Introduction: Homes for Good (Orphan) Books

Peter Brantley

In the last few years, books have energetically joined smaller niches of content in the digital uplift of the twenty-first century. Part of this sudden transition was the development of imaging systems that permitted individuals and companies to digitize at a pace never previously imagined. Technology, as it so often does, turned in part from facilitator to catalyst.

Yet technology never determines applications, and it was the release of Google's vision for a searchable online repository for the world's printed books that ignited the imagination of countless people and organizations, with a dollop of uncertainty on the part of every principal involved: authors, whose intellectual creations became suddenly transformed into a binary index; publishers, who wondered what role in distribution they would have in a visible future where Amazon and Google could surface their most valuable assets; libraries, who wondered if there would be a role for anything beyond warehousing print books gathering dust on shelves; and scholars, who worried about the quality of the book's online presence while exulting in the new capacities to find information, and combine it with a heterogeneity of sources never before imagined.

As libraries wonder which of their volumes they should digitize, and publishers contemplate the establishment of their own online repositories for digital content, only Amazon, and more importantly Google, combine an encompassing aggregation of content with a user-facing presence. This may yet change as backend providers of publishing services such as Ingram contemplate their role in a world increasingly revolving around media distribution, but their entrance into mindshare would lag behind the somber weight of established usage in online search (Google) and commerce (Amazon). Google, particularly, stands uniquely able to offer

LIBRARY TRENDS, Vol. 57, No. 1, Summer 2008 ("Digital Books and the Impact on Libraries," edited by Peter Brantley), pp. 1–7

a combined mass of public domain, out-of-print, and in-print volumes, combined with scholarly journal material and overlays of temporal and geographic information that bring previously high-end supercomputer-class data-mining to any device that is able to join the world's increasingly pervasive network.

It is worthwhile noting that it is this—the growing presence of the Internet in people's daily lives—that stands to profoundly disrupt scholarship and reading. It is not the issue merely of having content digitized, but rather the near-constant availability of that information combined with the means to find it with search, that enables disruptions in how we use books, and ultimately how we write books. Network ubiquity makes Amazon's Kindle ebook reader distinctive enough to prove that content acquisition can be serendipitous; that Googling on a powerful mobile phone or computing device can be deeply utilitarian rather than merely demonstrative of cool technology; and that research and education are being profoundly reshaped by the ability of users to engage with information from anyplace, and at anytime, where they can provide sufficient attention to the task.

Libraries, publishers, authors, scholars, and readers all hitherto were located in a multifaceted milieu in which time for learning and enjoyment were limited not merely by the constraints of daily living, but by the availability of acquiring the means by which those goals could be accomplished. The lessening of those barriers warps a set of organizational relationships, economic environments, social practices, and legal frameworks to the point where they all must gently yet persuasively lean like bamboo in the wind, or snap like brittle kindling.

Like many, I've recently been thinking a lot about the availability of books in online searchable repositories, and the likely outcomes for publishers, libraries, and the public. I have even been considering the impact of a rapprochement between publishers, authors, and Google over books whose availability is most savagely contested, largely because their legal status has been brought into a hazy dawn of uncertainty by the startling recent shifts in availability that catch them stranded between public goods and private property.

A significant portion of these implicated works are likely to be out-of-print, of uncertain copyright status, and no longer present in any publisher's archive—available only in the less-visited shelves of the largest research libraries. This substantial category, numbering in the millions of books, would undoubtedly incorporate a large number of what are called "orphan works," where the presence of any identifiable copyright owner in the work, or its constituent parts, is not known and is resistant to easy resolution as a result of poorly recorded mergers and acquisitions, lost archival contracts, publisher insolvency, and myriad other reasons. In turn, some of this orphan material is almost certainly domain; the original copyright never renewed, and long since expired.

What might break the logjam of access to these works, and frustrate a solitary reshaping of access that Google might obtain through a private agreement with publishers and authors? A digitization agreement involving libraries, either public or academic, and a suitable hosting service that would make this lost material broadly available on reasonable terms, with clear benefits facilitating research and education, could potentially present a strong counterpoint.

The content could be made available through various monetization arrangements, including subscription-based individual access that would support features such as print on demand or digital lending, and licensed access with payment tiers for universities, high school libraries, and similar institutions, which might also be willing to pay a premium for a renewable, local-hosting option. (In fact, if this material would be provided through a charitable nonprofit organization, hosting fees could be quite low.) Alternative arrangements, such as those pursued by the high-energy physics community's SCOAP3 journals project, might also be feasible, depending on the nature of interested parties.

A portion of fees could be escrowed in a common fund for allocation to rights holders should any come forth with the necessary proof of copyright retention. A basic access level, sans advanced features, to orphans and proven public domain books could be extended to registered cardholders of public libraries as a free public service (this would have the secondary benefit of driving use of a trusted OpenID through library participation at a community level).

