

OpenOffice.org : l'avantage XML !

***OpenOffice.org et l'utilisation du XML
sur le poste de travail***

Metadonnées

Author : Éric Barroca, Directeur des Opérations, Nuxeo

Othercredit : Sophie Clapier

Othercredit : Stéphane Fermigier, PDG, Nuxeo

Othercredit : Laurent Godard, Expert OpenOffice.org, Indesko

Othercredit : Jean-Philippe Gonthier, Chargé de mission TIC Préfecture de Région Haute-Normandie

Othercredit : Robert Dargaud, Bobiciel Informatique

Web : <http://www.indesko.com>

Copyright © 2004-2006 Indesko

Date 2006/06/15

Version 2.0

Abstract

Analyse des impacts de l'utilisation du format XML dans OpenOffice.org : intérêts, possibilités, perspectives, etc.

Licence

Vous avez la permission de copier, de distribuer et/ou de modifier ce document selon les termes de la GNU Free Documentation License, Version 1.1 telle que publiée par la Free Software Foundation (<http://www.gnu.org/copyleft/fdl.html>) sans texte de couverture (*Front Cover Text*), avec comme texte de 4ème de couverture (*Back Cover Text*) « Vous êtes libres de copier et de modifier ce manuel, comme tout le logiciel et la documentation libres produits par la société Indesko, selon les termes de la GNU Free Documentation licence », et comme section invariante (*Invariant Section*) la section « Support commercial ».

Table des matières

1 Introduction.....	4
2 Survol.....	5
2.1 Pourquoi XML ?.....	5
2.2 XML ?.....	5
2.3 Pour en savoir plus.....	6
3 Intérêt d'un format bureautique XML.....	7
3.1 XML sur le poste de travail.....	7
3.2 OpenOffice.org : éditeur XML natif.....	7
4 Le format OpenDocument.....	9
4.1 XML généralisé au sein de la suite.....	9
4.2 Les standards, un choix confirmé.....	9
4.3 Une normalisation ISO 26300.....	10
4.4 Structure d'un fichier OpenOffice.org.....	10
4.5 Pour en savoir plus	11
5 OpenOffice.org XML vs Microsoft Office 2003 XML.....	12
5.1 Une intégration partielle.....	12
5.2 Des choix de format XML discutables.....	12
5.3 Des formats « ouverts »... bien verrouillés !.....	13
5.4 Une réaction tardive inadaptée.....	13
6 Quelques cas d'utilisation.....	15
6.1 Application de support aux audits financiers.....	15
6.1.1 Synthèse du besoin.....	15
6.1.2 Solution proposée.....	15
6.2 Chaîne de traitement XML.....	16
6.2.1 Synthèse du besoin.....	16
6.2.2 Solution Indesko.....	16
7 Conclusion.....	18
8 Support commercial.....	19
8.1 À propos d'inDesko.....	19
8.2 Pour nous contacter.....	19

1 Introduction

La suite bureautique OpenOffice.org est actuellement un des logiciels libres les plus utilisés dans le monde, et dont le développement est le plus actif : elle est devenue en effet une alternative crédible aux offres des acteurs dominants du marché. Un de ses avantages historiques mis régulièrement en avant, mais souvent mal connu ou compris, est son support natif du XML.

Nous allons étudier en détail dans ce document l'intérêt du XML pour le poste de travail, l'utilisation faite du XML dans OpenOffice.org, les opportunités qu'elle offre aux DSI et aux développeurs d'applications, et les différences entre l'utilisation du XML par Microsoft Office 2003 et celle d'OpenOffice.org. Les évolutions induites par le processus de normalisation d'OpenDocument seront également abordées.

2 Survol

2.1 Pourquoi XML ?

Aujourd'hui, les organisations doivent faire face à des besoins de plus en plus importants d'échanges de données entre des plate-formes hétérogènes. Par exemple : l'application de CRM doit garder la trace de toutes les correspondances avec les clients ; le système de gestion documentaire doit communiquer avec la suite bureautique pour valider les rapports et les propositions commerciales ; l'ERP doit savoir retrouver, extraire et synthétiser les données des analyses financières et des statistiques de la CRM pour aider au mieux les dirigeants à orienter la stratégie de l'organisation, etc.

