

# **Comment récupérer un document OOo corrompu**

*Distribué par*

*Le projet [OpenOffice.org](http://OpenOffice.org)*

## Table des Matières

<u>1 Introduction</u>	<u>3</u>
<u>2 Bibliographie</u>	<u>3</u>
<u>3 Installation des logiciels</u>	<u>3</u>
<u>3.1 Logiciels nécessaires</u>	<u>3</u>
<u>3.2 Sous windows</u>	<u>4</u>
<u>3.3 Sous linux</u>	<u>4</u>
<u>4 Structure d'un fichier OOo</u>	<u>4</u>
<u>5 Réparation du fichier</u>	<u>5</u>
<u>5.1 étape 1 : repérage de l'erreur</u>	<u>5</u>
<u>5.2 étape 2 : décompression du fichier</u>	<u>6</u>
<u>5.3 étape 3 : correction de l'erreur et réparation du fichier</u>	<u>7</u>
<u>5.4 étape 4 : fermeture du fichier (et compression)</u>	<u>7</u>
<u>6 Remarques</u>	<u>8</u>
<u>7 Crédits</u>	<u>8</u>
<u>8 Licence</u>	<u>8</u>

## 1 Introduction

Ce How-to a pour but de montrer comment récupérer des fichiers corrompus. Pour suivre ce document et parvenir à une récupération de votre fichier, vous aurez besoin de quelques logiciels. Ceux-ci sont donnés à titre indicatif et d'autres peuvent exister. Ces logiciels sont listés plus bas.

Ce document est constitué comme suit : une partie qui permet d'installer les logiciels (assez succincte car ce n'est pas l'objet du présent document), une partie qui présente les balises xml et enfin la réparation du fichier proprement dite (la partie la plus importante).

OOo version 1.1 sera noté dans ce document OOo 1.1, OOo version inférieure à la 1.0.3 sera noté OOo 1.0.x

Un exemple permettra d'illustrer mon propos tout au long de ce document. Le fichier corrompu s'appelle mon\_fichier.sxw

Il est important de noter que la plupart des fichiers corrompus sont dûs à un problème sur plusieurs lignes. Ceci complique la récupération dudit fichier. Ce how-to vous permettra de récupérer vos fichiers dans certain cas seulement.

## 2 Bibliographie

La liste user-fr d'OOo-fr (<http://fr.openoffice.org>)

L'espace xml francophone : <http://xmlfr.org/>

FAQ XML (En) : <http://www.ucc.ie:8080/cocoon/xmlfaq>

Apprendre le XML : [http://babel.alis.com/web\\_ml/xml/REC-xml.fr.html](http://babel.alis.com/web_ml/xml/REC-xml.fr.html)

## 3 Installation des logiciels

### 3.1 Logiciels nécessaires

Il faudra un logiciel pour compacter/décompacter. Pour cela vous avez Unzip, Ark, FileRoller sous Linux, WinZip (shareware, limité dans le temps) et 7-Zip (GPL, sans limitation).

Vous aurez besoin d'un logiciel permettant de lire du texte. Vous pouvez utiliser , K Hex Edit (GPL, Linux), Nedit (GPL, Linux), (\*ne fonctionne pas, à confirmer). Ce dernier permet comme UltraEdit de reconnaître un grand nombre de langage (LaTeX, C++, html, php, XML, ...).

Eventuellement, vous avez Jaxe (GPL, fonctionne avec Java), Chilkat XML (**Windows**), KxmlEditor (GPL, Linux) ou XXE (fonctionne avec Java) qui interprètent le XML.

Il est important de noter que bien que les fichiers à éditer sont au format XML, il peut être utile de disposer d'un éditeur capable de se positionner à un caractère donné du fichier (par exemple le caractère numéro 64378). Tous les éditeurs texte n'offrent pas cette possibilité et il est dans ce cas souhaitable d'utiliser un éditeur hexadécimal.

Et bien sûr OOo pour lire vos documents !

**Remarque** : Sous Windows, choisir un éditeur de texte évolué (éviter Notepad) afin que les sauts de lignes soient interprétés correctement (issues du monde Unix). Si ce n'est pas le cas, la lecture et l'exploration d'un tel fichier pourront être difficiles et la localisation de l'erreur d'autant plus ardue.

