6
The Digital Disseminators

Ephemeralization, a law of human progress formulated by Buckminster Fuller, is the progressive tendency to do more with less. Satellites transmitting phone calls, replacing tons of seabed copper wire, was an example. Digitization would have delighted Fuller. We have downloads, not bookstores. Rather than the libraries' vast cathedrals, information now hangs ethereally in the cloud. The cloud is not as ephemeral as its name suggests—quite the contrary. The countless server farms that form its leaden feet swallow electricity voraciously, although they are becoming more abstemious.[1] Yet, the cloud's ubiquity, expandability, and speed are little short of miraculous.

Distributed globally, cloud storage will not go up in flames simultaneously, but it depends wholly on maintaining the voltage. We must fear solar flares or other electromagnetic disasters wiping our cultural memory. And we should ponder how to transmit into the deep future. It is sobering to consider how briefly Homo sapiens have been able to communicate culture so far. Clay tokens used for accounting purposes have been dated to 8000 BCE.[2] Beyond that, artifacts from the Blombos Cave contain markings that appear symbolic, going back 70,000 to 100,000 years.[3] But that is it.

How can we pass what we know forward over the edge of our current event horizon? Digitality is still too fiddly and dependent on electricity to be reliable, especially to bridge possible disasters. "Digital documents last forever," in Rothenberg's dictum, "or five years, whichever comes first."[4] Our most prescient catastrophists are therefore sensibly wedded to analog technologies. The Arch Mission, for example, aims to print the core of human knowledge microscopically on durable media, placed strategically around the globe and possibly extra-planetarily.[5] To read it, resurgent post-catastrophe humans would need to have advanced technologically no further than the seventeenth-century optics of Huygens. More mundanely, though PDFs may not have vellum's durability, they are versatile and enjoy a hybrid status, being present in both digital and analog worlds, at home on paper and screen.[6]

What does digitality mean for the three main avenues of dissemination: bookstores, libraries, and publishers? Digitality fulfills the promise latent in the first printing presses of making authors their own disseminators. Will that undermine the intermediaries? Acting as a global bulletin board, where all are welcome to post, the internet threatens to swallow up the once distinct acts of publishing, disseminating, and storing.

Bookstores

Bookstores in their current incarnation as purveyors of physical tomes will likely continue to disappear. Some works will remain in tangible form. The book as an aesthetic object is no more apt to disappear than lithography was to end oil painting. A durable market for the bibliographic craft is revealed by the success of Taschen books. Its output is less books than works of art with pages. And its aesthetic *Gesamtgestalt* continues back up the distribution chain to its stores, each an architectural temple of nacre, dispensing pearls.

Whether conventional books are bought in physical locations or online depends on consumer preference. As their dissemination function diminishes, bookstores will increasingly become coffee shops with more elaborate inventory. The survival of independent bookstores, exemplified by the franchising of Shakespeare & Co., suggests a fate akin to coffee shops, with small one-off outlets in city centers coexisting with serviceable chains everywhere. Despite gnashing of teeth, the number of independent bookstores in the US grew from 1,600 in 2009 to 2,500 in 2019.[7] The Covid pandemic will not likely have increased that number. And having peaked in 2007, sales in bookstores slumped by 40% to 2019.[8]

Used bookstores as physical retail destinations are probably doomed, victims of the internet's ability to match long-tail consumers with wildly various and dispersed products. Few used bookstores have bothered to offer consumers temptations other than browsing. Browsing in a conventional bookstore is one thing. Customers select among the latest products, some recently reviewed in newspapers, mentioned online, or even TV. Serendipity may bring them into contact with books they had not heard of. The experience of a used bookstore is quite different. They are graveyards of the human spirit. Only a tiny fraction of what is published each year remains of interest a short while later. Knowing the labor invested in even incidental works, all authors turn glum when entering a used bookstore to survey the endless ranks of books-no-longer-read.

Before the internet, the thrill of the hunt remained. Customers rarely located a specific volume, but intellectual omnivores stood at least a chance of devouring some prey. Browsing the shelves, they stumbled across an occasional nugget at a bargain price. But today, with inventory digitized, the long tail is searchable. We know instantly where to find the work we seek. Since the market is transparent to both sellers and buyers, arbitrage is seldom possible. Prices cluster around what the market will bear, and bargains are rare. Scholars may derive some solace from knowing that the cheapest

used books are the flash-in-the-pan bestsellers of a few years ago. They often cost less than to ship them, just a few cents. In contrast, scholarly tomes, published in modest print runs, are harder to find but often hold their value. Indeed, the most expensive used books tend to be still-read but out-of-print university press monographs—solid bricks in the wall of scholarship.