Si l'on sait, de nos jours, facilement stocker et gérer un certain nombre de données structurées et définies (nombres, relations, etc.), il est encore relativement complexe d'appliquer ces mêmes méthodes aux données hétérogènes et peu structurées que sont les données bureautiques (emails, rapports, documentation technique, documentation projets, etc.). Ces données, qui peuvent s'avérer critiques et vitales, sont peu ou pas valorisées par les organisations qui les produisent. Ainsi, il est très difficile, voire impossible, pour une personne de retrouver facilement les emails envoyés par un de ses collègues à un client dans le cadre d'un projet auquel tous deux participent. Il en va de même pour les rapports, les notes, les analyses métiers, etc. Il est donc de plus en plus critique pour les organisations de pouvoir capturer, gérer, partager et valoriser ces flux d'informations peu structurées et hétérogènes.

XML, standard ouvert¹ largement accepté et déployé par l'industrie, peut aider à résoudre ce défi : permettre aux données hétérogènes et peu structurées d'être utilisées par le système d'information de l'organisation dans son ensemble.

2.2 XML ?

XML est un un méta-format de données. Il sert à définir des formats de données et fournit des outils pour formaliser cette définition (DTD, XML Schemas, Relax NG, ...), pour transformer un document XML en un autre document XML, faire des recherches dans un document, etc.

On peut ainsi facilement créer des formats de données adaptés aux informations à représenter en suivant des règles de description communes (pour utiliser des outils de lectures, de transformation et d'exploitation communs). Les formats diffèrent mais les outils de manipulation sont communs. Les formats deviennent donc facilement interopérables : ils sont lisibles grâce à des outils et des techniques standards, répandus et maîtrisés. Le seul point à acquérir pour lire et interpréter un document XML est la description du dialecte XML utilisé (techniquement, son schéma ou sa DTD).

XML représente donc un grand pas en avant dans l'interopérabilité entre systèmes, mais à une condition : que les fournisseurs et créateurs de formats de données jouent le jeu, i.e. publient et autorisent l'utilisation par des tiers de leurs définitions de types de documents (schémas ou DTD).

Pour mieux comprendre, prenons une analogie : XML permet de définir une langue propre à un besoin d'expression donné en respectant un alphabet et une prononciation communs tout en imposant un formalisme standard pour décrire la grammaire et le vocabulaire de la langue.

2.3 Pour en savoir plus...

Pour en savoir plus sur XML :

¹ « On entend par *standard ouvert* tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en oeuvre. », article 4 de la loi sur la confiance en l'économie numérique du 21 juin 2004 (Journal Officiel n° 143 du 22 juin 2004, page 11168).

◆ un site de ressources francophones : <<http://www.xmlfr.org>>.

◆ la page de ressources du W3C : <<http://www.w3c.org/XML>>.

Pour en savoir plus sur les formats ouverts : <<http://www.formats-ouverts.org>>.

3 Intérêt d'un format bureautique XML

3.1 XML sur le poste de travail

La suite bureautique est au centre de la production d'information de la large majorité des travailleurs de la connaissance : ils rédigent des compte-rendus, des rapports, des propositions, des contrats, des analyses, des notes, etc. Le traitement de texte en premier lieu, suivi du tableur et de l'outil de présentation, sont donc les outils essentiels des collaborateurs utilisateurs d'informatique.

Ces informations représentent une valeur très importante pour l'entreprise et constituent le patrimoine immatériel de l'organisation. Ceci fait de la suite bureautique l'outil de production principal dans de nombreuses structures. Il est facile de comprendre l'apport productif et concurrentiel qu'apporte la maîtrise de cette information : le marketing peut facilement consulter et exploiter les documents produits par le service veille, la direction des finances les rapports d'avancement des projets majeurs en cours, et la direction générale, les synthèses financières du directeur financier. Chacun peut réutiliser tout ou partie des documents consultés pour produire de nouveaux documents, en sachant immédiatement le niveau d'approbation et de validité du document consulté et en évitant de perdre du temps à remettre en forme les données copiées-collées.

Aujourd'hui la maîtrise de la bureautique et des outils Internet (navigateur et client de messagerie) est largement répandue au sein des organisations. Cela en fait donc des outils de choix pour l'accès aux applications de l'entreprise. Le navigateur est déjà couramment utilisé pour ce faire (la plupart des applications d'entreprises sont d'ores et déjà accessibles au travers d'un navigateur web, qui devient le terminal de consultation et de manipulation standard). En revanche, il n'existe, pour l'heure, aucune solution simple pour l'édition de documentation ou de formulaires complexes. En effet, il est difficile d'imaginer l'édition d'un document de plusieurs pages via un navigateur web. De même un formulaire web complexe est d'une utilisation peu pratique depuis un navigateur. La suite bureautique pourrait donc devenir un terminal applicatif lourd de choix (parce que maîtrisée et largement déployée) pour gérer les données complexes ou les documents longs. Il manque seulement un mécanisme d'échange et des possibilités suffisantes en terme de programmation.

