# Information wants someone else to pay for it: Laws of information economics and scholarly publishing

Micah Altman [a,*] and Marguerite Avery [b,c]

[a] *Director of Research, MIT Libraries, Massachusetts Institute of Technology, Cambridge, MA, USA*
E-mail: *escience@mit.edu*

[b] *Director of Scholarly Communication at Hypothes.is, San Francisco, CA, USA*

[c] *Research Affiliate at Program on Information Science, MIT Libraries, Massachusetts Institute of Technology, Cambridge, MA, USA*
E-mail: *mavery@mit.edu*

**Abstract.** The increasing volume and complexity of research, scholarly publication, and research information puts an added strain on traditional methods of scholarly communication and evaluation. Information goods and networks are not standard market goods – and so we should not rely on markets alone to develop new forms of scholarly publishing. The affordances of digital information and networks create many opportunities to unbundle the functions of scholarly communication – the central challenge is to create a range of new forms of publication that effectively promote both market and collaborative ecosystems.

Keywords: Information, economics, scholarly publishing, affordances

## 1. The state of scholarly communication: More and less

This year marks the sesquarcentenial anniversary of *Philosophical Transactions of the Royal Society*, a notable milestone upon which to reflect upon the evolution of the scholarly communication ecosystem. In the beginning, three hundred and fifty years ago, scholarly communication was a relatively straight-forward proposition that entailed the collecting of individual research writings into a volume that was reproduced or published. This new method of communicating research findings to a dedicated audience was a vast improvement over the circulation of individual letters written by individual scholars on individual topics. This shift from the communicating of research findings from a one-to-one model, to one-to-many model (more accurately 'some' in the original instance) represented a major step forward in the evolution of what we now identify as scholarship. Centuries later, this is still the standard for authoritative scholarly communication, with nearly all developments and innovations replicating this original print model.

Based on its present state and trajectory, the future of research and of scholarly communications is "more". Some of the components of "more" are readily apparent – there is more data being produced,

---

*Corresponding author: Micah Altman, Director of Research, MIT Libraries, Massachusetts Institute of Technology, E25-131, 77 Massachusetts Ave, Cambridge, MA 02139. E-mail: escience@mit.edu.

disseminated and accessed than ever before with ninety percent of digital data being produced in the previous two years [35]. By some accounts, scientific publication output doubles every nine years, with one analysis stretching back to 1650 [42]. In some areas, the effect of the new data is so substantial that the entire research evidence base of fields appears to be shifting: in astrophysics which relies upon massive data sets for its collaborative global research output, CERN alone generates one petabyte of data daily; in genomics, where efforts to synthesize the human genome are reliant upon massive data sets, we know that the human body contains sixty zettabytes of data [11,47]. This doesn't begin to address the firehose of data spilling out from commerce and industry that has become the lifeblood of disciplinary studies in business and management, in the social sciences such as sociology and economics, and law and government regularly engage with digital research objects such as big data.[1] Even the more traditionally analog humanities disciplines like history and literature, which tended towards print artifacts as the only primary sources worthy of consideration, have (re)assembled under the mantle of Digital Humanities to directly engage with data-driven analysis of their respective corpuses. A more recent development is the distinction between Big Data Digital Humanities and Small Data Digital Humanities, by which the former is distinguished by one researcher as "large or dense cultural datasets, which call for new processing and interpretation methods", in comparison to the latter, "regroup more-focused works that do not use massive data processing methods and explore other interdisciplinary dimensions linking computer science and humanities research". However both are a departure from conventional humanities pursuits [18].

The digital impact on the scholarly communication ecosystem, which started with a spark and flash in ARPAnet, has been slowly burning and only now, decades later, is rapidly accelerating. With the increase in scholarly content produced and available, growth in the volume of publications seems inevitable, and this number has grown consistently at about three percent annually for over two centuries. A 2015 report, which identified twenty-eight thousand and one hundred active peer-reviewed journals publishing about 2.5 million articles a year, attributed such steady growth to a comparable increase in the number of researchers [43]. As of May 2014, over 114 million English-language scholarly papers were available on the Web [20] from circulating scholarly journals (and mega journals) with over ten thousand of these publications self-identifying as open access (with APC – article processing charges) in some capacity.[2] And if we expand content from publications in traditional channels to a more capacious definition of scholarly communication, the impact of the digital space is much more apparent.

## 1.1. Diversification in publication

A sharp increase in the type and number of publication outlets appears to be the most significant change in the digital scholarly communication landscape. Whereas publication formerly conferred status and authority upon a work, the ease of digital publishing has diluted this status on the one hand, while opening up the channels for communicating scholarship on the other.

In part, as a result, the models for "filtering" or selecting scholarly content have also changed. Preprint servers such as the Social Science Research Network (SSRN) and ArXiv (thematic) or institutional repositories, which publish with minimal filtering, are the standard means of communicating new research in many fields – with subsequent formal journal publishing used as a means of certification.

---

[1]Some sobering statistics circa 2014 include million search queries to Google and 204 million e-mails sent each minute, or annually, over 107 and over 1 trillion, respectively. More generally on the shift in the evidence base of the social sciences, see [21].

[2]https://doaj.org/ [retrieved 15 May 2015].

Journals such as *PLoS* have explicitly adopted a non-traditional filtering model, where submissions are reviewed and published based on scientific competence, not expected impact.[3] Journals such as *Faculty of 1000* and *PeerJ* represent different approaches to post-publication review [34]. The current scholarly communication ecosystem could have many instances of a single article for publication available via many different avenues. Preprint sites provide visibility for early drafts seeking reader feedback and commentary as the works are continually revised; and institutional repositories collect not-quite-final publication drafts for open access to smaller communities.

