

# A Brief Guide to Scholarly Self-Republishing

by Jeff Hecht

We've heard a lot about self-publishing new books. But you also can self-republish previously published books that have gone out of print. This includes both textbooks, which you may want to charge for, and scholarly research, which you may want to make available for free. I have succeeded in self-republishing a textbook, *Understanding Fiber Optics*, a investigative book with extensive referencing, and a young adult science book. This article explains the process, warns about potential pitfalls, and describes some useful tricks for scholarly self-republishing. My goal is to make the process as simple and easy as possible, not to produce a new edition. The bottom line is that you can do a lot with self-republishing, but that there are limits to what is practical.

This article describes the basics. I have self-published a more extensive 13,000+ word electronic version of this guide that covers all types of books in more detail. Called "Adventures in Self-Republishing". It's available as a PDF file for \$2.99 from Payhip at <https://payhip.com/b/AX9h> or in e-reader format from your favorite on-line retailer.

## The Starting Point - The Old Book

Your first step is to check your book's publication status and your contract with the publisher. If the book is out of print and the publisher has not formally reverted rights to you, request reversion of rights. If the publisher has gone out of business or been sold, you need to track down who -- if anyone -- holds publication rights. It may not be obvious, so search the web to see if any publishers are offering your book. I was amazed to that a major publisher had my investigative book in its data base, despite the fact that the original publisher had reverted rights to me many years ago. Mistakes can happen; in that case, the previous publisher evidently did not check what rights it still held when it sold publication rights to an entire line of books to the major publisher.

Once you resolve the rights question, your course depends on what form you have the book in. You almost certainly will have print copies of the final version, and they can be scanned for reproduction. In principle, having a copy of an electronic edition should make republishing easier. However, publishers tend to be reluctant to revert rights if they have electronic version available, and many electronic versions have copy protection that may prevent you from using it in republishing. Digital copies of your original manuscript will not include any editorial or layout changes, and normally do not include illustrations, so at best they require further processing. If you did fully typeset your original book, you may be able to use that directly in republishing, but you may not be able to convert it fully to e-book format, depending on the complexity of your book, as described below.

## Book Scanning and Scannos

Printed books can be scanned, which is a job for a professional service that has special equipment to scan pages quickly, carefully, and cleanly. I used BlueLeaf Book Scanning

(blueleaf-book-scanning.com). They generate PDF page images and run them through OCR (optical character recognition) software that recognizes the printed characters, generating both a plain text file and a separate Word file that includes fonts and formatting such as bold, italics, superscripts, and subscripts. You also get a searchable PDF (which includes an embedded text file from the OCR process). Options include a PDF configured for online publishing, a file of images in the book, and a version of the text formatted for e-readers. The extras are not expensive, and are worth getting for possible future use.

Although OCR is remarkably good in many ways, errors that I call “scannos” are inevitable. Text recognition is good, but some letter pairs cause problems, notably “rn” which the OCR often mistook for the letter “m,” converting “burn” into “bum.” A spellchecker can catch some spelling errors, but won’t spot the wrong font or legitimate words, like “bum.” Errors in the font and type size and style are much more common. The OCR built into Adobe Acrobat Pro made quite a mess of the numbered end notes in my investigative book; superscripts became 6-point type raised three points above the baseline, which caused serious problems in formatting e-books for electronic readers.

Illustrations are a problem. Black-and-white scanning gives clean sharp line art and letters, but makes an ugly mess of photos or other grayscale art. Scanning in grayscale reproduces photos better, but blurs line art and some lettering. Look carefully at your book before you choose which to use. In theory you can scan photos separately and paste them into the black-and-white scan, but that will cost you time and/or money. I have not tried color scanning, but any scanning inevitably degrades image quality to some extent.

### **Self-Republishing Options**

The three main options for self-republishing are print-on-demand (POD) of paper copies, electronic PDF files, and e-books in the HTML-like code used for electronic readers.

*POD books* are paperbacks printed on demand from image files, usually stored in PDF format. Many conventional publishers use POD to keep backlist books in print. These are essentially photocopies of the original printed book and may be taken from original digital layout files. But as in any photocopying, the print quality suffers, and if the book has been scanned, sharpness of lines and letters often suffers. The costs of printing and distribution also cut into your share of sales income, and make free distribution impossible unless you pay for the printing yourself. POD has been around since about 2000, and is a well-established part of the print publishing and distribution system.