**Remarque 2** : si votre fichier est volumineux, vous pouvez

rencontrer des problèmes selon l'éditeur de fichier.

### 3.2 Sous windows

Télécharger les logiciels.

#### 7-zip

#### Chilkat XML 1.01

“La fonction d’analyse ne valide pas mais Chilkat XML possède une propriété AutoFix qui lorsqu’elle est activée, corrige automatiquement certaines erreurs XML comme la suppression de caractères binaires incorrects, par exemple.”<sup>1</sup>

Il suffit de cliquer sur l'exécutable (fichier dont l'extension est 'exe') et de suivre les étapes.

### 3.3 Sous linux

Télécharger les logiciels

#### Unzip

#### Ark

Site officiel : <http://www.kde.org>

#### FileRoller

Site officiel : <http://fileroller.sourceforge.net/>

#### Nedit

Site officiel : <http://www.nedit.org>

#### Kedit

Site officiel : <http://www.kde.org>

#### K Hex Edit

#### KxmlEditor

Site officiel : <http://kxmleditor.sourceforge.net/>

Selon votre distribution, vous devez peut-être les compiler. Généralement cela se fait de cette manière :

```
$ ./configure
```

```
$ make
```

```
# make install
```

Soit les installer à partir des rpm ou deb :

```
# rpm -ivh nomdulogiciel.rpm
```

**Note** : le # et le \$ signifient que vous êtes en mode console et respectivement logué en root, ou sous votre identifiant.

## 4 Structure d'un fichier OOo

Une fois dézippé, un fichier OOo est constitué de plusieurs fichiers xml. XML est un langage de balisage extensible (Extensible Markup Language, XML).

**Remarque** : ce langage est un sous-ensemble de SGML. Son but est de permettre au SGML générique d'être transmis, reçu et traité sur le Web de la même manière que l'est HTML aujourd'hui. XML a

---

1 Zdnet [Fiche Chilkat XML](#)

été conçu pour être facile à mettre en œuvre et interopérable avec SGML et HTML.<sup>2</sup>Les objectifs de conception de XML sont les suivants :

1. XML devrait pouvoir être utilisé sans difficulté sur Internet ;
2. XML devrait soutenir une grande variété d'applications ;
3. XML devra être compatible avec SGML ;
4. Il devrait être facile d'écrire des programmes traitant les documents XML ;
5. Le nombre d'options dans XML doit être réduit au minimum, idéalement à aucune ;
6. Les documents XML devraient être lisibles par l'homme et raisonnablement clairs ;
7. La conception de XML devrait être préparée rapidement ;
8. La conception de XML sera formelle et concise ;
9. Il devrait être facile de créer des documents XML ;
10. La concision dans le balisage de XML est de peu d'importance.

Ces fichiers sont :

1. content.xml (fichier qui contient les en-têtes et le texte du document)
2. meta.xml (fichier qui contient les méta-informations sur le document, par exemple le nom de l'auteur ou la date de dernière modification)
3. settings.xml (configuration de l'imprimante entre autre)
4. styles.xml (vos styles personnalisés)
5. layout-cache (fichier binaire)
6. META-INF (répertoire contenant le fichier manifest.xml, qui décrit le contenu du zip, c'est-à-dire la liste des fichiers le constituant)
7. Pictures (répertoire présent si vous avez inséré des images)
8. un répertoire pour les macros si le document en contient. Les fichiers sont également au format XML

Les fichiers .xml contiennent des balises ouvrantes et fermantes et des commentaires peuvent être insérés comme suit :<!-- Ceci est un commentaire -->

## 5 Réparation du fichier

### 5.1 Étape 1 : repérage de l'erreur

Lorsque vous ouvrez votre fichier corrompu avec OOo 1.1, celui-ci intègre un analyseur XML qui vous permettra de localiser l'erreur (voir illustration 1 Message d'erreur d'OOo 1.1)

Ainsi l'erreur se situe dans le fichier content.xml à la ligne 2, colonne 68 357.

<sup>2</sup> Tiré de "Langage de balisage extensible (XML) 1.0 - Recommandation du W3C, 10 février 1998", traduit par P. Andries, S. Cuny et F. Yergeau.

