



SELL IT OR GIVE IT AWAY? OPEN ACCESS OR CLOSED?

Content Distribution Skills for Professional Genealogists in the Open Access Publishing Era


by Ian Watson

There's a town in upstate New York with an old clapboard church where one of my ancestors was a member. Luckily, the church has excellent records. They go all the way back to the early nineteenth century. Like other towns in New York State, this one has an official historian. In the 1980s, the elderly woman who then held the office arranged and copied the church records. About ten years ago, after her death, the records were published by a well-meaning friend.

That friend chose to publish the church records as a paper book. We can understand why. He probably thought that's just what you do with a manuscript. He found a publisher who turned the records into a book and put the volume up for sale. The book, still available, costs more than forty dollars on the publisher's website. Unfortunately, a glance at WorldCat (worldcat.org) shows only about a half-dozen libraries have purchased the book.

This is a case where, in hindsight, it might have been better to release the church records for free, probably in PDF format. They would have been more widely read and would have helped more people, which is what the original compiler wanted. (In her role as town historian, she had already spent hours corresponding with people and sending them extracts from the old church records at no charge.)

If you think about it, many content creators and compilers face dilemmas similar to that of this town historian and her friend. Let's say that the genealogical society you belong to asks you to



**In the old days,
there was only one way
to distribute knowledge:
printing it on paper.**

oversee a publishing project. Or a client hires you to transcribe a collection of old family letters. You're asked for your advice on the economics of the project—in particular, whether to print on paper or not, and whether to distribute electronic versions for free or behind a paywall. What do you say? Before you can answer, you need to understand the differences between traditional content distribution (content that's sold) and open-access content distribution (content that's given away).

Traditional Distribution

In the old days, there was only one way to distribute knowledge: printing it on paper (which cost money upfront) and then, usually, selling it (to recoup the printing costs and perhaps to make a profit). It didn't matter whether you were trying to make money off your book or not. Even if you were a selfless scholar who would have been happy to give your work away,

someone had to pay to print it on paper in order to distribute it, and readers usually had to share the cost. This was true even if your topic was nineteenth-century underwater basket weaving in the Ozarks.

Still today, authors who produce what some call “books written to be sold” need to make money off content. So, your Jeff Kinneys or Danielle Steels still have to put their work in a format that’s hard to reproduce and sell it. They make it hard to reproduce by printing it on paper or by controlling electronic access, thus creating a so-called paywall.

Open Access Distribution

When the internet arrived, it reduced the marginal cost of distributing knowledge to zero. Emailing another PDF of a book or article costs nothing. When one more person reads your work electronically, there are no extra printing expenses. The first people to really discover this were academics, who started to share their writing with colleagues for free. In 2001, a group of academic publishing professionals met in Budapest and decided to use the term “open access” to describe this new way in which scholars were working. Terms for the opposite of open access vary—you’ll variously hear people speak of *paywalled*, *toll-access*, or *closed content*.

At first, releasing new content in open access posed a challenge to the existing business models in scholarly publishing and the traditional (and sometimes unjust) ways that the academic world confers prestige on scholars. However, open access also saved millions of dollars in library and student budgets and helped fresh scholarly voices be heard. Over the last couple of decades, the idea of open access to new content has steadily established itself in academia. And as genealogists know, a lot of historical content is now open too.



Ian Watson at Bókasafn Dagsbrúnar, the research library that he once managed in Iceland. Photo by Róbert Ágústsson.

Today, people who produce “books written to be read” generally cover the expenses of producing content on their own. They want it to be as widely consumed as possible. Since distributing it digitally is free, they can just give their work away. Their books and articles are released openly. They may ask readers for donations, but they don’t force them to pay.

Closed and Open Content Coexist

Step into a bookstore and you’ll see that the advent of open-access publishing hasn’t made the old paywall model die out. Rather, the two models now coexist. Toll access still works for those kinds of content that people are willing to buy. And there are many genealogical books and databases which are financially viable only because readers are paying for them.

However, selling on the market for paywalled content is growing tougher, with more and more competition from free, substitutable products. For example, you’ve probably experienced how it becomes harder to justify buying a paper reprint of a classic genealogy book when the original is available at the Internet Archive (archive.org) for free. Library purchasing budgets, which used to be a strong source of support for paywalled content, are less dependable. Content providers who publish paywalled work today have to work extra diligently to figure out what will be saleable and profitable amidst rapidly changing market conditions.

