Digital Publishing's Four Challenges

DAVID W. LEWIS

Introduction

Publishing bits on a worldwide network is obviously different from publishing print on paper. But as is often the case with all new technologies, established institutions cram digital technology into their established practices and procedures. Then slowly, and often painfully, new ways of applying the technology with new business models emerge, and only then do new practices and institutions coalesce. Digital technologies have been applied to publishing in one way or another for over fifty years. This journal has been documenting the process for half of that time. Many practices have been developed and some new institutions have formed, but it would be premature to say that digital publishing has coalesced.

I would assert that digital technology presents four challenges to publishers that coalesced practices and institutions need to account for and resolve if they are to take full advantage of the opportunities digital technologies offer.

Challenge One: Free, Perfect, and Instant

As Andrew McAfee and Erik Brynjolfsson put it in their book, *Machine, Platform, Crowd*, digital content is "free, perfect, and instant." That is, a copy can be instantaneously delivered anywhere in the world, a copy is the same as the original, and a copy can be made at zero marginal cost. They go on to explain, "Free, perfect, and instant make a powerful combination, worth more than each of these characteristics separately. Thus, it is very difficult to compete with. . . . For most of history, few, if any, goods and services have been free, perfect, and instant. But with digital, networked goods, these three properties are automatic."[1]

1. Andrew McAfee and Erik Brynjolfsson, *Machine, Platform, Crowd: Harnessing Our Digital Future* (New York: W. W. Norton, 2017), 135–137.

https://doi.org/10.3998/jep.2012

From the economic perspective, the efficient price for a good, that is, the price that will generate the most societal benefit, is its marginal cost. For a digital good, that cost is zero. This means that open access publishing models should become dominant, at least in sectors where maximizing societal benefit is or should be a goal. Of course, we all understand that maximizing societal benefit is not the primary goal of all publishers. For-profit publishers have a fiduciary responsibility to another goal, maximizing investor return. In some cases, this might be alright. We may not care that someone who would benefit by being entertained by a recent novel is denied that benefit because they cannot afford the novel's price, but we should care when researchers, students, and the public are denied the benefit of scholarship. To date, much scholarship remains locked behind expensive paywalls, and so society suffers. For a long time, no one with leverage cared about this loss, but lately research funders have started to care, thus Plan S and other mandates. It is too soon to know how much this will change the market, but some change is inevitable. This is the first problem.

A related problem is that content that is instant and perfect can easily be pirated, as Sci-Hub so clearly demonstrates. You simply cannot contain easily duplicated files that can instantly be sent anywhere. The more expensive and difficult it is to access the content, the more likely a black market in that content is to develop. The only way to combat piracy, as the music industry learned the hard way, is to provide content that is easy to use and cheap.

Publishers will have to learn to make the economics of "free, perfect, and instant" work. This will require efficiencies that are difficult to achieve with legacy systems and thinking. What is clear is that locking digital content down and charging high prices will not work.

Finding a way out of this dilemma is the first challenge of digital publishing. Progress on this challenge has been made in open access publishing with a combination of means, from article processing charges (APCs) to institutional subsidies. Some low-cost subscription publishers have also made this work.

What is important to understand is that in publishing sectors in which societal benefit matters, one way or another, open access must become the primary publishing business model. No pay-for-use business model, whether subscription or one-time purchase, will ultimately be able to maximize the benefit to society.

Challenge Two: Anyone Can Be a Publisher

As Clay Shirky put it, "It makes increasingly less sense even to talk about a publishing industry, because the core problem publishing solves—the incredible difficulty, complexity, and expense of making something available to the public—has stopped being a

problem."[2] For all practical purposes, anyone with an internet connection and a computer (including a phone) can be a publisher if what that means is to make their work public. Beyond the time and energy needed to create the work, the costs are trivial. Put up a website or use Medium or Substack. Deposit a scholarly work in a preprint server or an institutional repository. Self-publish your book with BookBaby, Lulu, or Amazon. YouTube is available for video content, and for less formal work there is Twitter, Facebook, or TikTok. All are free or low cost to use, as they traffic in attention and personal data. The barrier to entry in publishing, at least at the low end, is almost nonexistent.

While brand still matters and scale can bring efficiency advantages, the competition for attention is intense, and players at the margin are always a threat. The growth of preprints and Twitter as a means of rapidly communicating COVID-19 research, though still controversial, is potentially such a threat. The fact that a small team at OurResearch could create OpenAlex, a replacement for the Microsoft Academic Graph, a comprehensive database of the world's scholarly research, in less than a year indicates what is possible with currently available technology.