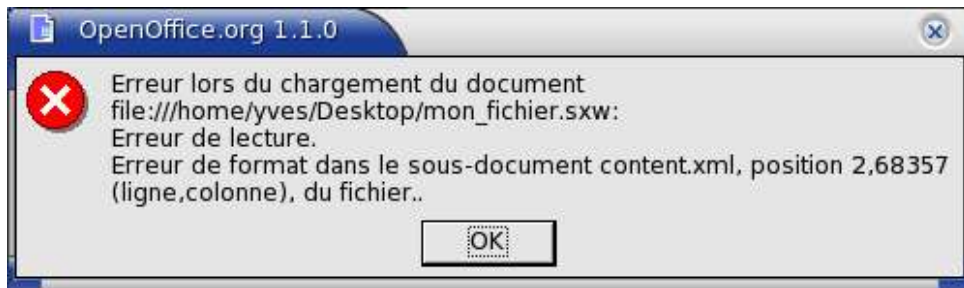


Illustration 1 Message d'erreur d'OOo 1.1

Deux possibilités peuvent expliquer cette erreur. Soit le fichier XML est invalide en soi, c'est-à-dire qu'il ne respecte pas les règles générales du format (caractères interdits dans les balises ou duplication d'attributs dans celles-ci), soit le fichier ne respecte pas la structure attendue d'un document OOo.

Si vous n'avez pas OOo 1.1 (autrement dit une version antérieure), vous pouvez utiliser KxmlEditor sous Linux. Celui-ci vous donnera également un message d'erreur. Cependant il vous faudra ouvrir tous les fichiers dont l'extension est xml après avoir dézippé le fichier (voir section suivante).

Il est à noter que cette méthode ne fonctionne que si l'erreur est du premier type décrit précédemment, à savoir que le fichier n'est pas conforme aux spécifications du format XML. Les erreurs du second type ne sont détectables (hors OOo) qu'en utilisant un lecteur XML capable de validation, qui va alors utiliser des fichiers .dtd (une DTD est une description formelle du contenu acceptable d'un fichier XML pour une application particulière) pour contrôler la cohérence du document. Les fichiers .dtd requis pour cette opération se trouvent dans le sous-répertoire share/dtd/officedocument/1\_0 d'une installation OOo.

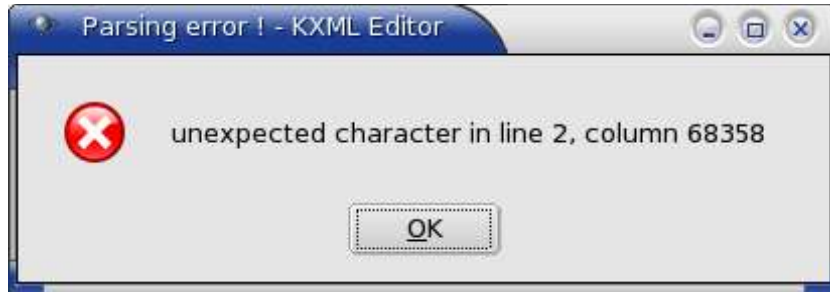


Illustration 2 Message d'erreur de KxmlEditor

## 5.2 Étape 2 : décompression du fichier

Vous devez décompresser votre fichier après en avoir modifié l'extension. Dans l'explorateur de fichier, sélectionnez votre fichier, appuyez sur la touche F2 et changez l'extension « sxw » par « zip ». Par exemple « mon\_fichier.sxw » donne « mon\_fichier.zip ».

On peut également simplement rajouter .zip ce qui donne mon\_fichier.sxw.zip.

Certains logiciels peuvent directement vous proposer la décompression à partir du menu contextuel (clic droit) sous Windows ou Linux (comme 7-Zip par exemple).

## 5.3 Étape 3 : correction de l'erreur et réparation du fichier

Maintenant ouvrez le fichier qui est corrompu avec un éditeur de texte. Allez à l'endroit où se trouve l'erreur. Dans notre exemple elle se trouve à la ligne 2, colonne 68 357. Les fichiers XML générés par OOo sont, pour des raisons de gain de place, en format

compact. Ceci signifie qu'hormis une ligne d'entête, le reste du contenu est sur une seule ligne (ce qui peut compliquer la tâche puisque toutes les erreurs surviennent sur la ligne 2).