Thanks to digitality, the secondhand book business has consolidated. Mega-operations, such as Better World Books, with inventory deep into seven figures, outcompete mom-and-pop operations. Such secondhand bookstores on steroids have, in effect, become adjuncts to libraries as they deaccession little-used holdings. Often, they rival even the largest collections in size.

Libraries

Digitality has undermined libraries, too. Like bookstores, libraries have no rights to content to exploit and defend. They are better positioned in the dissemination process than bookstores only because they also house our collective patrimony even after publishers have lost track of their backlist or vanished altogether. Only thanks to the first-sale principle do libraries exist at all—along with used bookstores, video rental institutions in their day, musical score services, and the like. Once the physical object has been sold, rights-holders can forbid its copying but cannot prevent lending or renting. But that is all libraries may do—lend, store, and preserve works that otherwise, outside the public domain, belong to others.

Some countries, starting with Denmark in 1946, introduced public lending rights. These mediate relations between rights-holders and libraries, taking the edge off sales forgone by paying authors fees proportional to how often their works are lent. No such compensation has been attempted for loss of sales through secondhand bookstores. Since books almost invariably depreciate on the

secondary market, losses here are less than from pirated works. For successful artworks, in contrast, subsequent prices sometimes dwarf initial sales, but artists do not profit from their burgeoning reputations. That unfairness helped persuade many nations—eventually the EU—to introduce a *droit de suite,* or resale right, that gives artists a cut of profits on the secondary market.[9]

Digital media change this landscape. For in-copyright trade print books, libraries will continue to play much their current role. There is no moral argument why such works should be opened up, and readers will have to negotiate access via libraries if they do not buy copies. Once books lose their commercial value, publishers may be willing to allow controlled digital lending, as discussed in Chapter 7, permitting patrons anywhere to read works on screen.

For e-books, however, the terrain has changed. The first-sale principle does not apply to digital media. They remain the property of publishers or intermediary vendors (such as OverDrive, the largest) even as they are licensed for reading in libraries. Rather than selling them, publishers lease their e-books to libraries for a specified time and/or a limited number of uses. The prices for e-books are generally higher than for the paper versions, and they vary depending on the kind of access allowed—one, many, or unlimited readers for few, many, or unrestricted uses, either time-limited or perpetual. While publishers cannot prevent libraries from buying their paper books at the same price as other customers, they can set library prices for e-books. Publishers have feared that leasing e-books to libraries would cannibalize print sales. Some, such as Macmillan, therefore embargoed new e-titles to exploit the retail market before allowing library lends.[10]

The situation is especially dire in the UK. Libraries there may lend only e-books licensed by their publishers, not scanned versions of paper books.[11] Worse, research libraries may not buy retail e-books, only versions specially licensed for university use.[12] Publishers can thus charge more for e-editions of works also published

on paper—as much as ten times.[13] They sometimes force libraries to buy an entire list of their e-books, rather than selecting them individually.[14] To counter such issues, Maryland and New York have recently attempted to pass laws requiring publishers to offer e-books to libraries on similar terms as for retail customers.[15] In return, the Association of American Publishers has sued to defend their right to charge what they please.[16] Congress has taken an interest in the problem.[17] A Dutch case has insisted on treating library acquisition of books equally, regardless of the precise format.[18]

On the other hand, variable licensing gives libraries some control over costs, unlike the one-book-forever pricing of physical volumes. An Australian study found the cost of lends for e-book licenses premised on an ongoing one-copy-one-user model was often cheaper than for physical copies. In contrast, the price per lend of metered use (X number of lends before the license had to be renewed) depended on the work's popularity and on assumptions of how many lends physical copies can withstand before needing replacement. But, where licenses were time-limited, the library no longer had the work in its collection once they expired.[19] That, of course, was not true with physical copies unless they were deaccessioned.

