

## Métadonnées et XML – une initiation

*Les mots ne sont pas les choses qu'ils représentent.*

(Alfred Korzybski, *Une carte n'est pas le territoire*,  
1933 à 1950. Editions de l'Éclat, 1998, p.64)

### Résumé

Dans le traitement des données, qui est l'une des tâches les plus difficiles du processus de recherche, on distingue deux étapes principales : la classification de l'information et l'analyse proprement dite des données.

Les outils de recherche actuels offrent des possibilités de recherche basées essentiellement sur les mots-clés (une première description des métadonnées). Les critères saisis par l'utilisateur n'ont pas de sémantique associée. C'est pourquoi une analyse des concepts présents dans chaque document doit être réalisée en permettant aux outils de recherche d'associer à la requête de l'utilisateur une sémantique précise et bien définie. La capacité de coder des structures de connaissance donne aux métadonnées un rôle principal dans la gestion de connaissance car selon ([Ruggles 1997](#)), "la gestion de la connaissance couvre trois activités de connaissance principales : la génération, la codification et le transfert".

Une application très intéressante on trouve dans le domaine des « objets pédagogiques » où la documentation pédagogique numérisée est au début. Les objets pédagogique peuvent être, par exemple, des transparents, des notes de cours, des pages Web, des logiciels de simulation, des programmes d'enseignement, des objectifs pédagogiques qui peuvent être utilisés, réutilisés ou référencés dans tout l'activité liée à l'enseignement ou à l'apprentissage. Pour obtenir une description plus précise sur les documents, sur les auteurs, sur leurs champs d'intérêt, leurs idées, leurs objectifs pédagogiques, on peut inclure ces métadonnées comme ressources elles-mêmes qui sont structurées suivant des catégories ou champs sémantiques. Pour structurer logiquement le contenu des documents de manière à pouvoir séparer le fond et la forme, XML (métalangages de balisage spécialisés) apparaît comme le moyen le plus adéquat pour décrire à la fois les métadonnées et la structure des documents afin d'assurer la pérennité de la documentation électronique.

## Les métadonnées et le web sémantique

### Introduction

Je me propose de décrire brièvement les métadonnées liées à l'idée comment on peut ajouter une information de nature sémantique aux documents afin d'obtenir une description plus précise de ceux-ci. Essentiellement, les métadonnées sont des données sur une donnée, comme un ensemble structuré d'informations décrivant une ressources quelconque et elles peuvent être utilisées à des fins diverses :

- la description et la recherche de ressources ;

- la gestion de collections de ressources ;
- la préservation des ressources.

Les métadonnées sont en général constituées de mot-clès ou de texte libre, elles sont importantes pour les ressources visuelles et servent aussi, par les informations ajoutées à une ressource, à évaluer la ressource, à porter un jugement sur celle-ci, et à la comparer à d'autres ressources.

Les métadonnées sont utilisées dans les systèmes de gestion de contenu (CMS) pour éditer, gérer, rechercher, réutiliser, diffuser (imprimé ou en ligne sur le Web), publier de multiples contenus (textes, images, vidéos, etc.)

Les métadonnées d'ordre technique et administratif (comme l'appartenance à une collection, les informations de copyright, les informations sur l'acquisition, le format de fichier, la résolution, etc.) permettent de gérer, maintenir et préserver des collections digitales.

