

## TD 1b — Un .odt est une archive ZIP (facultatif)

### Guide détaillé, étape par étape

Le TD ouvre un document LibreOffice, raven.odt, comme une archive ZIP, modifie son texte et sa mise en forme dans un éditeur de texte, puis recomprime l'ensemble : LibreOffice rouvre le document modifié, sans qu'aucun traitement de texte soit intervenu. Il dure une douzaine de minutes, et il est facultatif : il se fait en séance si le temps le permet, ou seul ensuite.

Il demande l'explorateur de fichiers, qui sait extraire et créer une archive ZIP, Notepad++ (ou, à défaut, le Bloc-notes) et LibreOffice Writer pour rouvrir le résultat.

| Étape | Objectif                                          | Ce qu'on fait                             |
|-------|---------------------------------------------------|-------------------------------------------|
| 1     | voir de quoi un fichier .odt est fait             | ouvrir le .odt comme une archive          |
| 2     | repérer le texte et ses styles dans le XML        | lire content.xml                          |
| 3     | modifier un document sans logiciel de bureautique | modifier le texte et la mise en forme     |
| 4     | reconstruire un document que LibreOffice ouvre    | recompresser, et rouvrir dans LibreOffice |

Ce que le TD fait constater est expliqué à la fin du guide, dans « Ce que le TD fait constater ».

## 1 · Ouvrir le .odt comme une archive

**À faire :** copier `depart\raven.odt` dans `travail\` sous le nom `raven.zip`, puis extraire l'archive.

**À obtenir :** un dossier `travail\raven\` qui contient six fichiers, dont `content.xml` et `styles.xml`.

Ouvrir le dossier `info01\cours1\1b_archive_odt\`. Comme au TD 1a, `depart\` contient le fichier fourni, `raven.odt`, et `travail\` reçoit les copies. Les extensions doivent être affichées dans l'explorateur (TD 1a, étape 1).

1. Copier `depart\raven.odt` dans `travail\` (Ctrl + C, puis Ctrl + V dans `travail\`).
2. Renommer la copie `raven.zip` (F2), et confirmer le changement d'extension.
3. Clic droit sur `raven.zip`, « Extraire tout... ». Le dossier proposé est `travail\raven` : le garder, et cliquer sur « Extraire ».

**Vérification :** `travail\raven\` contient :

```
raven\  
├─ META-INF\  
│   └─ manifest.xml  
├─ content.xml  
├─ manifest.rdf  
├─ meta.xml  
├─ mimetype  
└─ styles.xml
```

| Fichier                            | Ce qu'il contient                           |
|------------------------------------|---------------------------------------------|
| <code>mimetype</code>              | une ligne, le type du document              |
| <code>content.xml</code>           | le texte, entouré de balises                |
| <code>styles.xml</code>            | la mise en forme : polices, tailles, titres |
| <code>meta.xml</code>              | l'auteur, les dates, le nombre de mots      |
| <code>META-INF\manifest.xml</code> | la liste de ce que contient l'archive       |
| <code>manifest.rdf</code>          | des métadonnées complémentaires             |

## 2 · Lire content.xml

**À faire :** ouvrir travail\raven\content.xml dans Notepad++ ; y chercher Text\_20\_body et The Raven avec Ctrl + F.

**À obtenir :** le poème, lisible au milieu des balises.

Clic droit sur content.xml, Ouvrir avec, Notepad++. Notepad++ colore les balises, ce qui les rend plus faciles à lire que dans le Bloc-notes. Le fichier fait 21 lignes, dont une de plus de 1 300 caractères : chercher avec Ctrl + F plutôt que de le parcourir. Dans le Bloc-notes, activer le retour automatique à la ligne (menu Affichage) pour voir la ligne longue en entier.

1. Chercher The Raven : c'est le titre du poème.
2. Chercher Text\_20\_body : chaque paragraphe du poème commence par cette balise.

```
<text:p text:style-name="Text_20_body">Once upon a midnight dreary,  
while I pondered, weak and weary, ...
```

**À noter :** ce qui, dans la balise, désigne la mise en forme du paragraphe.

### 3 · Modifier le texte et la mise en forme

**À faire :** trois modifications, deux dans `content.xml`, une dans `styles.xml` ; enregistrer les deux fichiers.

**À obtenir :** les deux fichiers enregistrés, avec leur nom et leur extension d'origine.

| Ce qu'on change          | Où                       | La modification                                                                                                      |
|--------------------------|--------------------------|----------------------------------------------------------------------------------------------------------------------|
| le titre                 | <code>content.xml</code> | remplacer <code>&gt;The Raven&lt;</code> par <code>&gt;Le Corbeau&lt;</code>                                         |
| le style d'un paragraphe | <code>content.xml</code> | sur le premier paragraphe du poème, remplacer <code>Text_20_body</code> par <code>Heading_20_1</code>                |
| la taille des titres     | <code>styles.xml</code>  | dans le style <code>Heading_20_1</code> , passer <code>fo:font-size="115%"</code> à <code>fo:font-size="220%"</code> |

Pour la troisième modification, chercher `style:name="Heading_20_1"` dans `styles.xml` : la ligne suivante contient `fo:font-size="115%"`. Ne changer que cette valeur-là ; les attributs `style:font-size-asian` et `style:font-size-complex` voisins concernent d'autres écritures.

