utilisé) ainsi que la métadonnée (un fichier texte) décrivant cette donnée.

Conscient du fait que les métadonnées restent purement syntaxiques, d'autres approches ont été proposées pour y adjoindre la sémantique, en particulier la définition d'une ontologie géographique [4,5]. Cette proposition est présentée comme une réponse au besoin croissant de partage d'information sur le Web qui passe par la compréhension des informations mises à disposition, l'automatisation des processus de navigation, d'extraction ou de mises à jour de ces informations. Cependant, l'implémentation des ontologies géographiques reste un défi dans la mesure où il y a plusieurs propositions qui a notre avis souffrent d'une absence de normalisation.

La définition d'un schéma commun (une norme) pour toutes les ontologies serait une piste de solution. L'usage des standards différents dans des catalogues de

métadonnées requiert une stratégie où les utilisateurs ont seulement besoin de comprendre le schéma unifié et la syntaxe de requête utilisée pour exprimer des requêtes. Le schéma unifié ainsi que la syntaxe de requête sont regroupés au sein d'un module qui implémente une ontologie. Néanmoins, l'ajout d'un nouvel standard de métadonnée implique la modification de l'ontologie implémentée [7].

L'approche des métadonnées reste prédominante. Elles constituent la base de l'infrastructure INSPIRE qui est la directive européenne dont l'objectif central est de rendre l'information géographique accessible via internet et réutilisable au niveau européen [6]. Face à la multitude des standards se pose la question d'unification voire d'harmonisation de la recherche dans ce domaine. Cette préoccupation semble être un préalable à l'automatisation des dictionnaires de métadonnées [3].

L'expérience de Ramroop & Pascoe [1, 7] démontre la faisabilité de l'usage du protocole LDAP pour le catalogage des métadonnées géographiques. Quelques bénéfices de l'usage de la Directory Information Tree (DIT) de LDAP est la réplication entre les serveurs, la structure hiérarchique des DITs reflétant la structure des standards de métadonnée. Aussi, le *dn* (distinguished name) est l'unique identificateur pour les entrées. La représentation de la structure des standards de métadonnées en LDAP n'est pas restreinte par le niveau de détail car il peut y avoir un nombre infini de branche dans la définition de l'arborescence d'un catalogue dans LDAP.

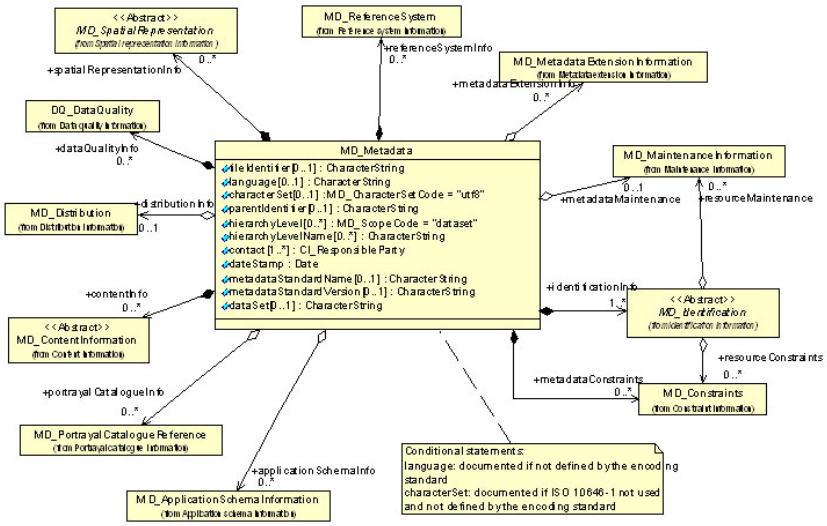


Figure 1 : schéma conceptuel de la norme ISO 19115

Les principaux standards en matière de métadonnées que l'on rencontre sont :

- CSDG qui est une norme mise au point par la FGDC, le comité fédéral de normalisation des métadonnées des États-Unis en 1998,

- ENV 12657 qui a été produite par le CEN/TC 287 a servi de référence à un certains nombres de mises en œuvre d'outils de catalogage à l'échelon européen,

- ISO 19115. Cette dernière a le statut de norme internationale (IS) depuis 2003. La norme ISO 19115 permet l'organisation par classes d'un grand nombre d'informations pouvant décrire une donnée.

