

# Construction d'un entrepôt de métadonnées - LOM

## Application: E-learning

*Nawel Iles*, Azzeddine Chikh, Sidi Mohammed Chouiti  
Faculté des sciences de l'ingénieur  
Université de Tlemcen Algérie  
([n\\_iles/ az\\_chikh /chouitism@mail.univ-tlemcen.dz](mailto:n_iles/az_chikh/chouitism@mail.univ-tlemcen.dz))

**Résumé:** Face à l'évolution rapide des moyens d'information et de communication dans le domaine du E-Learning, une masse importante de documents pédagogiques est produite chaque jour à travers de nombreuses universités. Cette augmentation sans cesse croissante des documents a engendré plusieurs difficultés liées aux volumes à manipuler, à leurs coûts élevés, mais également à leur hétérogénéité. Nous proposons dans cet article, un modèle d'entrepôt qui repose sur l'indexation des documents pédagogiques grâce aux métadonnées de LOM visant à faciliter leur réutilisabilité. Ce travail rentre dans le cadre d'un projet « EPOIS » Entrepôt d'Objet d'apprentissage et Indexation Sémantique.

**Mots clés:** E-Learning, ressource pédagogique, entrepôt, métadonnée « LOM », XML

### I- Introduction

Les nouvelles Technologies de l'Information et de la Communication "TIC" améliorent profondément nos façons de nous informer, de communiquer et de nous former. Cette émergence technologique fait apparaître un nouveau mode d'apprentissage connu sous le nom de e-learning. Celui-ci est basé sur l'accès à des formations en ligne, interactives et parfois personnalisées, diffusées par l'intermédiaire d'un réseau (Internet ou Intranet) ou d'un autre média électronique.

De nombreuses ressources pédagogiques, peuvent être utilisées dans le cadre du e-learning. Certaines sont produites par les différents acteurs impliqués dans la formation, d'autres sont accessibles à partir du web : cours en ligne, supports de cours, etc. Ces ressources sont difficilement réutilisables car elles n'ont pas été conçues à cet effet. Il serait alors indispensable de les intégrer en vue de les rendre facilement accessibles, exploitables et réutilisables. Le présent travail, qui rentre dans le cadre du projet EPOIS (EntrePôt d'Objets d'apprentissage et Indexation Sémantique) qui permet de contribuer à la résolution de ces problématiques et qui vise à rendre chaque composant d'un objet pédagogique « retrouvable, réutilisable, indexable », vise à construire un entrepôt pédagogique qui sera utilisé à la fois par l'enseignant et l'apprenant. Trois aspects de cet entrepôt doivent être pris en charge :

- 1- Description avec « LOM » des ressources pédagogiques
- 2- Représentation SCORM des objets pédagogiques
- 3- Spécification avec IMS-LD des scénarios d'apprentissage

Les trois représentations précédentes seront complétées par des ontologies de domaine en vue d'explicitier davantage les connaissances qu'elles contiennent.

Dans cet article, nous nous intéressons au premier aspect de description avec les métadonnées de LOM des ressources pédagogiques contenues dans l'entrepôt et représentées en XML.

L'originalité de cette approche réside dans sa capacité à intégrer tout type de documents pédagogiques complexes, représentés dans des formats différents (textes, images, sons, vidéos, bases de données, etc.) et issus de sources diverses (données de production, scanners, satellites, enregistrements vidéos, résultats d'analyse, Web, etc.).

La première partie de cet article est consacrée à la présentation des domaines clés intervenant dans notre problématique à savoir les entrepôts de données, les métadonnées et le langage XML. La deuxième partie est consacrée à notre modèle d'entrepôt pédagogique.

## II- Domaines d'intérêt

### 1- Les entrepôts de données

De nombreuses définitions ont été proposées cherchant à orienter ces définitions dans un sens mettant en valeur leur produit. La définition la plus appropriée est :

« Un entrepôt est défini comme "une collection de données intégrées, orientées sujet, non volatiles, historisées, résumées et disponibles pour l'interrogation et l'analyse"[DEVL97].

Les données de l'entrepôt doivent respecter les caractéristiques suivantes :

- **Orientées sujet** : Les données de l'entrepôt sont organisées par sujets ou thèmes d'analyse, ce qui permettra de rassembler toutes les données pertinentes par rapport à un sujet et nécessaires aux besoins d'analyse.
- **Intégrées** : les données d'un entrepôt sont le résultat de l'intégration de données en provenance de sources différentes et hétérogènes
- **Historisées** : l'entrepôt doit permettre l'historisation des données c.-à-d. garder leurs différentes versions.
- **Filtrées** : l'entrepôt ne doit contenir que les données pertinentes et utiles pour l'avenir.
- **non volatile** : les données sont matérialisées et ne sont pas modifiables par l'utilisateur. L'entrepôt est mis à jour de façon constante.