*PDF E-books* are electronic files generated by printing a word-processing document to a PDF file or by scanning pages to generate images. PDFs are widely used for distribution of scholarly papers, but little used for publishing of non-academic books. PDFs generated directly from word-processing files, including specialty typesetting software such as LaTeX are sharp and clean and tend to be modest in size. PDFs generated by scanners are much larger files because they display each page as an image. Commercial scanning services run scanned page images through OCR software to produce text, but the OCR text is prone to scannos. A commercial

drawback is that PDF books generally are not sold through major online bookstores, but they are available through other services such as Payhip, and you can distribute them freely from your own web site from services that distribute PDF ebooks.

*E-reader format books* are electronic files formatted to be read on e-book readers like Kindle and Nook. Like the HTML code used on web pages, they combine text, formatting codes, and images and flow the combination into the available screen space. The major main formats are MOBI (used by Kindle) and ePUB (used in most other e-readers). Books in those format benefit from an extensive marketing and distribution infrastructure. However, accessing that infrastructure requires converting the scanned book into a word-processing file, and that can be a major project if your book includes illustrations, references, or formatted text. E-reader books have become an important market in recent years, but their success has been uneven, with sales high in genre fiction but low in illustrated, professional and scholarly books.

### **Processing, Pitfalls and Tradeoffs**

The easiest way to self-publish a scanned book or one you already have in digital PDF format is to post the PDF at a service that specializes in electronic distribution or on your own site. Sometimes you can post the scanned copy without change, but you may have to make minor changes, such as removing the name of the previous publisher from the copyright page.

Directly editing PDF pages is difficult, but you can make changes on individual pages from the text version in a word processor, then print the pages to PDF and replace individual pages with the new versions. Apple's Preview program is easier to use than Adobe Acrobat Pro, and is available for free. (I don't do Windows.) Be careful in looking for PDF editors; many programs called "editors" are designed for creating or filling out forms, and can't edit page contents.

The major practical challenge is finding a suitable distribution service because Amazon and other big e-book companies prefer selling e-reader formats. I chose Payhip.com and Google Books. Payhip is easier to use and offers higher royalties, Google has better name recognition, but is awkward to use, and at this writing is not accepting new "partners" to its publishing program. Payhip has little search-engine presence, but my website has a page devoted to the book with a link to Payhip, and it ranks second on a Google search for my best-selling book by my name and its title. (Google Books is in fifth place.)

Self-publishing in POD is almost as easy because it also copies scanned PDF images. POD services store digital PDF images of the book, and print copies when needed. Two basic classes of service are available: Do-it-yourself where posting is nominally free or full-service companies which charge for posting and formatting. I chose do-it-yourself, but looking back I think paying for full-service would have saved considerable time and trouble for my large and complex textbook.

The two major do-it-yourself options are Createspace (an Amazon subsidiary) and Ingram Spark (ingramspark.com/Portal). You post the files on their site, their software checks that the files meet quality and content standards, and you revise as necessary until they're satisfied. Then they

sell copies through on-line and brick-and-mortar bookstores, and pay you royalties based on sliding scales that depend on the price you select. I chose Createspace because it offered a simpler process.

Preparation was simple and easy for a 160-page young adult science book. Createspace's quality scan caught a couple of minor problems that were easy to fix, and with minimal effort the book was on sale.

In contrast, preparation of my 800-page introductory textbook on fiber optics was difficult and frustrating. The first problems I encountered were scannos that introduced bogus fonts into the PDF file. I was able to fix those with an evaluation copy of Adobe Acrobat Pro, but the process was very tedious and time-consuming. However, I had to pay \$149 for Createspace to adjust the position of the printed area on the pages to allow for the wide gutter (binding margin) needed for an 800-page book, because Acrobat Pro could not fix that problem. To be fair, 800 pages is close to the maximum size for Createspace's POD, but Acrobat Pro was very disappointing.

Self-republishing in the HTML-like format used by e-readers is a far more difficult problem because you have to convert the text and graphics in the book into the new format, and essentially lay the whole thing out again. That means you have to check the entire text for scannos, reformat it to make the fonts uniform, lay out the art so it falls between paragraphs and within page boundaries, clean up any scannos introduced into the art when the OCR tried to read labels, and clean up references and formats such as subscripts and superscripts. That also means you have to fight Microsoft Word's default and Autoformat settings, which do very odd things to bulleted lists or numbered lists of references. It also means you have to learn about other arcane formatting tools in Word, such as "styles" for paragraphs, headings, references, and so forth. Essentially, it turns you into a production manager charged with typesetting a book submitted in very messy format.