Consumption of research publications has also grown substantially. Notably, open access (OA) publishing has grown rapidly, increasing the audience for scholarly content. There are now thousands of active OA journals, according to the *Directory of Open Access Journals* (DOAJ) and the published list of Open Journal Systems (OJS) journals.[4] Although the majority of the debate and discussion about OA pertains to journal publishing, there are emerging initiatives to adapt OA to monographs and other forms of publication. And OA makes content available to readers worldwide, including developing countries with limited access to current research outputs (see [45]).[5] Furthermore, this content is made more accessible through channels not exclusively reserved for scholarship and research; e.g., Web-based avenues such as blogs and social media sites like Twitter and Facebook. As a result it is used (or reused) by more people. Social citation tools such as Mendeley, Academia.edu and Zotero offer further visibility to works across the spectrum of completion.

### 1.2. Diversification in content

But it is not just more stuff[6] being produced in research, but more types of stuff that is being produced by larger, more complex research teams; and these outputs are published and/or shared in more ways outside of the traditional distribution channel – the book–journal binary. Although scholarship and research have never been limited to textual and numerical representations on a page (consider Darwin's study of bird beaks and Mendel's peas), the age of digital research objects and methods has only complicated the impulse to distill scholarship onto the page. Video clips of research processes and results (e.g. fMRI sequence, demonstration of a process) are tackled head-on by *The Journal of Visual Experiments*, though generally relegated to ancillary status within traditional publications.[7] Data visualization, which has grown in importance alongside the rise of big data as it is one of the few ways to distill, capture, and consume the vast data sets, is still not a regular feature of publications using this research due to its challenge to existing formats.

Researchers are increasingly citing and publishing scholarly outputs that are not in the form of traditional articles and monographs. Some recent examples of this include the online-only digital journal *Scientific Data* that focuses on the publication of "Data descriptors".[8] *SoftwareX*,[9] an open access digital publication from Elsevier, seeks to validate software as a mode of scholarly inquiry. It joins a list

---

[3] See *PLoS*'s criteria for publication: http://journals.plos.org/plosone/s/criteria-for-publication; more generally see [8].

[4] https://doaj.org/; https://pkp.sfu.ca/ojs/.

[5] It is important to note that while OA publications remove barriers to access, one requires digital access at the outset. In his article "Promoting open access to research in academic libraries", Priti Jain (Department of Library and Information Studies at University of Botswana) discusses the challenges of adopting OA in developing countries; the primary obstacles are lack of critical ICT infrastructure and awareness of OA collections.

[6] Or more formally, "scholarly outputs".

[7] *Journal of Visual Experiments*, available at: http://www.jove.com/.

[8] *Scientific Data*, available at: http://www.nature.com/sdata/.

[9] *SoftwareX*, available at: http://www.journals.elsevier.com/softwarex/.

of more than sixty scholarly publications across subject areas, including Engineering, Humanities and Social Sciences, Image Processing, Informatics/Mathematics/Statistics, Life Sciences and the Physical Sciences/Geosciences.[10] And in some scholarly fields, such as law, non-traditional sources such as blogs and Wikipedia entries are cited in formal publications (and even legal decisions) [30,31].

Collaborative research is enabled by expedient, inexpensive and immediate qualities of the digital, and this has proven to be more of a socio-cultural issue rather than a technology challenge [41]. While scholarship has always relied upon its connection to existing publications, the digital space engenders a new level of participation. The number collaborating on scholarly outputs has increased dramatically over the last twenty years. The complexity of collaboration, and the diversity of roles requires new approaches to documenting authorship, such as Project CRediT (Contributor Roles Taxonomy), which identifies fourteen high-level roles in journal publications.[11]

The explosion of students in MOOCs and rapidly-increasing participation in citizen science dramatically broadens the audience for scholarly content. For example, is not unusual for tens of thousands of people to participate in significant portions of an online course; and citizen-science projects such as Galaxy Zoo, which provided the opportunity for astronomy enthusiasts to classify images collected by the Sloan Digital Sky Survey, attracted over one hundred and fifty thousand participants in its inaugural year.[12] These opportunities for participation beyond the traditional research community continues to grow in terms of public participation and with an increasing array of subject matter. The growth of MOOCs doubled in 2014, with over four hundred universities offering upwards of twenty-five hundred courses to an estimated sixteen to eighteen million students.[13] Zooniverse, a citizen science platform with twenty-nine projects within the topics of Space, Climate, Humanities, Nature, Biology and Physics boasts of over 1.3 million participants worldwide [49].

Scholars are diverging from traditional channels of established publishers and journals with almost twenty-five percent communicating research findings via social media [16]. A 2014 study in *Nature* found almost fifty percent of scientists have posted original research on social media, with Twitter being the favorite [40].

The upsurge in volume and complexity of stuff, use, and audiences has an impact on both the evaluation of scholarship as well as its infrastructure. This is exemplified in the increasing variety of evaluation metrics and systems yet to be codified into a standard,[14] as well as the growing complexity of information flows within systems of scholarly communication and information management – even in the case of traditional services, such as naming authority provision.[15] A recent OCLC report located ten different types of researcher IDs at play in a number of systems across scales, ranging from institutional and disciplinary, to national and international, and beyond at web-scale [36]. Within the scholarly communication ecosystem (as with all spaces), the systems of evaluation are informed by its infrastructure, and these must be considered in concert.