3.2 OpenOffice.org : éditeur XML natif

OpenOffice.org produit nativement du XML et cette particularité est totalement transparente pour l'utilisateur. En effet, le format de stockage d'OpenOffice.org est un format XML tout à fait classique, qui permet de faire ainsi cohabiter simplicité d'utilisation, formats bureautiques (hétérogènes, souvent peu structurés) et fortes possibilités d'échanges. Le format XML d'OpenOffice.org 1.0 a d'ailleurs été pris comme base de travail par l'OASIS pour créer le standard "OpenDocument", qui a été adopté comme standard le 23 mai 2005. OpenOffice.org 2.0, supporte bien entendu le format OpenDocument de façon native et sert, en quelque sorte, d'implémentation de référence.

En plus de son format initial, OpenOffice.org est capable d'éditer tout type de format XML, tant qu'on lui fournit la DTD (définition du format) et les feuilles de transformations correspondantes. La distribution d'OpenOffice.org inclut d'ailleurs des définitions pour éditer nativement du XHTML, des fichiers Word 2003 XML et du DocBook.

Ces possibilités lui confèrent un large potentiel : édition directe de documents XML, intégration du format OpenDocument à une chaîne de traitement XML complète², génération de documents, manipulation de documents dans des applications métiers, génération de modèles de documents,

² Voir exemple ci-dessous (OpenOffice.org et DocBook XML)

etc.

Nous allons voir plus en détail le format bureautique d'OpenOffice.org, la différence entre l'utilisation du XML dans Microsoft Office 2003 et OpenOffice.org, puis étudier quelques exemples d'applications concrètes utilisant OpenOffice.org et le XML.

4 Le format OpenDocument

4.1 XML généralisé au sein de la suite

Tout d'abord, il est important de noter que tous les composants d'OpenOffice.org utilisent XML comme format natif de stockage. Le format XML est partagé par tous les composants de la suite. Ce format XML a été standardisé par l'OASIS³ sous le nom "OpenDocument". L'OASIS souhaite avec ce format définir le standard d'échange des documents bureautiques. Il est donc fort souhaitable que l'industrie se rassemble derrière ce nouveau standard qui permettrait d'ouvrir la voie vers l'interopérabilité optimale des applications bureautiques. Ce format désormais normalisé par l'ISO⁴ sous la référence 26300, processus entamé en 2005 suite à l'annonce de Mathias Bauer lors de la conférence OpenOffice.org 2004..

4.2 Les standards, un choix confirmé

Le format OpenDocument définit une structure de données cadre commune et utilisée par tous les composants de la suite bureautique. Une structure commune implique que ses objets identiques mais utilisés dans des circonstances différentes sont décrits de la même manière. Le jeu de caractère est lui-même standard puisque le choix s'est porté sur UTF-8, permettant une couverture unique des différentes langues.

Voici la liste des schémas utilisés par les documents OpenDocument :

- ◆ texte
- ◆ style
- ◆ tableau
- ◆ nombre
- ◆ diagramme
- ◆ DR3D
- ◆ formulaire
- ◆ scripts / macros
- ◆ paramètres de visualisation, impression et édition
- ◆ mise en forme de texte (W3C Formatting Output :FO)
- ◆ formule mathématique (W3C MathML)
- ◆ graphique vectoriel (W3C Scalable Vector Graphics : SVG)
- ◆ lien interne ou externe (W3C XLink)
- ◆ méta-données (DublinCore)

Il est important de noter l'effort de modularisation et de mutualisation des schémas des différents éléments utilisés pour créer des documents bureautiques. Par exemple, le schéma de tableau est utilisé autant par le tableur que par le traitement de texte pour décrire les tableaux insérés dans un document texte (hors tableur). Réciproquement, le tableur utilise le schéma de texte pour décrire le texte inséré dans les cellules et le format de mise en forme pour décrire la mise en forme. De même, les styles sont décrits en utilisant W3C FO et des éléments propres pour les paramètres

3 L'OASIS (Organization for the Advancement of Structured Information Standards) est un consortium international à but non lucratif dont l'objectif est de conduire le développement, la convergence et l'adoption de standards dans le domaine du e-business.

4 Les travaux de la commission de normalisation se poursuivent et peuvent être suivis à l'URL <http://www.iso.org/iso/fr/CatalogueDetailPage.CatalogueDetail?CSNUMBER=43485>

non couverts par FO.