With time-limited licenses, libraries are cut out of curating their collections. True, flexible licensing permits libraries to tailor their offerings to demand: a few permanent copies of enduring classics, many cheaper time-limited ones of flash-in-the-pan bestsellers.[20] While this may make sense for neighborhood public libraries, it wreaks havoc on major research collections and their preservation function. Penguin Random House claims that time-limited licenses at lower prices are what librarians want, given the short shelf life of many new works, especially fiction. For academic libraries, it offers perpetual e-books at a much higher cost.[21] In British university collections, the price of the most requested e-books on an unlimited-use license is two and a half times that of a one-user-at-a-time plan.[22]

Digital books are fiddly. Unlike robust cellulose products—foldable, droppable, shippable—e-content is finicky, corruptible, and subject to mishaps, and it works only on specific devices and platforms. E-books are not housed in libraries and are read with proprietary apps and software that belong to others. The curatorial, lending, preservation, and storage functions that libraries once served are leached away. Library budgets increasingly go not to buying curated content, then lending and preserving it, but to paying publishers for use of collections assembled and stored elsewhere. Access, not control, is the new mantra.[23] If things continue in this direction, libraries may end up serving only as the conduit for payments from funding agencies to e-book publishers.[24]

Digitality will change libraries. As content migrates to the cloud, it seems quaint to think it must be collected physically, sitting in a particular place awaiting readers. Eventually, Google Books and similar projects will have digitized the public-domain and grey literature. Increasingly, the most recent, in-copyright scholarly literature will be published via open access of one stripe or another. What role, then, will libraries play? For content unlikely to become open—recent in-copyright fiction and trade nonfiction—both bookstores and libraries will continue to be disseminators. Small public libraries will therefore least notice this shift. Their patrons will continue to demand books that are still sold and lent.

Large research collections, in contrast, will become almost redundant. Their vast precopyright holdings are rapidly being digitized, and their grey works will eventually follow suit. Their current scholarly work will increasingly be born digital and open, whether green or gold.

Google Books has already scanned 25 million books. As a collection, that brings it abreast of the British Library, with about the same number. Only the Library of Congress, with its 39 million books, is still bigger. In 2013, five years after its founding, the HathiTrust, which holds the Google Books, had more works on

marine biology than the Library of Congress and 60% more Russian periodicals than its nearest competitor, the University of Chicago library.[25] Inventory, as logistics experts put it, is being consolidated. Digitality ends the need for physical libraries to store and lend content. We used to go to the village well to draw water, now we open the tap at home. So, too, will the digital flows be managed.

Nor will we need the research libraries' duplication of works. The dirty little secret of analog libraries is that they broadly replicate each other, at least within nations and cultures. They were, after all, created to solve a topographical problem by depositing content at reasonable distances. From the librarian's vantage, a simpler solution would have been one central, national library in each country. In smaller nations, that is broadly the outcome. When they organize into consortia to pool resources and expand holdings, libraries duplicate even more.[26]

Having said that, it is also true that individual volumes are often rare. Three-quarters of the titles of the Big Ten Academic Alliance's collections are held by three or fewer of its libraries.[27] In other words, on average, every one of these ten institutions has a core collection duplicated in the others that is 25% of total holdings. In turn, each library then holds another 25% of the entire collection, shared with only a few of its peer institutions, and would have to borrow a volume from the other half of total holdings via interlibrary loan. Similarly, 88% of the Research Libraries UK's collection is held by fewer than five of its constituent partners.[28] Sixty percent of the aggregate Google 5 (Harvard, Michigan, Stanford, Oxford, and the New York Public Library) library collection is held by only one institution. Among any two of these libraries, common holdings account for only 20% of their collections.[29]

Major research collections thus consist of a substantial common core, trailing thick and very long tails of rare works. Fully a quarter of holdings (in the case of the Big Ten) are duplicates everywhere, with a researcher pursuing a particular volume outside that

core having a 30% chance of finding it in their home institution. Interlibrary loans make up less than 5% of all circulation in major research institutions (1.7% among all libraries).[30] That indicates the relative rarity of scholars reaching beyond the core duplicative collection at their home institutions (though it could, in theory, reveal a lot of travel by scholars to other institutions, sidestepping interlibrary loan). Put differently, 95% of scholarly work is accomplished within the duplicative core collection of the local research library.

Some 10% of all books at major US research libraries account for 90% of all circulation (this figure does not measure books consulted but not checked out).[31] As research libraries come online, the underconsulted bulk of their physical holdings will become superfluous. The core collection may remain present, absorbing scholars' energies. Meanwhile, the marginal interlibrary loan aspect of their function can readily be turned over to common digital storage of little-used books. Increasingly, interlibrary loan materials are digitized and sent over the internet rather than entrusted to the tender mercies of the postal system. Once the long tail has been digitized, even that will no longer be necessary.