Ce travail est basé sur la réflexion autour de la proposition d'un langage canonique pour la construction des dictionnaires de métadonnées distribuées [3]. Ce langage constitue une approche pour construire un modèle canonique de métadonnée qui résulterait de l'intersection des standards du domaine. Pour l'implémentation du composant chargé de gérer le catalogue issu de cette vision, nous préconisons une architecture plus légère inspirée de la SOA [8], à la différence de l'architecture CORBA, une infrastructure plus complexe à administrer, utilisée dans [1] pour les accès distants à de tels catalogues.

3. Construction d'un langage canonique de métadonnées

Les métadonnées en matière d'information géographique se composent d'éléments relatifs à l'identification, la représentation spatiale, la qualité, le contenu, les modalités d'accès et de diffusion. La définition d'une norme en matière de métadonnée peut être représentée par des sections qui sont en relations avec d'autres sections ou sous-sections qui n'ont qu'une valeur. Elles apparaissent dès lors comme une structuration d'un ensemble de règles établies. Ces règles respectent une certaine syntaxe et une sémantique, elles peuvent faire l'objet de l'utilisation d'un interpréteur pour analyser si toute occurrence peut oui ou non être considérée comme faisant partie de cette norme.

$Metadata \coloneqq Metadata\ section^{*} | \varepsilon$

$Section \coloneqq NomSection\ ElementComposé^{+}$

$ElementComposé \coloneqq ElementComposé^{*}\ ElementDeDonnée^{+}$

$ElementDeDonnée \coloneqq Nombre | Date | Texte$

$NomSection \coloneqq Idinfo \mid dataqual \mid spdoinfo \mid spref \mid distinfo \mid metainfo$

Il existe un certain nombre de standards ou normes de métadonnée en usage pour la structuration et la distribution des données à référence spatiale. Par exemple, CSDGM, ENV 12657, ISO 19115 [1, 6]. Chaque norme est constituée d'un ensemble varié d'éléments qui décrivent les données géolocalisées à partager. Ces normes sont basées sur une structure hiérarchique et arborescente avec plusieurs éléments (nœuds ou attributs) communs [1]. De même, les éléments terminaux se trouvent sur le même embranchement et les embranchements d'une arborescence qui découle des nœuds de même sorte sont identiques. De telles observations permettent de représenter la

spécification d'une norme de métadonnée géographique en un arbre de dérivation (syntaxique) généré à partir d'une grammaire algébrique, et la construction d'une métadonnée canonique est possible par intersection de plusieurs normes de métadonnée géographique précédée éventuellement d'un renommage de certaines sortes pour les juguler les similarités. La norme canonique obtenue comporte moins d'éléments et est suffisante pour permettre le catalogage dans un dictionnaire commun.

La structure hiérarchique et arborescente d'une norme de métadonnée géographique présente une racine, des nœuds et des feuilles. Des règles de production sont générées à partir de la racine de l'élément « metadata ». La figure 2 présente la structuration d'une norme en matière de métadonnée.

