

Partager les modèles XML : quel intérêt ?

Abderrazak Mkadmi

Université Paris VIII
mkadmi@hymedia.univ-paris8.fr

Nasreddine Bouhai

Université Paris VIII
nasreddine.bouhai@univ-paris8.fr

Marc Langlois

Edifrance
langloismarc@compuserve.com

Apparu en réponse au besoin d'interaction et de coopération entre des systèmes d'information hétérogènes utilisant jusqu'alors des structures de données largement incompatibles entre elles, XML (Extensible Markup Language, langage de balisage extensible) a pour objectif de devenir le format privilégié de l'échange de documents – et plus généralement d'informations – en milieu ouvert.

En effet, les travaux de normalisation menés par le World Wide Web Consortium (W3C), les développements des éditeurs de logiciels et les préconisations de différents groupes et consortiums (ebXML, OASIS) conjuguent leurs efforts pour définir, promouvoir et utiliser XML dans différentes situations.

Un langage d'avenir pour les échanges électroniques

Ce métalangage est utilisé aujourd'hui par tous : les fournisseurs d'ERP (Enterprise Resource Planning ou

logiciels de gestion intégrée), les éditeurs de *middleware* (logiciel des couches intermédiaires), les fournisseurs de bases de données, etc. Les raisons de ce consensus sont à chercher du côté de la simplicité et de la richesse d'expression d'XML.

Cependant, pour que les objectifs d'XML, en l'occurrence permettre l'échange généralisé intersectoriel quel que soit le type d'acteur, soient réellement atteints, il paraît très intéressant que les modèles de ces documents structurés échangés, ainsi que toutes les informations associées, soient partagés. Ces informations associées servent à mieux comprendre

Titulaire d'un DEA Sciences de l'information et de la communication et d'un master en ingénierie des systèmes informatiques ouverts, **Abderrazak Mkadmi** est doctorant chercheur à l'université de Paris VIII.

Nasreddine Bouhai est maître de conférence à l'université de Paris VIII (département Hypermédia).

Titulaire d'un master en organisation, **Marc Langlois** est directeur de projet pour le GIP Modernisation des déclarations sociales. Il a publié *L'euro et les systèmes d'information* (Dunod, 1998), *Le commerce électronique B to B* (Dunod, 2^e éd. 2001) et a collaboré à plusieurs ouvrages de gestion de projets informatiques.

les modèles pour pouvoir les exploiter de façon rapide et efficace. Il devient alors possible de recevoir n'importe quel document issu d'un modèle particulier afin d'être en mesure de l'exploiter avec des logiciels génériques. On pourra, par exemple, éditer un document avec des environnements standard et disponibles, mais aussi le visualiser sur un système de consultation standard du web.

Associant aux données une structure sémantique (sous forme d'éléments et attributs) et permettant de séparer cette structure du contenu, ainsi que la présentation de ce contenu, XML a été retenu comme le langage d'avenir pour la génération des échanges électroniques, que ce soit entre les grandes entreprises, entre les PME ou entre les grandes entreprises et les PME. Comparé à l'EDI (Electronic Data Interchange) conventionnel, XML offre la capacité de transmettre des données multimédias (image et vidéo pour illustrer un catalogue en ligne par exemple) ; d'afficher les données sous une forme humaine par l'utilisation des feuilles de style ; de convertir facilement une structure de message en une autre structure de message (ce qui facilite l'intégration des données dans des applications existantes).

Cependant, pour permettre un traitement automatique de documents XML provenant d'autres partenaires, il est nécessaire que les différents acteurs impliqués se mettent d'accord sur un formalisme de structuration

des informations. Cette structuration est définie selon un modèle sous formats de schémas ou de DTD (Document Type Definition) qui donne les règles d'assemblage et d'ordonnement des données.

De ce fait, il devient alors intéressant de pouvoir partager ces modèles. C'est dans ce contexte que notre projet de création d'un répertoire de modèles XML a vu le jour pour pouvoir identifier, partager et réutiliser les modèles de documents pour les différentes applications XML.