L'entrepôt de données offre à l'entreprise les avantages suivants [ROUS02] :

- il constitue une collection de données centralisée disponible pour l'aide à la décision (OLAP, datamining,...),
- les évolutions des données de l'entrepôt sont conservées (historisation des données),
- il contient un ensemble de données consolidées (données homogènes et fiables),
- il contient des données agrégées permettant une analyse à différents niveaux de détails,

- il permet de développer différents thèmes d'analyse (réorganisation en fonction des sujets à analyser).
- L'accès direct aux données : facile, rapide, sécurisé
- Plus de contrôle sur les données
- Possibilité d'ajouter facilement des annotations

Dans notre travail, les données représentent essentiellement des documents pédagogiques. Le problème majeur des entrepôts pédagogiques est dû à l'hétérogénéité sémantique des documents issus de bases différentes. Une Solution à ces problèmes repose sur l'intégration de ces documents à travers leurs métadonnées.

## **2- Les métadonnées**

Les métadonnées sont définies comme étant des données relatives à d'autres données (data about data : données sur des données). Par conséquent, une notice catalographique classique peut-être considérée comme une métadonnée. Ces métadonnées peuvent être incluses dans les ressources elles-mêmes ou enregistrées dans un fichier séparé selon le type du contenu. Les métadonnées sont structurées suivant des catégories ou champs sémantiques. Chaque champ représente une caractéristique particulière sur la ressource, exemple, son titre ou son résumé. Les métadonnées peuvent décrire des ensembles plus petits qu'un document, par exemple, des images, ou des fichiers sonores, à l'intérieur d'un document [BOUR 00].

Les métadonnées ont pour objectifs de :

1. Permettre une description plus ou moins détaillée des ressources.
2. Faciliter le repérage de l'information : elles permettent une facilité de recherche dans la masse informationnelle du web,
3. Permettre une évaluation rapide de la pertinence du contenu d'un document
4. Permettre la gestion des droits d'accès ;
5. Faciliter l'organisation et la gestion de collections de ressources
6. Certifier une certaine autorité intellectuelle du contenu (métadonnées sur l'auteur, la date de création, l'organisme responsable, la date de mise à jour...)

En e\_learning les métadonnées nous permettent une recherche sémantique efficace des ressources pédagogiques sur le Web et leur traitement automatique. Différentes normes ont été définies pour aider l'élaboration de systèmes d'apprentissage, des objets pédagogiques associés, leur représentation et leur inter-relation. L'application de ces normes, qui sont considérées comme des langages communs de description des ressources éducatives numérisées, garantit non seulement l'interopérabilité mais également la qualité du système [HERN07]. Parmi les normes de la formation en ligne, on peut citer LOM, SCORM et IMS-LD. LOM s'intéresse à la description des ressources pédagogiques, SCORM à la structure du contenu des objets, et IMS-LD au scénario d'apprentissage. Nous présentons ci-dessous le langage de description LOM [GOIT01].

**Le langage de description LOM** spécifie la syntaxe et la sémantique des métadonnées décrivant des ressources pédagogiques numériques ou non et définit les attributs nécessaires à une description complète des ressources pédagogiques à partir des 78 attributs divisés en neuf catégories :

1. **Généralité** : caractéristiques indépendantes du contexte comme l'identifiant, le titre, la langue de la ressource...
2. **Cycle de vie** : caractéristiques relatives au cycle de vie, comme la version ou l'état (brouillon, final, révisé, non disponible) ;
3. **Méta-métadonnées** : caractéristiques de la description elle-même comme : la contribution (personnes ayant participé à l'élaboration), catalogue, ...
4. **Technique** : caractéristiques techniques comme le format (du logiciel nécessaire pour accéder à la ressource), taille...
5. **Pédagogie** : caractéristiques pédagogiques : type d'interactivité, type de ressource; niveau d'interactivité, densité sémantique, rôle présumé de l'utilisateur final, contexte, tranche d'âge, difficulté, temps d'apprentissage moyen, description, langage ;
6. **Droits** : coûts, copyrights, description...
7. **Relation** : caractéristiques exprimant les liens avec d'autres ressources ;
8. **Annotation** : commentaires sur l'utilisation pédagogique de la ressource
9. **Classification** : caractéristiques de la ressource décrites par des entrées dans des systèmes de classification : but, classification de références, chemin...