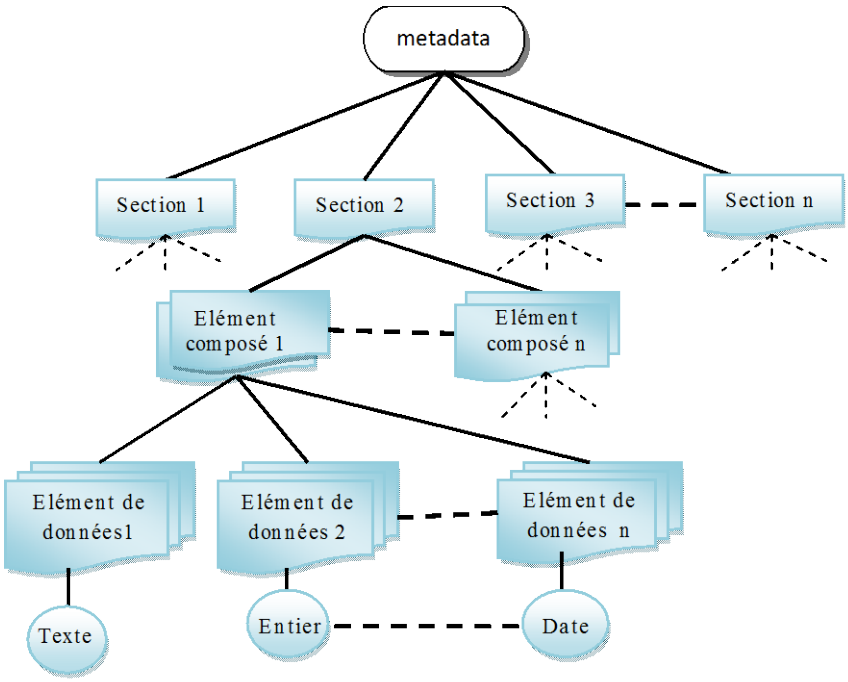


Figure 2. *Structuration hiérarchique*

En définitive, si un élément est présent dans les trois normes, cela signifie qu'il est significatif dans la caractérisation des métadonnées. Ainsi, l'intersection de ces derniers ensembles (normes) fournit tous les terminaux nécessaires pour la caractérisation de la métadonnée. La nouvelle norme obtenue est plus représentative et a moins d'éléments : puisque si $A = B \cap C \cap D$ alors Card $A \leq$ Card B et Card $A \leq$ Card C et Card $A \leq$ Card D. De l'ensemble canonique Ω des symboles non terminaux, résultant de l'intersection

des normes en présence, nous définissons une grammaire G={ Ω, Γ, R, metadata} où Γ est l'ensemble des terminaux, R l'ensemble des règles de la grammaire. Metadata∈Ω est l'axiome de la grammaire.

La figure suivante présente un extrait d'une hiérarchisation de la norme ISO 19115.

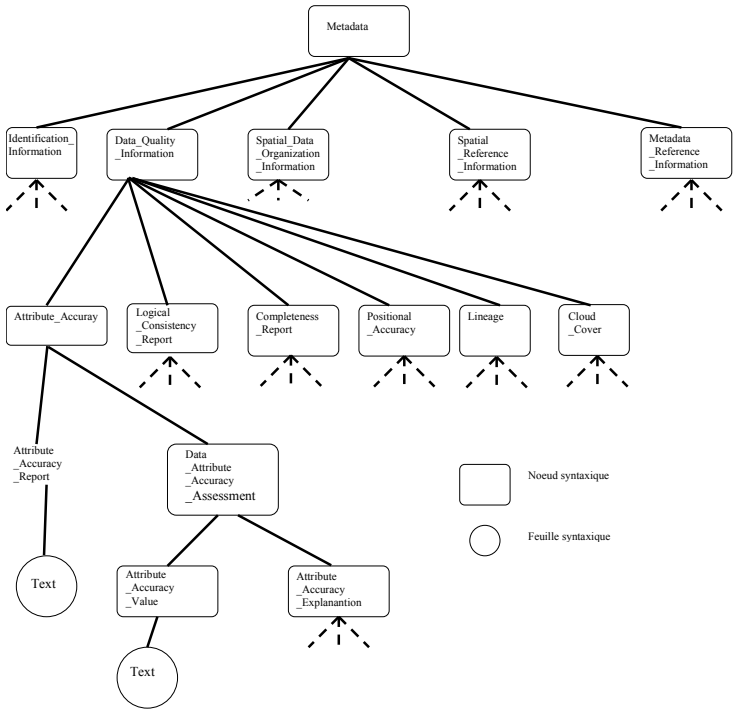


Figure 3 : Implémentation d'un arbre syntaxique hiérarchique

L'usage d'une métadonnée canonique pour la structuration des métadonnées géographiques est une réponse prometteuse au défi que soulève l'interopérabilité syntaxique. Quelques actions peuvent être ajoutées aux règles pour accroitre la sémantique des échanges. Cet usage favorise la construction d'un unique catalogue (dictionnaire) de métadonnées, pré requis pour la construction des plateformes de fédération de données géographiques.