Major research libraries in North America hold almost a billion physical volumes, but only 59 million distinct titles.[32] That is an average duplication ratio of 17:1. In 2011, WorldCat libraries across the globe had 1.238 billion books, but 128 million separate print book publications, a duplication ratio broadly the same.[33] A single institution approximately twice the Library of Congress's size could thus hold a single copy of every book in the US. A global library would have to be at least triple that.

Conversely, the extension of the long tail in major research collections, the commonness of rarity, suggests two other implications for digitality. The corncrake is considered an exotic and endangered species in Britain, with enthusiastic birders lying in wait on its remaining redoubts, the Western Isles. In Russia and Kazakhstan,

in contrast, it is common. Seen globally and sought in its customary habitat, the corncrake is neither rare nor endangered. So, too, a book that is rare in the collections of one city, region, or nation may not be when the world is surveyed. Approached broadly, there may be no reason to fuss over an allegedly rare volume.[34]

Only 34% of the University of Chicago's books are also in Ohio State's library. But three-quarters of them are duplicated across the broader CIC library consortium in the upper Midwest.[35] Similarly, 88% of books in the UK's research libraries are held by fewer than five libraries, but across the whole WorldCat system, that drops to 56%.[36] Once works are digitized, available anywhere, and geography's artificial scarcity has been eliminated, rarity will emerge only in its proper worldwide context. Much that today appears exotic will prove to be common. And in any case, it will not matter, since even the rarest tomes will be readable anywhere.

Second, using the many works held in only a few collections involves much friction. Among the Google 5 libraries, almost 40% of works are owned by one institution only.[37] A third of the holdings of libraries in the BosWash corridor can be found in no other US region.[38] Even if the long-tail books are rarely in demand, once sought, they are unlikely to be nearby. To be useful, they must then be shipped via interlibrary loan, or the patron must travel to them. Digitization costs money, of course, but only once. Indeed, as libraries have discovered, the expense of digitizing a book is usually less than sending it off for a singular use via interlibrary loan. Once the long tail has been digitized, the price of using any particular volume in the global holdings approaches zero, and the cost distinction between core and peripheral collections vanishes. If works are digitized as they are requested via interlibrary loan, current budgets—now wasted on postage and packing—will, over time, pay for digitizing the long tail of holdings.

Once the 60 million distinct titles in the major North American libraries have been digitized and put online, existing scholarly

content will have been made accessible. Conventional libraries have been needed less for the diversity of their holdings than for their geographical dispersion. When geography and distance have been eliminated, what is left? Some items that each collection alone might possess will remain, motivating the occasional visit. But in that sense, libraries will have become like archives. And ultimately, the same fate awaits archives as will overtake libraries—their content in the cloud, their earthly purpose unclear.

The emergence of collective collections testifies to the forces behind such changes. Collective catalogs have long been assembled, allowing single searches among multiple holdings—WorldCat, the British Union Catalog, UC's Melvyl catalog. Institutions have long been fusing and coordinating their physical collections—the University of California's Northern and Southern Regional Library Facilities, those of the Five College Consortium in western Massachusetts, or the UK Research Reserve. Large collections are amalgamating their long tails, assembling least-used items in joint storage and deaccessioning duplicates. ReCAP (Research Collections Access and Preservation consortium) unites the Columbia, New York Public, Princeton, and Harvard libraries. Why not do this nationally or globally? Once digitized, the physical holdings will be merely for backup. HathiTrust's collection of Google Books and other digitization efforts are far down this road. By 2011, HathiTrust held in digital form about one-third of the books physically present in the Association of Research Libraries' collections.[39]

Libraries will continue to play roles. But they will be less needed for what was once their primary activity: lending or reading in situ. Publishers have taken over many of the functions once theirs. Digital publishing leaves libraries no longer owning physical copies of journals and books. Storage is left to the publishers. Big-deal bundles mean that librarians no longer select which content to acquire. Cataloging, metadata management, search services, and the like are also increasingly supplied by publishers.[40]

Libraries may still add value through curation: selecting, organizing, sorting, presenting, contextualizing, and the like. And they may become social centers, as on university campuses, assuming some of the functions that once belonged to bookstores. Conversely, today's internet café, with every patron busy at a laptop and the sound level at what Amtrak mandates in its quiet cars as a "library-like atmosphere," have taken on functions akin to study halls. That would make libraries into pedagogical facilities more like schools or adult education centers than content repositories. We hear much about libraries' new functions as social centers. Sites of connection, not collection, says Frances Pinter.[41]

When Stanford redesigned its engineering library in 2010, only 10,000 of its 80,000 books were not consigned to off-site storage. Half as big as its predecessor, much of the new library's space was given over to private and collaborative study areas.[42] All well and good, but deep down, are libraries part of the hospitality industry? Other than perhaps undergraduates, most patrons go there not to socialize but for reasons analogous to why Willie Sutton robbed banks, because that is where the content is. And when it isn't, they won't.

