

PDFTron DocPub User Manual

Version 1.1

LEGAL STATEMENT AND COPYRIGHT NOTICE

PDFTron DocPub™ Command-Line Application User Manual
 Part number: PDFTRON-1-DocPubCMD
 Last Updated: June 26, 2013

© 2001 – 2013 PDFTron Systems, Inc. All Rights Reserved.

All information contained herein is the property of PDFTron Systems, Inc. ("PDFTron"). No part of this publication (whether in hardcopy or electronic form) may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of PDFTron Systems, Inc.

The information in this publication is provided for informational use only, is subject to change without notice, and should not be construed as a commitment by PDFTron. PDFTron assumes no responsibility or liability for any loss or damage that may arise from the use of any information in this publication. The software described in this user manual is furnished under License (enclosed in the software package) and may only be used or copied in accordance with the terms of that License.

PDFTron and the names of PDFTron products referenced herein are either trademarks and/or service marks and/or registered trademarks of PDFTron Systems, Inc. PDFTron, PDFNet SDK, WebViewer, PDF/A Manager, PDF2Image, PDF2SVG, PDF2Text, PDF2XPS, DocPub, XPSCovert, PDFTron PDFSecure, PDF PageMaster, CosEdit, PDF2Image SDK, PDF2SVG SDK, PDF2Text SDK, XPSCovert SDK, PDFSecure SDK, PDF PageMaster SDK and associated Logos are either trademarks and/or service marks and/or registered trademarks of PDFTron Systems, Inc.

Any other brand or product names mentioned in this publication are the registered trademarks or trademarks of their respective holders. Mention of a product in this document does not necessarily imply endorsement of the product.

TABLE OF CONTENTS

LEGAL STATEMENT AND COPYRIGHT NOTICE	2
1. Introduction	5
1.1 An Introduction to PDFTron DocPub	5
1.1.1 Key Functions	5
1.1.2 Common Use Case Scenarios	5
1.1.3 Operating Systems Supported	5
1.1.4 System Requirements	5
2. Installing and Uninstalling DocPub	6
2.1 DocPub Installation	6
2.2 Demo Version	6
2.3 Registering DocPub on your Server(s)	6
2.4 Uninstalling DocPub	6
3. Overview	7
3.1 Basic Syntax	7
3.2 Command-Line Summary	7
3.3 Basic Usage	11
3.3.1 How do I save converted files in a given folder?	11
3.3.2 How can I control the output name for the converted files?	11
3.3.3 How do I specify which files to convert?	11
3.3.4 How do I batch convert files?	11
3.3.5 How do I convert a password protected PDF?	11
3.3.6 How do I create an encrypted XOD file?	12
3.3.7 What quality can I expect from the output XOD document?	12
3.4 General Usage Examples	13
Example 1. The simplest command line: Convert a file to PDF or XOD format.	13
Example 2. Convert XOD file to PDF.	13
Example 3. Batch convert all files to XOD.	13
3.5 Batch Processing and the Use of Wildcards	13
3.6 Exit Codes	14
4. Frequently Asked Questions	15
4.1 General FAQ	15
4.1.1 Is DocPub available as a toolkit (SDK) for integration with third party applications?	15
4.1.2 What is XPS?	15
4.1.3 What is XOD?	16
4.1.4 Does DocPub have any dependencies on third party components/software?	16
4.1.5 How do I convert Microsoft Office formats on Windows?	16
4.1.6 How do I convert other file formats not listed on Windows?	16

4

3. Overview

PDFTron DocPub is a command-line application designed to convert documents to PDF, XPS, XOD, HTML, or EPUB format files, while presenting several options to control the conversion process. This section covers the basic usage of DocPub explaining all of the available options.

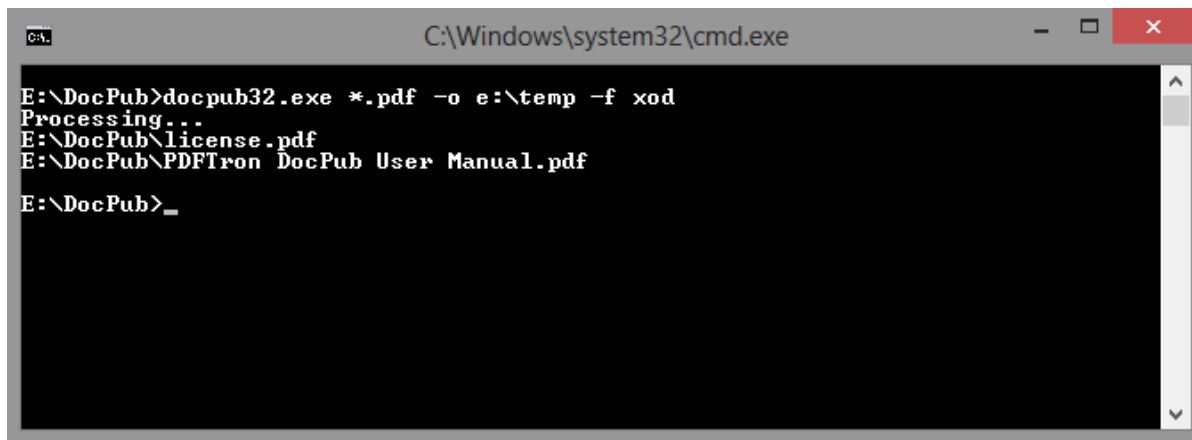


Figure 3.0 DocPub Command-line Application.