Partage des modèles XML : état de l'art

Qu'est ce qu'un répertoire de modèles XML ?

Un répertoire de modèles XML est une base de données permettant à tout utilisateur de prendre connaissance des modèles existants dans un domaine d'activité particulier pour un besoin particulier, ainsi que des modèles permettant d'échanger avec ses partenaires. Ce répertoire favorise l'échange ouvert entre professionnels qui, avec cet outil, seront capables de se définir par rapport à l'existant et d'avoir une cohérence des méthodes de travail dans leurs domaines d'activités.

Un répertoire de modèles XML doit alors proposer des accès à un ensemble de modèles bien documentés donnant une idée des initiatives normatives prises dans un domaine d'activité bien déterminé. Il doit représenter aussi un espace de travail collaboratif dans le cadre des échanges électroniques professionnels.

Exemples de modèles XML partagés

DocBook

Il existe dans le domaine industriel un modèle dénommé DocBook. Ce modèle de documentation technique, actuellement maintenu par un co-

mité technique du consortium OASIS, DocBook Technical Committee, permet de créer et d'échanger des documents techniques ou des articles grâce à une sémantique plutôt généraliste, qui la rend utilisable dans toute l'industrie de l'électronique et de l'informatique. Il est disponible et extrêmement bien documenté sur Internet, avec des feuilles de style, utilisant XSLT et XSL, destinées à faciliter la diffusion de ce type d'information sur le web, sur WAP (Wireless Application Protocol), et sur papier.

Mis au point pendant plusieurs années par de nombreux utilisateurs, DocBook a rapidement fait figure de modèle parmi les DTD à vocation technique. Beaucoup d'industriels s'en sont emparés et l'ont adapté à leurs propres besoins. Cependant, ce modèle, étant conçu au départ pour gérer du SGML, est jugé par certains trop difficile à paramétrer et à utiliser. Les concepteurs travaillent aujourd'hui sur une vision simplifiée dénommée « Simplified DocBook ».

Répertoire de l'Atica

Dans le domaine de l'administration, il existe un projet similaire mené par l'Atica¹, visant à publier dans un répertoire tous les schémas et DTD issus des domaines documentaire et juridique présentant un intérêt général. Ce répertoire constitue un outil mutualisé pour favoriser les échanges au sein des administrations et avec leurs partenaires.

Ce projet a été déclenché par une circulaire du 21 janvier 2002 relative à la mise en œuvre d'un cadre commun d'interopérabilité pour les échanges et la compatibilité des systèmes d'information des administrations.

« Enfin, il sera bon que chaque nouveau projet de système comportant des échanges d'informations (au sein de l'administration ou avec les tiers) soit l'occasion de

1. Atica : Agence pour les technologies de l'information et de la communication dans l'administration.

poursuivre, et même d'intensifier, l'élaboration de schémas XML, dont on connaît l'importance pour faciliter les échanges. Ils seront conçus de manière à faire clairement apparaître leur définition, ainsi que celle des éléments qui les composent, par application de la méthode dite des "espaces nominatifs", conforme aux standards de l'Internet. Ils seront publiés, d'abord à l'état de projet, puis sous leur forme définitive, dans le répertoire des schémas XML de l'administration.»

Cependant, malgré l'intérêt que présente ce projet quant à la dématérialisation et l'interopérabilité des échanges au sein de l'administration et entre l'administration et ses partenaires, il n'est pas précisé jusqu'à aujourd'hui comment les modèles seront présentés au sein du répertoire, ni quelles documentations doivent les accompagner. Ces lacunes n'ont pas encouragé les administrations à se lancer dans ce projet pour enrichir le répertoire par leurs modèles.