As publishers become repositories of their own content and libraries merely funnel revenues in one direction and content in the other, issues arise. Libraries were once curated collections, differing as they served various clienteles. But increasingly, the content rests with the publishers, as the libraries cut deals for large swaths of their lists. Where eager readers once used the library catalog as the main integrated source of knowledge, multiple search engines now take researchers to the publishers' individual content silos.

No single library can own everything, but at least librarians thought about what to acquire using criteria other than what motivates the presses. Thanks to digital balkanization, we now need bridges to span a fragmented content landscape.[43] Kopernio is one example. It is now owned by Clarivate, one of the main science

indexers. A browser extension, Kopernio provides access to articles via various routes—primarily through the publisher for content licensed to the user's library, but otherwise to an open version, so long as it is legal.[44]

Research libraries might find useful roles as dependable storage, at least for our predigital cultural patrimony. They could become more akin to symphony orchestras and museums: preservers of a canon. But they are also well-positioned to serve a crucial backup function for digital content. Although magical, digitizing does not solve all problems. To insure against catastrophe, paper copies of predigital works must be kept, and our cultural inheritance would be more secure if we had a printout of born-digital materials. Perhaps not every last e-mail or utility bill, but of content whose loss would be painful.

The Internet Archive stores its books in containers in the Bay Area and elsewhere. Nicholson Baker once crusaded to preserve decommissioned periodicals.[45] As Google Books has shown, kinks remain to be worked out in scanning content before copies become wholly satisfactory. Some readers will still need paper or original versions for reasons of their own. Content remains partly artifact.[46] Sometimes copies reveal their information more easily than originals, other times less so. Physical editions of past publications will not vanish altogether. But these are comparatively minor issues.

More important, someone has to be the long-term archivist. When libraries license e-books, they are stored on the publishers' servers. What happens when they go bankrupt, cease issuing a periodical, or otherwise stop being responsible stewards of their content? The materials in the cloud need to be preserved, kept up to date, made compatible with new software, and otherwise usable in perpetuity. No more than we could rely on conventional publishers to caretake their output forever can we with the digital. Amazon keeps its library of self-published works online for as long as authors wish. But, although ubiquitous, is Amazon forever?

One failure is on the supply side, the disappearance of publishers or their output. The other is on the demand side, the cancellation of subscriptions. What happens if a library ends a subscription to a digital journal? Can its patrons still read back issues, as in the paper days? Or is the new model akin to a lease, where all access is lost once the subscription lapses? The question also applies to publishers' bundled collections of monographs.

Preserving born-digital work becomes complicated the further it moves from a simple electronic mirror-copy of print version. With more bells and whistles, greater embeddedness in the web, connections to other materials, ongoing updating, fluidity of content, and multiple formats, e-books are tricky to preserve. The publishers know and control their software better than any would-be archivist. How likely are others, such as libraries acting as archives, to reproduce the original look and feel? Will they have what it takes to keep proprietary digital content up to date and retroactively compatible? Archiving digital materials is more complicated than the fairly rudimentary storage of the paper era. Yes, the fiddliness of keeping paper copies intact over centuries is often underestimated. Even so, heating, cooling, and moisture control have been the most technically challenging aspects, apart perhaps from the intricacies of compact shelving. Digital storage and preservation require more attention.

For digital content, libraries retain many of the virtues as institutions of storage and preservation that they had for paper. They are numerous, functional, solvent, long-lived, popular, and motivated by public service and the greater good. All these factors make them excellent stewards of our inheritance. The big difference now is that instead of maintaining hundreds of thousands of collections, we will have just a few, with suitably robust and durable backups. Another looming question is whether, as content is made available at the expense of the producers, taxpayers will as readily pay for libraries whose main function no longer is access, but storage and preservation.