Avec K Hex Edit, vous pouvez vous simplifier la vie en tapant ctrl+G ou bien dans le menu "Edition", choisissez "allez à l'offset", tapez "68 357" (pour notre exemple). Il est préférable de temporairement supprimer la première ligne avant d'effectuer cette opération de manière à ce que le caractère dont l'offset a été fourni par le message d'erreur soit celui sur lequel l'éditeur pointe.

Mais attention ! L'erreur se trouve en fait à proximité, normalement juste avant ou juste après. Dans notre exemple, l'erreur était une modification de l'attribut "draw:shadow-color" en "draw:s(adow-color". Lorsque l'erreur est trouvée et que le fichier est réparé, il vous faut le sauvegarder.

#### 5.4 Étape 4 : fermeture du fichier (et compression)

Compressez l'ensemble de vos fichiers en mon\_fichier\_recup.zip (lire la remarque 2 ci-dessous), remplacez l'extension zip par celle de votre fichier d'origine (mon\_fichier\_recup.sxw dans notre exemple).

Ouvrez-le avec OOo, ça marche ? Bravo ! vous avez récupéré votre fichier. Ça ne marche pas ? Est-ce que l'erreur est la même que la première ? Si oui, vous n'avez pas corrigé l'erreur, ou bien en avez rajouté une ! Sinon, votre fichier est corrompu par une autre erreur. Recommencez les étapes !

**Remarque** : si le fichier corrompu est un des fichiers styles.xml, meta.xml ou settings.xml, vous pouvez les effacer directement et compresser vos fichiers ensemble (moins celui que vous avez effacé bien-sûr). Bien entendu vous perdrez soit vos styles, soit votre configuration, soit vos balises méta. Rien de bien grave, mais n'avez-vous pas utilisé ce document pour en faire un autre ? Dans ce cas copiez le fichier manquant de l'un vers l'autre.

**Remarque 2** : Il est important de compresser les fichiers, et non le répertoire dans lequel vous les avez décompressés. Il faut garder la même structure que dans le fichier originel.

## 6 Remarques

Ainsi que mentionné plus haut, les fichiers XML générés par OOo sont optimisés pour prendre moins de place et sont de ce fait plus difficilement lisibles autrement que par un logiciel adapté. Pour forcer OOo à générer des fichiers lisibles à l'oeil nu (c'est-à-dire utilisant une ligne par balise et des tabulations pour souligner la hiérarchie des éléments XML), il est nécessaire de modifier une option accessible par Outils/Options, section Chargement/Sauvegarde, sous-section Général. Sur cette page il faut alors désélectionner l'option « Optimiser la taille pour le format XML ».

## 7 Crédits

**Auteur** : Yves Jacolin

**Remerciements** : Tony Galmiche, Laurent Godart, Cyrille Moureaux, Stéphane Purnelle, Christophe Fichot, Sophie Gautier.

**Intégré par** : Sophie Gautier

**Dernière modification** : 10/12/2003

**Contacts**: Projet Documentation OpenOffice.org <http://fr.openoffice.org>

**Traduction** :

## 8 Licence

### Appendix

#### Public Documentation License Notice

The contents of this Documentation are subject to the Public Documentation License Version 1.0 (the "License"); you may only use this Documentation if you comply with the terms of this License. A copy of the License is available at <http://www.openoffice.org/licenses/PDL.html>.

The Original Documentation is Comment récupérer un document OOo corrompu. The Initial Writer of the Original Documentation is Yves Jacolin Copyright (C) 2003. All Rights Reserved. (Initial Writer contact(s):yjacolin@free.fr).

Contributor(s): \_\_\_\_\_.  
Portions created by \_\_\_\_\_ are Copyright (C) \_\_\_\_\_ [Insert year(s)]. All Rights Reserved.  
(Contributor contact(s): \_\_\_\_\_ [Insert hyperlink/alias]).

NOTE: The text of this **Appendix** may differ slightly from the text of the notices in the files of the Original Documentation. You should use the text of this **Appendix** rather than the text found in the Original Documentation for Your Modifications.