What You Need to Know about Open Access

Typically, you’ll find the following new and historical content in open access:

- Some current academic journals (perhaps a third of them) and some monographs
- Most blogs and some newspapers
- Many government publications and databases
- Much reference material (that was once published in handbooks and almanacs)
- Most books and journals published before 1923
- Some books published in the United States from 1923 to 1963
- Some newspaper archives
- More and more genealogical records

Genealogists are particularly interested in historical content. The largest open-access repositories include:

- Internet Archive (archive.org)
- HathiTrust (hathitrust.org)
- FamilySearch (familysearch.org)

You’ll notice Google Books (books.google.com) isn’t on this list. Google Books was one of the pioneer sources of open-access

content, and I see many genealogists still citing it, but I use it only rarely. For years Google has back-burnered the project, which didn't work out to their advantage. (They had hoped to work out a deal with rightsholder representatives to charge for in-copyright content, but their proposed agreement was ruled unfair by a United States federal court in 2011.) Most of the out-of-copyright content in Google Books is also viewable at the Internet Archive or HathiTrust, which have more robust search and viewing interfaces, reliable permanent URLs, and a clear public-interest mission.

How Does Copyright for Open-Access Content Work?

In traditional publishing, authors and publishers rely on copyright laws to ensure that no one else can distribute their work and that all the sales revenue goes to them. Copyright helps maximize authors' monetary profit by making it illegal to duplicate their work.

In open-access publishing, there is no sales revenue. Authors often *want* other people to copy and redistribute their work. They deliberately forgo some of their rights under copyright law. They include a statement with their work that they are doing so. It's basically a license that gives users permission to recopy and redistribute the content.

An organization called Creative Commons (creativecommons.org) has developed a set of standard licenses for open content. Creative Commons licenses use varying combinations of the following four elements to set conditions on redistribution:

- **BY** (by attribution; author must be credited for the original creation)
- **SA** (share-alike, meaning that any derivative works must be similarly licensed; this basically implements the concept of "copyleft," used in software licensing)
- **NC** (noncommercial use only; applies to any derivative works or redistributions)
- **ND** (no derivative works permitted; redistribution OK if the work is kept unchanged and whole)

How to Fund Open-Access Content

If you publish content openly, you can't depend on readers to fund your writing. You have to be able to fund the work on what's sometimes called the "author side." That means you have to cover any costs on your own. Here are some of the possible sources of author-side funding. You may use some of them already, while others may not have occurred to you:

- An author donates his or her own time
- A client pays a professional genealogist to research and write an article or book



Advice and Support for Opening Access to Older Content

Cabrera, Nicole, Jordyn Ostroff, and Brianna Schofield. *Understanding Rights Reversion*. Berkeley, CA: Authors Alliance, 2015.

This free, informative 124-page book answers questions about how authors can ensure they have the right to release their own content for free. The Alliance, which is loosely connected with the law school at the University of California–Berkeley, is basically an advocacy group for authors who "wrote to be read" and want better tools for open access. The book is available (as a free PDF, naturally) at authorsalliance.org > Resources > Rights Reversion.

Rubow, Lexi, Rachael Shen, and Brianna Schofield. *Understanding Open Access: When, Why, & How to Make Your Work Openly Accessible*. Berkeley, CA: Authors Alliance, 2015.

This book addresses open access more broadly. Find the PDF at authorsalliance.org > Resource > Open Access. While this book and the one mentioned above are freely available, you can also purchase a printed version, or donate to the organization as a gesture of your support.

Suber, Peter. *Open Access*. Cambridge, MA: MIT Press, 2012.

Not only is the content worth reading, visiting the home page for the book illustrates the many ways in which the author has "opened access" to his book. See [cyber.harvard.edu/hoap/Open_Access_\(the_book\)](http://cyber.harvard.edu/hoap/Open_Access_(the_book)).

- A prospective author solicits donations for an upcoming project using an online crowdfunding site (kind of like old-style book subscriptions)

- A scholar writes a book or article as part of salaried research duties at a university
- An institution pays the cost of creating a publication or some other kind of content that aligns with the institution's mission
- An organization gets a private or government grant to compile data or digitize documents
- A group of volunteers works together to create content (for example, members of a genealogical society who transcribe a cemetery, or collaborative indexing at FamilySearch)
- A nonprofit organization hosts crowdsourced content (for example, Wikipedia, WikiTree, or the FamilySearch family tree). There are costs for hosting the website, but they are much lower than the value of the crowd's donated time.

More Issues to Consider about Electronic Content

Print-on-demand for Electronically Released Content

You might think that no one will buy a paper book if it's available electronically for free, but it turns out that some people are happy to pay specifically to have a physical book that they can hold in their hands. In fact, there are successful scholarly publishers that publish simultaneously on paper and in free PDF, such as Athabasca University Press in Alberta (aupress.ca). Print-on-demand can complement free electronic distribution if there are enough people who want a physical book for reasons of convenience, sentiment, or vanity (another book to add to their impressive collection).