---

[10]For a thorough though not exhaustive list of these publications, please see the Software Sustainability Institute, posted 28 May 2015, available at: http://www.software.ac.uk/resources/guides/which-journals-should-i-publish-my-software [retrieved 28 May 2015].

[11]http://www.nature.com/news/publishing-credit-where-credit-is-due-1.15033, http://credit.casrai.org/.

[12]http://www.galaxyzoo.org/#/story.

[13]https://www.edsurge.com/n/2014-12-26-moocs-in-2014-breaking-down-the-numbers.

[14]http://www.niso.org/topics/tl/altmetrics_initiative/.

[15]https://www.oclc.org/developer/develop/web-services/lc-name-authority-file-lcnaf.en.html, https://www.oclc.org/developer/news/2015/oclc-lc-name-authority-file-now-available.en.html [published 21 January 2015, retrieved 30 May 2015].

And finally, to attempt measuring the impact of this new content in these historically ignored or unacknowledged verticals, new methods of assessment have spawned the analysis of *altmetrics* – a system of metrics measuring impact in non-traditional spaces [32]. For example, article Level Metrics (ALM) is one example of the new methods to find more granular and current measures of impact. ALM, which is being developed by *PLoS*, seeks to measure the impact of an article as independent of the ranking of the journal within which it is published [26]. Unlike journals, books do not have an impact factor, however this possibility is beginning to be explored (see [50]).

And plenty more stuff is coming soon. Experts predict substantially-increasing use of social media, mobile devices, learning analytics, sensor data (some from the Internet of things), linked data, and much, more to occur in the next five years [2,17].

## 2. More stuff, but less competition – What's wrong with the market?

In general, the trends in scholarly communication are more, more and more. In 2011, the value of the market in 2011 was estimated to be $23.5 billion [44]. But there is one area in which the trend is "less" and that is in market competition. Although the number of publications and journals is expanding at approximately three percent a year, and the market is expanding at four percent, the number of mergers and acquisitions over the past three decades have dramatically decreased the diversity of and competition among publishers. Today, following the recent merger of MacMillan and Springer, the market is dominated by a handful of companies: Pearson, Reed Elsevier, Springer, Taylor & Francis, Thomson/Reuters and Wolters Kluwer [15]. These companies happen to be the top four publishing companies globally as well [33]. And this is the culmination of a long-term trend: over the last three decades, there has been dramatic consolidation in the scholarly publishing industry (see for example [6,24]). Profit margins are commensurately high, with some credible estimates of Elsevier's profit margin as high as thirty-seven percent.[16]

Thus far, there are no signs that the general expansion of the content, contributors, and audience for scholarly outputs has countered this decline in competition. In fact, preliminary studies of "gold" open access publishing[17] indicate that the vast majority of subsidies for OA publishing are going to the dominant commercial publisher.[18] There is, however, an increase in the participation levels with the variety of methods and mediums for transmitting research in real-time as it develops. This span the research lifecycle, from posting initial ideas and observations via social media and blogs, through the sharing of draft articles on preprint repositories, to the archiving of the post-print in a repository – all of this amidst continuous efforts to keep abreast of developing research via social media, listservs, and other channels of informal discourse. And this doesn't begin to address the development of new – often digital – methods to capture research objects and results for which very few avenues exist for adjudication and publication. The current scholarly communication ecosystem has more producers, more consumers and more goods, but dramatically less competition. From an economic view, this presents a

---

[16]Reed Elsevier 2014 Annual Report http://www.relxgroup.com/investorcentre/reports%202007/Documents/2014/relxgroup_ar_2014.pdf. Press release with summary of 2014 AR http://www.relxgroup.com/mediacentre/pressreleases/2015/Documents/reed-elsevier-results-2014-pressrelease.pdf. For an interpretation of these numbers, please see Alex Holcombe's blog: https://alexholcombe.wordpress.com/2015/05/21/scholarly-publisher-profit-update/.

[17]Open Access publications fall into two categories – *gold* and *green*. Gold publications are made available from the publisher whereas green publications are available via repositories, and often not quite the final published version.

[18]http://blog.wellcome.ac.uk/2015/03/03/the-reckoning-an-analysis-of-wellcome-trust-open-access-spend-2013-14/.

puzzle – in a well-functioning[19] market we would expect to see plenty of competition whenever there are lots of goods and producers. What could explain the concurring trends of expanding scholarly content and contraction in competition?

## 3. Rules of practical scholarly information economics

One part of what happened seems obvious: scholarly communication became digital. But, how is digital different, and how is digital disruptive?

For the majority of its existence, scholarly publishing has been conducted, more or less, as a market-activity. One could argue that it was an accident of history that allowed scholarly publishers, specifically university presses, to operate as self-sustaining businesses through the sale of scholarship. It was possible due to the primary distribution channel for information being printed books and journals, which constituting a fixed, excludable product – a book couldn't be on two library shelves simultaneously, nor could it be read by two readers. Publishers openly sold journals and scholarly monographs, and Universities (through their libraries) bought them. This industry was, however, tightly bound with higher education and research. In part, a market for scholarly publications existed because scholars needed to publish scholarly works in order to advance in their careers; scholars needed access to substantive collections to advance; and universities were able to support scholars through the income produced by teaching – all simultaneously.