Digital backups already exist, at least for some periodicals. Most (68%) diamond open-access journals have no preservation policy.[47] But other institutions help plug such gaps. The felicitously acronymed LOCKSS (Lots Of Copies Keeps Stuff Safe), its variant, CLOCKSS (Controlled LOCKSS), and Portico are examples. These enterprises synchronize digital journals with data from publishers, updating content and migrating it when software changes. As dark archives, they are not normally accessible. But with certain trigger events—if a publisher does not respond to information requests or fails altogether—the backup supplies the data instead.[48]

All well and good, and libraries, including the New York Public Library (NYPL) and Indiana, are involved. But how exhaustive is the digital archives' coverage? In 2007, only some 50 journals participated in LOCKSS, and 30 in Portico.[49] In 2011, Columbia and Cornell's libraries discovered that LOCKSS and Portico preserved only 15% of their e-journals.[50] But a decade is an eternity in the digital world. By 2021, that had mushroomed to almost 15,000 journals for LOCKSS, 26,000 for CLOCKSS, and 35,000 for Portico.[51] Nonetheless, however worthy these institutions, they do not have the heft, likely duration, and reliability of something like the Library of Congress or the NYPL.

So far, we have ignored most of what an institution to preserve our digital patrimony must include. As born-digital content becomes the bulk of human output, protecting it is urgent. Start with the elephant in the room, the internet itself. The Internet Archive's Wayback Machine has been archiving web pages since the 1990s and now contains half a trillion. But even it cannot be omnivorous. Sites can opt out, and it provides only episodic snapshots.

Other agencies also preserve the web, but they are often country-specific in accord with national electronic legal deposit laws, and often bizarrely restricted in terms of who may access, where, and how, all while imposing restrictions that outlive copyright itself.[52] As legal deposit institutions, libraries do not do digital materials

justice. Hemmed about with regulatory limits modeled on print-on-paper, they are prevented, for example, from allowing viewers to access content from anywhere other than reading rooms or from having more than one person use material simultaneously. Based on fair use, private organizations, such as the Wayback Machine, have proven more capacious. But, they are also more impermanent than government institutions.[53] More is needed.

Further down the ranks of what must be archived come the predatory journals. In the most flagrant cases, having collected fees, predatory publishers have no interest in incurring further expenses, and minimal concern with durable archiving. Some other entity must shoulder that task, as it must also for well-intentioned journals whose long-term survival is questionable.

The same holds for the vast ocean of self-published books washing in via Amazon and other outlets. (We touch on the quantities shortly.) As with the predatory journals, it is easy to be snobbish and dismiss self-published tomes as rubbish whose loss means little. Granted, the value of some self-published books can be hard to spot. Many are personal testimonies to transformative events, whether religion, war, or encounters with aliens. They often unmask nefarious occult conspiracies. Others deal with topics dear to the author's heart, ranging from manuals on peripheral technologies to histories of localities and their sports teams.[54]

For historians, such ephemeral material often delivers fruitful insights. Speaking from the *Volksmund*, giving voice to the collective id, plodding novels, sentimental poems, pedestrian memoirs, and confessional autobiographies tap directly into rich historical veins. Historians are less interested in those who supposedly transcend their era than in those who exemplify it.

Norbert Elias based his influential idea of the civilizing process, the behavioral self-domestication underpinning much of early modern history, on etiquette manuals housed at the British Museum.[55] *The Other Victorians*, Steven Marcus's classic account of

the far-from-prudish sexual habits of our much-maligned predecessors, took evidence from a treasure trove of pornography, carefully amassed by Alfred Kinsey at his Indiana University institute.[56] Robert Darnton's studies of pamphlet literature during Old Regime France were possible thanks to the fortuitous preservation of the papers of the Société typographique de Neuchâtel in the municipal library.[57] To the past's nitty-gritty, ephemera are a well-paved route.

Future historians, cultural sociologists, evolutionary psychologists, and literary theorists will have a field day mining the unmediated outpouring of digital self-publication. In the past, when storage was expensive, selecting what to keep was undertaken at the time by people steeped in the culture given voice by the material they were judging. Culturally blinkered by the conventional prejudices of their era, the archivists of any given moment are never the best to know what to save or discard.

In digitality, storage is all but costless. Everything can be saved, and no selection needs to be made now. That can be left to future scholars, who will have a more clear-eyed vantage to judge from. Yes, they will suffer an embarrassment of riches. But they will also enjoy search and filtering tools unavailable to us. They will be savvier in navigating the data cornucopia dumped in their laps. Much better to leave decisions of what is important about today to our descendants.

