



John VanDyk has been innovating with information technology for more than 20 years. [Read more...](#)

[John's Blog](#)  
[Index of All Posts](#)  
[Index by Topic](#)

### John's Senu Book



This Senu book gets you up to speed with [Senu monitoring and metrics!](#)

### John's Drupal Books

## Creating Open Packaging Conventions files with zip on OS X 10.6

Submitted by John on Fri, 2011-02-04 15:07

The .docx format used by Microsoft is actually a [collection of XML files](#) (and other files) and can be opened with zip.

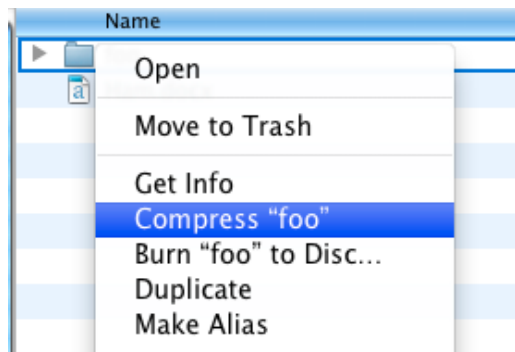
For example, on OS X you can use unzip at the command line to look at the OPC file like this (we'll unzip it into a directory called foo):

```
$ ls -l
-rw-r--r--@ 1 john  staff   13678 Feb  3 11:52 Ham.docx

$ unzip Ham.docx -d foo
Archive:  Ham.docx
  inflating: foo/[Content_Types].xml
  inflating: foo/_rels/.rels
  inflating: foo/word/_rels/document.xml.rels
  inflating: foo/word/document.xml
  inflating: foo/word/theme/theme1.xml
  inflating: foo/word/settings.xml
  inflating: foo/word/webSettings.xml
  inflating: foo/word/stylesWithEffects.xml
  inflating: foo/docProps/core.xml
  inflating: foo/word/styles.xml
  inflating: foo/word/fontTable.xml
  inflating: foo/docProps/app.xml
```

Now you've extracted the .docx file into a new directory called foo. You can go in there and poke around at the .xml files.

But what happens when you want to put the whole thing back together? "No problem," you might think. "I'll just use the handy-dandy Compress service to do that."



Then just rename it to foo.docx and we're all done! But...wait.



*Pro Drupal  
Development,  
Second  
Edition  
(and now  
Third  
Edition)*



Now you're ready to take the gloves off. "I'll just do it at the command line!"

```
$ zip -r bar foo/*
  adding: foo/[Content_Types].xml (deflated 75%)
  adding: foo/_rels/ (stored 0%)
  adding: foo/_rels/.rels (deflated 61%)
  adding: foo/docProps/ (stored 0%)
  adding: foo/docProps/app.xml (deflated 48%)
  adding: foo/docProps/core.xml (deflated 52%)
  adding: foo/word/ (stored 0%)
  adding: foo/word/_rels/ (stored 0%)
  adding: foo/word/_rels/document.xml.rels (deflated 71%)
  adding: foo/word/document.xml (deflated 66%)
  adding: foo/word/fontTable.xml (deflated 77%)
  adding: foo/word/settings.xml (deflated 62%)
  adding: foo/word/styles.xml (deflated 89%)
  adding: foo/word/stylesWithEffects.xml (deflated 89%)
  adding: foo/word/theme/ (stored 0%)
  adding: foo/word/theme/theme1.xml (deflated 79%)
  adding: foo/word/webSettings.xml (deflated 42%)
$ mv bar.zip bar.docx
$ open bar.docx
```

Nope!



Aargh! So what's the secret? Here's the secret: **cd into the directory first, then build your zip archive outside the directory.** Here we cd

into foo and make bar.zip in the directory above, then rename it to bar.docx.

```
$ cd foo
$ zip -r ../bar *
  adding: [Content_Types].xml (deflated 75%)
  adding: _rels/ (stored 0%)
  adding: _rels/.rels (deflated 61%)
  adding: docProps/ (stored 0%)
  adding: docProps/app.xml (deflated 48%)
  adding: docProps/core.xml (deflated 52%)
  adding: word/ (stored 0%)
  adding: word/_rels/ (stored 0%)
  adding: word/_rels/document.xml.rels (deflated 71%)
  adding: word/document.xml (deflated 66%)
  adding: word/fontTable.xml (deflated 77%)
  adding: word/settings.xml (deflated 62%)
  adding: word/styles.xml (deflated 89%)
  adding: word/stylesWithEffects.xml (deflated 89%)
  adding: word/theme/ (stored 0%)
  adding: word/theme/theme1.xml (deflated 79%)
  adding: word/webSettings.xml (deflated 42%)
$ cd ..
$ mv bar.zip bar.docx
$ open bar.docx
```

Presto! The .docx file opens up in Word with no corruption.

#### References:

- [Open Packaging Format creator/manager for OS X](#)
- [docx "File is corrupt" error in Microsoft Word](#)

Topic: [Mac OS X](#) [Microsoft Office](#)

[Log in](#) to post comments

## Comments

### [Additional query](#)

[Permalink](#) Submitted by vinay (not verified) on Mon, 2012-01-16 08:06.

Hi John,

I just got introduced to OpenXML, this is great information for beginners like me who like to experiment with the structure.

I've an additional question, is there any specific compression level we need to use with the DEFLATE compression mode. I mean, if I am correct, the specification does not talk about the compression level, so does this mean that such applications(MS Word 2010) should support all available levels of compression for this mode(DEFLATE)?

I am not sure whether you are going to look at this, just hoping you might have some information on this if you take a look at this.

Thanks,  
Vinay

[Log in](#) to post comments

[Awesome! Thank you.](#)

[Permalink](#) Submitted by Anonymous (not verified) on Mon, 2012-11-05 17:51.

Awesome! Thank you.

[Log in](#) to post comments