Furthermore, scholarly publishing was never a pure market economy. From its inception, a researcher freely contributed articles to be published in scientific journals and also contributed labor for the peer review of articles and books, which was (and is) an integral, essential quality control mechanism within scholarly publication. And although a researcher was oftentimes paid modest royalties on the sale of scholarly monographs, this take is not comparable to the effort invested. However when one considers that the author is supported by the University with a regular salary, sometimes even a direct subvention to defray publication costs, as well as being given time to research and write for career advancement, it seems more reasonable.

Private philanthropists sometimes contributed to scholarship, as it was viewed as a public good. In the mid-nineteenth century, higher education, and indirectly scholarship, received a huge boost from the subsidization of "land-grant" colleges, initiated by the Morrill Acts of 1862 and 1890. And in the mid-twentieth century, governments began to substantially support Universities both through research grants, and indirectly through scholarships to students. Research funding has stimulated an increase in scholars and researchers, particularly in the hard sciences and life sciences [14,37].

Nonetheless, the historical scholarly communications quasi-market was relatively predictable: For centuries, to a first approximation, scholars contributed writing and review; publishers selected, bundled, distributed and sold content; and libraries bought. The shift to digital changes this. Advances in information and communication technology, resulting in the presence of robust and widespread digital storage and digital communications networks, radically changed the cost structure of knowledge production and communications.

Specifically, the marginal cost of taking a wide variety of actions is radically lower for digital objects than it was for their physical analogues. Digital objects are cheaper to access, cheaper to replicate,

---

[19]In a well-functioning market the equilibrium outcome is Pareto-optimal, and the price of goods acts as an unbiased estimator of the true value of the good. This is a minimal form of economic optimality, since it neither maximizes the size of the economy nor guarantees an equitable distribution.

cheaper to transmit, cheaper to computer over, cheaper to modify, and cheaper to divide and cheaper to recombine. Moreover, in the case of scholarly communications, there are strong reasons, based on both history and economic thought, to believe that the market will not fix itself. We call these reasons, perhaps provocatively, the "rules" of practical scholarly information economics.

The first rule, which sums up the disruptive nature of technology [22], was proposed by historian Melvin Kranzberg:

**Rule 1:** Kranzberg's First Law of Technology.
"Technology is neither good nor bad; nor is it neutral" – Melvin Kranzberg.

Kranzberg explains his rule as follows *"technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances"*. And although this explanation was not made with scholarly publishing or communications in mind, it could not fit better: We are witnessing changes not only in the scale, volume, inputs and forms of scholarly communication, but also to the audiences and uses as well.

Disruptions shake things up, but what about the long run? Classical economic theory addresses the 'long run'. Given enough time, and the right conditions, markets will converge to an equilibrium. However, as Keynes pointed out, even if all the appropriate conditions are satisfied, such equilibration can take so long that it is irrelevant to forming public policy [19]. Keynes' rule can be best summarized in his own words:

**Rule 2:** Keynes's Theory of Long Run Economic Equilibrium.
"In the long run we are all dead" – John Maynard Keynes.

Frequent changes may keep the system out of equilibrium. And in the area of information and communication technology, the rate of change is so rapid, that frequent changes may keep the ecosystem for scholarly publication out of equilibrium for an extended period.

Furthermore, even in the long run, for markets to function properly, many conditions have to apply to buyers, sellers, institutions and goods: Francis Bator coined the term "market failure" [7] to denote the failure of a system of price-based market institutions to yield a desirable equilibrium.[20] Some violations of economic assumption are minor: they lead only to marginal losses of efficiency in outcomes. Others issues however, are more serious, because they can interfere the markets reaching good outcomes – they undermine the rationale for using market mechanisms. Thus our third rule:

**Rule 3:** Bator's Observation:
Many things in the real world violate market assumptions and – and some of these matter.

Since Bator's provocative article, social scientists have investigated many potential micro-level causes of market failure in the real world (see for example [46]), including: structural information asymmetries; unstable or time-inconsistent preferences; externalities; transaction costs; bounded rationality; lack of common knowledge;[21] path dependencies; increasing returns to scale; and non-consumptive goods. It could be argued that almost all of these issues have the potential to affect the current market for digital

---

[20]Bator noted that many things in the real world violated market assumptions, but focused on conditions on which the market would fail with even with perfect rational actors with perfect information and foresight.

[21]Here, common knowledge is used in the game-theoretic sense, as used in [5].

information. However, we believe that three of these potential causes are likely to be critical going forward: information contagion, network effects, and non-consumptive goods.

*Information contagion*, was coined by Brian Arthur and David Lane [9]. They note that when products are difficult to evaluate, buyers may rely on the experiences of other previous buyers – either through informal channels or more formal mechanisms. This feedback system can create path dependence, providing a strong and persistent advantage to the products that develop an early mass of positive feedback. This leads us to formulate a corollary to Bator's Observation.

> **Rule 3.1:** Arthur's corollary of information contagion:
> Reputation matters, and the first product to reach "good enough" often wins.

We conjecture that three attributes of scholarly communication makes information contagion likely and important: First, there are many choices of publications and publication outlets, which makes a complete evaluation of the options infeasible; second, evaluating the quality of scholarly publications without consuming them is hard, forcing a reliance on reviews; and third, mechanisms for review and commentary on digital publications are widespread.