Enregistrer les deux fichiers (Ctrl + S), sans changer leur nom.

**Vérification** : dans `content.xml`, le premier paragraphe commence par `<text:p text:style-name="Heading_20_1">`.

## 4 · Recompresser, et rouvrir dans LibreOffice

**À faire :** compresser les six fichiers de travail\raven\ en raven2.zip ; renommer en raven2.odt ; l'ouvrir.

**À obtenir :** LibreOffice ouvre raven2.odt ; le titre est « Le Corbeau », et le premier paragraphe est en grand.

1. Ouvrir le dossier travail\raven\, et sélectionner tout son contenu (Ctrl + A) : les cinq fichiers et le dossier META-INF.
2. Clic droit sur la sélection, « Compresser dans » (ou « Envoyer vers »), « Dossier compressé (zip) ». Windows crée l'archive dans travail\raven\.
3. Renommer l'archive raven2.zip, puis raven2.odt, et confirmer le changement d'extension.
4. Double-cliquer sur raven2.odt.

**Attention :** compresser le contenu du dossier raven\, depuis l'intérieur, et non le dossier raven\ lui-même. Sinon, l'archive contient raven\content.xml au lieu de content.xml, et LibreOffice refuse de l'ouvrir (« source file could not be loaded »). Dans ce cas, supprimer l'archive et recommencer à l'étape 4.1.

**Vérification :** LibreOffice ouvre le document ; le titre est « Le Corbeau », et le premier paragraphe du poème est un titre, en grand.

## Ce que le TD fait constater

Cette section se lit après avoir fait les étapes.

### Un .odt est une archive de fichiers texte

Un .odt est une archive ZIP qui contient des fichiers XML. Le texte est dans `content.xml`, la mise en forme dans `styles.xml` : c'est la séparation du contenu et de la présentation déjà vue au TD 1a avec HTML et CSS. Les formats .docx, .xlsx et .epub sont construits de la même façon.

Le format est ouvert et documenté : il se manipule avec des outils quelconques, un gestionnaire d'archives et un éditeur de texte, sans le logiciel qui l'a produit. Le TD a changé le texte, le style d'un paragraphe et la taille des titres sans ouvrir de traitement de texte.

Les fichiers qu'il contient sont du texte, mais le .odt lui-même est compressé, donc binaire : c'est pourquoi il se versionne mal, un outil comme git ne pouvant pas en comparer deux versions ligne à ligne (cours 2).

### \_20\_ dans les noms de style

Le paragraphe modifié à l'étape 3 n'a pas changé de texte : seul le nom de son style est passé de `Text_20_body` à `Heading_20_1`. Ces noms ne sont pas faits de trois morceaux : ce sont « Text body » et « Heading 1 », dont l'espace est écrit `_20_`.

Un nom de style est un nom XML, et l'espace y est interdit. Le format OpenDocument écrit donc chaque caractère interdit sous la forme de son code hexadécimal entouré de tirets bas, et l'espace vaut 20 dans la table ASCII du TD 1a. Le nom lisible est rangé à part, dans l'attribut `style:display-name` de `styles.xml` (`style:display-name="Heading 1"`) ; c'est lui que LibreOffice affiche, traduit en français, dans son panneau des styles.

| Dans le panneau des styles de LibreOffice | Dans <code>content.xml</code> |
|-------------------------------------------|-------------------------------|
| Corps de texte                            | <code>Text_20_body</code>     |
| Titre 1                                   | <code>Heading_20_1</code>     |
| Titre 2                                   | <code>Heading_20_2</code>     |

C'est le même principe que le `%20` d'une adresse web, vu au TD 1a : les deux interdisent l'espace, et l'écrivent par son code, chacun avec sa marque.

|                     | Ce qui interdit l'espace | L'espace s'y écrit |
|---------------------|--------------------------|--------------------|
| Une adresse web     | la syntaxe des URL       | <code>%20</code>   |
| Un nom de style ODF | la syntaxe des noms XML  | <code>_20_</code>  |

La règle est fixée par la spécification OpenDocument 1.3, partie 3, aux sections sur `style:name` et `style:display-name` : <https://docs.oasis-open.org/office/OpenDocument/v1.3/OpenDocument-v1.3-part3-schema.html>.

### Les couleurs dans ODF et dans CSS

Pour aller plus loin, une couleur peut s'ajouter au style `Heading_20_1`, dans `styles.xml`, par l'attribut `fo:color`. ODF n'accepte que le code hexadécimal : `fo:color="#c0392b"` colore les titres, `fo:color="red"` est ignoré et le titre reste noir. CSS, au TD 1a, accepte les deux écritures.