Books that have newly apparent IP holders could be taken down through a simple, authenticated request mechanism, or alternatively retained in the delivery system with a different share of income returned to the identified author and/or corporate parties. The escrow fund would provide a modest, yet reasonable compensation for the works' past use, partially offset by the virtue of the hosting service's implicit discovery fee. Easily accessed lists of available works, for example, through publication of OpenSearch RSS feeds, would assist possible copyright owners in finding bereft works; transparency would increase trust for all parties.

What might be most challenging in this scenario would be finding an appropriate mass of books that would be both coherent and compelling; that would include a significant enough number of out-of-print orphans to be of use to readers; and where large libraries might hold sufficient numbers of these books to be able to mobilize for their digitization. Perhaps a subject with an accumulation of desirable material might best meet these parameters: for example, works of U.S. history, or autobiographies, or American literature. Alternatively, a discipline with a long history, such as anthropology or economics, might embolden a tribe of scholars and interested amateurs to make organization for online access compelling.

And yet, would such a service work? While this might be an attractive vision, it calls for the quiet eye of an otherwise vertiginous storm to hover endlessly and benevolently over an island where there may be few inhabitants left living.

Any new online digital library for books will have to demonstrate a sustainability that will sink its roots in a very different economy than the one in which its traditional brethren are situated, where a social understanding among universities, governments, and communities underwrites the cost of delivering information to others. This silent but pervasive subsidization of access is reminiscent of the ubiquity of the Internet—a network whose low barriers of access has dramatically hamstrung libraries' participation in the cultural marathon of preservation and access that began only a few short centuries ago.

Google and Amazon have made their own economies, provided their own subsidies, all calculated to a point where the larger enterprises find a growing whole against which profit yields attract dividends. These economies, based on advertising and commerce at astronomical volume, are out of bounds for a new generation of online public services, which must ask their users to license a right of access.

We are at a great point of note in history, where the direction of change has not yet been made solid. If Google provides an unfettered, rich, and satisfying access to books and the ties among them, even if their quality is at best adequate and never culminating, what compulsion will be sufficient to underwrite the foundation of a new public service for information? What new form of subsidy would arise to populate the endeavors of those who have acquired the word, image, and frame, and made it accessible as a public good? Is there a counterpoint to Google's plebian and pervasive search? Google's search traces patterns between texts and tracks the paths between people and texts that are never forgotten.

These are but nothing but specific instances of the larger questions affecting the aging industries of content production, discovery, and use. What role will publishers have in their quest to make that information as widely available as possible? Will writing itself be transformed through novel ways of accessing information, distributing creation through acts of search like so many seeds from windswept flowers? Will reading turn into a peripatetic and jarring tryst with commentary instead of solitary devoted contemplation with a well-spun inquisition?

Of course none of the writers in this volume can provide an answer to these issues. Their voices are scattered, and populated across differently imagined ecosystems of thought, yet they are not discordant. Rather, and shockingly so for this editor, they are echoes of the same questions, perhaps posed with different inflection and of different audience, but of the same form.

Jason Epstein, a cofounder of the *New York Review of Books*, writes pas-

sionately about how access to digital access will transform access to books, while providing reinforcement to the life of the book, as a physical product, that will be as revolutionary in its implications as the original printing press. While certain categories of content—reference, travel, cooking—might be entering into a nativistic digital environment where information can be more quickly and easily obtained, with contemporaneousness assured, long-form narratives will continue to be preeminent for epistolary arguments in fiction and nonfiction realms. Digital access also, in other words, encourages the formation of digital printing presses that can be nearly as distributed as network access to a library of infinite volumes.

Juliet Sutherland of Distributed Proofreaders takes us through the many layers of the contents of a book as it is converted to digital form. First, page images are often attractive high-fidelity replicas of a book's physical appearance, but usually offer poor usability profiles when accessed outside of sufficiently large computer displays on high speed networks. For users of portable devices, or those accessing digital libraries from areas with restricted network speeds, obtaining the actual text of a book, as opposed to the image of the text, permits a far more useful and immediate application. Text can be resized, spoken, and flowed in a fashion that escapes images. Textual representations also enable powerful content discovery through Internet search, and subsequent multisource content integration, that Google compellingly provides for the open Web. Sutherland discusses the difficulties in transforming image to text, and of then eliciting from it useful semantic information that can assist in the fundamental reinterpretation of how books are read and used.