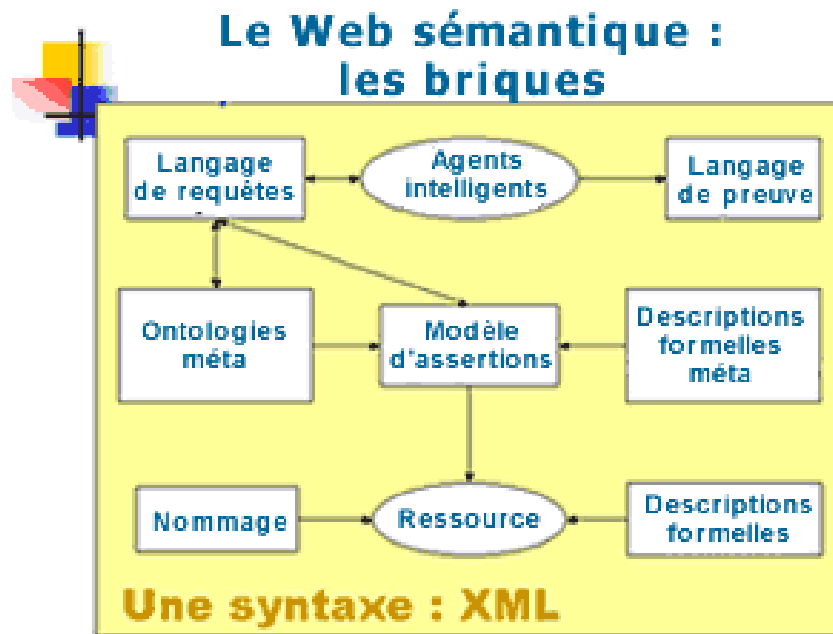
Dans le domaine de l'éducation, les informations relatives aux documents pédagogiques peuvent être structurées suivant des catégories ou champs sémantiques. Chaque champ représente une caractéristique particulière de la ressource, par exemple, son titre ou son résumé.

On trouve aussi implicites ou explicites des métadonnées utilisées à l'égard des objets informatiques, par exemple : les champs <title> et <meta> des pages HTML, dans les documents MS Office, PDF, dans les fichiers des images ou des sons, dans les plates-formes (soit Macintosh ou Windows), dans l'estampillage électronique (qui permet d'authentifier un document et de prouver l'appartenance d'une œuvre à son propriétaire au moyen de tatouages – insertion d'informations numériques dans les fichiers binaires que sont les images, sons, vidéo).

## ***Normes et standard de métadonnées***

La capacité des métadonnées à faciliter l'accès aux ressources en ligne dépend grandement de l'existence d'un standard, doté de propriétés. Les ressources sont partagées par différentes institutions et collectivités. Les métadonnées attribuées sauvagement aux ressources, sans règles établies et sans principes directeurs, ne seront pas interopérables entre différentes collectivités. Il est donc absolument nécessaire d'adopter des standards de description à l'aide des métadonnées. Les normes et les standards, concernant en principal les ressources électroniques sur le Web, sont très motivés par la décentralisation de l'archivage et de la recherche de ressources électroniques qui induisent un besoin de cohérence et d'uniformité.

Le web sémantique, avec une standardisation minimale, permet la création d'une structure sémantiquement riche, de type base de données, orientée ressources, tout en conservant les métadonnées dans les documents. Une série de standards est en cours d'élaboration où l'essentiel est à définir les ontologies (qui sont basés sur les mêmes principes que les systèmes experts) et comment les relier entre-elles. Nicolas Chung et Marie-Noelle Cormenier (<http://www.educnet.education.fr/tech/normes/0402.htm>) on fait un schéma intéressant d'une structure du Web sémantique :



Dans le domaine de la documentation et du e-learning, le monde de l'Internet est dominé par des acteurs des ressources électroniques qui travaillent sur des descriptions formelles de métadonnées :

Le **Dublin Core** (DC), créé en 1995, rassemble des bibliothécaires et des documentalistes au niveau international, il définit des métadonnées génériques et développe des outils pour implémenter des métadonnées dans les ressources. Plus récemment, des groupes de travail sectoriels ont été mis en place, en particulier le Dublin Core Education pour ce domaine

Le **Dublin Core Education** est composé de métadonnées génériques et reprend des métadonnées de la partie « Éducation » du LOM : type d'utilisateur, niveau et type d'interactivité, durée, type de médiateur, conformité à un standard éducatif, réflexion sur type de ressources, qualité, niveau, type de pédagogie

**L'ARIADNE** (Alliance of Remote Instructional Authoring and Distribution Networks for Europe), créé en 1996, rassemble des acteurs du secteur public et universitaire européen et suisse. Cette alliance est axée sur la FOAD (Formation Ouverte A Distance). Elle part des métadonnées du Dublin Core en les étendant.