4. Modélisation

Le dictionnaire de métadonnée que nous présentons est conçu et implémenté sous la forme d'un composant réutilisable. Une instance déployée dans une architecture donnée devra être accessible à ses clients grâce au protocole SOAP. Le composant obtenu, appelé GS-DIRECTORY, comprend :

- une façade qui offre aux clients les méthodes pour insérer, supprimer, rechercher et modifier les métadonnées ;
- et un contenu qui se structure selon la hiérarchie des métadonnées. Dans ce sens, elle se conforme naturellement à l'organisation des données dans un annuaire électronique.

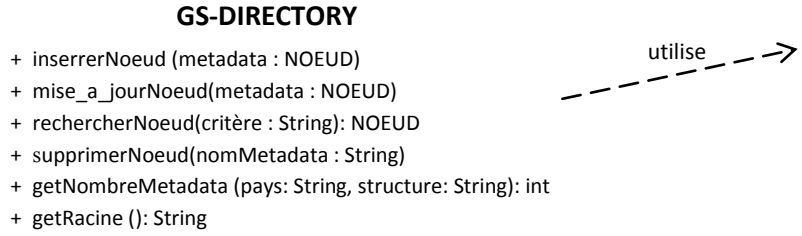


Figure 4. *Architecture à composant du dictionnaire*

En plus des méthodes publiques exposées par la facade, GS-DIRECTORY possède des méthodes privées qui lui permettent de se connecter à LDAP et de lui déléguer certains traitements. Lorsqu'un client invoque un service, GS-DIRECTORY se connecte à l'annuaire LDAP sous jacent, puis l'authentification est effectuée avant l'invocation du service LDAP correspondant. la séquence pour la recherche est illustrée sur la figure 5.

GS-DIRECTORY est une façade au sens des patrons de conception GoF [9] dont l'architecture interne repose sur une structure informationnelle aborescente. A la racine se trouve un nœud correspondant au type « Domain Component » de LDAP. La classe Metadata rassemble un ensemble d'informations pour représenter une métadonnée. Cette classe est fille de la classe Nœud. En plus des attributs hérités, les attributs titre, idinfo, dataqual, spdoinfo, spref, metainfo découlent directement de la structure informationnelle de l'annuaire GS-DIRECTORY.