We conjecture that a second critical cause of market failure in scholarly publishing are network externalities (demand-side economies of scale). According to classical microeconomic theory, economies of scale create monopolies. At the beginning of the 20th century, Theodore Vail recognized how a telephone network could lead to economies of scale, and successfully applied this insight to build the extensive and durable AT&T monopoly (see [48]).[22] We state this principle in modern terms:

> **Rule 3.2:** Vail's corollary on networked information:
> Privatized information networks favor monopolies.

In scholarly publishing network externalities are created because much of the value of scholarly publication is not solely in the content of the publication itself, but in its meaning in context with a body of work, and in contributing to a larger collection of publications. Scholarly works are embedded in networks of citation, evidence and meaning that give them value, and that contribute to the value of the network. Moreover, the value of the scholarly information network to potential users has greatly increased as information becomes digitized, because it is now possible to compute over the entire network. For example, scholarly information networks are now used as subjects of text, topic, content and sentiment mining to extract new findings or concepts; to analyze the relationships among publications, funding, and evidence to evaluate theories, actors, institutions and policies; and to identify potential collaborators, and antecedents of successful collaboration.

The last critical source of market failure in scholarly communication merits its own category, because it has to do with the nature of scholarly publications as market goods. Economic theory predicts that markets will work well when the goods being sold are both "consumptive", meaning that they are consumed with use; and "excludable", meaning that it is possible for the owner of the good to control who uses it. These are sometimes known as "private" goods.

As Ostrom and Hess, Foray, Stiglitz [13,29,38] and have pointed out, knowledge goods are not private goods. Some knowledge goods, such as licensed databases or journal subscriptions, are non-consumptive but excludable, "toll goods"; others, such as public libraries are partially consumptive but non-excludable; and fundamental knowledge, such as that produced about the world by some basic research, is neither consumptive nor non-excludable – a "pure public good".

---

[22]On network effects in depth see [23].

Further, Ostrom's career-long Nobel Prize-winning work focused on identifying the institutional design principles that enable the successful management of local common pool goods: She found that for local-pool commons, such as fisheries, to be successfully managed by a local community required that clear boundaries be in place, rules are matched to local context; actors participate in rulemaking; self-monitoring and graduated sanctions are available; low-cost conflict-resolution is available; external authorities respect community rules; and enterprises are nested with respect to the terms of governance, rules, etc. [27,28]. This is summarized in rule 4:

**Rule 4:** Ostrom's Rule of the Information Commons:
Information goods should not be managed like private goods.

A corollary to Ostrom's rule was articulated by the National Digital Stewardship Alliance leadership [3]. The report notes that because the digitization of information lowers the cost to access, almost every institution now relies for its business, operation, and mission on large amounts of information that go beyond the institutional boundaries. The amount of information is so great, and the risks so diverse, that no single organization can effectively ensure long-term access to all the information it needs. At the same time, for many pools of digital information, multiple institutions value it. Together, these factors imply the following:

**Rule 4.1:** The NDSA Stewardship Corollary:
Stewardship of information requires collaboration.

An implication of Ostrom's argument is that information goods, if left to the private market, will be under-provisioned – the amount of knowledge produced and used will be less than is good for everyone in aggregate. This is a one of the core rationales for government support of basic research.

### 3.1. Summing up the rules. . .

Information technology has disrupted the economics of scholarly communication. This discussion is ongoing, and so the market is unlikely to sort it out in the short term. Moreover, the characteristics of information goods and networks do not lend themselves to pure market solutions. Even in the long run, economic theory itself predicts that left to the market, too little knowledge will be created, too little used, and access to too much of what is available will be controlled by a small group of distributors.

Due to the rapid rate of change in the scholarly ecosystem and the relative paucity of research on this issue, many of the responses to this disruption fail to address the crux of the issue. Academic libraries focus on the price of journal subscriptions, implicating publishers; publishers lament the reduced purchases from academic libraries resulting in higher prices; scholars resent the reduction in publishing output from scholarly presses and the increased competition for possible publications. And much of the recent focus of debate and proposed interventions is around open access and open data – both sort of interventions we generally support. However, lowering barriers to access will not by itself guarantee robust, sustainable scholarly communication. The stakeholders in the scholarly communication process need to thoughtfully engage with new opportunities to shape the future of intellectual communication.

The failure of the market does not necessarily imply that "information should be free": Knowledge production, distribution, and long-term access (preservation) has a cost – and relying on uncoordinated individual self-interest and altruism is also practically certain to produce and share too little. Reaching the "right" level of knowledge creation and distribution requires a coordinated effort – which could be implemented through government taxes and subsidies, direct regulation of the market, or community

governance of commonly-valued knowledge resources. In other words, as MacCleod[23] humorously, but aptly points out, it's not that information always wants to be free, but that...

*. . . (MacCleod's observation) Information want's someone else to pay for it.*

## 4. Rebundling scholarly publications for markets and commons

It isn't news to anyone that systems of higher education are experiencing challenges

The structures and roles within scholarly communication, including university presses, archives, repositories, academic libraries, authors, journal articles, monographs, datasets, authors, editors and reviewers are interconnected in complex ways.

In their current form and for the most part the digital scholarly monograph and journal article duplicate the form and function of their print ancestors. However, this is not a necessity – advances in digital production and communication make it possible to unbundle the content and semantics of scholarly publication, and to change the relationships among stakeholders and scholarly works. There are a handful of publishing experiments innovating on container and content in this area such as University of Minnesota Press's Forerunners, a groundbreaking departure from the book–journal binary.[24] However, none of these so far has achieved widespread adoption.

