

PDFTron DocPub User Manual

Version 1.1

www.pdftron.com

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

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, XPSConvert, PDFTron PDFSecure, PDF PageMaster, CosEdit, PDF2Image SDK, PDF2SVG SDK, PDF2Text SDK, XPSConvert 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

1. Introduction

1.1 An Introduction to PDFTron DocPub

PDFTron's DocPub is an easy-to-use, multi-platform executable that provides users with a high-quality and efficient way to convert wide range of documents to PDF, XPS, XOD, HTML and EPUB formats.

Like other PDFTron products, DocPub does not rely on any third party components. The technology is suitable for use in high-throughput server environments.

1.1.1 Key Functions

- Fast, high-quality conversion from practically any document to PDF (Portable Document Format), XPS (XML Paper Specification) or XOD (web optimized OpenXPS standard), HTML or EPUB format, that maintains the original document quality and layout and preserves hyperlinks, colors and fonts.
- Create XOD documents for use with PDFTron WebViewer (HTML5, Silverlight, and Flash), including streaming conversions on-demand.
- The conversion process preserves the original document's meta-data as well as other non-graphical information such as bookmarks, logical structure, and articles.
- Thumbnail generation option for fast navigation through multi-page documents.
- Automatic repair of broken PDF documents.
- Support for all versions of the PDF Language Standard, including Acrobat 9 documents as well as ISO PDF (ISO 32000).
- Batch conversion.
- Wild card and subfolder processing.
- 100% conversion accuracy.
- Optional flattening to create PDF's that render faster on mobile devices with lower memory and speeds.

1.1.2 Common Use Case Scenarios

- Developers may want to use DocPub to quickly add batch conversion support to any application or workflow that requires PDF, XPS, XOD, HTML or EPUB as output formats.
- Server-based, on-demand conversion of documents to PDF, XPS, XOD, HTML and EPUB files.
- Stream XOD conversions on-demand to a WebViewer client.
- Batch processing of large collections of files with the same conversion options.
- Extending existing applications to take advantage of the new XPS Print API and XPS print path available in Windows 8, 7, and Vista.

1.1.3 Operating Systems Supported

- Windows, Linux and Mac.

1.1.4 System Requirements

- At least 12 MB of free disk space.
- Memory requirement is dependent on source document being converted.

2. Installing and Uninstalling DocPub

2.1 DocPub Installation

DocPub Command-line Application is supplied as a download from a distributor or directly from www.pdftron.com. The release is packaged as a .zip file (DocPub.zip). To install the software, simply unzip the archive in the desired location and make sure to preserve the directory/folder structure during this process. To register the software, copy the license file provided to you into the “DocPub” folder.

2.2 Demo Version

If you wish to evaluate the product, you can download the demo version of the product without any serial number or license key. The limitation of the evaluation version is that all output pages will have demo stamps.

2.3 Registering DocPub on your Server(s)

When you are ready to switch to production, and you want to get rid of the watermarking feature of the demo version, follow these steps to register DocPub and sign up for an On-Premise PDFTron Web Services (PWS) pay-as-you-go account.

1. Go to <https://api.pdftron.com>
2. Click on the “Sign up” button and fill in the registration details.
3. Shortly after the registration you will receive a confirmation email with an activation link and your API ID and API Secret, which is required to connect to Cloud API from your app.
4. To activate subscription, log into your account and proceed to “Subscription” page to select your account plan.
5. Select the plan that meets your projected conversion volume and fill in your billing info.
6. Open ‘docpub.lic’ in a text editor.
7. Fill-in API_ID and API_Secret that you obtained via email (in step 3). You can also find the same information under the ‘Account’ section in your PDFTron Cloud account. For example:

```

#-----
# License Information:
#-----
API_ID = your_API_ID_here
API_Secret = your_API_Secret_here
#-----
# To register the Software, save the attachment in the 'docpub' folder.
# When extracting the archive, please make sure to preserve the directory (folder)
  
```

8. Once registered, any conversions you perform will be free from demo stamps and will be reported in your PDFTron Cloud management console (<https://api.pdftron.com/v2/console>).

2.4 Uninstalling DocPub

To remove DocPub from your computer, simply delete the “DocPub” folder.

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.