Ces deux exemples de projets (qui ne sont pas exclusifs) montrent bien l'intérêt de la création d'un répertoire de schémas XML. En effet, le fait de partager des modèles participe à l'évidence à l'acceptation de ceux-ci et représente, de plus, un facteur de montée en compétences des organisations confrontées à l'explosion des applications XML. Plus les modèles sont accessibles de manière facile, plus l'échange entre partenaires devient aisé.

Intérêt des modèles de documents

Un modèle de document est une structure permettant de donner les règles d'assemblage des données. Cette structure permet à un système d'information de comprendre que `<Code_Postal>96100</Code_Postal>` représente bien un code postal d'une ville et non pas une quelconque suite de chiffres. Ce système peut aussi comprendre, d'un point de vue struc-

turel, que ce code fait partie d'une adresse d'un client. À défaut d'un tel modèle, ce numéro peut être codé de différentes manières dans des différents documents, exemple : `<code-postal>`, `<CodePostal>`, et par conséquent, il peut être interprété comme une suite de chiffres.

Un modèle de documents sert donc, dans un premier temps, à définir tous les éléments utilisés dans un document, et, secondement, à définir les relations et l'ordonnement entre ces éléments.

Pour créer des modèles de documents, deux recommandations existent aujourd'hui, concernant les DTD et les schémas XML. La première sur les DTD est historique et est désormais reconnue dans tous les domaines applicatifs. La deuxième est récente et vient pallier les déficiences de la première se rapportant notamment au typage de données, au langage utilisé et au support des espaces de noms. En effet, XML Schema est un nouveau langage proposé par le W3C qui offre, en plus des fonctionnalités fournies par les DTD, plusieurs nouveautés :

- un grand nombre de types de données intégrées comme les booléens, les entiers, les intervalles de temps, etc. De plus, il est possible de créer de nouveaux types par ajout de contraintes sur un type existant ;
- des types de données utilisateurs qui nous permettent de créer notre propre type de données nommé ;
- la notion d'héritage : les éléments peuvent hériter du contenu et des attributs d'un autre élément. C'est sans aucun doute l'innovation la plus intéressante de XML Schema ;
- le support des espaces de nom ;
- les indicateurs d'occurrences des éléments peuvent être tout nombre non négatif ;
- une grande facilité de conception modulaire de schémas.

Les modèles de documents servent donc à définir la cohérence d'un ensemble de documents, lesquels peuvent être utilisés par n'importe quelle application informatique en

ne se définissant que par rapport au modèle sous-tendu. Ceci permet évidemment de gagner beaucoup de temps, d'argent et de fiabilité dans les travaux coopératifs.

Présentation du répertoire de modèles XML

Principes de base

Le répertoire de modèles est accessible librement à tout le monde et il n'y a aucune limite dans la consultation des schémas et des DTD qui y sont stockés. Quant à la sécurité, il n'y a pour le moment aucune notion de confidentialité au niveau accès des données. La recherche des modèles, comme nous allons la présenter par la suite, se fait à travers plusieurs critères définissant le contexte et l'appartenance de chaque modèle. Quant à la modification et la mise à jour, elles se font par le propriétaire des modèles, seul responsable des structures de ses documents.

Cependant, la publication de ces modèles passe par un « comité éditorial » qui vérifie la forme de ce qui est proposé à publication, c'est-à-dire la cohérence des données contextuelles, la pertinence de la définition par rapport au contexte, la bonne syntaxe des schémas et des DTD, ainsi que la cohérence des documents d'exemple au regard des modèles.