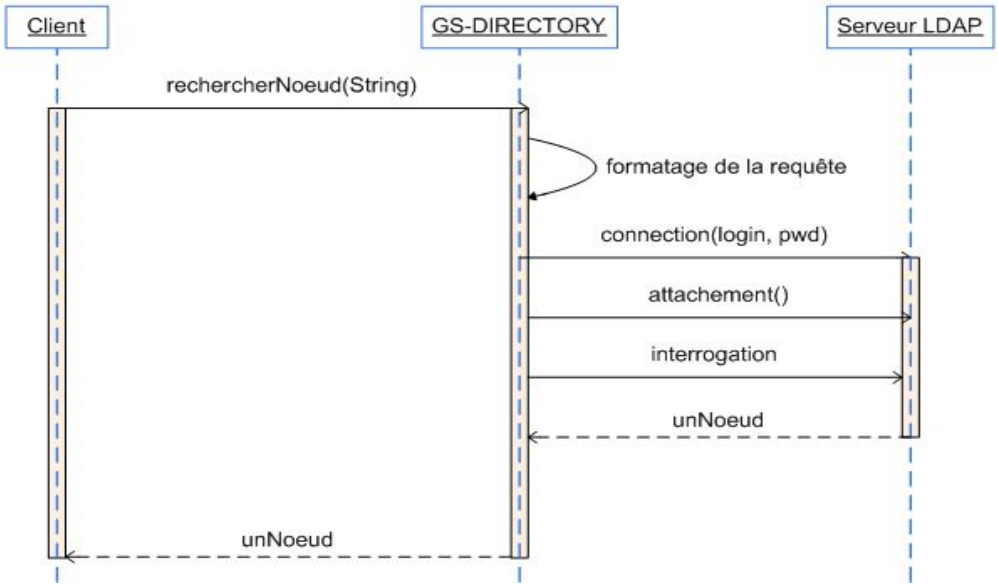


Figure 5. *Séquence d'interaction entre les différents constituants*

De nouvelles entrées sont définies dans LDAP pour représenter les concepts de NŒUD et de FEUILLE conformes à la spécification du langage des métadonnées. La classe Noeud rassemble des informations pour l'implémentation d'un nœud de l'arbre informationnel de GS-DIRECTORY. Les attributs de cette classe sont: le nom, l'abréviation, la description, le distinguished name et la liste des "enfants" de ce nœud (descendants directs). La classe Nœud peut être composé d'autres nœuds (descendants directs) ou des feuilles (champ de saisie).

La classe Feuille est similaire à la classe Noeud à la seule différence que l'attribut valeur représente une valeur opérationnelle d'un attribut, et ses instances sont des terminaux [3].

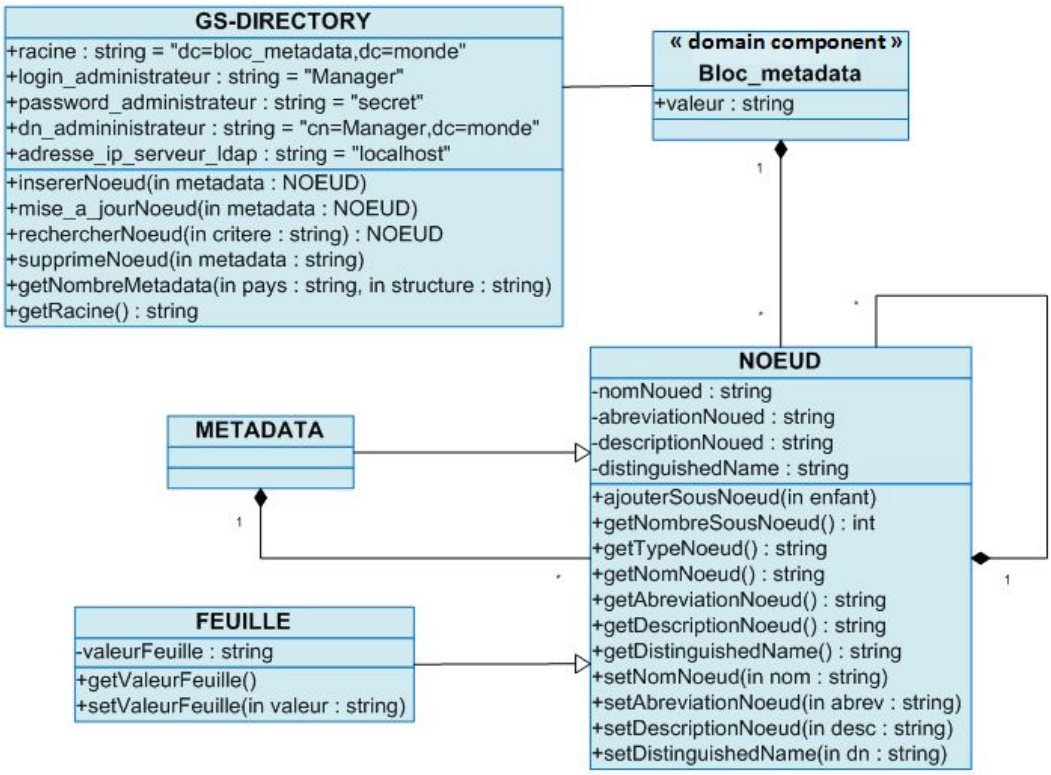


Figure 6 : *Structure Interne de l'annuaire*

Cette représentation, inspirée du patron composite [9] permet d'implémenter les structures de données hiérarchiques. Cette approche améliore la flexibilité et la réutilisation de la composante « Annuaire » qui est le pilier de notre plateforme de fédération.