This also means that traditional gatekeepers and authoritative voices no longer enjoy a uniquely privileged position. This clouds notions of authority and expertise. "Truth" has become slipperier.

Does the consolidation of traditional publishers into a small group of large players indicate their mastery of the digital arena or is it a sign of their inability to do so? It is possible to make both cases.

Challenge Three: Putting the Pieces Together Is Complicated

The *New York Times* didn't put a color photograph on its front page until 1997. Today it routinely publishes not only text but also photos, video clips, dynamic charts and graphs, maps, interactive quizzes, and a host of podcasts and audio versions of many of its stories. The tools for explaining the news have become more interesting and more powerful. And immersive technologies are coming soon.

The ability to integrate multiple media formats to create a stronger, more compelling argument is one of the great promises of digital content, but it is rarely used consistently or well. In part this is because few organizations have the scale of the *New York Times* that allows them to bring both the technology and talent to the task. There is also a lack of standards, and the technology has yet to become cheap and easy to use,

2. Clay Shirky, "Newspapers and Thinking the Unthinkable," *Clay Shirky* (blog), http://web.archive.org/web/20090405130459/http://www.shirky.com/weblog/2009/03/newspapers-and-thinking-the-unthinkable/.

though, given Moore's Law, it is likely that in time, probably sooner than we might expect, these problems will be solved.

There is a further complication. Works printed on paper are distinct objects. Other works might be referenced, but each work stands on its own. When Ted Nelson first conceived of hypertext in the early 1960s, he saw it as a tightly controlled structure of links connecting works in whole or in part. That is not how it worked out. The web is open and chaotic. Lots of things are behind paywalls. Links rot. Copyright, a legal response to the printing press, often constrains what is possible rather than, as it was intended, encouraging creativity. In theory there should be interesting ways to use the ability to link many works together to create different stories and more powerful arguments, and while this is currently done to some degree, mainstream publications are little changed.

The interesting question is, What will come of these new powers? Print birthed the scholarly journal and the novel. Is some version of the multiplayer video game fiction's next venue? Is a data rich simulation a compelling way to make a scholarly argument? These developments will be fun to watch. But who publishes them and how are still open questions.

Challenge Four: Nothing Is Fixed

Perhaps the most fundamental trait of a work printed on paper is that it is fixed with the production of multiple unchangeable copies of the same text. Bits on the network are different. They are by their nature fungible.

The long-term preservation of print relies on multiple dispersed collections of long-lasting originals. While not without its own challenges, such as the "slow fires" of nineteenth-century high acid paper, the preservation of print is understood, and for most content in many parts of the world, institutions are committed to the task. The additional protective layer of digitizing printed texts provides alternative means of access, which saves wear and tear and stores the content in a second format, which provides additional protection from loss.

On the upside, the fungible nature of digital content makes collaboration and commentary easy. Wikipedia is an example of how this can work. Many contributors change it constantly, while the changes and the dialogue that produce them are recorded and archived. Unfortunately, Wikipedia is the exception. For the most part, versions and commentary are spread across multiple systems in an uncoordinated fashion.

Digital texts are unstable, and preservation is a significant challenge. The Clay Shirky blog post that I quoted above is a good news, bad news story. The blog is gone, but the Wayback Machine has saved it. For too much content, only the bad part of the

story holds, and the content is lost. Several studies have demonstrated what is familiar to all of us: "404 Not Found" is an everyday event.[3] As Jill Lepore put it in a *New Yorker* article discussing scholarly impact of this phenomenon, "It's like trying to stand on quicksand."[4]

Progress has been made, especially with flat text files, but more complex content is far from secure. Much has already been lost, and much more will be.

Conclusion

Digital content is a very different beast from print on paper. The development of practices that take advantage of the differences is an ongoing process that is still underway. Whether or not established institutions will prove able to change to be successful in the market is yet to be seen, but if history is any guide, the odds are probably against them. Apple and Spotify, not Universal or Sony, dominate digital music. The Internet Archive, not the Library of Congress, is archiving the web.

Changes of this magnitude usually bring new players who are unencumbered by legacy practices and organizational structures to the fore, but not always. What is most likely true is that what we think is certain is not, and what we think is unlikely could easily happen.

3. Probably the most extensive study is Martin Klein, Herbert Van de Sompel, Robert Sanderson, Harihar Shankar, Lyudmila Balakireva, Ke Zhou, and Richard Tobin, "Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot," *PLoS ONE* 9, no. 12 (December 26, 2014): e115253, https://doi.org/10.1371/journal.pone.0115253.
4. Jill Lepore, "The Cobweb: Can the Internet Be Archived?" *New Yorker*, January 26, 2015, https://www.newyorker.com/magazine/2015/01/26/cobweb.