Pour cela, trois acteurs au moins se distinguent : l'utilisateur, le participant au groupe de travail et l'administrateur. L'utilisateur devra avoir accès aux modèles par le biais d'une interface web à partir de n'importe quel navigateur. Un système de session personnel est mis en place. Cet utilisateur pourra être n'importe qui, mais il pourra être aussi un soumissionnaire d'un modèle dans la base. Le participant au groupe de travail est une personne qui se dote d'un mot de passe pour accéder à un espace de travail bien particulier. L'administrateur devra pouvoir gérer cette base de données (ajouter, modifier ou sup-

Figure 1. Structure de données simple et générique

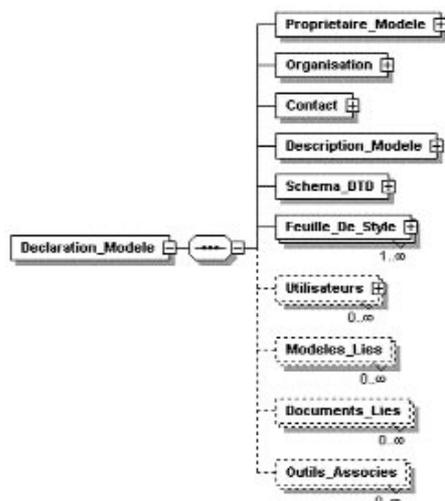
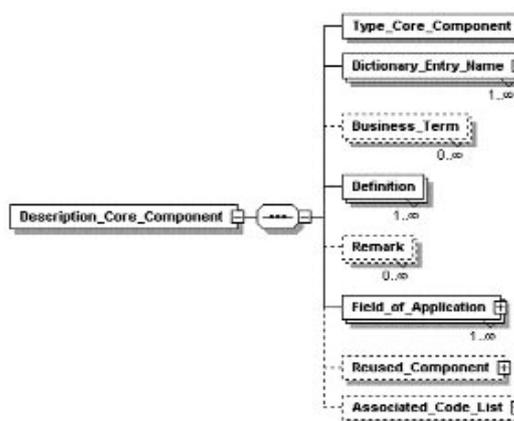


Figure 2. Structure de données liée aux « core components ebXML »



primer un champ), mais il devra aussi pouvoir modifier les différents modules du serveur.

Parmi les contraintes définies, on peut noter que :

- aucun utilisateur et/ou propriétaire du modèle ne peut modifier un modèle qui ne lui appartient pas ;
- les propriétaires des modèles doivent être munis de mots de passe pour accéder à leurs espaces de travail ;
- les modèles déclarés doivent subir une opération de révision et de contrôle de la part d'un comité spécial pour assurer la cohérence des données ;
- l'application doit être ouverte et évolutive : facile à mettre à jour et à enrichir par d'autres fonctionnalités répondant à des nouveaux besoins ;
- ni le logiciel serveur, ni le logiciel client ne doivent être d'une technologie propriétaire.

Structure des données

Avant de concevoir la structure de l'application, nous avons essayé de définir la structure des données qui permet de préciser tous les éléments, les entités, les attributs, les relations entre eux, ainsi que les différentes caractéristiques de ceux-ci. Cependant, vu la

nature de notre projet qui se veut à la fois générique (touchant le maximum possible de domaines) et spécifique à l'échange électronique des données d'affaires (commerce électronique), nous avons opté pour deux structures de données. La première répond au premier objectif : c'est une structure générale permettant à n'importe qui de l'utiliser pour déclarer son modèle de données. La deuxième structure est spécifique à la déclaration des « Core Components² ebXML³ » (composants élémentaires ebXML). La définition de ces deux structures s'est faite en utilisant les schémas XML.

La première structure se présente brièvement selon le schéma de la figure 1.

2. Traduit en français « composants élémentaires », les *core components* sont des pièces d'assemblage, ayant chacune une définition unique en sémantique d'affaire. Pour plus d'informations, veuillez voir : <http://www.autoroute.gouv.qc.ca/publica/normes/norme111.htm>

3. EbXML (e-business XML) est le nouveau standard des échanges électroniques professionnels basé sur l'utilisation des technologies de l'Internet et plus particulièrement du langage XML. Il s'agit d'une initiative des Nations unies et d'OASIS pour créer une infrastructure basée sur XML et l'Internet ouvrant le commerce électronique à toutes les entreprises, petites ou grandes, sur tous les continents. Une première plate-forme a été implémentée en mai 2001.