De nombreux experts XML (notamment Normal Walsh, leader du projet DocBook) déclarent qu'OpenDocument est un format techniquement exemplaire : il allie astucieusement des formats existants et des formats propres (pour les éléments pour lesquels il n'existe pas de standard défini). C'est, au dire de tous, un archétype de l'utilisation de la puissance du XML en terme d'interopérabilité et de réutilisation. À la fois génialement simple et puissant !

James Clark (membre actif du groupe de travail sur XML 1.0 et expert XML international incontesté) a d'ailleurs activement participé à l'élaboration de ce format.

Enfin, le standard OpenDocument est décrit en utilisant RELAX NG comme format de schémas (Le format initial d'OpenOffice.org utilisait des DTD), et les fichiers XML produits respectent le [6ème but du XML](#), à savoir être lisibles par un humain et raisonnablement clairs.

4.3 Une normalisation ISO 26300

La genèse du format OpenDocument se situe en 1999 lors de la création du nouveau format de fichier StarOffice. En 2000, le projet OpenOffice.org naît sur ces bases et généralise l'utilisation du XML au sein des documents bureautiques.

La standardisation a été menée par un consortium indépendant regroupant les acteurs majeurs de la bureautique, visant à l'établissement d'un véritable ouvert pour les fichiers bureautiques.

La normalisation ISO qui aboutit, après toutes ces années d'élaboration cohérente au service de la rigueur du format, est donc l'étape logique et naturelle de l'évolution, concrétisant ainsi un travail de définition et de structuration du format entamé voilà plusieurs années.

La normalisation ISO d'un format de document XML est une avancée majeure dans la manière de considérer le fonds documentaire. C'est une reconnaissance aux rédacteurs et administrateurs, possesseurs de documents, d'un droit à l'indépendance vis-à-vis d'un éditeur logiciel, à l'interopérabilité de leur fonds documentaire, à la pérennité de leur information, tout en offrant les garanties technologiques de traitements ultérieurs.

4.4 Structure d'un fichier OpenOffice.org

OpenOffice.org reste l'implémentation de référence d'OpenDocument. Un fichier OpenOffice.org est en fait une ressource JAR (une archive ZIP au sens Java) contenant divers éléments dont les fichiers texte XML. Le format est donc totalement ouvert et pérenne puisque un simple éditeur de texte permet de consulter les informations des fichiers XML, l'algorithme de compression zip étant lui même libre et ouvert⁵

Lorsque l'on décompresse un fichier OpenOffice.org (à l'aide d'un simple décompresseur ZIP comme WinZip sous Windows ou unzip sous UNIX), on obtient entre autres les fichiers suivants :

- ◆ *META-INF/manifest.xml* : liste et description des fichiers contenu dans le JAR
- ◆ *meta.xml* : méta-données du document
- ◆ *mimetype* : type MIME du fichier
- ◆ *settings.xml* : paramètres d'édition, impression et visualisation d'un document
- ◆ *styles.xml* : description des styles du document
- ◆ *content.xml* : Contenu sémantique du document
- ◆ *Pictures/* : répertoire contenant les images insérées dans le document
- ◆ *Basic/* : répertoire contenant les macros contenues dans le document

5 Il est d'ailleurs couvert par la RFC 1950 : <ftp://ftp.uu.net/pub/archiving/zip/zlib/rfc1950.txt>

- ◆ *Thumbnails/* : répertoire contenant les vignettes du document pour l'intégration au bureau
- ◆ *Configurations2/* : répertoire stockant les fichiers de paramétrages propre à l'interface graphique d'OpenOffice.org (comme certaines barres d'outils par exemple)

On constate ici encore la modularité du format OpenOffice.org : un document XML pour les méta-données, un pour les styles, un pour le contenu du document et un pour les paramètres d'impression, d'édition et d'affichage. Les images et les macros, elles, sont tout simplement stockées dans leur répertoire dédié.

Cette organisation facilite vraiment l'utilisation du format par d'autres logiciels. Ainsi, il est possible de ne travailler que sur le contenu d'un document sans être perturbé par les styles (ce qui est très utile lors de la génération de document ou de l'extraction de données, par exemple). De même, la synchronisation des méta-données de documents OpenOffice.org dans un système de gestion documentaire est facilité : il suffit de ne travailler que sur le document XML de méta-données (l'implémentation du standard DublinCore apporte un autre niveau d'interopérabilité dans ce domaine).

