



---

## 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.

L'information géographique n'échappe pas à cette réalité. Les SIGs s'intègrent maintenant dans les systèmes d'informations interconnectés, transversaux et ouverts sur l'extérieur, appelés Infrastructure de Données Spatiales (IDS) ou géoportails. Cependant, cette idée de fédération doit reposer sur des plateformes logicielles qui assurent l'organisation et l'acheminement effectif des données à référence spatiale. Ces plateformes de distribution doivent disposer d'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 ces types 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 en se basant sur la nécessité des catalogues de métadonnées dans le contexte d'automatisation de la

fédération des données géographiques. La section 3 présente le langage canonique pour la construction des dictionnaires de métadonnées distribuées qui est le point de départ de ce papier. La section 4 décrit la structuration des concepts inhérents aux métadonnées géographiques et leur mise en œuvre dans LDAP. Une conclusion clôture le papier en précisant le cadre de validation qui a été mis en place.

---

## 2. Motivation

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 à référence spatiale. Le support d'échange contiendra au minimum la géodonnée échangée (au format du logiciel utilisé) ainsi que la métadonnée (un fichier texte) décrivant cette donnée.

D'autres approches ont été proposées à cet effet, en particulier la définition d'une ontologie géographique [4,5]. Cette proposition est présentée comme une réponse prometteuse au besoin croissant de partage d'informations 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 standards à prendre en compte. La définition d'un schéma commun pour tous les standards 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 implémentant 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 [1, 3,7].

L'expérience de Ramroop & Pascoe [1] 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 y a un nombre « *infini* » de branche dans la définition de l'arborescence d'un catalogue dans LDAP.

Ce travail est le prolongement d'une réflexion autour de la proposition d'un langage canonique pour la construction des dictionnaires de métadonnées distribuées [3]. Ce langage canonique constitue une approche pour construire un modèle canonique de métadonnée résultant de l'intersection des standards du domaine. Cette démarche rend possible la construction d'un catalogue unique de métadonnée. A la différence de l'architecture CORBA utilisée dans [1] pour les accès distants, nous préconisons une architecture plus légère inspirée de la SOA [8].

---

### 3. Un langage canonique pour les dictionnaires de métadonnées géographiques

Il existe un certain nombre de standards ou normes de métadonnée géographique en usage. Par exemple, CSDGM, ENV 12657, ISO 19115 [1, 3, 6]. Chaque norme est constituée d'un ensemble varié d'éléments qui décrivent les données géolocalisées à stocker. Ces normes sont basées sur une structure hiérarchique avec plusieurs éléments (nœuds ou attributs) communs. En outre, les éléments terminaux se trouvent sur le même embranchement et les embranchements de l'arborescence qui découlent des nœuds communs sont identiques.

Le constat suivant se pose. Si un élément est présent dans les trois normes, cela signifie qu'il est significatif dans la caractérisation d'une métadonnée. Ainsi, l'intersection de ces derniers ensembles (normes) fournit tous les non-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  $\text{Card } A \leq \text{Card } B$  ;  $\text{Card } A \leq \text{Card } C$  et  $\text{Card } A \leq \text{Card } D$ .

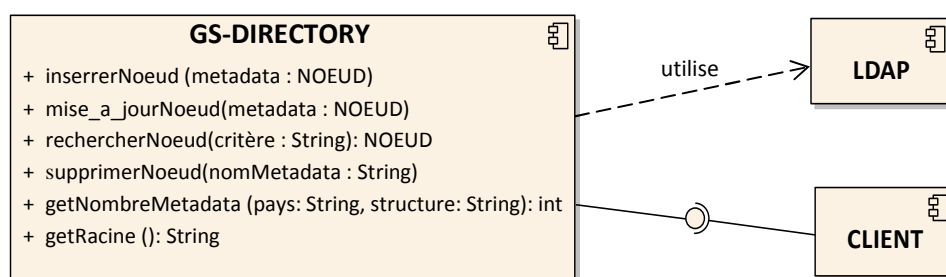
De l'ensemble canonique  $\Omega = \{\text{idinfo}, \text{dataqual}, \text{spdoinfo}, \text{spref}, \text{distinfo}, \text{metadata}\}$  qui résulte de l'intersection de l'ensemble des non terminaux provenant des normes en présence, une grammaire formelle  $\Theta = \{\Omega, \Gamma, \mathbf{R}, \text{metadata}\}$  est défini où  $\Gamma$  est l'ensemble des terminaux et  $\mathbf{R}$  l'ensemble des règles de grammaire.  $\text{Metadata} \in \Omega$  est l'axiome de la grammaire. De ce fait,  $\text{metadata} := \text{idinfo} \mid \text{dataqual} \mid \text{spdoinfo} \mid \text{spref} \mid \text{distinfo} \mid \varepsilon$ . La liste complète des éléments de  $\mathbf{R}$  permet d'aboutir à l'arbre de dérivation syntaxique de la métadonnée canonique décrite dans [3].