Already now, with vast and growing research output, complaints are heard that the average article is uncited, with the implication that it is also unread.[58] That follows from the power-law distribution of attention—a few hog it, most receive none. Ever since we have had more content than could be consumed in one lifetime, that has been true. A small fraction of library holdings makes up the most circulated works.[59] Pop songs as well as scholarly articles follow such distributions. It is inherent in how culture is consumed. We would not have bestsellers without leastsellers. Our increased storage capacity then multiplies the effects of surfeit. With free and

unlimited storage, everything can be saved. From that, it follows that the average content will never be seen by anyone except its author and a typical bit of information never by anyone at all.[60]

Publishers

Digitality has weakened all disseminators. Publishers, the recording industry, and film studios have been sidestepped—by illegal downloading, but more broadly by the leaching-away of their once-exclusive role as content purveyors. Yet, what looked dire just a few years ago has become a shift in their roles, not the end. With Pirate Bay, Megaupload, Napster, and other pirate sites shuttered, rights-holders have cut deals with new disseminators willing to license and pay royalties—Spotify, Apple Music, Netflix, Amazon Prime, and the like, some of which also produce content themselves.

Corporate rights-holders have tamed the frontier conditions that prevailed two decades ago. Clunky, overpriced physical media (VHS, Betamax, DVDs, CDs) have given way to streaming's seamless cornucopia. Rather than paying $20 for a CD with a couple of listenable tracks or for a movie DVD, customers sign up for a celestial gorge-yourself buffet—unlimited streaming of tens of millions of songs and thousands of films and shows, all for the price of a couple of lattés monthly. Who would have expected bingeing to have become something good? If there is one annoyance in this cornucopia, it is the music streamers' insistence that every chunk of every musical form is a song—whether the latest K-pop ditty or the last movement of Beethoven's Opus 111.

We finally come to the publishers. One might have expected them to suffer like bookstores and libraries. But they are doing better than expected. Indeed, as seen, scientific journals' juicy profits have exacerbated the libraries' affordability problem. Some distinctions

are needed, first and foremost between books and journals. We start with books.

For conventional books by established publishers, the industry has flatlined rather than nosediving. In music and film, new competitors (Spotify, Netflix) operating in novel media have eaten the established industries' lunch. For books, in contrast, the publishers have managed to keep control over e- and audiobooks, exploiting the efficiencies of digitality. Overall, their revenues have remained stable, while profits increased.[61] Inveterate complainers, publishers are less badly off than they claim. Having shuttered bookstores and libraries, the Covid pandemic left publishers' mail-order departments and the online sellers an uncontested field.[62] Sales spiked as readers huddled at home, occasionally tiring of television. Educational sales were down, digital up, and print slumped, then rebounded. In Sweden, where no lockdown shut bookstores, Bonnier did so well that it decided to pay back government pandemic assistance.[63] University presses that opened their lists to online reading for free were gratified to discover that this sparked increased printed book sales.[64]

Even though other media have become fiercer competitors for the average eyeball, never have more titles been published. In France, while the number of books sold declined, titles increased from 63,000 to 68,000.[65] Italian titles grew 18% from 2010 to 2017.[66] In the Spanish-reading world, new titles climbed steadily during the millennium's first decade, plateauing thereafter.[67] Even the oft-forecast demise of the academic monograph appears to have been exaggerated. The numbers of US humanities monographs have either slightly declined or slightly increased, depending on whom one asks.[68] Among the four largest British academic presses, the trend was steadily upward for the twenty-first century's first decade.[69]

Scientific publishing has done very well, and that includes above all the academic journals. Among the top ten publishing

conglomerates globally, seven are either scientific or educational, only three trade. In 2019, these seven houses' revenues were thrice those of the trade presses in the top ten. Put another way, of the 55 largest publishing conglomerates globally, almost 60% of their revenue came from professional and academic content, with trade books contributing only about 20%.[70]

Taking a step back to encompass the entire publishing universe, new vistas open up. As digitality has unleashed an outpouring of creativity beyond the conventional arenas, the legacy publishers no longer dominate. While the number of traditional books has remained steady, self-published editions—mostly print, but also e-books—have avalanched. These are the kind of works brought out by Amazon (CreateSpace and Kindle Direct, amalgamated in 2018), Smashwords, Author Solutions, Lulu Press, Blurb, and similar ventures.

