

Le XML dans Microsoft Office

par Lebeau Olivier ([Heureux-oli sur DVP](#))

Lebeau Olivier

Date de publication : 15-04-2007

Dernière mise à jour :

Depuis quelques temps, Microsoft offre la possibilité d'utiliser le XML pour la sauvegarde des documents Office, et avec la dernière version de la suite bureautique, on peut désormais utiliser le format Open XML. L'article qui suit ne traite en rien le XML ni celui contenu dans les fichiers Office Open XML.

Lebeau Olivier

I - Le Format Open XML

II - Office et le XML

III - Office et l'Open XML

III-A - Office 2003 et l'Open XML

III-B - Office 2007 et l'Open XML

III-B-1 - Types de fichiers disponibles

III-B-2 - Document Word (*.docx)

III-B-3 - Document Word avec Macro (*.docm)

III-B-4 - Protection des documents Word

III-B-5 - Document Excel (*.xlsx)

IV - Conclusions

Liens

VI - Remerciements

I - Le Format Open XML

Office Open XML est un schéma de fichiers XML qui forment un ensemble structuré pour répondre aux exigences de logiciels de traitement de textes, tableur et présentation.

Dans l'élaboration de la définition de l'Open XML, Microsoft n'a pas joué cavalier seul. On retrouve entre autre des sociétés comme Apple, BP, Essilor, Intel, Novell, ...

Le but final étant de développer un format de fichier libre de droit et universel.

Office Open XML n'est pas le seul format à briguer la place de format unique. L'OpenDocument est en lice pour être également reconnu comme le format universel de fichier pour les suites bureautiques.

Le principe utilisé par OpenDocument est le même que celui utilisé dans Office Open. OpenDocument est antérieur à Office Open XML et est le type de document utilisé principalement par la suite OpenOffice.org.

II - Office et le XML

Le XML fut le précurseur de l'Open XML dans la suite Office. Dans la version 2003 de la suite bureautique de Microsoft, une approche du XML a été faite. C'est ainsi qu'il est possible de sauvegarder des documents Word, Excel et PowerPoint en format XML. On retrouve dans ce cas un seul document qui contient toutes les informations du fichier. Pour vous en convaincre, je vous invite à créer un nouveau fichier Word qui ne contient que quelques mots. Donnez un nom à ce fichier et sauvegardez le en XML.

Ouvrez ensuite ce fichier à l'aide du bloc note. Vous verrez que ce fichier contient des balises de début et des balises de fin, et perdu dans ces balises, le texte que vous avez écrit.

Si vous tentez la même opération avec un fichier .doc, avec un peu de chance, vous trouverez votre texte perdu au milieu de signes inintelligibles. Ce format est un format binaire propre à Word.

Les autres programmes de la suite Office sont également capables de traiter les documents en XML.

Lebeau Olivier

III - Office et l'Open XML

Open XML est le nouveau format de fichier utilisé par défaut par la dernière suite bureautique de Microsoft. Les logiciels de la suite utilisent dorénavant ce format de fichier. Si le détail de ce format vous intéresse, vous trouverez au bas de ce document un lien vers un article publié par un membre de la rédaction de DVP. Le format Open XML y est décrit en long et en large avec comme fichier d'exemple un document Word 2007.

Ce nouveau format de fichier est défini dans une norme ISO (ISO-8859-1)

III-A - Office 2003 et l'Open XML

Les utilisateurs de la suite Office 2003 ne sont pas oubliés par Microsoft. Un fichier est librement mis à leur disposition pour assurer la compatibilité avec les nouveaux produits.

Le fichier de conversion est disponible sur le site de téléchargement de Microsoft.

Ce fichier est un convertisseur de fichier, mais on pourrait le comparer à un Add On pour Office.

Une fois installé, il est alors possible d'ouvrir des fichier *.docx, *.xlsx, ...

Pour se rendre compte que ce n'est pas un simple convertisseur de fichier, il suffit de créer un nouveau fichier Word, et dans la boîte de dialogue de sauvegarde du fichier, on trouve l'extension .docx avec comme description Word 2007 document.

C'est une avancée notable. C'est la première fois que pour une suite office Microsoft, on retrouve une compatibilité ascendante.

Il est désormais possible de profiter de la stabilité des fichiers en Open XML avec une version qui au départ n'était absolument pas prévue pour son utilisation.