4.5 Pour en savoir plus ...

- ◆ L'Alliance ODF (OpenDocument Format), regroupement international d'institutions soutenant le format OpenDocument
<http://www.odfalliance.org/>
- ◆ Lettre ouverte de l'Alliance au Directeur de l'Agence de modernisation de l'Etat (Français) suite à l'appel à commentaire sur le Référentiel Général d'interopérabilité
http://www.odfalliance.org/resources/FranceRGI_20060614.pdf

5 OpenOffice.org XML vs Microsoft Office 2003 XML

Depuis l'annonce d'Office 2003, Microsoft se fait le chantre de l'interopérabilité des systèmes et de l'ouverture, utilisant comme fer de lance l'intégration du XML dans les logiciels de sa suite MS-Office (notamment les trois plus utilisés, Word, Excel et PowerPoint). Il convient donc d'analyser plus en détail et sans *a priori* les modalités de l'intégration du XML dans Office 2003.

5.1 Une intégration partielle...

Après un rapide tour d'horizon, on peut noter que tous les logiciels de la suite MS-Office ne bénéficient pas du même traitement. En effet, seuls Word, Excel, Frontpage, et Visio en bénéficient. PowerPoint et Access sont donc exclus de cet effort d'ouverture, alors que Microsoft promettait que tous les logiciels de la suite seraient fondés sur XML.

Comparons avec l'usage de ces différents logiciels : Word et Excel sont incontestablement les plus utilisés, suivi de près par PowerPoint. FrontPage est très largement devancé par ses concurrents (comme Macromedia Dreamweaver) et en tant qu'éditeur HTML, on peut penser raisonnablement que le support du XML est un peu tardif. Pour ce qui est de Visio et d'Infopath (nouveaux logiciels de la suite), ils ne connaissent pas l'audience de leurs aînés : Visio, récemment acquis par Microsoft est un éditeur de diagrammes et Infopath un outil de gestion de formulaires.

Un dernier point à remarquer : au-delà des possibilités techniques offertes par un outil bureautique, on sait que ce sont celles, proposées par défaut aux utilisateurs qui sont les plus utilisées. Or, par défaut, les logiciels de la suite MS-Office persistent à proposer aux utilisateurs de sauvegarder leurs documents dans le format binaire, propriétaire, non complètement documenté, de la suite Office 97/2000/XP. Et seuls les utilisateurs des versions les plus onéreuses de MS-Office ont la possibilité de tirer partie de toutes les fonctionnalités que Microsoft leur propose dans le domaine du XML. Nous pensons que ces deux faits en disent long sur la volonté auto-proclamée de Microsoft de démocratiser l'usage du XML sur le poste de travail.

5.2 Des choix de format XML discutables...

Au premier abord, on peut être tenté de saluer l'effort d'ouverture de Microsoft avec la nouvelle version de sa suite — ce qui est, il faut bien le dire, une pratique toute récente de Microsoft vis-à-vis de sa poule aux oeufs d'or. La tendance précédente était plutôt l'incompatibilité descendante et la dissimulation du format de fichier, ce qui a empêché pendant de nombreuses années ses concurrents ou confrères de réaliser des logiciels de lecture pour les formats MS-Office. Ces pratiques anti-concurrentielles ont d'ailleurs été reconnues à de multiples reprises par les cours de justice américaines et européennes.

En analysant un peu plus l'intégration effectuée, on s'aperçoit tout d'abord qu'Infopath (né Xdocs), dont la fonction est de gérer des formulaires, ne repose pas sur le standard XForms du W3C. C'est d'autant plus surprenant que W3C XForms a été conçu pour la description complète de formulaires complexes, palliant ainsi les nombreuses lacunes du (X)HTML dans ce domaine. Or c'est exactement à ce besoin que semble répondre Infopath. Il est donc particulièrement étonnant qu'il ne supporte pas le standard du W3C mais apporte son propre format.

Dans un registre connexe, le manque de lisibilité du format XML édité va à l'encontre du [6ème but du XML](#). Les abréviations utilisées rendent le décryptage du fichier difficile et, là où un objectif de XML est de rester compréhensible, on peut mettre en évidence une certaine opacité des termes employés accrue par la non-utilisation des standards en cours.⁶

6 Un article révélateur en anglais de Carrera, D'Arcus, Eisenberg et Hudson
<http://www.groklaw.net/article.php?story=20051125144611543>

Pour les formats de Word et Excel, lors de leur création, il n'existait pas de format XML pour les fichiers issus de traitement de texte et de tableur. En revanche on ne peut que regretter que Microsoft n'ait pas pris part à l'effort conjoint de plusieurs ténors du marché (Sun Microsystems, IBM, Oracle...) pour créer un format bureautique interopérable et standard, au sein du consortium "neutre" OASIS.