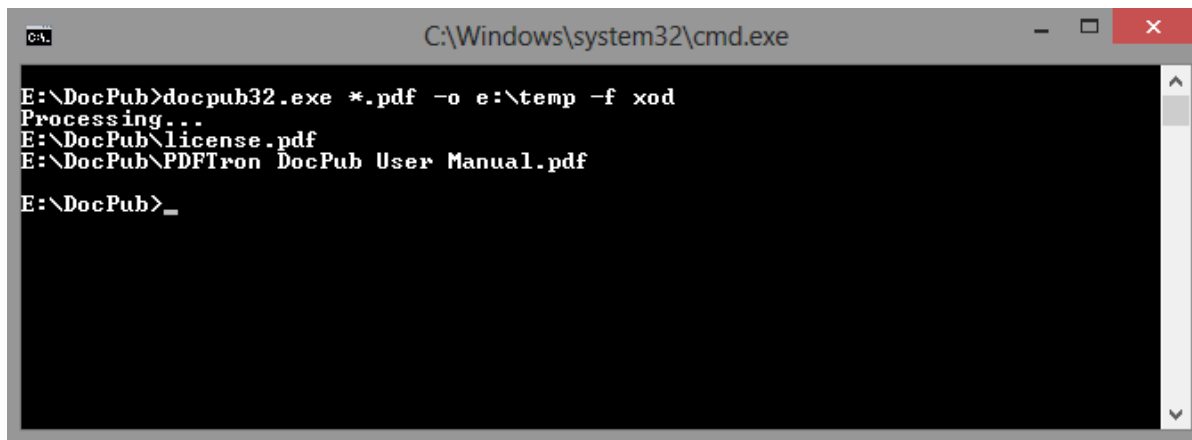


Figure 3.0 DocPub Command-line Application.

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.

-u or --uninstall	(This option is currently valid only in Windows environment).	Uninstalls PDFTron PDFNet virtual printer.
-fname	-fname c:\myfolder\myfile.pdf	The explicit name for the converted file. The parameter is used only when converting one file at a time. The output file will be stored in the 'outdir' folder.

--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	Specifies the maximum image size in pixels. Default is 2 Megapixels.

		layer and an overlapping vector text layer. By default, FAST is used for conversion to .xod and OFF is used for conversion to .pdf.
--flatten_threshold	--flatten_threshold keep_most For XOD and PDF conversions that use --flatten [simple fast].	Used to control how precise or relaxed text flattening is. Some text can be preserved (not flattened to image) at the expense that the output might not be exactly the same as the input. VERY_STRICT, render (flatten) any text that is clipped or occluded. STRICT, render text that are marginally clipped or occluded. DEFAULT, render text that are somewhat clipped or occluded. KEEP_MOST, only render text that are seriously clipped or occluded. KEEP_ALL, only render text that are completely occluded, or used as a clipping path.
--prefer_jpg	--prefer_jpg false (XOD, HTML, EPUB and PDF (with flattening) conversions only).	Where possible, output JPG files rather than PNG. This will apply to both thumbnails and document images. Default is true.
--external_parts	(Currently XOD conversions only).	Output XOD as a collection of loose files rather than a zip archive. This option should be used when using the external part retriever in Webviewer.
--xod_encrypt_password	--xod_encrypt_password secret (Currently XOD conversions only).	Encrypt XOD parts with AES 128 encryption using the supplied password.
--mask_workaround	(Currently XOD conversions only).	Enable rendering of pages containing opacity masks. This option is used as a workaround to a bug in Silverlight where opacity masks are transformed incorrectly.
--silverlight_text_workaround	(Currently XOD conversions only).	Outputs rotated text as paths. This option is used as a workaround to a bug in Silverlight where pages with rotated text could cause the plugin to crash.

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.

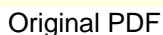
```
DocPub -f xod --norender doc.pdf
```

This option will preserve resolution independent properties of input documents (including fonts, paths, and shadings) as well as text selection and extraction capability in the processing software.

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.

```
DocPub -f xod --flatten off doc.pdf
```

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.



3.4 General Usage Examples

Example 1. The simplest command line: Convert a file to PDF or XOD format.