**L'IEEE** (Institute of Electrical and Electronics Engineers), créé en 1998, basé aux Etats-Unis, rassemble des entreprises et des universités. Le sous-comité "technologies éducatives" de l'IEEE travaille sur la modélisation de tout objet d'apprentissage, il développe des métadonnées pour décrire les objets d'apprentissage y compris non numériques, il standardise aussi des vocabulaires (valeurs de métadonnées), les travaux de l'IEEE sont quasiment finalisés. Cet institut est à l'origine du LOM (Learning Object Model). Le LOM reprend toutes les métadonnées du Dublin Core en les détaillant davantage : structure de l'objet (structure, niveau d'agrégation), gestion de versions (version, état), gestion plus fine des contributeurs (rôle), description fine des métadonnées utilisées, description de l'implémentation technique au-delà du format : taille, localisation, prérequis techniques, gestion plus fine des droits (coûts...) et des relations entre objets, gestion de commentaires (annotations), gestion plus fine du sujet (par mots-clés ou classification). Pour la partie éducation, le LOM décrit les caractéristiques pédagogiques essentielles de l'objet concerné : type et niveau d'interactivité, type de ressource : exercice, QCM, examen, diagramme, figure, etc. « densité sémantique », type d'utilisateur : prof, élève, & contexte (primaire, secondaire, supérieur, formation continue &), classe d'âge typique & difficulté, durée, description, langue.

Le document LOM va probablement être repris par l'ISO. Si l'on souhaite faire de la veille sur les métadonnées et le web sémantique, l'IEEE est l'institut dont il faut suivre les travaux.

Tous ces acteurs implémentent leurs métas en HTML et en XML/RDF. Plusieurs standards ont été proposés pour formaliser cette sémantique, par exemple sous forme de métadonnées: c'est le cas du langage RDF (Ressource Description Framework), un idiome XML développé par le W3C, comme moyen d'encoder, échanger et réutiliser des métadonnées structurées et du formalisme des Topic Maps, standard ISO développé au départ dans le monde de l'édition pour constituer un équivalent électronique des index traditionnels qui utilise les réseaux sémantiques. Les métadonnées permettent de décrire les ressources, ainsi que les relations entre ces ressources. Une comparaison HTML-RDF-Topic Maps découvre brièvement:

- HTML relie des données de pages Web entre elles ;
- RDF relie des ressources quelconques entre elles, qu'elles soient des données, des concepts ou des objets ou des concepts Web ;
- Topic Maps structurent et organisent des connaissances, associent des sujets et des occurrences d'objets ou des concepts.

Une autre initiative d'un schéma de métadonnées est le Dublin Core destiné à faciliter la recherche des ressources peu complexes où les travaux sont en cours dans le domaine éducatif. Le Dublin Core, développé par une commission internationale et interdisciplinaire, est en ensemble de 15 éléments de métadonnées qui peuvent être encodés dans les balises HTML <meta> :

- au Contenu : Title, Description, Subject, Source, Coverage, Type, Relation
- à la Propriété intellectuelle : Creator (comme auteur), Contributeur, Publisher, Rights
- à la Version : Date, Format, Identifiant, Language

Chaque élément est optionnel, il peut être répété et associé à un nombre limité de qualificatifs afin d'affiner la description sémantique. Des terminologies précises sont utilisées pour permettre l'interopérabilité.

## ***La puissance des métadonnées et de XML***

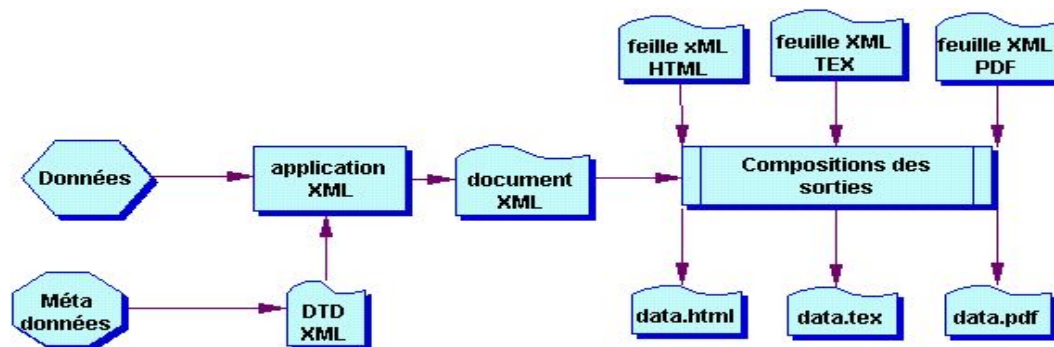
Pour diminuer le temps de réponse pour la mise en place de formations continues à la demande, pour développer des enseignements à distance réactualisés sur le Web et pour avoir une meilleure productivité dans l'élaboration des objets pédagogiques il est nécessaire de disposer d'un ensemble de standards généraux afin d'accomplir certains objectifs parmi lesquels on pourrait noter:

- faciliter le travail de groupe, la réutilisation et les mises à jour ;
- éditer, sous de multiples formats, (document imprimé, Web, présentation assistée par l'ordinateur) ;
- faire des recherches avec des critères sémantiques ;
- constituer des bases de données relatives aux « objets pédagogiques ».

Nous avons besoin des outils pour atteindre ces objectifs comme des outils « auteurs », des moteurs de recherches efficaces, des convertisseurs pour exporter les documents en HTML ou des applications XML, pour importer des documents depuis d'autres formats, des agents intelligents qui soit capable d'assembler automatiquement et dynamiquement des cours sur mesure, etc.

Pour structurer d'une manière logique le contenu des documents, XML devient de plus en plus important car HTML, le langage du Web, a trop de limitations pour les documents complexes.

La DTD (Définition de Type de Document) est la grammaire décrivant la structure logique d'une classe particulière de documents (par exemple, on peut considérer HTML comme un DTD de SGML). Elle permet la différenciation entre le contenu du document et sa mise en page. En XML, on peut toujours ajouter des nouveaux éléments à la grammaire, on peut faire référence à plusieurs DTD au sein du même document et on peut analyser des documents sans leur imposer une structure. La syntaxe XML est simple, elle contient des balises dont le nom est une description symbolique du contenu et la DTD associée permet de définir la structure de ces éléments et donne le caractère valide aux documents XML. Une autre caractéristique de XML est XSLT, une feuille de style associé au document XML qui permet des mises en pages différentes, par exemple, on peut associé une feuille de style pour produire un fichier HTML et u autre pour produire un fichier PDF. L'unicité du document source est une caractéristique fondamentale améliorant le processus de mise à jour en évitant d'avoir des versions multiples et désynchronisées des mêmes informations. Yolande Bourda et Marc Hélier, proposent dans leur article « Métadonné et XML : applications aux Objets Pédagogiques » (<http://www.mutu-xml.org/xml-base/shared/SHAREDDBASE-METADATA-FR.html>) un schéma de mises en page des données à partir d'un document source XML unique :



Vu que XML permet de différencier strictement le contenu de la mise en page et qu'il décrit la structure logique des documents, il est, pour l'instant, une solution satisfaisante pour simplifier le maintien en bon état et les mises à jour des documents électroniques.

## Conclusion

Dans le domaine des métadonnées pédagogiques plusieurs applications sont en train d'être développées et XML soutient la manière de produire, de gérer et d'utiliser les objets pédagogiques. Le standard qui se dégage va permettre dans le domaine de l'éducation de répondre à tous les besoins des enseignants comme producteurs d'objets pédagogiques, de décrire de façon unique un exercice, de produire différents types de sorties (imprimées, en ligne sur le Web...) avec des mises en pages différenciées et des contenus différents.

## Ressources

Bourda Yolaine, Hélier Marc, *Métadonnées et XML.Applications aux "Objets pédagogiques"*, Supélec, Plateau de Moulon, F-91192 Gif-sur-Yvette cedex France (<http://www.mutu-xml.org/xml-base/shared/SHAREDDBASE-METADATA-FR.html>)

Chung Nicolas, Cormenier Marie-Noelle, *Normes et standard de métadonnées* (avril 2002), (<http://www.educnet.education.fr/tech/normes/0402.htm>)

Peccatte Patrik, *Métadonnées : une initiation. Dublin Core, IPTC, EXIF, RDF, XMP, etc.*, (mars 2003), Soft Experience, (<http://peccatte.karefil.com/softqare/Metadata.htm>)

Ruggles, Rudy L., (ed.) *Knowledge management tools* (Butterworth-Heinemann, Boston 1997)