5.3 Des formats « ouverts »... bien verrouillés !

Microsoft a récemment déposé un brevet (accepté par l'OEB, l'Office Européen des Brevets) sur "l'utilisation d'un schéma XML pour valider le document XML de sortie d'un traitement de texte" (référence : EP-1376387). Tout logiciel de traitement de texte utilisant un schéma XML pour lire ou écrire un document XML (dans un autre format que WordML) est donc contrefacteur de ce brevet ! On peut se demander où se trouve l'innovation technique dans ce procédé : à quoi sert à un schémas XML, si ce n'est à valider un document XML ?

D'un autre côté, Microsoft publie les formats WordML et ExcelML sur son site web⁷, et autorise l'implémentation de ces formats dans tous les logiciels sans le versement de royalties à Microsoft pour l'utilisation de leur brevet (voir la *Patent License* des formats en question). En d'autres termes, Microsoft interdit aux traitements de texte d'utiliser un schéma XML pour valider les documents produits *sauf s'il s'agit du format WordML*. Astucieuse technique de standardisation par la force...

Cette initiative contredit largement les dires de Microsoft quant à sa volonté d'ouverture et de respect des standards (en tout cas, selon la loi française, les formats WordML et ExcelML ne peuvent être considérés comme des formats ouverts). Heureusement, dans l'Union Européenne, nous ne sommes pour l'instant pas soumis à ce traitement, les procédés logiciels étant exclu du périmètre de brevetabilité. Si ce n'était plus le cas, il y aurait alors une contradiction manifeste entre les formats WordML et ExcelML et la définition de standard ouvert dans la Loi sur la confiance en l'Économie Numérique (voir note de bas de page 1, page 5).

De plus, la Commission Européenne, dans le cadre du programme IDA Interchange of Data between Administrations, a mis en concurrence les formats XML de Microsoft et ceux d'OpenOffice.org. La commission a choisi, après une analyse détaillée, le format OpenDocument de l'OASIS et a demandé à Sun Microsystems et à l'OASIS de soumettre le format à l'ISO en vue d'une normalisation formelle internationale.⁸

5.4 Une réaction tardive inadaptée

Devant la normalisation ISO 26300 démarrée pour OpenDocument après l'annonce de la standardisation de l'OASIS, Microsoft réagit en annonçant le lancement de son coté d'une standardisation auprès de l'ECMA, autre organisme de standardisation (alors que Microsoft est également membre de l'OASIS).

L'objectif affiché est une normalisation ISO pour début 2007, soit faire en un peu plus d'un an le travail de fond de définition d'un format bureautique ouvert en interopérable que les acteurs du consortium OASIS ont mis plus de 7 ans à définir, implémenter et normaliser.

Maintenant, à la lecture des objectifs de la standardisation⁹, il apparaît que la tâche est tronquée

7 Il est d'ailleurs intéressant de noter que les formats sont diffusés uniquement dans un installateur Windows dont on peut douter de l'intérêt réel à part l'interdiction imposée aux utilisateurs d'autres systèmes d'accéder à ces formats.

8 Pour plus d'information, il est possible de consulter le [texte des recommandations](#) et un extrait de [la lettre qu'à reçu l'équipe d'OpenOffice.org](#).

9 Les objectifs du comité technique 45 de l'ECMA couvrant la standardisation du format Office Open <http://www.ecma-international.org/memento/TC45.htm>

puisque l'objectif affiché est de définir un standard « interopérable » avec le format Office Open¹⁰ défini par Microsoft. Toute la démarche est centrée autour de ce format. C'est lui qui guide le processus de l'ECMA là où la recherche de l'intérêt commun et la définition d'un format interopérable dédié à la bureautique sont menées par les différents intervenants du format OpenDocument, eux même éditeurs bureautiques mutualisant leur expérience.

La marche forcée vers la standardisation et processus de normalisation de la part de Microsoft valide les choix technologiques qu'OpenDocument a fait dès ses débuts : XML et zip ne sont pas nouveaux en bureautique. Leur utilisation n'est qu'un préalable à l'établissement d'un format ouvert, interopérable, puisque l'utilisation des standards et une démarche globale centrée sur la structure du document sont nécessaires pour y arriver.

10 L'attention du lecteur sur l'ambiguïté du nom du format Microsoft est attirée : Office Open n'est pas OpenOffice.org, ni OpenDocument.