3.1 Basic Syntax

The basic command-line syntax is:

```
DocPub [options] file1 file2 folder1 file3 ...
```

3.2 Command-Line Summary

The following is a list of available command-line options for DocPub:

Option	Parameter	Description
-h or --help		Print a listing of available options.
-v or --version		Print the version information.
-o or --outdir	-o myfolder -o c:\myfolder	The output directory. The folder can be relative to the current working folder. If the folder does not exist DocPub will attempt to create the required path. If this parameter is not specified all files will be saved relative to the current working folder.
-s or --subfolders		Process all sub-directories for every directory specified in the argument list. By default, sub-directories are not processed.

8

--enhance_thin_lines	--enhance_thin_lines false (Currently XPS and XOD conversions only).	If converting from PDF makes thin lines appear thicker in the output document. This is useful because some thin lines can appear invisible in XPS/XOD, but not in the original. Default is true.
--printmode	(Currently XOD conversions only).	Converts annotations in print mode. This option can be used to convert 'Print Only' annotations and to hide 'Screen Only' annotations.
--verb	--verb 2	Set the verbosity level. Valid parameter values are 0, 1, and 2. The higher number results in more feedback. The default is 1.
--op	--op on (Currently XPS and XOD conversions only).	Specifies if overprint simulation is used. Overprint is a device dependent feature and the results will vary depending on the output color space and supported colorants (i.e. CMYK, CMYK+spot, RGB, etc). Possible values are: off, on, and pdfx (overprint is turned on only for PDF/X files). The default is pdfx.
--nothumbs	(Currently XOD conversions only).	Exclude thumbnails from the document.
--thumbsize	--thumbsize 1000 (Currently XOD conversions only).	The width and height of a square in which all thumbnails will be contained. The default value is 400 pixels.
--console_out	(Currently XOD conversions only).	Exactly one input file may be specified. Write the output to the console rather than a file which allows output files to be streamed to the consumer.
--max_image_pixels	--max_image_pixels 5000000 (Currently XOD, PDF (with flattening), HTML and EPUB conversions only).	Specifies the maximum image size in pixels. Default is 2 Megapixels.
--scale	--scale 2.5 (Currently HTML and EPUB conversions only).	Adjusts the generated html page size. Default is 1.
--epub_reuse_cover	(EPUB conversions only).	The first EPUB page uses the cover image, and will have no selectable text.
--flatten	--flatten off (Currently XOD and PDF conversions only).	Used to reduce some PDF content to a simple background image. While flattening tries to preserve vector text, some text might be flattened, especially in simple mode. Options are: OFF, disable flattening. FAST, will convert content deemed complex to a background image, while trying to preserve vector text, and keeping file size down. SIMPLE, reduces the PDF to two layers; a RGB background image

10

Note: DocPub supports all standard security options available in PDF, including 40 and 128 bit RC4 encryption, Crypt filters, and AES (Advanced Encryption Standard) encryption.

To create a XOD file where the individual parts are encrypted with AES 128.

The generated XOD file can be used with one of the PDFTron WebViewer's to decrypt using the password 'secret'.

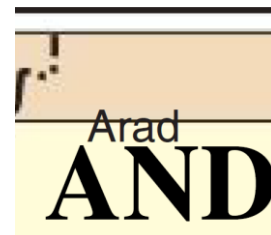
Since DocPub always attempts to maintain the original document appearance, the vast majority of output files will successfully preserve the appearance and quality of the original documents. Occasionally, there will be elements that can't be accurately converted. In these cases DocPub will by default render the necessary elements at 150 dpi or at the resolution given by the `--dpi` parameter. Note though that there is a cap on how large an image can be, controlled by `--max_image_pixels` parameter, which by default is 2 megapixels. For this example we will set the cap at 5 megapixels:

would render those elements at 200 dots per inch.

```
DocPub -f xod --norender doc.pdf
```

Finally, there is the option to 'flatten' content. The default when converting to XOD is 'fast' flatten mode, which try to generate a PDF that renders faster on limited speed/memory devices. This can include changing color spaces, to converting complex paths to an image. You can also turn this off to help ensure that as much content is preserved as is.