LOM est très complet et permet d'identifier tout type de contenu, et plus spécifiquement les contenus éducatifs, avec un format commun décrivant le type de contenu, son auteur, la meilleure manière de l'utiliser, etc. Les profils d'applications ont été introduits afin de répondre aux besoins spécifiques et concrets des utilisateurs. Un profil d'application est une instance d'un modèle comme LOM dans un contexte particulier. De fait, cela signifie interpréter, raffiner, étendre ou parfois même simplifier les syntaxes et les sémantiques des méta-données LOM.

### 3- Le langage XML

« XML » eXtensible Markup est un langage de description et d'échange de documents structurés. Il permet de décrire la structure logique des documents. A l'aide d'un système de balisage, XML permet de marquer les éléments qui composent la structure et les relations entre ces éléments [AMAN 01]. XML apparaît actuellement comme le moyen le plus adéquat pour décrire, dans la perspective d'une diffusion sur l'intranet, à la fois métadonnées et structure des documents. Le rapprochement entre XML et les entrepôts de données est très prometteur et le passage par XML semble être une voie privilégiée pour entreposer des données complexes. XML permet d'assurer les avantages suivants :

- **Vues multiples des données**- XML a la possibilité d'afficher les documents de différentes façons
- **Séparation du contenu, de la structure et de la présentation** : l'idée centrale d'XML est qu'il permet d'apporter de la valeur ajoutée si les trois aspects fondamentaux d'un document que sont son contenu, sa présentation et sa structure sont séparés pour celui qui le rédige.

- **Modularité et réutilisation des structures types** : XML permet de définir librement la structure type d'un document.
- **Recherches plus faciles** : la recherche des documents devient très facile et plus puissantes
- **Interopérabilité**- Les documents provenant de plusieurs sources peuvent être intégrées et manipulées par différentes applications.

### III- Modélisation de l'entrepôt de métadonnée

Le projet EPOIS vise à proposer un modèle d'entrepôt d'objets pédagogiques qui soit structuré de sorte que les composants puissent être facilement retrouvés par les concepteurs (enseignants), puis agrégés afin de recomposer de nouveaux objets pédagogiques. En effet, un concepteur va réaliser un cours en agrégeant différents composants élémentaires (images, fragment de textes, etc) qui vont correspondre à des sections, à leur tour agrégées pour réaliser des modules puis des formations.

L'indexation des composants doit répondre aux exigences et besoins des concepteurs et devra pour cette raison prendre en compte plusieurs facettes : les contenus des composants en lien avec les connaissances à acquérir, les métadonnées (comme celles préconisées par LOM), la structuration des objets (comme le préconise SCORM), la théorie éducative utilisée lors d'un scénario, etc. C'est cette indexation qui permettra aux concepteurs de retrouver les objets existants qui répondent à leurs besoins pédagogiques. L'objectif du projet EPOIS est que le concepteur puisse interroger l'entrepôt d'objets au travers d'un langage d'interrogation naturel et de façon flexible par rapport à ses besoins. Pour mener à bien ce projet, nous présentons dans cet article, la modélisation de l'entrepôt dans le but d'indexer les documents pédagogiques afin de faciliter la recherche et la réutilisation.

Un document pédagogique est composé d'objets pédagogiques, qui peuvent être des exercices, des définitions, des exemples, etc... Chaque objet pédagogique peut rassembler des composants élémentaires qui peuvent être de format numérique (.DOC, .PDF, .JPG etc) [HERN07]. Un objet pédagogique peut par ailleurs être composé d'autres objets pédagogiques (norme scorm).

Les objets pédagogiques ont les propriétés suivantes:

- **autonomie** : chaque objet pédagogique peut être utilisé indépendamment des autres;
- **réutilisabilité** : un objet pédagogique élémentaire peut être utilisé dans des contextes et dans des buts multiples;
- **agrégation** : les objets pédagogiques peuvent être regroupés pour constituer d'autres objets pédagogiques;
- **indexation**: chaque objet pédagogique est muni d'une description permettant de le retrouver facilement.
- **au moment voulu** : comme les objets pédagogiques sont indexés, on peut les retrouver instantanément;
- **sur mesure** : les objets pédagogiques permettent une personnalisation des cours à l'échelle d'une organisation ou de chaque personne.

