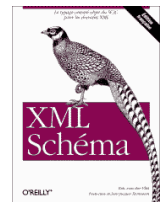


Vers la “babelisation” de XML?

Eric van der Vlist
The XML Day 2003
Dyomedea / XMLfr

Eric van der Vlist

- Consultant, formateur, conférencier, auteur indépendant
- Expert XML de niveau international
- Spécialiste des langages de schémas, transformations XSLT et web sémantique
- Membre de l'ISO (ISO/IES JTC1/SC34 WG1 DSDL)
- Membre d'OASIS et de l'ISUG
- Créateur et animateur de XMLfr.org
- Gérant, fondateur, consultant Dyomedea SARL
- Auteur des livres « XML Schema » et « Relax NG » aux éditions O'Reilly.



Vers la “babelisation” de XML?

- Les moteurs de cette “babelisation” sont bien connus :
 - Editeurs de logiciels cherchant à bloquer leurs clients.
 - Editeurs de logiciels vendant des outils de conversion.
 - Intégrateurs développant et maintenant des chaînes de conversion.
 - Fournisseurs de matériel intéressés par la puissance de calcul nécessaire à ces conversions.
 - Partage des développements que représente un vocabulaire XML pas toujours perçu comme naturel.
- Voir l'excellente présentation d'Hervé Crespel à ce sujet (<http://xmlfr.org/documentations/articles/011011-0001>)

Nous sommes tous responsables

- En dehors de ces moteurs, nous avons tous une part de responsabilité dans cette babelisation.
- Quel informaticien n'a jamais
 - préféré réinventer "son" vocabulaire plutôt que de rechercher ce qui existait et faire l'effort de comprendre ce qui avait été fait ?
 - créé un vocabulaire sans le documenter ?
 - documenté un vocabulaire sans le « publier » ?
 - publié un vocabulaire sans en faire la promotion ?
- Tout cela participe à la « babelisation » de XML.
 - Nous sommes tous responsables!

Vocabulaire XML

- A propos, qu'est ce qu'un vocabulaire XML?
- Je propose la définition suivante :
 - « Un vocabulaire XML est l'ensemble des règles et documentations nécessaires à ce que des documents XML soient compréhensibles par tous les programmes et développeurs participant à une application de XML .»
- Cette définition peut englober :
 - un ou des schémas XML,
 - et/ou un ou des espaces de noms.
- Espace de noms et schémas n'en sont que des composantes techniques.
 - Parler de répertoires de schémas ou d'espaces de noms donne une vision étriquée des vocabulaires XML.

Espace de noms

- Recommandation W3C (Namespaces in XML)
 - Identifiés de manière unique par une URI.
 - Peuvent cohabiter dans un même document.
- Vocabulaires XML et espaces de noms :
 - Un vocabulaire peut utiliser plusieurs espaces de noms (exemple d'OpenOffice).
 - Un espace de noms peut être utilisé par plusieurs vocabulaires.

Schémas XML

- Règne de la diversité.
 - Deux langages de schémas W3C (DTD et W3C XML Schema).
 - Plusieurs autres propositions (RELAX NG notamment).
- Pluralité des objectifs.
 - Modélisation mais également validation.
- Schémas XML et vocabulaires :
 - Un vocabulaire peut être décrit par plusieurs schémas.
 - Un schéma peut décrire la combinaison de plusieurs vocabulaires.

Combattre la babelisation?

- En combattant les quatre points déjà évoqués :
 - Utilisation de vocabulaires existants dès que possible.
 - Documentation de tout nouveau vocabulaire.
 - Publication de toute documentation.
 - Promotion de toute publication.
- Nous nous attacherons à approfondir les trois derniers points dans cette présentation.

Documentation

- Problème à peine effleuré par le W3C
- Plusieurs approches sont possibles, notamment :
 - l'approche « sémantique »
 - l'approche « modélisation »
 - l'approche « documentaliste »

La vision sémantique

- Attribution de définitions formelles aux éléments et attributs composant un vocabulaire XML.
 - En pratique, passe par l'attribution d'identificateurs uniques aux notions abstraites.
 - Relie les éléments et attributs d'un vocabulaires XML à des identificateurs représentant des notions de base (« core components » par exemple).
- Ajoute un niveau d'indirection.
 - On ne se met plus d'accord sur la définition d'éléments et d'attributs, mais sur la signification des notions de base reliées aux éléments et attributs.
- Repousse le problème fondamental de la définition de ces notions.
 - Core components universels.
 - Ou ontologies définissant des équivalences.

Les niveaux de la modélisation

- La modélisation est une des manières de relier des notions de base à des éléments et attributs.
 - Devrait donc permettre de réaliser la vision sémantique.
- La modélisation de vocabulaires XML peut se faire à trois niveaux
 - Niveau logique (UML)
 - Niveau physique (documents eux-mêmes, examplotron).
 - Niveau intermédiaire (schémas XML)
- Cela pose des problèmes d'interopérabilité.
 - Un modèle UML peut être utilisé pour générer des schémas XML.
 - L'inverse (reverse-engineering) est difficilement généralisable.