Michael Jensen discusses the profound changes ripping publishing and libraries apart, and notes the important cultural and organizational stresses resisting the movement of these organizations toward a future that is more user focused. Reimagining publishing, and reimagining libraries, will inevitably coincide with their rebirth as bundles of functions and services quite unlike those they have presented historically. An increasing need for "speciation"—growth in the exploration of new habitats for publication and collection—will be required to adapt to changing worlds of access, distribution, and creation. Jensen posits a world where the best and most unique assets of libraries—curation and selection—are turned inside out and distributed across the network as a pervasive reader's guide of pervasiveness and subtlety, refined through use and commentary. In turn, this might drive a new role for publishers that acquire, cultivate, and polish those samples of thought and communication, which are most worthy of endorsement and repackaging in print and electronic forms.

Turning to authorship, Laura Dawson explores the opportunities for individuals and organizations to publish directly without the mediation of traditional publishers. The growth of self-published literature—fiction, poetry, cookbooks, travel memoirs, photography, and nonfiction—has ex-

ploded as the ease of assembling content packages through simple Web services has deskilled the construction of publishing supply and distribution chains. Due to its construction outside existing means of production, the integration of self-published literature into traditional product discovery venues—bookstores and libraries—has proceeded awkwardly. However, the flood of books will no doubt find a more comfortable home on the Internet as full-text searchable, native digital editions rather than carbon-based tomes. The discovery and recommending of works that otherwise would be assumed amateurish and deprecated will increasingly be evaluated on par with those from mainstream publishers. Indeed mainstream publishers may eventually endorse self-published literature as an accompaniment to their own processed products in the same fashion that academic journal publishers have begun to explore the benefits of a wider market of preprints and postprints.

In a deeply complementary essay, Sara Lloyd of Pan Macmillan UK delves into the new roles that publishers must investigate. Publishers' traditional tender of artistic creation and distribution are becoming more easily and fully accessible, supporting layers of participation unimaginable in publishing houses short years ago. The book, she notes, will inevitably be more porous both in its creation and its usage, and how publishing supports that porosity through the development of frameworks and communities where readers and writers can participate will shape publishing's contours in the future. In an era where Google provides discovery and becomes the preeminent house for content distribution, the publishing endeavor must shape and enable creativity and engagement that perforce shatters their age-old alignment along a linear path diverging scarcely from creation to editing to publication to sale, and there to be done with it. On a landlocked continent no longer, in concert with Jensen she notes the urgency of developing specializations around competencies that foster use and acquisition for a wider range of practices surrounding new media.

Lehmberg et al. provide a glimpse of the complicated legal framework that will ultimately have to surround digital books that can be torn apart and reassembled in various sized chunks, pursuant to whatever license terms govern their use. Publishers are long used to complex rights situations: even a simple trade book will have rights associated with the primary author, perhaps a jacket photographer; there may be illustrations, maps, and other content inside. When a world of content bricolage is embraced through tools that permit readers and scholars to lace and corset new works, it is less clear whether these new texts are works of creation, or databases of content. Further, when books "know" when and how they have been read or browsed, and by whom, what are the privacy ramifications for societies that are increasingly under constant surveillance. Will professors

soon be able to learn which pages or sections of textbooks their students actually read?

Sukovic discusses the inherent migration from hierarchical to "rhizomatic" or network-based topologies for information modeling and access. Using a representation of a "lemon" as an example, she asks how we might imagine provenance, authenticity, and multiplicity of interpretations as we enter a world where content is more easily modified and recontextualized. To "know" a book will not, therefore, be simply a matter of reading its contents and a few scholarly commentaries, but sensing those aspects of its use, its interrelationships with others, and its creative progeny and their own contexts of interpretation that are relevant to the consideration of the scholar, or even a casual reader. Island hopping from one domain to another through a network of connections that are both explicitly available through metadata descriptions, and implicitly available through user-generated content and usage paths, will ultimately make shelf browsing look as antiquated as perusing through card catalogs in a preautomation library.

These perspectives are unique but intertwined with fascinating surety; we are reaching a point where horizons can be seen from the decks of our vessels, but whether they indicate islands or continents we are not yet close enough to tell, nor whether they are hospitable to the efforts so deeply inquired of here. In a poor echo of Thoreau, we must be the Mungo Parks, Frobishers, and Lewis and Clarks of our own latitudes yet unexplored. Readers, authors, publishers, libraries—let us all pile our new ventures sky-high as a sign of our explorations. This collection is but one vessel, but I am pleased that it is a swift and sturdy one.

Peter Brantley is the executive director for the Digital Library Federation (DLF), a not-for-profit international association of libraries and allied institutions. His background includes significant experience with research libraries and digital library development programs. Before joining the DLF, he was director of technology at the University of California's California Digital Library; he served previously as the director of IT services for the Libraries and Press at New York University. He blogs at http://toc.oreilly.com and http://pubfrontier.com.