Apr 29, 2016, 12:42pm

The Future Of Open Access: Why Has Academia Not Embraced The Internet Revolution?

Kalev Leetaru Contributor ⓘ

I write about the broad intersection of data and society.

Card catalogs in the Main Reading Room of the Library of Congress during a 2012 open house. (Alex Wong/Getty Images)

More than any other technology, the web has revolutionized access to the world's information, putting everything from recipes to encyclopedias to books to news at the fingertips of anyone with an internet connection anywhere on the planet. The web's role in democratizing access to global information has made it a poster child for the power of technology to advance society. Yet, given the modern web's

roots as a platform for sharing academic research, it might be surprising to learn that the vast majority of the knowledge and advances generated by the world's universities has been left out of this information revolution.

One of the most remarkable aspects of the story of the web's evolution is that the collective output of the world's universities has remained largely absent from the open online world, even as most other forms of information have shifted to some form of open online access. In the case of encyclopedias, entirely new forms of collaborative knowledge documentation like Wikipedia have emerged, while journalism has shifted to free advertising-supported distribution and even music and videos are increasingly legally available through ad-supported streaming services or affordable licensed download services.

Academic papers, the lifeblood of the scholarly world of academia, have resisted this transition. To those outside academia it might be surprising that most universities don't publish all of their books, papers, presentations and course materials on their websites for the world to access. Even at public universities, where the salaries of faculty and staff and the operating costs of the institution are often heavily subsidized by taxpayer money, either directly by their states or indirectly through grants from NSF, NIH and other federal agencies, the majority of the research output of the institution is not publicly accessible.

Instead, much of the world's scholarly knowledge is owned and controlled by commercial enterprises that operate the journals that academic researchers publish in. Many journals even require researchers to sign a legal agreement transferring the copyright of their paper to the journal and prohibiting the researcher from sharing the paper with others or reposting on their own personal website (though a growing number of journals are permitting so-called "preprint" reposting on personal websites after a particular embargo period).

These commercially-owned journals recruit editors (who may be salaried professionals, academics paid an honorarium, or frequently simply unpaid volunteers) to coordinate the journal's operation, while legions of unpaid volunteers recruited from academia review each submission and provide detailed commentary on whether to accept for publication or reject. They also hire

proofreaders, typesetters and others to publish each issue and manage the final content delivery platforms.

In return, journals charge very high subscription fees for access to those publications, charging tens or even hundreds of dollars in some cases to purchase access to a single article, with typical fees of $20-50 to read an article if your university does not subscribe to that particular journal. The extremely high cost of journal subscriptions means that even top universities must be very selective in the materials they acquire for their researchers. As the Chronicle put it, this leads to a situation in which "when academics use privately held databases ... the texts that they'll find depend on the terms of their subscription. If two academics run an identical search, they might get different results." As the director of the Digital Library Federation put it, "Your database subscription level, and all the invisible machinery between you and what should be public-domain information, determines what you can see, and therefore what you can say about the past."

The high cost of journal access and the cumbersome and complex interfaces that libraries provide to their subscription holdings, has fed an underground movement to pirate academic literature. While news headlines about online piracy tend to focus on illegal downloading of music tracks or streaming of videos, the academic community is facing its own pirating crisis.

Yesterday Science published a fascinating behind-the-scenes look at Sci-Hub, one of the most infamous academic pirating sites, which provides free access to more than 50 million illegally acquired papers. One of the most fascinating findings is that its download traffic comes not exclusively from the developing world for which journal subscriptions are often claimed to be inaccessible, but also extensively from major Western universities which likely have legal subscriptions to the journals already. One of the reasons for this, the article claims, is the cumbersome and difficult-to-use web portals that university libraries provide to their holdings, making it incredibly difficult to locate a paper even if the university has a legal subscription to the journal. Having spent more than a decade and a half in academia at multiple institutions from public to private, I

can personally attest to just how difficult it can be to navigate library portal systems to locate a particular paper.