Quant à la deuxième structure (figure 2), elle contient toutes les données de la première structure, mais elle en englobe d'autres, liées à la définition d'un *core component*. Ces données sont extraites de la spécification ebXML.

Structure de l'application

L'application est structurée en six parties :

- une partie contenant les éléments de définition du modèle ;
- une partie définissant toutes les informations associées au modèle ;
- un module définissant tous les critères de recherche ;
- un module pour gérer le contrôle et la révision des modèles proposés à publication ;
- un module pour la consultation des données du répertoire ;
- un module pour la soumission des modèles et documents XML.

Cette structure a été définie après une étude des besoins des principaux partenaires du projet, à savoir Edifrance (association pour le développement des échanges électroniques professionnels), Mutu-XML et GFII (Groupement français de l'industrie de l'information), Fing (Fondation Internet nouvelle génération),

ainsi que l’UIC (Union internationale de chemins de fer) à Bruxelles.

Cette structure peut être schématisée selon la figure 3.

Le module de recherche peut utiliser plusieurs critères : secteur d’activité, domaine d’usage, rôle dans le processus d’affaire, classification du produit, processus d’affaire, nom du modèle et mots clés. Ces critères peuvent être utilisés conjointement avec les opérateurs logiques « et » et « ou ». Le module de recherche est accessible par consultation à distance. Les trois premiers critères constituent une liste fermée maintenue par le comité éditorial. Les autres sont ouverts.

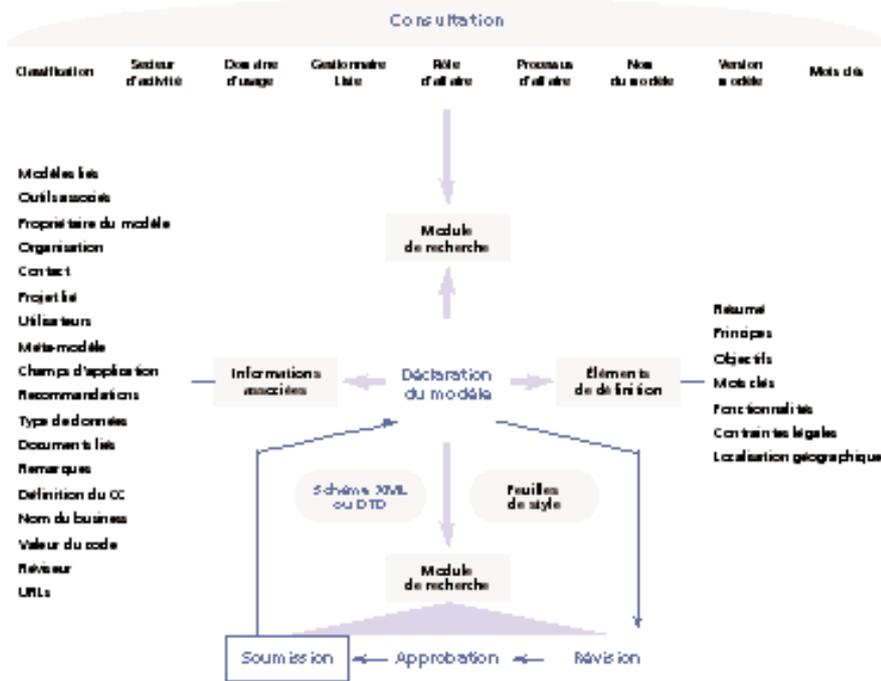
Le module de consultation donne accès, à partir d’une recherche, aux données du répertoire. Il permet soit une consultation sur navigateur, soit un téléchargement du résultat de recherche.

Quant au module de révision, il représente une antichambre pour assurer la meilleure diffusion des modèles XML et, donc, pour favoriser la meilleure compréhension possible de l’objectif et des conditions d’utilisation d’un schéma particulier.

Le module de soumission d’une nouvelle entrée utilise le même modèle que celui qui sert à la restitution, à partir du module de recherche. La soumission est stockée en attente d’approbation. L’approbation permet de mettre à jour le répertoire.