6 Quelques cas d'utilisation

Voici maintenant quelques illustrations de l'utilisation du format XML d'OpenOffice.org dans la réalisation d'applications métiers et la mise en place d'une interopérabilité entre les logiciels bureautiques et d'autres applications du système d'information d'une organisation.

6.1 Application de support aux audits financiers

Dans le contexte d'un projet global d'Extranet de services pour la [Compagnie Nationale des Commissaires aux Comptes](#), Indesko a réalisé une application de gestion d'audits financiers fondée sur OpenOffice.org et sur [Nuxeo CPS](#).

OpenOffice.org est utilisé comme client lourd pour la production des documents et données liés aux audits financiers. Nuxeo CPS, pour sa part, prend en charge les fonctionnalités de travail collaboratif nécessaires à la coopération d'une équipe pour la réalisation d'un audit financier.

6.1.1 Synthèse du besoin

Dans leur mission, les Commissaires aux Comptes doivent, pour certifier la validité des comptes des entreprises, réaliser des audits financiers. Chaque année, il doivent produire un « dossier d'audit » qui rassemble un ensemble d'informations sur les modes de gestion et les modes opératoires de l'entreprise.

Un dossier d'audit financier est un ensemble de formulaires et de documents que le Commissaire aux Comptes (CAC) doit remplir et/ou produire dans le cadre de sa mission d'audit.

La CNCC souhaite donc proposer à ses membres une application complète permettant de dématérialiser autant que possible les dossiers d'audits et de gérer la collaboration de plusieurs CAC sur un même dossier (entre 2 et 20). Les principales fonctions nécessaires sont : la comparaison de documents, la fusion, la copie d'information d'un document à l'autre, etc.

6.1.2 Solution proposée

Un dossier d'audit est un ensemble de classeurs¹¹ Calc d'OpenOffice.org. Chaque classeur (nommé intercalaire dans le vocabulaire métier de l'application) contient entre 5 et 100 feuilles. Chaque membre d'une équipe d'audit est mandaté pour une partie d'intercalaire et, en pratique, une même feuille d'un intercalaire ne peut-être modifiée que par une personne.

Chaque CAC, en début de mission, emporte donc les classeurs Calc (les intercalaires) sur son portable pour aller travailler chez son client. Chaque jour ou chaque semaine, il poste sur le portail collaboratif CPS le résultat de ses travaux (dans une zone nommée « Contributions »).

Le chef de mission consulte régulièrement les contributions et peut demander une comparaison entre plusieurs classeurs issus des contributions (les éléments de comparaison) et un classeur utilisé comme classeur de référence (la référence de comparaison) pour fusion. Le portail compare alors les documents et présente les différences feuille par feuille. L'utilisateur peut alors sélectionner les modifications à reporter dans le classeur de référence. Le portail fusionne les modifications sélectionnées dans le classeur de référence et stocke ce classeur modifié en tant que nouvelle version du document.

De même, les utilisateurs peuvent sélectionner deux classeurs OpenOffice.org et demander à réaliser des copies de feuilles entre classeurs. Le portail propose alors à l'utilisateur la liste des feuilles pour chaque classeur et une interface pour déplacer ou copier des feuilles d'un classeur à l'autre.

¹¹ Un « classeur » est ce qu'on appelle parfois aussi « feuille de calcul », ou communément et improprement « tableau Excel ».

L'application dans le portail permet également de générer des rapports en ligne sur des données issues des classeurs OpenOffice.org (données liés au bilan comptable de l'entreprise).

Enfin, Indesko a réalisé des extensions logicielles à OpenOffice.org permettant de générer, depuis OpenOffice.org, des feuilles de calculs à partir de documents comptables (bilan comptable et balance).

La solution implémentée manipule directement les documents XML présents (ici : metadata.xml et content.xml) dans les fichiers OpenOffice.org en toute transparence pour l'utilisateur et en validant les documents XML modifiés par rapport aux DTD du format.

6.2 Chaîne de traitement XML

6.2.1 Synthèse du besoin

CNCC Éditions, éditeur de nombreux ouvrages sur les règles de gestion et les normes d'audit souhaitait :

- ◆ Refondre complètement sa chaîne de production des publications et ouvrages de la CNCC (normes comptables, normes de gestion, commentaires de lois, textes juridiques, livres complets, etc.).
- ◆ Structurer la production de contenu pour mettre en place une chaîne fluide et continue de la production de documents à l'imprimeur (en passant par les contributeurs, les relecteurs, les correcteurs, les valideurs, etc.).
- ◆ Stocker les documents produits dans un format XML pérenne et reconnu, indépendamment de la forme initiale, pour capitaliser et générer facilement des rendus différents des documents (HTML pour publication web, HTML pour CDROM, PDF pour diffusion électronique, PDF pour impression d'ouvrages, etc.).