Working with Rightsholders to Open

The dilemma of how to distribute content is not just limited to new releases. It's also something that the authors of books published long ago are facing anew as the internet opens up new distribution possibilities. Authors now have the chance to revisit the way they distribute books published two, three, or four decades ago. While they can continue to try to sell them if there is a market for the work, they often do well to release older works for free. Here's how.

Case 1: Changing Rights Status at HathiTrust

Recently I was working on an ahnentafel project as a subcontractor for a genealogy publisher. I found that part of the client's ancestry had been covered in a fairly good "ancestors-of-X" genealogy published in 1991. Within the last few years, the now elderly author of the book had helpfully made scans of each sketch in the book, cropped them down, and posted each one on his Ancestry tree. Unfortunately, you couldn't flip between them and he hadn't posted the bibliography. I noticed, though, that the book was in HathiTrust. Because

its copyright was still in force, access to it was closed. It was marked "Limited Access" and "Protected by copyright law." But from running a project in Iceland that helped authors open access to their own work, I knew something I could do.

HathiTrust has a permission form (hathitrust.org/permissions_agreement) which allows rightsholders to change the rights status of a work and release it under a Creative Commons license of their choice. (Among the rightsholders might be the author, an author's heirs, or a publisher.) In this case, since the author had already posted part of the book on Ancestry, I suspected he would appreciate the chance to release it at HathiTrust (in an even more convenient format). I emailed him through Ancestry and explained how to do this. He responded positively, sent in the form, and a few weeks later the book was open. You can do this too!

You can also alert HathiTrust when they list the rights status of nineteenth- and early twentieth-century volumes incorrectly. This happens particularly often with older serials. Newer volumes are still in copyright and are accurately

classed that way, but volumes that have entered the public domain get classed as "limited view" when they should be "full view." If you see this happening, do a good deed and ask HathiTrust to fix it (email: feedback@issues.hathitrust.org).

Many people wonder about HathiTrust's unusual name. The short answer is that *hathi* (pronounced HA-tee) means "elephant" in Hindi, and elephants never forget. (Remember the 1980s floppy-disk brand?) I've heard people say that the name honors an anonymous Indian philanthropist who helped underwrite the project, but that could be a myth. The story behind HathiTrust is that it was founded by a consortium of major universities to archive digital copies of all the volumes that they contributed to the Google Books project. Its headquarters are at the University of Michigan.

Case 2: Uploading Content to the Internet Archive

Perhaps you know of a genealogical work which has never been scanned and which its rightsholders would like to make

With print-on-demand, there are no upfront costs to recoup and only the number of copies actually sold are printed and bound (one at a time). This reduces publishers' risk, but at some cost to print quality. Authors can build in a small profit margin so they make a little money off each paper sale and thereby offset their costs. There are several good commercial platforms for print-on-demand books, including CreateSpace (createspace.com), BookBaby (bookbaby.com), Lulu (lulu.com), and others.

Citation Considerations

Changes in citation technology—changes in how we refer and link to our sources—are making it harder and harder to publish

on paper. Increasingly, our sources are online. A large percentage of citations are web addresses or URLs (ideally long-lived, durable ones).¹ Long URLs on paper are unattractive, and clumsy because you have to physically type them into your browser to follow them. Many people now do a lot of reading on electronic devices which let them call up a URL with one click.

Do you want to make it as easy as possible for readers to follow your sources and links? Then electronic publication (whether open or closed) will maximize their convenience. There are even, if you wish, ways to include URLs in the document without forcing them on the reader's visual experience. And this is making it harder and harder to justify paper-only publication for citation-rich content.

¹ For more on durable URLs see Ian Watson, "Citing Web Addresses that Last," *APGQ* 32, no. 4 (December 2016), 168–73.

Access to Older Content



available openly. It could be anything from a published genealogy to a typescript cemetery transcription. It could be your work or the work of others. For this kind of content, the Internet Archive is usually a good place. I've given several works a home there (see archive.org/details/@iwatson). Ideally, all this should be done by the rightsholder(s) in their name(s). Alternatively, get the rightsholders' written permission.

To upload your work, you'll first need to scan it. You can do this yourself or have it professionally done. Or if it's a word-processed document, just save a PDF version. The Internet Archive can also help with scanning, for a small fee.

Then, create a login at the Internet Archive and upload the document. It will be OCR'd and indexed. Jeff Sharpe, who runs the Internet Archive's scanning center at the Allen County Public Library, is a good contact for genealogists with advanced questions about contributing content. For contact details, see archive.org/scanning.