The extreme cost and paywalled access to information of traditional journal publishing has led to the open access revolution in which grant funding agencies (such as NIH) are increasingly requiring publications stemming from research they support to be made publically accessible. More and more journals offer open access publishing models, but most transfer the cost of publishing to the author. In the traditional journal world, publishing a paper in most journals was free, with the costs of publication paid for by journal subscribers. In the open access world, journal articles are free to read, so authors must instead pay to publish their work, traditionally costing hundreds or even thousands of dollars per article. Publishing in PLOS journals ranges from $1,495 an article to $2,900 an article, while the average open access fee for Elsevier journals according to their current journal pricing list is $2,200 an article with at least 23 journals charging $5,000 per article.

Alternatively, many academic disciplines permit preprint publication of works in arXiv and similar scholarly repositories, as well as institutional archives. The arXiv repository in particular is often cited as a model for academic publishing, hosting just under 1.14 million papers that have been collectively downloaded nearly 720 million times. The site is hosted by Cornell University with an annual operating budget of around $826,000 and is funded through support from Cornell, the Simons Foundation and institutional memberships. Unlike traditional journals arXiv does not provide a full peer review process and submitting to the site is considered a "preprint" rather than a fully peer reviewed publication counting towards promotion.

Academia has been on the forefront of the revolution to open the world's information, from open source software to open access journals, so why is the majority of their professional output still absent from the open revolution more than 20 years after the dawn of the modern web? Perhaps the simplest answer is that in academia promotion and tenure are still tightly linked with publishing in top-tier journals, which are largely well-established commercial venues. A young

professor just starting out in their academic career who chooses to publish exclusively in open access venues will find it difficult in most disciplines to match the impact factors and prestige of the commercial journals of their field. The high cost of paying to publish each article will also necessitate allocating precious research funds to publication fees rather than graduate student salaries or purchasing equipment, placing the researcher at a competitive disadvantage with their peers.

Faculty who choose to simply post their research on their personal or institutional websites or in archives like arXiv will find that those papers do not count towards their performance reviews for tenure and promotion. In academia, only publications in peer reviewed venues are traditionally counted as true publications and it is difficult when creating new fully open access journals to convince faculty to submit their valuable publications to a venue without an established impact factor.

In the end, someone has to pay for research to be published online, whether it is potential authors or members of the general public wishing to read and benefit from the collective learned output of the world's universities. Moreover, how do we ensure that what is published today in open access journals is preserved for decades and centuries to come? Where will the funding come from to ensure that a paper published today in 2016 is still accessible in 2116 if there if the only source of revenue to support its accessibility is a single one-time payment of a few thousand dollars at the time of publication?

While nearly every other form of informational output has been reinvented in some fashion in the internet era, academic literature has remained steadfastly locked in the centuries-old subscription format, paywalled away from all but those who can afford to purchase access (though some journals and publishers offer discounts for authors/researchers in low-income countries).

One possible solution is to transfer the burden of subscription costs to the national government as a service to its citizens. Last year the government of Egypt launched an ambitious initiative called the Egyptian Knowledge Bank, described as "the biggest digital library in the world, housing contents of the most

prominent publishing houses all over the world such as National Geographic, Discovery Cambridge, Oxford, Reuters, Britannica and others." Instead of the American model in which each individual university purchases its own journal subscriptions and ordinary citizens have no access at all, the Egyptian government purchased national site licenses to a wide array of content, making it freely available to its citizens. However, even this dream has encountered the reality that simply placing content in the hands of citizens is not enough if 70-80% of the material is published in a language they cannot read (an issue for the Internet at large).

As the drumbeat of open access continues to grow, the fierce debate over the future of how academic research is published and distributed will only rage louder. In parallel, as the trend towards open access expands to data sharing and replication, the pressure to change how academia does business will reach a breaking point where change will become inevitable. In the end, it is a fascinating commentary that the world of academia, from which the modern web sprung, has been among the most resistant to change and one of the last to embrace the internet revolution.

Based in Washington, DC, I founded my first internet startup the year after the Mosaic web browser debuted, while still in eighth grade, and have spent the last 20 years working to reimagine how we use data to understand the world around us at scales and in ways never before... **MORE**