En définitive, l'usage d'une métadonnée canonique issue du langage canonique pour la structuration des métadonnées géographiques est une réponse prometteuse au défi que soulève l'interopérabilité sémantique et syntaxique. 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

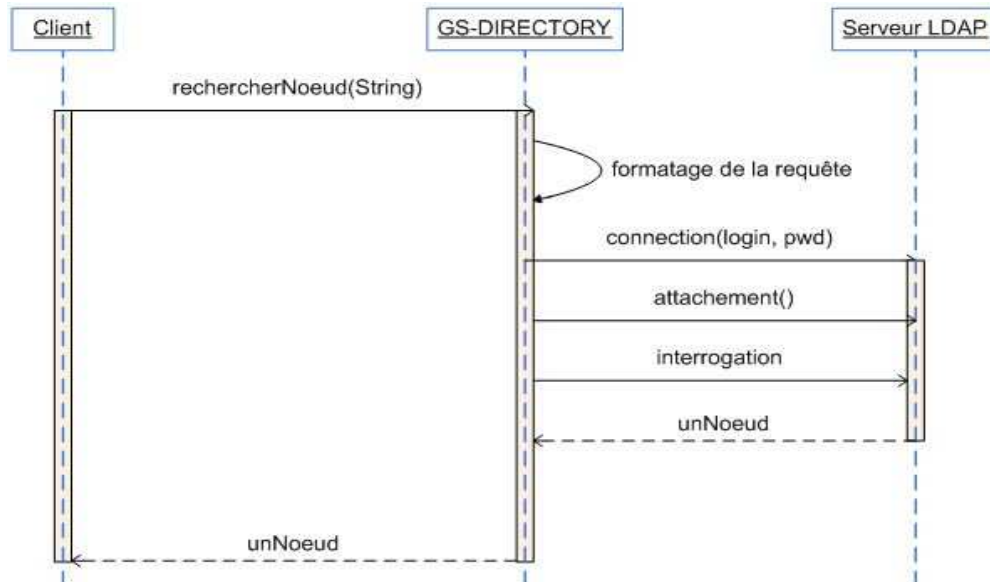
Le dictionnaire de métadonnée que nous présentons est conçu et implémenté sous la forme d'un composant réutilisable. Une instance déployée dans une architecture donnée devra être accessible à ses clients grâce au protocole SOAP. Le composant obtenu, appelé GS-DIRECTORY, comprend :

- 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 1.** Structure composite de GS-DIRECTORY

En plus des méthodes publiques exposées par la façade, GS-DIRECTORY possède des méthodes privées qui lui permettent de se connecter à LDAP (composant fournissant des services d'annuaire) 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. Un exemple est illustré sur la figure suivante où le client souhaite faire une recherche dans l'annuaire.



**Figure 2.** Séquence d'invocation d'un service de GS-DIRECTORY

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 NOEUD. En plus des attributs hérités, les attributs *titre*, *idinfo*, *dataqual*, *spdoinfo*, *spref*, *metainfo* découlent directement de la structure informationnelle de GS-DIRECTORY.

De nouvelles entrées sont définies dans LDAP pour représenter les concepts de NOEUD et de FEUILLE conformes à la spécification du langage des métadonnées [3]. 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 NOEUD 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].