Though tricky to pin down, the numbers of self-published works are huge. The US publishes slightly more than 300,000 conventional new titles annually, but over four million nontraditional works were issued in 2010. This quieted somewhat after 2010, but the number has remained well over a million annually.[71] Self-published works surpassed conventional output first in 2008.[72] They then more than tripled, from 461,000 in 2013 to 1.6 million in 2018, and that is counting only those issued ISBN numbers.[73] Since one title may have several IBSNs, one for each format, this may overcount new books, but not every work receives an ISBN.[74] In 2018, the US had 1.55 million print self-published books and 130,000 e-books, a total of 1.68 million.[75] CreateSpace was the largest venue, with 1.4 million works in 2018, almost 20 times the next, Smashwords.[76] E-book growth has been similar. In the first five years of digital self-publishing, up to 2015, that sector went from zero to one-third of all American e-book sales.[77]

Yet many such works are not new. Some self-publishing houses specialize in bringing forth reproductions of public-domain works

as print on demand.[78] In 2010, BiblioBazaar issued 1.4 million such items. Together with General Books and Kessinger Publishing, they dominated the nontraditional market in 2010 with 2.7 million works.[79] The supposed mystery of why more books from 1850 are available on Amazon than books from 1950 turns on this.[80] Without these cheap and cheerful reprints, a trek to the nearest research library would be the only way to read obscure nineteenth-century works.

Whatever the precise figures, the point is that self-publishing continues to dwarf its conventional peers, with well over triple its output. This describes the US situation. Equivalent outpourings are not found elsewhere, or possibly non-American output is channeled via Amazon regardless of its origin, showing up in the US figures. The disparity is evident in the volume of ISBNs issued. For 2018, this was 3.4 million in the US, but only 186,000 in the UK, 140,000 in Germany, and even fewer elsewhere.[81] Other nations have also been less inundated by the self-publishing tsunami. In Chile recently, no more than 13% of books have been *autoediciones*.[82] In Latin America more generally, 12%.[83]

Under totalitarian censorship, authors wrote privately for their desk drawer—works that might one day be issued after a regime change or perhaps smuggled abroad. The self-publishing wave shows that everyone, not just dissidents in autocracies, has manuscripts squirreled away, awaiting new outlets to channel their creative urges. Who is writing, for whom, about what? Such questions remain largely unexplored. And how will these works be preserved for posterity?

Most self-published original works are fiction. More precisely, they are what the industry calls genre fiction, output that can be classified into easily identifiable categories so that consumers know what to expect—romance, mystery, horror, science fiction, and fantasy. Such genres dominate self-published books. Of Smashwords' works, 87% are fiction. By some measures, erotica alone makes up

almost a third of e-books, with no less than one in ten dealing with specialized tastes, such as bestiality or incest, not to mention sex with billionaires and adult diaper eroticism.[84]

And erotica is to be distinguished from romance, another elephant in the room. On Smashwords, 11% is erotica, another 49% romance,[85] while 40% of Kindle books are romance.[86] One segment of the genre market where conventional publishers still hold the upper hand is thrillers—home of well-established brands such as James Patterson, John Grisham, Dan Brown, John Connolly, and their ilk.[87]

Conventional publishers' output, in contrast, is less fiction, either literary or genre. In 2018, fiction accounted for 27% of UK publishers' works, 25% of the French, 19% of the Italian, and 25% in Latin America.[88] Self-publishing's output is even more at odds with the libraries that collect conventional works. In North American major research libraries, of post-1923 holdings, only 8% are fiction.[89] Whether that reflects publishers' output or library collecting habits is unknown. Very few scholarly books are self-published.[90] The lack of peer review casts a pall over ambitious academics' hopes of clawing their way up the hierarchy. But on the other hand, blogs, tweets, and other auto-productions—self-publishing by the day or minute—are becoming ever more common among scholars. So possibly the trend will accelerate and broaden to new media.

Athena Unbound

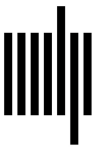
Why and How Scholarly Knowledge Should Be Free for All

By: Peter Baldwin

Citation:

Athena Unbound: Why and How Scholarly Knowledge Should Be Free for All
By: Peter Baldwin
DOI: 10.7551/mitpress/14887.001.0001
ISBN (electronic): 9780262373968
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from the author



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Baldwin, Peter, 1956– author.
Title: Athena unbound : why and how scholarly knowledge should be free for all / Peter Baldwin.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022027103 (print) | LCCN 2022027104 (ebook) | ISBN 9780262048002 (hardcover) | ISBN 9780262373951 (epub) | ISBN 9780262373968 (pdf)
Subjects: LCSH: Open access publishing. | Scholarly electronic publishing.
Classification: LCC Z286.O63 B35 2023 (print) | LCC Z286.O63 (ebook) | DDC 070.5/7973—dc23/eng/20220628
LC record available at https://lccn.loc.gov/2022027103
LC ebook record available at https://lccn.loc.gov/2022027104

10 9 8 7 6 5 4 3 2 1