The Internet Archive was established by tech entrepreneur Brewster Kahle in

Access Issues for International Users

Users outside of the United States should note that some works in HathiTrust which, at first glance, appear to be limited-access, are actually full-access in the United States and limited-access for users with an IP address in any other country. You can tell these volumes because HathiTrust lists their rights status as "Public Domain in the United States." If you connect to the HathiTrust website through an American VPN (virtual private network), you can access these works from abroad too.

1996 and is now a well-established non-profit institution. It looks like it will be around for a long time and that makes it a reliable place to store digital content. The Internet Archive also takes content in other formats, including audio and video.

Case 3: Opening Access to Material at the Family History Library

The Family History Library (FHL) in Salt Lake City encourages rightsholders to submit a "Permission to Duplicate" form, which is available via a link at familysearch.org/wiki/en/Donations.

This form allows rightsholders to ask the library to scan their work and make it available digitally. Patrons will be able to download the content at home, not just in a Family History Center. The actual physical work will then be moved to off-site storage, saving shelf space in the FHL. The form can be used when making a new donation, or by rightsholders for paper works already at the FHL.

The form basically contains a Creative Commons license (CC-BY), although it doesn't call it by that name, and uses its own wording. For questions, you can contact books@familysearch.org.



Should you advise selling content or giving it away?

Can I Get Ancestry to Sell My Content?

I wonder sometimes whether it might become possible for individual genealogical authors to fund their work by contributing compiled content to large, pay-access genealogical portals such as Ancestry (ancestry.com), American Ancestors (americanancestors.org), or findmypast (findmypast.com). The idea would work like this: Say you've put a lot of work into abstracting the first five volumes of land records from a county in North Carolina. You've created a nice little database with the names of grantees, grantors, abutters, and so forth. In the old days, you would have published it on paper. Today, what about asking Ancestry or American Ancestors if they'd like to house the database in return for a tiny fee each time someone accesses it? Could this be a win-win publishing model in the internet age?

As far as I know, no such option has materialized to date. The big commercial portals don't seem to want these small datasets, or if they do, they don't want to pay for them. However, it's worth keeping tabs on whether something like this becomes an option in the future. It could become a viable way of funding genealogical projects that might not otherwise get done. After all, independent genealogists often have a good sense of what records are valuable, an entrepreneurial spirit, and a desire to contribute to the community, but don't generally have the capital or the visibility to develop viable for-pay internet-access portals. The big portals, in contrast, could, in theory, add small databases at a reasonable marginal cost. There would be a number of obstacles to overcome, however, such as data quality, payment models, and also the stiff competition from volunteer abstracting and indexing efforts.

When to Sell It and When to Give It Away?

It's a pity to see content for sale that should have been free, as in the case of the upstate New York church records. I've often seen disappointment strike authors who publish in the traditional way. Imagine a genealogist who publishes a book only on paper. It doesn't sell well enough to cover the costs. It also fails to reach the many potential readers who are interested in some of the material but don't want to buy the entire book. The author should have released the book for free. Instead, the book is walled off from its audience.

It's also a pity to see content given away that should have been sold. Imagine a genealogist who invests time and expertise in creating a manuscript which could easily sell two thousand copies at fifty dollars each. On principle, the genealogist gives the work away. Generous, yes, except that the \$100,000 in gross proceeds could have been a nice reward for the work and perhaps funded a subsequent project. If content has real sales potential, take advantage of it. Don't martyr yourself to open access. Reap your rewards and consider reinvesting them.

Back to the question posed at the beginning. Now that you know more about open access, should you advise selling content or giving it away? Not sure what to do? I suggest these basic guidelines:

Advise selling if:

- You truly foresee strong demand for the content.
- The content is novel, of high quality, and marketable (whether directly by the author or with the help of a traditional publisher).
- The content is not substitutable by something else.

Advise giving the content away if:

- Author-side funding is available.
- The author doesn't need to make money off the content.
- The author wants to maximize the number of readers.
- There is low demand for the content (such as with some family genealogies).

We can all see that open access is definitely the most sensible model for a significant number of publishing cases. At one time, if you had to give your content away, it meant it wasn't good enough to sell. Today, open content is more often a sign of a successfully managed project. At its best, publishing in open access means that you covered your expenses, distributed your work to everyone who wants to see it, and archived it safely for future generations.

Whatever you do, stay pragmatic, not dogmatic. Neither open nor closed access is the one right way.



Ian Watson, PhD, has worked both on the open and the paywalled sides of the publishing business, both within and outside genealogy. He also teaches in the media design department at a Norwegian university. He first came into contact with open-access publishing ten years ago, while living in Iceland, where he edited a small open-access academic journal and worked on several other open-access projects. For more information and links, see www.ianwatson.org. This article is loosely based on his presentation at the 2017 PMC.