It's not hard to think of other possible desirable features for the scholarly work of the future. One description of the ideal "article" might be as follows: a high-quality, authoritative work; expertly-selected and thoughtfully-curated; one that participates actively in the information network; connects seamlessly to primary sources; is expertly-marketed and widely-discoverable; can be accessed on and off line; offers interactivity; contains multiple pathways of "reading" and understanding; supports reader services such as annotation and institutional services like usage evaluation; is regularly-updated and systematically-versioned; is included in tenure files and wins disciplinary prizes; is durable and easy to preserve; and is produced, maintained, and disseminated through an economically-sustainable process. A contrasting model of the future scholarly "article" is a minimalistic nano-publication: an assertion – perhaps concerning evidence, theory, analysis (and possibly an annotation on another work); preferably expressed in a machine-understandable form; and is associated with a persistent identifier for the work and accompanied by the identifier of the author.

How to generate, compare and evaluate these two proposed "bundles" of features is challenging. One potential frame of reference is the literature on the scholarly publishing process: Morris Barnas and Lefrenier [25] define the primary functions of a journal as selection, preparation, collection, navigation and collection. Their guidebook further emphasizes the importance of the functions of editing, production, *ex-post* evaluation metrics, marketing, sales, and resolution of ethical issues through policy and process. Campbell [10] describes the core lifecycle of scholarly publishing as comprising the following phases: soliciting and managing submissions, managing peer review; editing and preparing the manuscript; producing the final form of output; publishing and disseminating; and promoting and archiving. They also provide a list of ways in which the publishers add value to scholarship: sorting and assessing research outputs; publishing primary literature, supplementary data and patents; aggregating content; distilling evidence through reference works and meta review; creating standards and seeking consensus; aggregation of content; granularization, tagging and prioritization of content; identification; application of rules; systems integration; data structure and exchange standards; content maintenance;

---

[23]http://gapingvoid.com/2012/04/20/information-wants/.
[24]https://www.upress.umn.edu/book-division/series/forerunners-ideas-first [retrieved 28 May 2015].

integration of content from multiple sources; creating and monitoring behavior change; development of workflow analytics and best practice benchmarking.

We propose a complementary framework for comparing and evaluating new forms of scholarly publication. Rather than focus directly on what publishing does, the value it adds, or on the "ideal" model of publication, we propose to focus on the features relevant to how publications are, and could be *used* by the various actors and stakeholders in the scholarly evaluation, e.g. researchers, in their roles as readers, writers and reviewers; publishers; administrators; and scholarly intermediaries, such as libraries; funders, etc. Framing scholarly communications in terms of the actors involved and the affordances provided suggests a number of questions: *What are the critical actions that stakeholders apply to scholarly publications in these use-situations? What are the use-situations that are most relevant to each group of stakeholders?*

In the absence of in-depth analysis of a large sample of use case, any list of the affordances of scholarly publications should be considered preliminary – however, we can identify a subset of these: First, many of Unsworth's list of scholarly primitives [39] may be applied directly to publications: discovery, annotation, comparison, sampling, representing (formatting), and reference.

Second, reference, identification and citation have been the object of extensive recent study in other contexts, and are clearly important use-cases for scholarly publications; e.g., stakeholders frequently cite scholarly works, and this activity is central to the process of scholarship [4,12]. A study of the uses of citation reveals affordances that apply directly to scholarly publications: attribution, identification, access, persistence and verification. In other words, scholarly publications can be cited, and through citation the work can be attributed to creator, contributors and authors; identified uniquely as the object of the reference; an instance of the work located and accessed; and the object can be verified to correspond to the citing reference through versioning, fixity or provenance.

Based on our experience in research and in scholarly publishing, supported by a review of the operations of scholarly publishing (above) and of authoring handbooks, we propose seven additional affordances that are critical within the current system of scholarly communication: reviewing, brokering, marketing, distributing, organizing, aggregating and computing.

Table 1 sums up the inventory of eighteen affordances of current scholarly publications. More can certainly be added, however our conjecture is that a publication with these affordances (perhaps with minor refinements or extensions) would satisfy all primary uses of existing scholarly publications. Further, we note that the existing affordances can be nested; for example, it is common to annotate a specific sample (selection) of a publication, rather than the entire work. More generally, these affordances may be conceivably nested to an arbitrary depth.

The shift to digital information and communication technologies offers the opportunity to unbundle the current affordances of the scholarly journal article and monograph, remix these, and extend them. Future inventors of a new form of scholarly publication could develop publications that satisfy any combination of these affordances and which could add new ones (e.g. active alerting, interaction, automatic updating, etc.).

The possibilities are beyond easy enumeration. We suggest that a productive approach will be to identify mutually-dependent groups of stakeholders (such communities of practices, actors using a specific common pool resource, vertical markets) and focus on the affordances they most rely upon. And we believe that the central challenge is not simply to invent a unique remix, or add novel new affordances, but to produce scholarly publication "products" that will not only be acceptable within a market framework, but also will be sustainable through a knowledge commons.

