

SPARC*

Landscape Analysis

The Changing Academic Publishing Industry –
Implications for Academic Institutions

March 28 2019

© 2019 SPARC, subject to a Creative Commons Attribution 4.0 International License



About the Report

SPARC (the Scholarly Publishing and Academic Resources Coalition) is a global coalition that works to enable the open sharing of research outputs and educational materials in order to democratize access to knowledge, accelerate discovery, and increase the societal and economic return on our collective investment in research and education. As a catalyst for action, SPARC focuses on promoting changes to both infrastructure and culture needed to make open the default for research and education.

This report was commissioned in response to the growing trend of commercial acquisition of critical infrastructure in our institutions. It is intended to provide a comprehensive look at the current players in this arena, their strategies and potential actions, and the implications of these on the operations of our libraries and home institutions. It also outlines suggestions for an initial set of strategic responses for the community to evaluate in order to ensure it controls both this infrastructure and the data generated by/resident on it.

To produce this analysis we were fortunate to secure the services of Claudio Aspesi, a respected market analyst with more than a decade of experience covering the academic publishing market for international investors. Between 2004 and 2016, Aspesi was the Senior Research Analyst at Sanford C. Bernstein covering European Media Stocks. The academic publishing market – and Reed Elsevier and Pearson in specific – was a key area focus for him during his tenure. Previously he was Global Senior Vice President of Strategy at EMI Music and was responsible for defining EMI's business model as the music industry entered the digital age. Before joining EMI in 2002, Claudio was a member of the executive team at Airclic, an Internet infrastructure company and, prior to that, a partner at McKinsey and Co., working with many leading media and entertainment companies.

Aspesi produced this comprehensive analysis in close collaboration with the SPARC team, after conducting interviews with dozens of key stakeholders including provosts, CIO's, library leaders, students, and higher education administrators in a wide variety of North American institutions, as well as publishers, and other market experts.

We are at a critical juncture where there is a pressing need for the academic community – individually and collectively – to make thoughtful and deliberate decisions about what and whom to support – and under what terms and conditions. These decisions will determine who ultimately controls the research and education process; and whether we meaningfully address inequities created by legacy players or simply recreate them in new ways. These decisions will shape libraries' role in the scholarly enterprise, now and for the future.

Report authors

Claudio Aspesi (lead author)

Nicole Allen

Raym Crow

Shawn Daugherty

Heather Joseph

Joseph McArthur

Nick Shockey

Table of Contents

| | |
|--|-----------|
| Summary | 5 |
| Background | 6 |
| Research | 9 |
| The Companies | 10 |
| Elsevier | 10 |
| Springer Nature Group | 21 |
| Wiley | 25 |
| Clarivate | 26 |
| Other Publishers | 29 |
| The Products | 30 |
| Teaching and Student Activities | 34 |
| The State of the Textbook Market | 34 |
| The Rise of “Inclusive Access” | 39 |
| The Companies | 41 |
| Pearson | 41 |
| McGraw-Hill Education | 44 |
| Cengage | 45 |
| Other publishers | 47 |
| The Products | 48 |
| Preliminary Recommendations | 52 |

Summary

Academic publishing is undergoing a major transition as some of its leaders are moving from a content-provision to a data analytics business. This is evidenced by a change in the product mix that they are selling across higher education institutions, which is expanding beyond journals and textbooks to include research assessment systems, productivity tools, online learning management systems – complex infrastructure that is critical to conducting the end-to-end business of the university.

Through the seamless provision of these services, these companies can invisibly and strategically influence, and perhaps exert control, over key university decisions – ranging from student assessment to research integrity to financial planning. Data about students, faculty, research outputs, institutional productivity, and more has, potentially, enormous competitive value. It represents a potential multi-billion-dollar market (perhaps multi-trillion, when the value of intellectual property is factored in), but its capture and use could significantly reduce institutions' and scholars' rights to their data and related intellectual property. A set of companies is moving aggressively to capitalize on this data, often by exploiting the decentralized nature of academic institutions.

This shift is still in its early days. There are actions and strategies that institutions can consider adopting, both individually and collectively, to limit the potential harms posed by this trend, and to leverage potential benefits. These range from simple risk mitigation actions – such as revising existing data policies, establishing coordination mechanisms, negotiating to ensure institutional ownership of the data and infrastructure and establishing open terms and conditions – to larger, more strategic actions like re-thinking the institution's relationship to its data in terms of commercial exploitation, IP ownership, and research investment outcomes.