L'utilisation de la norme LOM permet outre une description pertinente des documents, la possibilité d'échange avec d'autres applications utilisant également cette norme. Le modèle LOM permet de décrire chaque objet pédagogique par ses neuf catégories comprenant au total 79 éléments comme le montre la figure 1.

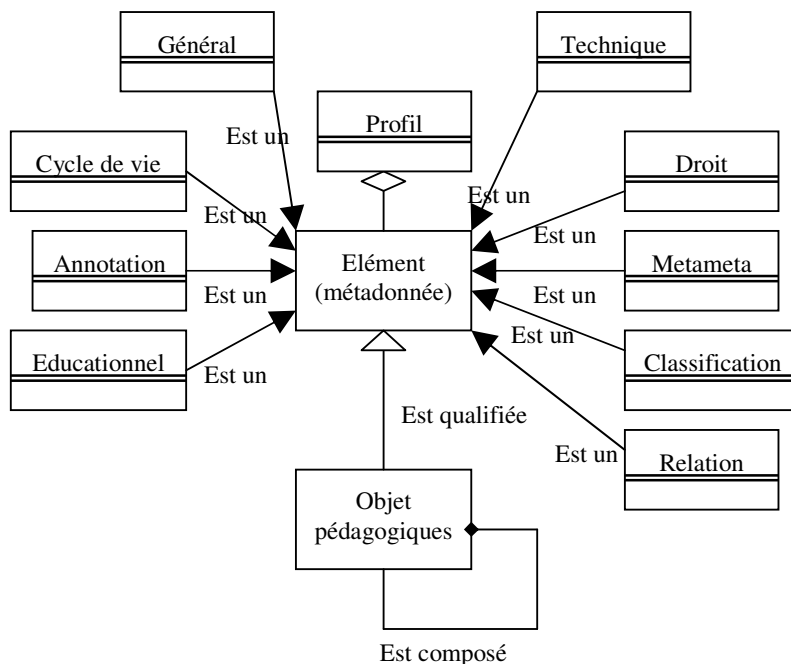


Figure 1 : Modèle de Description d'un objet pédagogique par LOM.

Notre étude requiert l'intervention de deux catégories d'acteurs, ceux qui sont chargés de gérer et qualifier les objets pédagogiques de l'entrepôt "experts", ainsi que ceux qui utilisent l'entrepôt à des fins de recherche multicritère "utilisateurs".

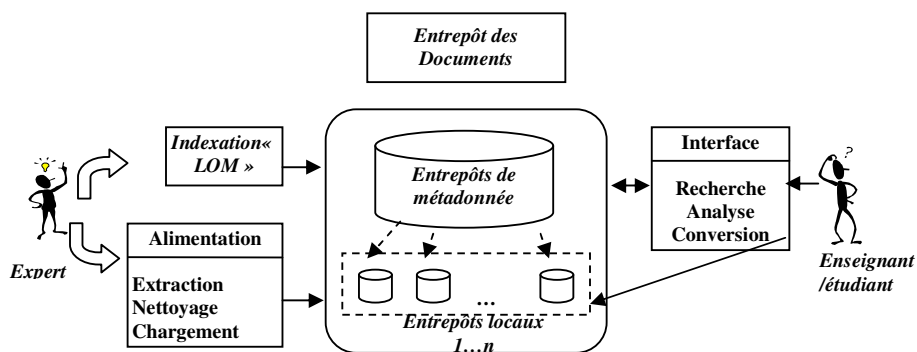


Figure 2 : représentation des acteurs dans l'entrepôt de métadonnées

- **Les experts** décomposent le document en objet pédagogique grâce à la norme SCORM, et les indexant avec LOM.

- **Les utilisateurs** peuvent être soit des enseignants ou des étudiants. Ils désirent réaliser des recherches multicritères sur l'entrepôt pédagogique, afin de repérer et d'acquérir des documents adaptés à leurs demandes d'une manière simple, rapide et efficace.

Les différents descripteurs résultant de l'indexation des objets pédagogiques avec LOM sont rassemblés dans un entrepôt de métadonnées. L'entrepôt pédagogique sera constitué d'un ensemble d'entrepôts locaux contenant les objets pédagogiques eux-mêmes et d'un entrepôt commun des métadonnées associées à ces objets.

Le module alimentation « Extraction Nettoyage Chargement »: est destiné aux experts. Son rôle consiste à l'alimentation des entrepôts locaux par les objets pédagogiques jugés nécessaires.