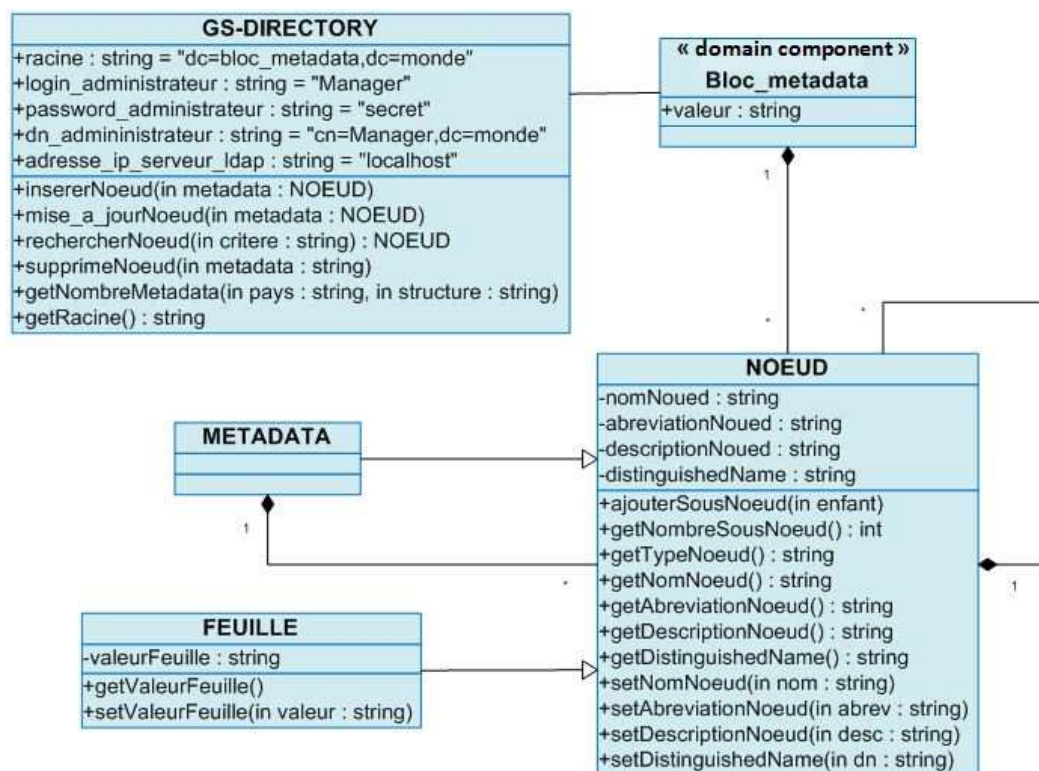


Figure 3. Diagramme de classe de GS-DIRECTORY

## 5. Conclusion

Notre préoccupation initiale consistait à structurer et mettre en œuvre une composante de mémorisation des métadonnées géographiques devant servir de support pour les plateformes de fédération de 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 à été faite.

Pour valider notre étude, une mise en œuvre d'un prototype fonctionnel de GS-DIRECTORY est 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.

---

## 6. Bibliographie

- [1] S. Ramroop and R. Pascoe, 2001. Use of spatial metadata in a LDAP/CORBA architecture. *The 12th Annual colloquium of the Spatial Information Research Center*. Dunedin, New Zealand: University of Otago.
- [2] E. Leclercq, D. Benslimane, & K. Yetongnon, (1998). ISIS : une architecture multi-agents pour l'interopérabilité des SIG. *Colloque SMAGETn Cemagref-Engref*. France: Université Clermont-Ferrand.
- [3] L. E. Tongo, G. E. Kouamou, 2009. Building a Canonical Language for Distributed System Repository. *Proceedings of the 4<sup>th</sup> ICSEA 2009*, pp. 173-178
- [4] N. Cullot, C. Parent et S. Spaccapietra, 2003. Des ontologies pour données géographiques. *Revue Internationale de Géomatique*, Vol 13/3, pp 285-306
- [5] H. Belhadef & M.-K. Kholadi, (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] M. Treguer, (2006). Techniques d'interopérabilité au service de l'intégration des données géographiques . *La lettre du Sismer, IFREMER de Brest Technopole Brest Iroise*.
- [7] S. Ramroop & R. Pascoe. Use of LDAP to partially implement the OGIS discovery service. *International Journal of Geographical Information Science*, Volume 15, Issue 5, 2001
- [8] J. Rouillard, T Vantroys, V. Chevrin, 2007. *Les architectures orientées services, une approche pragmatique des SOA*. Viubert, Paris.
- [9] E. Gamma, R. Helm, R. Johnson and J. Vlisside, 1994. *Design Patterns : Elements of reusable object-oriented software*, Addison Wesley.