5. Validation

Pour expérimenter notre modèle, une mise en œuvre d'un prototype fonctionnel de GS-DIRECTORY est réalisée en utilisant openLDAP qui est une implémentation du protocole LDAP. L'utilisation des standards favorisent l'hétérogénéité de l'environnement de test et la flexibilité de localisation des composants :

- la distribution des principaux constituants sur différents ordinateurs,

- l'indépendance des langages pour écrire chaque constituant,
- la réutilisation accrue de l'existant en terme d'API,
- chaque constituant peut être installé sur un ordinateur différent.

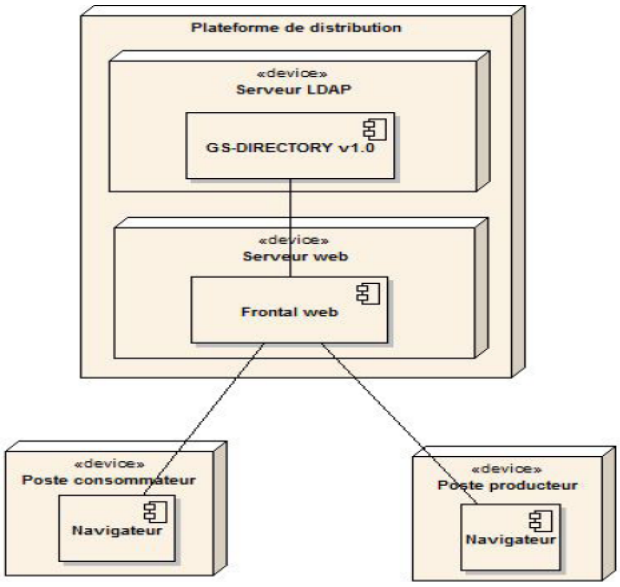


Figure 7. *Architecture de déploiement des constituants de test*

Pour tester le composant GS-DIRECTORY, un frontal a été développé en PHP dans un environnement dont l'architecture de déploiement est illustrée par la figure 7. Il est constitué d'une interface Web couplé à un module de visualisation des cartes. Cette combinaison offre la possibilité d'enregistrer de nouvelles entrées, d'en supprimer et de rechercher dans le catalogue. Les résultats d'une recherche sont proposés à l'utilisateur qui peut sélectionner les couches de données à visualiser (cf. figure 8).

Les critères de recherche sont représentés par l'onglet 1 suivant l'arborescence des métadonnées canoniques, puis l'onglet 2 affiche les résultats de la recherche et enfin le 3 permet de visualiser les géodonnées obtenues à l'issue de la recherche. Le frontal web de la plateforme donne l'accès aussi bien au consommateur qu'au producteur. Ce frontal permet de visualiser les données géographiques hébergées par les producteurs et obtenus par les critères de recherche proposer par la plateforme. Ainsi l'on peut mieux observer la qualité des données géographiques.