If you want to flatten, there is a way to control how much is flattened, or not, by adjusting the flatten threshold. The images below demonstrate how `--flatten_threshold` can affect flattening. Notice that the large text is never flattened, only the text occluded by the rectangle.



```
--flatten_threshold keep_all
```



```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

12 0 obj
<<
  /F2 1 Tf stream
  /Length 13
  /Filter /FlateDecode
  /Size 268
>>
endobj

```

```

:failed_create_err
echo Failed to create a directory.
goto exit

:othererror
echo An error encountered during processing.
goto exit

:exit

```

4. Frequently Asked Questions

4.1 General FAQ

4.1.1 Is DocPub available as a toolkit (SDK) for integration with third party applications?

For developers who are looking for more control over the conversion process than a command-line utility can provide, developers can also license DocPub functionality as part of PDFNet SDK and the applicable add-ons for the required conversions (such as WebViewer Add-on for conversion to XOD, or the PDF to HTML Add-on, etc.). PDFNet SDK is an industry-leading, high-quality document core technology platform powering mobile, server, desktop, web, and cloud-based apps. Using the PDFNet library, developers can create powerful PDF document solutions and applications that can convert, generate, manipulate, optimize, print, view, and markup PDF documents without any third-party software dependencies.

PDFNet SDK is available as a .NET component, and as a cross-platform C/C++, Java, Ruby, Python, PHP and Objective-C, PDF library, that is available on a wide range of platforms (i.e. Windows, Linux, Mac OS X, Android, iOS, Windows 8/RT, etc).

Alternatively, developers can also subscribe to [PDFTron Web Services \(PWS\)](#), a pay-as-you-go platform for licensing the DocPub, WebViewer or PDFNet conversion capabilities. PWS enables developers to sign up to a Cloud account, with the document conversion hosted by PDFTron, or to an On-Premise account with the conversion processing performed on their own servers.

For more details, please visit PDFTron's website at <http://www.pdftron.com> or contact a PDFTron representative via info@pdftron.com or sales@pdftron.com.

4.1.2 What is XPS?

"XPS" stands for "XML Paper Specification" and is a new document format as well as the native print spooler format in Microsoft's Windows Vista® and Windows® 7. The XPS document format consists of XML markup that defines the layout of a document and the visual appearance of each page along with rendering rules for distributing, archiving, rendering, processing and printing the documents. Just like PDF, the XPS document format enables users to view, print, and archive any type of documents without the original program that created them and without loss of fidelity.

To find out more about XPS, please visit Microsoft's website at:
<http://www.microsoft.com/whdc/xps/default.mspx>

4.1.3 What is XOD?

XOD is a web optimized Open XPS format developed for hosting documents online. It facilitates fast viewing of large size files while ensuring high quality display.

XOD documents can be stored locally as well, and can also be viewed by any XPS viewer. Any PDF, XPS, or any printable document, can be converted to XOD format with PDFTron's command-line DocPub utility.

To find out more about XOD, please visit PDFTron®'s website at:
<http://www.pdftron.com>

4.1.4 Does DocPub have any dependencies on third party components/software?

For most conversions, including PDF, SVG, XPS, XOD, PNG, JPG, etc, DocPub is a completely stand-alone application and does not include any dependencies on third-party components or software. For some other types of conversion, external programs are required, see below.

4.1.5 How do I convert Microsoft Office formats on Windows?

To convert Word, Excel, PowerPoint, Publisher and Visio files, you need either the respective Microsoft application (which we will call Office for short), or another program that can print them, installed on the computer doing the conversion. For best performance, and results, install Microsoft Office. Otherwise, see 4.1.6 below.

4.1.6 How do I convert other file formats not listed on Windows?

Finally, to convert any (proprietary or non-standard) document on Windows, you need to have an application that can print the file type. See the link below for more information regarding custom conversions:

<https://groups.google.com/forum/?fromgroups#!searchin/pdfnet-sdk/dwg/pdfnet-sdk/ovStpU-HGYk/s-PMo7sGMT8J>

4.2 Common Troubleshooting Issues

4.2.1 Why do conversions stop working after entering API key and secret (Error Code 4)?

DocPub returns with error code 4 if connection with PDFTron servers wasn't established. To help identifying what's the issue you can run DocPub with a "--verb 2" option, which would print additional information to the command line. These are the error messages you will see if something goes wrong:

- 1) Can't establish a connection due networking error; Check your connectivity to the internet and firewall settings.
- 2) Credentials provided for authentication are incorrect. Make sure you are subscribed to pay as you go plan. You can check your subscription plan and credentials at api.pdftron.com.
- 3) Server is not responding. If the error persists contact support@pdftron.com.
- 4) Server is not recognizing a conversion. Please contact support@pdftron.com.

4.2.2 Why is a white space separating neighboring pictures?

In some cases, XPS viewers that support anti-aliased rendering produce line/space artifacts at neighboring picture elements (e.g. for image tiles or polygons sharing common edges). These artifacts are not a byproduct of DocPub conversion, but are produced due to anti-aliased rendering in the XPS viewer. The same issue applies to the input PDF document however the current PDF

4.2.3 Why are some pages rasterized?

4.2.4 Why are some fonts in PDF not rendered consistently?

4.2.5 Why is the image quality so low?

4.2.6 Why does the Silverlight plugin crash?

4.2.7 Why do shadows (masks) not appear correctly in Silverlight?

4.2.8 How do I stream XOD conversions?

17