I did that for one book, using Acrobat Pro's OCR to produce a Word file from a PDF I already had. That book included end-note references, bulleted lists, line drawings, photos and other gray-scale illustrations, and lots of subscripts and superscripts. I spent long hours editing the Word file for submission to Createspace, which converted it to MOBI for Kindle. I also submitted the edited Word file to Smashwords and Draft2Digital to format for other e-readers.

Later I discovered a downloadable application called Kindle Textbook Creator that converts PDF page images into a form displayable on a Kindle reader, with each PDF turned into an image displayed on the screen. Amazon called them "print replicas" because they are replicas of a print book rather than a standard flowing MOBI format. The republishing process is quick and amazingly easy because it uses the page images you already have from scanning, including graphics and special formatting. It handled all 800 pages of my fiber-optics book without problem. It also should work well for other large textbooks, including those with full-color interiors.

However, that ease of generating the files comes with important tradeoffs for both readers and authors. The can be resized on the screen, but not flowed. Hold the screen vertically, and it shows the image full vertical height, which can make the type very small. Hold the screen horizontally, and it shows the image full width, making it easier to read, but only showing a part of the image, so it has to be moved to read the full page. It's not bad for a reasonable-sized tablet, but it's not nice to read on a small phone.

The downside for the author is economic. At this writing, only Amazon has such a service, limiting your market. More seriously, the Textbook Creator generates huge files which Amazon's royalty scheme penalizes. Amazon offers 70% royalty for e-books priced between \$2.99 and \$9.99, but charges a "transport" fee based on the file size. The transport fees for my "print replica" editions would come to almost my entire royalty for the smaller book I tried. The only practical alternative is to settle for the lower 35% royalty rate, which is not subject to transport fees.

### **Is Self-Republishing Worthwhile?**

Was the whole effort worthwhile? It depends what you want. PDF distribution is clearly worthwhile if you want to make your books more accessible, and POD lets you offer print copies for less than most publishers want. It's hard to tell how widely accepted "print replicas" will be, but they are much easier to produce than flowing-text e-reader formats, which take an inordinate amount of effort and have not added much to my audience.

Definitely it was worthwhile for my fiber-optics textbook. The former publisher had raised the price above \$150 before abandoning it. To make it more accessible for students. I cut the price for the print edition to \$39 and offered the PDF for \$9.95 -- selling a total of about 200 copies in 10 months. The PDF price is low enough for African students to afford. As a self-employed professional writer, I need to earn something from sales, but if you have other sources of income you can achieve broader distribution for free or a more nominal cost. The "print replica" edition was posted only recently and has not been promoted, so I don't know how well it will work. All in all, the reaction is positive enough to think about a new edition.

I also am happy with results for a 160-page young adult book. Originally published in 1987, it's old enough that I didn't feel right charging for the PDF, so I've posted it for free at Payhip. A few people have bought the POD edition and the "print replica" Kindle version, and I am thinking about an update at some point in the future.

My only real disappointment was the e-reader edition of a 30-year-old investigative book with extensive references and illustrations, *Beam Weapons*. It took an inordinate amount of effort to clean up the mess made by scanning and OCR. I'm glad to have the book available, but I could have done that with POD and PDF alone. Unless you have a complete digital file available or your book is text-only, it isn't worth the time trying to convert an existing scholarly or serious nonfiction book into e-reader format.

The bottom line is that you should take a long, hard look at the market before you self-publish. POD and PDF will keep your old books available with little effort, but don't expect to make much money. Unless your book is text-only or you start with a complete digital version, republication for e-readers will take much more effort, so is unlikely to be worth your time unless you try a "print replica" edition.

An extended version of this guide with a list of resources is available for \$2.99 from Payhip at <https://payhip.com/b/AX9h> or in e-reader format from your favorite on-line retailer.

*Jeff Hecht is science and technology writer based in the greater Boston area; his web site is <http://www.jeffhecht.com>. An earlier version of this article appeared in the Winter 2015-2016 issue of ScienceWriters, published by the National Association of Science Writers..*

Copyright 2016 Jeff Hecht, [jeff@jeffhecht.com](mailto:jeff@jeffhecht.com) <http://www.jeffhecht.com>