Table 1

Selected affordances of current scholarly publications

| | |
|---|---|
| Access | Obtaining the content of the publication (understanding requires access in a form that has recognized syntax and semantics) |
| Aggregation | Grouping multiple publications to create a new publication |
| Annotation | Attaching metadata to a publication (where the provenance of the metadata is independent that of the publication) |
| Attribution | Associating a publication with authors or contributors |
| Brokering | Acting as an intermediate for purchase or right to access and publication |
| Comparison | Assessing differences and similarities among the content/metadata of multiple publications |
| Computing | Applying automated functions or algorithms to the content of publications (this may include automated indexing, abstracting or topic analysis) |
| Discovery | Finding the publication using samples of or criteria for metadata/content |
| Distribution | Acting as an intermediate entity for access or discovery |
| Identification | Associating the publication with a unique persistent string, using a well-known naming authority |
| Location | Resolving an identifier to a specific instance of the publication that can be accessed |
| Marketing | Communicating the value of a publication to potential consumers |
| Organizing | Structuring and ordering the content of the publication |
| Persistence | Acting to ensure long-term access to the object |
| Representation | Presenting the content of the object in a different format |
| Reviewing | Evaluating the quality of the publication using some known rubric |
| Sampling | Extracting a subset of the object using criteria or rules |
| Verification | Checking that the content of a publication is identical to that associated with a specific identifier |

## Author statement and acknowledgments

## About the authors

Dr. Micah Altman is Director of Research and Head/Scientist, Program on Information Science (http://informatics.mit.edu) for the MIT Libraries, at the Massachusetts Institute of Technology. Dr. Altman is also a Non-Resident Senior Fellow at The Brookings Institution. He conducts work primarily in the fields of social science, information privacy, information science and research methods, and statistical computation – focusing on the intersections of information, technology, privacy and politics; and on the dissemination, preservation, reliability and governance of scientific knowledge. He has authored more than sixty-five scholarly books and articles and several open source software packages. His work has been published in leading journals, been the subject of national media coverage, and received numerous professional awards.

Marguerite Avery is currently the Director of Scholarly Communication at Hypothes.is. Prior to joining Hypothes.is she was Senior Acquisitions Editor at The MIT Press where she acquired scholarly, trade, and reference works in Science and Technology Studies, Information Science, Communications and Internet Studies. She is also currently a Research Affiliate at the Program on Information Science,

MIT Libraries where she focuses on the future of academic publications in academic libraries and topics such as the following: authoritative digital publications and the development of standards for works beyond the book–journal binary; the role and reception of collaborative scholarly publications (beyond journal articles); and the role of the academic research library in publication of content beyond the scope of the traditional university press. She serves as the Digital Publications Chair for the Society for the Social Studies of Science (4S), as a member of the Digital Public Library of America's Content & Scope group, on the Board of Directors for Anvil Academic Press, as member of the Academic Steering & Advocacy Committee for the Open Library of Humanities, and as a trustee for the Somerville Public Library.

## References

[1] L. Allen, A. Brand, J. Scott, M. Altman and M. Hlava, Credit where credit is due, *Nature* **508**(7496) (2014), 312–313.

[2] M. Altman et al., The 2015 national agenda for digital stewardship, NDSA, 2015.

[3] M. Altman, J. Bailey, K. Cariani, J. Corridan, J. Crabtree, M. Gallinger, A. Goethals, A. Grotke, C. Hartman, B. Lazorshak et al., 2015 National agenda for digital stewardship, National Digital Stewardship Alliance (NDSA), 2014.

[4] M. Altman and M. Crosas, The evolution of data citation: From principles to implementation, *IASSIST Quarterly* **37** (2013), 62–70.

[5] R. Aumann and A. Brandenburger, Epistemic conditions for Nash equilibrium, *Econometrica* **63**(5) (1995), 1161–1180.

[6] R.E. Baensch, Consolidation in publishing and allied industries, *Book Research Quarterly* **4**(4) (1988), 6–14.

[7] F.M. Bator, The anatomy of market failure, *The Quarterly Journal of Economics* **72**(3) (1958), 351–379.

[8] B.-C. Björk and T. Hedlund, Emerging new methods of peer review in scholarly journals, *Learned Publishing* **28**(2) (2015), 85–91.

[9] A.W. Brian and D.A. Lane, Information contagion, *Structural Change and Economic Dynamics* **4**(1) (1993), 81–104.

[10] R. Campbell, E. Pentz and I. Borthwick (eds), *Academic and Professional Publishing*, Elsevier, 2012.

[11] CERN Data Output Statistic, Computing, available at: http://home.web.cern.ch/about/computing [retrieved 15 May 2015].

[12] Co-Data Data Citation Task Group, Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data, Committee on Data for Science and Technology (CODATA) of the International Council for Science (ICSU), 2013.

[13] D. Foray, *Economics of Knowledge*, MIT Press, 2004.

[14] H.D. Graham and N. Diamond, *The Rise of American Research Universities: Elites and Challengers in the Postwar Era*, JHU Press, 1997.

[15] House of Commons Select Committee on Science and Technology, Scientific publications: Free for all? Tenth report of Session 2003-04, Vol. II: Oral and Written Evidence, Stationery Office, 2004.

[16] R. Housewright, R.C. Schonfeld and K. Wulfson, ITHAKA S+R, US Faculty Survey 2012, published 8 April 2013, available at: http://sr.ithaka.org/sites/default/files/reports/Ithaka_SR_US_Faculty_Survey_2012_FINAL.pdf [retrieved 29 May 2015].

[17] L. Johnson, S. Adams, M. Cummins, V. Estrada, A. Freeman and H. Ludgate, *The NMC Horizon Report: 2014 Higher Education Edition*, New Media Consortium, 2014.

[18] F. Kaplan, A map for big data research in digital humanities, *Frontiers in Digital Humanities* (2015). doi:10.3389/fdigh.2015.00001 [published 6 May 2015, retrieved 30 May 2015].