III-B - Office 2007 et l'Open XML

L'Open XML est le type de fichiers natifs de Office 2007.

Cette suite offre comme toutes les autres la compatibilité descendante des fichiers.

III-B-1 - Types de fichiers disponibles

On retrouve dans cette suite une multitude d'extension et ce pour chaque produit.

Extensions	Types de fichiers
*.docx	Documents Word par défaut
*.docm	Document Word autorisant les macro. Cet type de fichier est récent. Sur les version précédentes aucune différence n'était faite pour les fichiers avec macro.
*.dotx	Modèles de documents Word
*.dotm	

	Modèles de documents avec macro incorporée.
*.xlsx	Classeur Excel par défaut
*.xlsm	Classeur Excel par autorisant les macros
*.xltx	Modèle Excel par défaut
*.xltm	Modèle Excel autorisant les macros
*.xlsb	Classeur Excel binaire
*.xlam	Module d'extension autorisant les macros
*.pptx	Présentation Power Point par défaut
*.pptm	Présentation Power Point avec macro
*.potx	Modèle Power Point
*.potm	Modèle Power Point avec Macro
*.ppam	Module d'extension avec macro
*.ppsx	Diaporama
*.ppsm	Diaporama avec macro

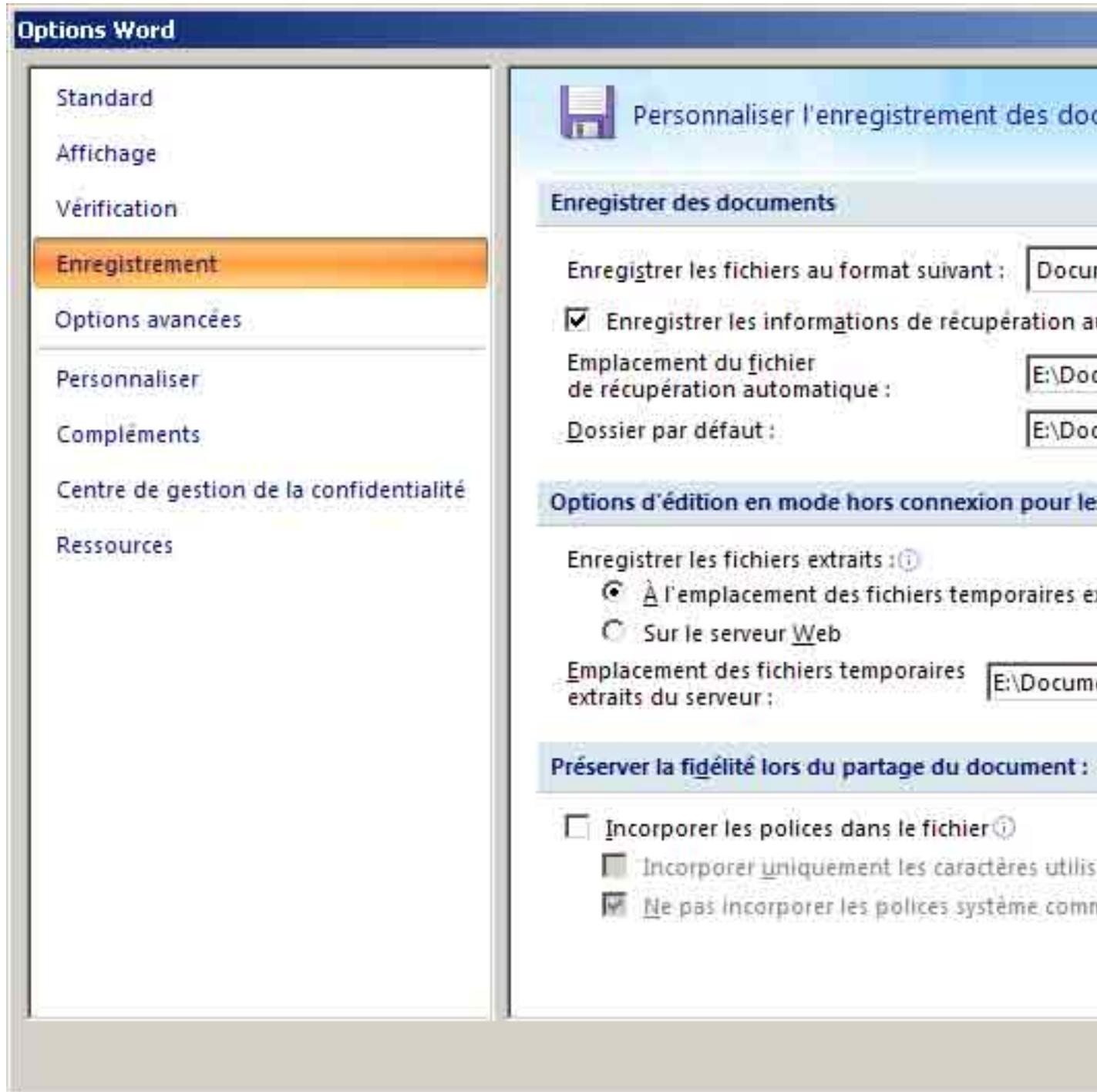
III-B-2 - Document Word (*.docx)