La partie se rapportant aux éléments de définition du modèle permet de définir le modèle à travers le résumé, les objectifs, les principes de base, les différentes fonctionnalités, les mots clés et la localisation géographique. Elle permet de définir aussi

Figure 3. Structure de la base de données du répertoire de modèles XML



les caractéristiques des situations d’affaires influencées par des exigences légales ou réglementaires (lois, règlements, conventions, traités...).

L’autre partie permet, quant à elle, de présenter toutes les informations associées au modèle, à savoir son propriétaire, l’organisation responsable, les modèles et documents liés, les outils utilisés, les principaux utilisateurs, etc.

Modes d’utilisation

L’utilisation de l’application « répertoire de modèles XML » diffère selon qu’on est soumissionnaire de modèle, utilisateur ou réviseur.

Un soumissionnaire doit passer par une identification (nom, prénom, login et mot de passe, etc.) s’il est nouveau. Sinon, il n’entre que son login et son mot de passe pour accéder à une interface qui lui permet de choisir le type de modèle (modèle simple ou modèle des core compo-

nents) qu’il souhaite enregistrer, ainsi que le mode de soumission (manuelle ou par exportation⁴). Il reçoit ainsi un formulaire qui lui permet de remplir toutes les données y compris le schéma XML. Ce système d’identification nous permet d’une part d’éviter les redondances des entrées au niveau des propriétaires de modèles, et d’autre part de contrôler et d’assurer que la modification d’un modèle ne se fait que par son propriétaire.

Un utilisateur du grand public, qui veut consulter un modèle ou tous les modèles existants dans la base, lance sa requête à partir de l’interface de recherche et reçoit une liste de mo-

4. Lors de la soumission, on peut, au lieu de remplir toutes les données manuellement, les exporter à partir d’un fichier Excel, et ceci représente un besoin déclaré par les partenaires à l’état actuel qui travaillent beaucoup avec des tableurs Excel pour stocker leurs core components. Lors de la consultation et téléchargement, on peut faire l’opération inverse (importer les données de la base dans un fichier Excel).

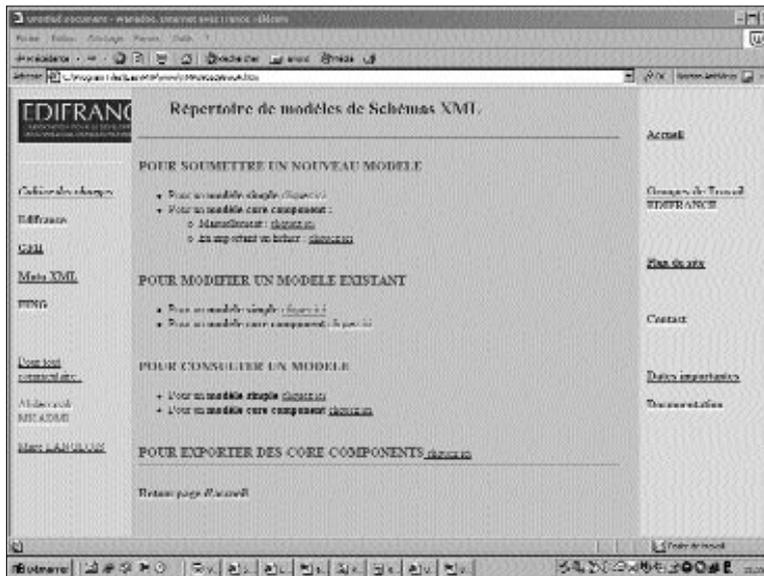


Figure 4. Page sommaire du répertoire de modèles XML

dèles, avec seulement le nom du modèle et un lien vers son contenu global.