Nous montrons ci-dessous un *exemple d'une instance XML qui décrit deux objets pédagogiques qualifiés avec des métadonnées LOM*

```
< ?xml version='1.0' encoding ='UTF-8' ? >
<results>
  <document q_id= "q1"
    <title> initiation à Microsoft Word </title>
    <uri> http://www.documentserveur1.dz/univ1/a.pdf </uri>
    <size> 4280 </size>
    <Educational.difficulty> facile </Educational.difficulty>
  </document>
  <document q_id= »q2 «
    <title> Architecture des ordinateurs </title>
    <uri> http://www.documentserveur1.dz/univ5/ordi.ppt </uri>
    <size> 5800 </size>
  </document>
</results >
```

La figure 3 montre la phase d'indexation d'un document pédagogique, on y voit les différents descripteurs de LOM.

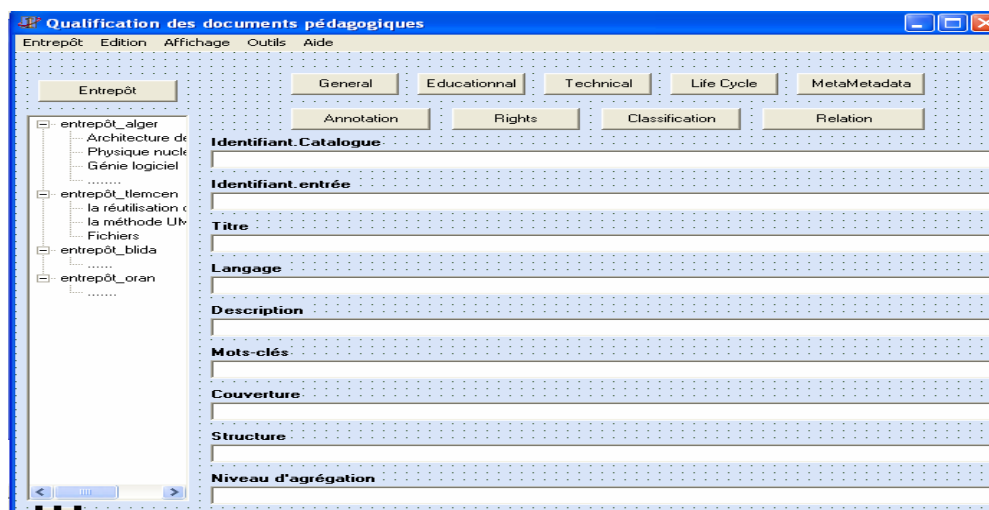


Figure 3 : Indexation d'un objet pédagogique

## **VI- Conclusion**

Dans l'enseignement, le recours aux métadonnées apparaît aujourd'hui comme l'une des meilleures solutions pour faciliter l'accès aux objets pédagogiques pertinents. L'association du concept des métadonnées avec le langage XML, est en train de métamorphoser la manière de stocker, de gérer et d'utiliser les ressources numériques. Le « LOM » spécifie une variété de propriétés techniques et bibliographiques pour les ressources pédagogiques, à travers laquelle on peut déduire l'objectif pédagogique « le Quoi ? », et les critères temporelles « Quand ? ».

Nous travaillons actuellement, dans le cadre du projet EPOIS, sur la réalisation d'un prototype d'indexation automatique des objets pédagogiques. Nos futurs efforts porteront sur la structuration des contenus des ressources pédagogiques avec la norme SCORM, et puis la communication avec un environnement web (scénarios d'apprentissage) avec IMS-LD.

## Références bibliographiques

- [AMAN 01] Bernd Amann, « Données Semistructurées et XML », Conservatoire National des Arts et Métiers, Paris, February 8, 2001 ;
- [BOUR 00] Yolaine Bourda et Marc Hélier, « MÉTADONNÉES ET XML: APPLICATIONS AUX OBJETS PÉDAGOGIQUES », Plateau de Moulon, F-91192 Gif-sur-Yvette cedex France, 2000 ;
- [DEVL 97] Devlin, B. "Data Warehouse from Architecture to Implementation. Addison Wesley Longman", inc.432 p. 1997;
- [HERN07] Nathalie Hernandez , Josiane Mothe, « Indexation multi-facettes des ressources pédagogiques pour faciliter leur ré-utilisation », INRIA Toulouse, France
- [GOIT01] Yacouba GOITA, « Les applications de XML à la production d'objets pédagogiques interactifs », *DEA préparé au sein du Laboratoire CLIPS –IMAG, Université Josph Fourier*, 2001 ;
- [ROUS 02] David ROUSSE, « Entrepôt de données », Université Paul Sabatier., TOULOUSE 2001-2002 ;