Les documents Word par défaut contiennent les fichiers suivants :

```

DocxSansMacro.docx
[Content_Types].xml
DocxSansMacro\docProps
  app.xml
  core.xml
DocxSansMacro\word
  document.xml
  fontTable.xml
  settings.xml
  styles.xml
  webSettings.xml
DocxSansMacro\word\theme
  theme1.xml
DocxSansMacro\word\_rels
  document.xml.rels
DocxSansMacro\_rels
  .rels
  
```

Il est désormais possible d'enregistrer les polices de caractère dans le fichier. Par cette action, les destinataires des fichiers ne devront plus se procurer les polices pour afficher correctement vos documents. De plus, si ces documents sont retravaillés, la police reste utilisable.



III-B-3 - Document Word avec Macro (*.docm)

La composition des fichiers avec macro incorporée est presque identique aux fichiers sans macro. Deux fichiers se trouvent en plus dans l'archive. L'un de ces fichiers possède une extension *.bin et n'est pas éditable du moins pour un humain normal. Le second est un fichier *.xml avec une description des normes utilisées.

DocxAvecMacro\word

Lebeau Olivier

```
vbaData.xml  
vbaProject.bin
```

III-B-4 - Protection des documents Word

Cette fonctionnalité n'a pas été mise de côté avec Office 2007, il est toujours possible de protéger ses fichiers par mot de passe. J'ai fait l'expérience de protéger un fichier Word avec un mot de passe pour la lecture. Lors de l'ouverture, le mot de passe est demandé.

J'ai tenté la même opération que précédemment, et j'ai changé l'extension du fichier en *.zip pour l'ouvrir avec un programme d'archivage. Le fichier n'est pas reconnu comme étant valide.

Le seul inconvénient que j'ai rencontré, c'est la taille du fichier qui devient démesurée. Pour un fichier qui au départ fait moins de 20 Ko, elle passe à 800 Ko pour avoir simplement ajouté un mot de passe.

III-B-5 - Document Excel (*.xlsx)

Le document Excel possède une structure similaire au document Word.

Et l'édition des différents composants de ce fichier peut se faire au travers d'un simple éditeur de texte.

```
Classeur.xlsx  
[Content_Types].xml  
Classeur\_rels  
  .rels  
Classeur\docprops  
  app.xml  
  core.xml  
classeur\xl  
  SharedStrings.xml  
  Styles.xml  
  workbook.xml  
classeur\xl\_rels  
  workbook.xml.rels  
classeur\xl\theme  
  theme1.xml  
classeur\xl\Worksheets  
  sheet1.xml  
  sheet2.xml  
  sheet3.xml
```

IV - Conclusions

Avec son nouveau format de fichier, Microsoft franchit un pas vers la portabilité et l'universalité. Il est désormais possible de générer des fichiers compatibles Microsoft Office depuis n'importe quelle application.

De plus la stabilité des fichiers Open XML n'est pas à démontrer, elle coule de source. Terminée l'époque où pour récupérer des données importantes d'un fichier corrompu, il fallait user de ruses et de stratagèmes. Avec des fichiers qui sont enregistrés en texte clair et plus en binaire, n'importe quel éditeur de texte simple sera à même de récupérer les données.

Lebeau Olivier

Liens

Je vous conseille de visiter le lien ci-dessous.

 **Structure des fichiers Open XML**

Lebeau Olivier

VI - Remerciements

Je remercie toutes les personnes qui m'ont aidé de près ou de loin à la rédaction de cet article et plus particulièrement l'équipe Office, ainsi que Yogui, ZedroS et troumad pour leur touche finale.