[19] J.M. Keynes, *A Tract on Monetary Reform, Vol. 4*, Macmillan, London, 1923.

[20] M. Khabsa and C. Lee Giles, The number of scholarly documents on the public web, *PLoS One* (2014). doi:10.1371/journal.pone.0093949, available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093949 [published 9 May 2014, retrieved 15 May 2015].

[21] G. King, K. Schlozman and N. Nie, The changing evidence base of social science research, in: *The Future of Political Science: 100 Perspectives*, 2009, pp. 91–93.

[22] M. Kranzberg, Technology and history: "Kranzberg's laws", *Technology and Culture* **27**(3) (1986), 544–560.

[23] M.A. Lemley and D. McGowan, Legal implications of network economic effects, *California Law Review* **86**(3) (1998), 479–611.

[24] C.E. Lipscomb, Mergers in the publishing industry, *Bulletin of the Medical Library Association* **89**(3) (2001), 307–308.

[25] S. Morris, E. Barnas and D. LaFrenier, *The Handbook of Journal Publishing*, Cambridge Univ. Press, 2013.

[26] C. Neylon and S. Wu, Article-level metrics and the evolution of scientific impact, *PLoS Biol.* **7**(11) (2009), e1000242. doi:10.1371/journal.pbio.1000242.

[27] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge Univ. Press, 1990.

[28] E. Ostrom, *Understanding Institutional Diversity*, Princeton Univ. Press, 2009.

[29] E. Ostrom and C. Hess, A framework for analyzing the knowledge commons, in: *Understanding Knowledge as a Commons*, C. Hess and E. Ostrom, eds, MIT Press, Cambridge, MA, 2005.

[30] L.F. Peoples, The citation of Wikipedia in judicial opinions, *Yale Journal of Law & Technology* **12**(1) (2010), Article 1.

[31] L.F. Peoples, The citation of blogs in judicial opinions, *Tulane Journal of Technology and Intellectual Property* **13** (2010), 39–80.

[32] J. Priem et al., Altmetrics: A manifesto, 26 October 2010, available at: http://altmetrics.org/manifesto/ [retrieved 15 May 2015].

[33] Publishers Weekly, The world's 56 largest book publishers of 2014, published 27 June 2014, available at: http://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/63004-the-world-s-56-largest-book-publishers-2014.html [retrieved 30 May 2015].

[34] T. Rabesandratana, The seer of science publishing, *Science* **342**(6154) (2013), 66–67.

[35] SINTEF, Big Data, for better or worse: 90% of world's data generated over last two years, *ScienceDaily*, 22 May 2013, available at: www.sciencedaily.com/releases/2013/05/130522085217.htm [retrieved 14 May 2015].

[36] K. Smith-Yoshimura, M. Altman, M. Conlon, A.L. Cristán, L. Dawson, J. Dunham, T. Hickey, D. Hook, W. Horstmann, A. MacEwan, P. Schreur, L. Smart, M. Wacker and S. Woutersen, *Registering Researchers in Authority Files*, OCLC Research, Dublin, Ohio, 2014, available at: http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearchregistering-researchers-2014.pdf.

[37] E.P. St. John and M.D. Parsons (eds), *Public Funding of Higher Education: Changing Contexts and New Rationales*, JHU Press, 2005.

[38] J.E. Stiglitz, Knowledge as a global public good, in: *Global Public Goods: International Cooperation in the 21st Century*, 1999, pp. 308–325.

[39] J. Unsworth, Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this, in: *Humanities Computing, Formal Methods, Experimental Practice Symposium*, 2000, pp. 5–100.

[40] R. Van Noorden, Online collaboration: Scientists and the social network, *Nature* **512**(7513) (2014), 126–129.

[41] R. Van Noorden, Online collaboration: Scientists and the social network, *Nature* News Feature, posted 13 August 2014, available at: http://www.nature.com/news/online-collaboration-scientists-and-the-social-network-1.15711 [retrieved 29 May 2015].

[42] R. Van Noorden, Global scientific output doubles every nine years, *Nature* News Blog, posted 7 May 2014, available at: http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html [retrieved 25 May 2015].

[43] M. Ware and M. Mabe, *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*, 2015, p. 6, available at: http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf [retrieved 29 May 2015].

[44] M. Ware and M. Mabe, *The STM Report*, 4th edn, STM: International Association of Scientific, Technical and Medical Publishers, The Netherland, 2015.

[45] J. Willinsky, *The Access Principle: The Case for Open Access to Research and Scholarship*, 2006.

[46] C. Winston, *Government Failure Versus Market Failure: Microeconomics Policy Research and Government Performance*, Brookings Institution Press, 2007.

[47] V. Woollaston, How many gigabytes does it take to make a HUMAN? Physicians works out that genetic code is made up of just 1.5GB of data (genomics data from Derek Muller), available at: http://www.dailymail.co.uk/sciencetech/article-2680849/How-gigabytes-does-make-HUMAN-Physicians-works-genetic-code-just-1-5GB-data.html [retrieved 25 May 2015].

[48] T. Wu, *The Master Switch: The Rise and Fall of Information Empires*, Vintage, 2011.

[49] Zooniverse, available at: https://www.zooniverse.org/#about [retrieved 29 May 2015].

[50] A. Zuccala, R. Guns, R. Cornacchia and R. Bod, Can we rank scholarly book publishers? A bibliometric experiment with the field of history, *Journal of the Association for Information Science and Technology* (2014). doi:10.1002/asi.23267 (Impact Factor: 2.23).