Notes:

- Converts 'my.doc' to 'my.pdf' located in the current working folder.

```
DocPub my.doc
```

- Converts 'my.doc' to 'my.xod' located in the current working folder.

```
DocPub -f xod my.doc
```

Example 2. Convert XOD file to PDF.

Notes:

- The '-o' (or --output) parameter is used to specify the output folder. If this option was not specified, all images would be stored in the current working folder.
- The '-f pdf' specifies that the output should be an PDF file.
- The '--verb' option instructs DocPub to output more feedback in the console window.

```
DocPub -f pdf --verb 2 -o my_output_folder my.xod
```

Example 3. Batch convert all files to XOD.

Notes:

- The '-f xod' option instructs DocPub to convert all files in dir1 folder and its subfolders to XPS format.
- The '--subfolders' option is used to recursively process all PDF documents stored in subfolders of dir1 and dir2.

```
DocPub -f xod --subfolders dir1 dir2
```

3.5 Batch Processing and the Use of Wildcards

DocPub supports processing of multiple input documents in the same run. For example, it is possible to specify multiple PDF folders and DocPub will automatically process all documents matching a given file extension. For example, the following command-line will process all PDF documents in folders 'test1' and 'test2'

```
c:\> DocPub --extension pdf -o c:/output_folder c:/test1 c:/test2
```

Wildcard characters can also be used to process multiple input files.

For example, if a directory contains the following PDF documents:

```

C:\test1 >dir
Directory of C:\test1
01/04/2007  03:35 PM    <DIR>        .
01/04/2007  03:35 PM    <DIR>        ..

```

```
05/21/2004 02:27 PM A1.pdf
05/03/2005 09:38 AM A2.pdf
05/20/2003 08:46 AM B1.pdf
05/15/2003 12:50 PM B2.pdf
```

To process all PDF documents in this folder and convert them to XOD format, you could specify:

```
c:\>DocPub -f xod -o c:/output_folder c:/test1/*.pdf
```

To process all PDF documents starting with 'A', you could specify:

```
DocPub -f xod -o c:/output_folder c:/test1/A*.pdf
```

Or to process all PDF documents ending with '1' and convert them to XOD documents, you could specify:

```
DocPub -f xod -o c:/output_folder c:/test1/*1.pdf
```

You can use either of the two standard wildcards — the question mark (?) and the asterisk (*) — to specify filename and path arguments on the command line.

The wildcards are expanded in the same manner as operating system commands. (Please refer to your operating system user's guide if you are unfamiliar with wildcards). Enclosing an argument in double quotation marks (" ") suppresses the wildcard expansion. Within quoted arguments, you can represent quotation marks literally by preceding the double-quotation-mark character with a backslash (\). If no matches are found for the wildcard argument, the argument is passed literally.

3.6 Exit Codes

To provide additional feedback, DocPub returns exit codes after completing processing. The exit codes can be used to provide user feedback, for logging etc. This is particularly important for applications running in an unattended environment.

The following table lists possible exit codes and their description:

Exit Code	Description
0	All files converted successfully
1	Unspecified error
2	Bad license key
3	Failed to create or write the output file or directory
4	Failed to connect to server

All codes other than '0' indicate that there was an error during the conversion process.

To get detailed information on an error, set the `--verb` parameter to 2.

The following illustrates a sample Windows batch script that processes exit codes:

```
@echo off
rem convert all files in 'data' folder

DocPub ./data
if errorlevel 1 goto other_error
if errorlevel 3 goto failed_create_err
if errorlevel 0 goto exit
```

```

: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

5.1 Reporting Problems

When submitting a problem you will be asked to provide the following information:

- Contact details
- Product and Version of the product
- Detailed description of problem
- Problem file(s)
- Whether you have an AMS (Annual Maintenance Subscription)
- Any other information that may be related

5.2 Contact Information

Tel: 1-604-730-8989
Fax: 1-604-676-2477

Web site: www.pdftron.com

PDFNet SDK Forum: <http://groups.google.com/forum/#!forum/pdfnet-sdk>

WebViewer Forum: <http://groups.google.com/group/pdfnet-webviewer>

Email Contacts:

Sales & Licensing: sales@pdftron.com

General Business Inquiries: info@pdftron.com

Product Support: support@pdftron.com

Professional Services: services@pdftron.com

Website related questions: webmaster@pdftron.com

Press & News: press@pdftron.com