Quant au réviseur, il passe lui aussi par un mot de passe pour accéder aux modèles soumis, et, après vérification de leurs contenus, il note ses commentaires et remarques sur chaque modèle pour son propriétaire si le modèle manque d'informations ou s'il ne répond pas aux exigences prédéfinies. Sinon, il valide le modèle pour qu'il devienne accessible au grand public.

La première page de l'application nous permet donc de consulter la liste des modèles, de chercher un modèle par plusieurs critères ou de soumettre et/ou modifier un modèle en passant par une identification. Cette dernière utilisation (soumission et/ou modification) donne à l'utilisateur accès à une autre page sommaire qui lui propose quatre fonctionnalités (figure 4).

Comme le montre cette figure qui représente une des pages d'accès au répertoire de modèles XML, les principales fonctionnalités sont : la soumission, la consultation, la modification et l'exportation des modèles.

Conclusion

Nous avons présenté dans cet article le répertoire de modèles de schémas XML. Le contexte de création et les structures de données et de l'application ont également été exposés, ainsi que quelques intérêts de cet outil. Cependant, d'autres avantages peuvent être évoqués à travers cette application, comme le travail collaboratif ou la recherche sémantique, qui s'intègrent dans une problématique très large dépassant le cadre de cet article. Toutefois, nous pouvons dire que le répertoire de modèles XML constitue un pas exceptionnel dans le domaine de l'EDI. Un tel projet était, depuis un bon moment (en tout cas, depuis le développement et l'introduction du métalangage XML dans le domaine du commerce électronique et dans l'administration française), le souci de tous les organismes qui travaillent dans les domaines de la normalisation et de la standardisation. En effet, les partenaires du projet se trouvent déjà avec un outil qui leur permet de publier leurs modèles XML et leurs *core components*, et de profiter de l'exis-

tence de différents modèles liés à différentes activités. Ceci évitera certainement des travaux redondants et encouragera d'autres personnes à utiliser la technologie XML, tout en s'inspirant des modèles déjà publiés. Le répertoire offre, par ailleurs, un service d'aide à l'élaboration des schémas XML.

Mars 2003

BIBLIOGRAPHIE

ATTAR, Pierre, *Documentation technique du modèle permettant l'échange de descriptions de modèles, version 0.1*, 1^{er} janvier 2002.

CHAUVET, Jean-Marie, *Services web avec SOAP, WSDL, UDDI, ebXML*, Eyrolles, 2002.

« Circulaire du 21 janvier 2002 relative à la mise en œuvre d'un cadre commun d'interopérabilité pour les échanges et la compatibilité des systèmes d'information des administrations », *Journal officiel*, n° 30 du 5 février 2002, p. 2335, NOR : PRMX0205357C.

<http://www.legifrance.gouv.fr/WAspad/UnTexteDeJor?numjo=PRMX0205357C>

Le commerce électronique pour l'entreprise : guide de mise en œuvre des échanges électroniques professionnels, Edifrance, 2002.

Comprendre XML pour les échanges électroniques professionnels, cédérom, Edifrance, 2002.

Core Components Technical Specifications, Part 1 = Spécification technique des composants élémentaires, partie 1, UN/CEFACT, 2001.

<http://www.autoroute.gouv.qc.ca/publica/normes/norme111.htm>

LANGLOIS, Marc ; MKADMI, Abderrazak, *Création d'un répertoire de modèles XML : cahier des charges*, Edifrance, 2002.

http://www.edifrance.org/accueil_schemaxml.php

OASIS Registry/Repository Technical Specification, Working Draft 1.1, OASIS, 20, 2000.

PARENT, Richard, *Normes ouvertes en technologies de l'information : spécification technique des composants élémentaires, partie 1*.

<http://www.autoroute.gouv.qc.ca/publica/normes/liste.htm>

VAN DER VLIST, Eric, *Le répertoire de schémas XML de la MTIC ouvre ses portes*.

<http://xmlfr.org/actualites/decid/010413-0002>

WALSH, Norman ; MUELLNER, Leonard, *DocBook, la référence*, O'Reilly, 2001.