Modélisation UML

- Suffisamment souple pour modéliser des vocabulaires XML.
- Ne pas se tromper de cible :
 - Modélisation UML de données « exprimables » en XML.
 - Indépendante de XML (peut être réutilisée pour générer un schéma SQL, ASN.1, des classes Java ou C#, ...)
 - Les schéma XML éventuels sont des « sous produits » de la modélisation.
 - Le « style » de ces schémas sont fixés par l'outil de génération
 - Ou modélisation d'un schéma XML.
 - Utilisation de stéréotypes fortement dépendants du langage de schéma.
 - Permet de contrôler le schéma au niveau UML.
 - Importe la complexité du langage de schéma dans UML.

Modélisation par schéma XML

- Diversité des langages de schéma XML
 - DTD, W3C XML Schema, Relax NG, ... lequel choisir?
 - Des convertisseurs permettent de passer de l'un à l'autre.
- Modélisation se situant à un niveau technique intermédiaire entre « logique » (UML) et « physique » (documents).
 - Les langages de schéma XML introduisent leur propre complexité.
 - De plus en plus de projets génèrent des schémas.
- Egalement composants techniques ayant un rôle important à jour pour la validation des documents (projet ISO/DSDL - <http://dSDL.org>).

Intéropérabilité UML - schémas

- Les partisans de UML préconisent une modélisation des données indépendante de leur format physique.
 - Conventions et décisions d'implémentation généralement figées au niveau des générateurs de schémas.
 - Un schéma XML ne respectant pas ces conventions ne peut pas être « reverse-engineered » en modèle UML.
 - Il n'est pas intéropérable avec les modèles UML imposant ces conventions.
 - Il en est de même de deux modèles UML n'utilisant pas les mêmes règles de générations de schémas.
- C'est la raison pour laquelle certains partisans de UML font figure d'intégristes : pour que leur vision de l'intéropérabilité des systèmes se réalise, il faudrait que tout le monde respecte les mêmes règles.

L'enjeu n'est pas technique

- Un vocabulaire XML introduisant un balisage parfait est voué à l'échec s'il n'est pas accompagné d'une documentation claire.
- L'enjeu principal est donc pédagogique.
- La documentation qui constitue l'image du vocabulaire pour les développeurs.
- Elle doit être :
 - facile et agréable à lire
 - aussi complète que possible
 - Multi-points d'entrée
- Elle a un double rôle :
 - guide utilisateur pour apprendre le vocabulaire.
 - guide de référence.

La tactique documentaliste

- La documentation peut être générée par les approches sémantiques ou modélisation.
 - Un modèle UML constitue une documentation par lui-même.
 - Un schéma XML commenté peut également générer une documentation.
- Ces documentations sont en général insuffisantes pour utiliser les vocabulaires.
 - Un modèle UML décrit les données sans donner le détail de leur formalisation XML.
 - La documentation générée à partir d'un schéma XML documente en général le schéma plus que les documents.
- L'approche documentaliste consiste au contraire à faire de la documentation le principal livrable d'un vocabulaire XML.

Programmation littéraire

- Introduite en 1984 par Donald Knuth (TeX).
 - Place la documentation au centre du processus de développement.
 - Peut être utilisée quelque soit le format de la documentation ((X)HTML, DocBook, ...).
 - Pour tout langages de schéma ou de programmation.
- Fragments de code (ou de schéma) inclus dans la documentation.
 - Code source (ou schéma) généré à partir de la documentation.
- Bien adaptée à la documentation de vocabulaires XML.
 - Permet d'intégrer des ressources de nature variée (schémas, feuilles de style, scripts, ...) dans une même documentation.
 - (voir <http://xmlfr.org/actualites/tech/021218-0002>)

RDDL

- Pour concilier la vision sémantique et la tactique documentaires, on peut rendre la documentation lisible par des « agents sémantiques ».
 - RDDL (<http://rddl.org>) qui utilise XHTML et Xlink.
 - XHTML pour les utilisateurs humains.
 - Xlink pour les agents.
- Point de regroupement pour les différentes approches :
 - Compatible avec la programmation littéraire.
 - Rend la documentation accessible aux agents sémantiques.
 - Permet d'encapsuler des schémas XML et de faire référence à des modèles UML.

Démonstration

- Présentation de travaux réalisés pour l'INSEE.

Marketing des vocabulaires XML

- Sans un marketing adapté, un vocabulaire XML a peu de chances de s'imposer.
- Ressemble au marketing de sites Web :
 - Rôle important du « référencement » .
 - Sur les répertoires adaptés au domaine.
 - Sur les moteurs de recherche classiques.
 - Sur les listes de discussions.
 - Les méthodes traditionnelles doivent être utilisées en complément.

Un nouveau métier?

- La création, la documentation et la publication de vocabulaires XML fait appel à des compétences de nature différente :
 - Technique
 - Métier
 - Pédagogie
 - Rédaction
 - Présentation
 - Marketing
- Demande en forte augmentation.
- On peut donc penser que cette activité est en passe de devenir un « nouveau métier ».