6.2.2 Solution Indesko

Indesko a proposé et mis en place une chaîne de production XML fondée sur OpenOffice.org, le format Open Office, le format DocBook et Nuxeo CPS.

- ◆ OpenOffice.org est utilisé pour éditer les documents.
- ◆ Le format OpenOffice.org est utilisé lors de l'élaboration des documents et la transformation à la demande (via une feuille de style XSLT) en documents DocBook XML valides.
- ◆ DocBook est utilisé pour stocker le contenu de façon pérenne et indépendante de la forme, puis pour générer les formats de rendus finaux via des feuilles XSLT : PDF pour diffusion, PDF pour impression d'ouvrages, HTML simple, HTML complexe.
- ◆ Nuxeo CPS est utilisé pour gérer la coopération autour des documents, les transformations XML et la publication web sur les sites Internet et Extranet de la CNCC.

Les utilisateurs éditent les documents depuis OpenOffice.org et coopèrent à leur édition en utilisant les outils adéquats offerts par CPS. La qualification sémantique du contenu saisi est effectuée en utilisant les styles d'OpenOffice.org (ce qui, d'après les retours d'expérience, leur apparaît une solution d'usage naturel).

L'outil de travail collaboratif (Nuxeo CPS) transforme les documents dès leur modification en DocBook puis en HTML et en PDF. Les utilisateurs ont donc à tout moment la possibilité de visualiser le rendu « final » du document et d'exporter le document en format DocBook.

Le portail utilise DocBook XML comme format de stockage et comme pivot vers les rendus « finaux ».

Grâce à cette architecture, la CNCC bénéficie donc à la fois :

- ◆ D'un éditeur de contenu simple et convivial : OpenOffice.org. Les utilisateurs peuvent de plus tirer avantage de toutes ses fonctionnalités utiles à l'édition de contenu (suivi de changements, commentaires dans le texte, gestion d'index, etc.)
- ◆ D'un outil de travail collaboratif intégré profondément avec l'éditeur de contenu (ex : édition directe et workflow depuis OpenOffice.org) : Nuxeo CPS.
- ◆ D'un stockage complètement en XML de son contenu, dans un format standard créé et reconnu pour gérer de la documentation.

En bref, OpenOffice.org a permis la mise en place d'une chaîne intégrée de production XML, des auteurs jusqu'à l'imprimeur !

7 Conclusion

XML est une avancée importante dans l'informatique pour favoriser l'interopérabilité des systèmes et des formats.

Mais il ne suffit pas de lire ou produire du XML pour être ouvert et intéropérable. Il est nécessaire que les schémas soient diffusés, compréhensibles, documentés et légalement utilisables (sans brevets restreignant leur usage). Il est également primordial d'utiliser et d'intégrer, en exploitant le caractère modulaire de XML, les formats standards existants, pour tirer partie des applications ou des composants déjà réalisés autour de ces formats.

Lorsque l'on conçoit et réalise un format de document basé sur XML, il est indispensable de penser aux autres utilisateurs du format et d'impliquer l'ensemble des acteurs concernés. De même, verrouiller un format ouvert par des moyens directs ou détournés n'est pas acceptable, dans cette vision.

OpenOffice.org a fait du XML une utilisation exemplaire et a permis de définir un nouveau standard, qui est à n'en pas douter, une grande avancée pour la bureautique : OpenDocument.

OpenOffice.org apporte ainsi, directement sur le poste client, la possibilité de créer des documents structurés exploitables par l'ensemble du système d'information de l'entreprise. Cette possibilité est déjà exploitée dans des projets d'envergure.

8 Support commercial

8.1 À propos d'inDesko

inDesko est une division de Nuxeo, spécialisée dans les services autour de la bureautique libre et de la suite OpenOffice.org. inDesko conseille et accompagne ses clients dans leurs projets de migration vers OpenOffice.org et vers la bureautique libre, et développe des modules métiers et des applications dédiées leur permettant d'intégrer la suite logicielle au sein de leur système d'information tout en optimisant leurs processus de travail.

8.2 Pour nous contacter

inDesko

18-20, rue Soleillet

75020 Paris

Tél: +33 (0)1 40 33 71 56

Fax: +33 (0)1 43 58 14 15

Mél: contact@indesko.com

Web: <http://www.indesko.com/>