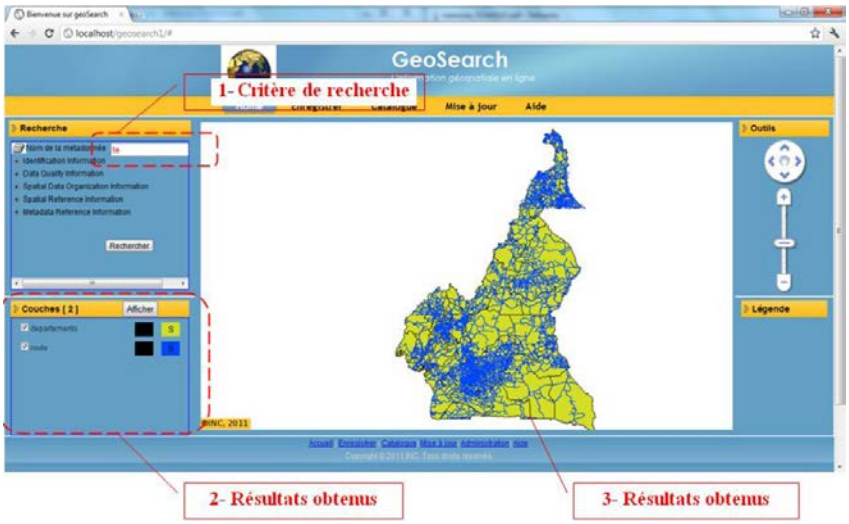


Figure 8. *Une vue de l'interface de navigation et de recherche*

6. Conclusion

Notre préoccupation initiale consistait à structurer et mettre en œuvre une composante de mémorisation des métadonnées devant servir de support pour une plateforme de fédération des données à référence spatiale d'une part, et d'autre part de permettre à un producteur de référencer ses données partagées dans un tel environnement.

Pour modéliser le dictionnaire de métadonnées géographiques GS-DIRECTORY, nous avons utilisé une métadonnée canonique qui résulte de l'intersection des standards de métadonnée utilisés dans le domaine. Ensuite, nous avons montré la possibilité d'utiliser le protocole LDAP pour la mise en œuvre de ce dictionnaire de métadonnées géographiques alors que celui-ci est traditionnellement utilisé par les annuaires de personnes et à l'authentification des utilisateurs dans une organisation. Pour cela, nous avons décrit deux structures de données propres aux annuaires LDAP pour la conception de l'arbre informationnel ou structure arborescente de base. Plus précisément, nous avons défini la structure d'un nœud et d'une feuille. Sur le plan technique, la description architecturale et fonctionnelle du modèle de classe de GS-DIRECTORY a été faite.

Pour valider notre étude, une mise en œuvre d'un prototype fonctionnel de GS-DIRECTORY a été réalisée en utilisant openLDAP qui est une implémentation du protocole LDAP. Pour tester le composant GS-DIRECTORY, un frontal a été

développé également en PHP. Il est constitué d'une interface Web couplé à un module de visualisation des cartes. Cette combinaison offre la possibilité d'enregistrer de nouvelles entrées, d'en supprimer et de rechercher dans le catalogue. Les résultats d'une recherche sont proposés à l'utilisateur qui peut sélectionner les couches de données à visualiser.

7. Bibliographie

[1]. Ramroop S., Pascoe R. (2001). Use of spatial metadata in a LDAP/CORBA architecture. *The 12th Annual colloquium of the Spatial In formation Research Center*. Dunedin, New Zealand: University of Otago.

[2]. Leclerq E., Benslimane D., Yetongnon K., (1998). ISIS : une architecture multi-agents pour l'interopérabilité des SIG. Colloque SMAGET Cemagref-Engref. France: Université Clermont-Ferrand.

[3]. Tongo L.-E., Kouamou G. E. (2009). Building a Canonical Language for Distributed System Repository. *Procedings of the 4th ICSEA,* 2009, pp. 173-178

[4]. Cullot N., Parent C., Spaccapietra S. (2003). Des ontologies pour données géographiques. *Revue Internationale de Géomatique*, Vol 13/3, pp 285-306

[5]. Belhadef H., Kholladi M.-K. (2007). Conception d'une nouvelle ontologie pour l'interopérabilité des systèmes d'informations géographique. *The 4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications* (SETIT 2007). Tunisia.

[6]. Treguer M. (2006). Techniques d'interopérabilité au service de l'intégration des données géographiques . *La lettre du Sismer*, IFREMER de Brest Technopole Brest Iroise.

[7]. Ramroop S., Pascoe R. (2001). Use of LDAP to partially implement the OGIS discovery service. *International Journal of Geographical Information Science*, Volume 15, Issue 5, 2001

[8]. Rouillard J., Vantroys T., Chevrin V. (2007). *Les architectures orientées services, une approche pragmatique des SOA*. Viubert, Paris.

[9]. Gamma E., Helm R., Johson R. and Vlisside J. (1994). *Design Patterns: Elements of reusable object-oriented software*, Addison Wesley.