This document is designed to provide higher education leaders with an analysis of the leading commercial players' strategies in this domain, the implications of those strategies, and a preliminary set of possible broad-stroke strategies that higher education institutions might consider taking to secure outcomes consistent with their own values and goals.

THE ACADEMIC PUBLISHING INDUSTRY IN 2018

Background

Twenty-plus years into online distribution, and thirty-plus years since the first digital products first became available commercially, the academic publishing industry is undergoing a massive adaptation process. At first glance, academic publishing should have shared the traits of most Business to Business (B2B) media businesses after the transition to digital products and services: concentrated supply (protected by scale economies in sales and distribution and high barriers to entry), strong pricing power (driven by oligopolistic supply and by inelastic demand), and rising profitability. While the industry has indeed experienced these trends, fault lines are appearing everywhere, forcing the publishers to pursue different strategies from the past, which have massive – and potentially negative – implications for the academic community.

These fault lines are driven by media usage behavior which is familiar to consumers. Very much like cable TV viewers “cutting the cord” on their subscriptions, college students are lowering their spending on textbooks by renting them or purchasing them from the second-hand market; librarians, for their part, are taking tougher stances when they negotiate the renewal of scholarly journal collections and are more willing to let subscriptions expire. At the same time, digital dissemination has enabled piracy to play an increasing role, in the form of gray market imports and counterfeit physical textbooks, as well as unlicensed downloads of digital copies.

The response of the publishing industry has been predictable. On one hand, it has put some effort into protecting its copyrights through legal action; on the other hand, it has quietly started to lower prices (or the rate of price increases) to reflect the changing elasticity of demand. Many publishers have started to also adapt to new business models, in the hope that the industry can settle into a new equilibrium

by embracing the equivalent of cable TV's "skinny bundles" or the music industry's streaming subscriptions. As a result, hybrid scholarly journals, which maintain their subscription model but accept Open Access (OA) publication fees (APCs), have taken a substantial share of APC spending, and textbook publishers are pushing a model they call "inclusive access" in the hope of recapturing student spending lost to the secondary market. Some publishers, however, understand that all these are actions are – essentially – palliative remedies.

Elsevier, Pearson and Cengage in particular are transforming themselves into data analytics companies built atop their content, effectively adding ways to monetize it. None of these companies shows any inclination to abandon its traditional content business, and for sound reasons. These businesses are very large relative to the overall size of either company, and failure to sustain their profitability would have severe consequences for their respective valuations. In addition, without content these companies would have a much harder task building credible data offerings. As a result, the traditional journal business of Elsevier and the higher education textbooks of Pearson and Cengage are likely to stay for a long time in their respective portfolios. But the management teams of these companies clearly view the future as driven by adding the provision of data and data analytics services to their respective customers, rather than by growing only the traditional core business.

By leading the shift of the publishing industry into supplying data services, however, Elsevier, Pearson and Cengage (as well as other entrants, coming both from the publishing industry and from the information industry) are posing challenges for the academic community. Until now, these companies were – at worst – seen by institutions as an annoyance for selected communities within academia. Librarians complained about the cost of periodicals and talked about a "serials crisis", but the impact on the overall budget of a university was well below half of a percentage point. Similarly, the high cost of textbooks was an issue for students, and in particular those coming from disadvantaged backgrounds, but scholarships and some forms of financial aid, as well as the used textbook market, tended to mitigate the problem.

The move by publishers into the core research and teaching missions of colleges and universities, with tools aimed at evaluating productivity and performance, means that the academic community could lose control over vast areas of its core activities. In addition, the collection of massive amounts of data about faculty and students poses

a significant legal and reputational risk for institutions, along with potential privacy and security threats for individuals.

It is important to underscore upfront that we are not opposing the use of data and data analytics in academic institutions. This project is aimed at ensuring that academic institutions retain control over the use of data and data analytics, that the use of data and data analytic tools is consistent with the goals of the academic community and that academic institutions are properly equipped to deal with the risks and implications posed by the rising amount of data being collected, analyzed, and used.

RESEARCH

Background - An Industry in Transition

After a window between the late nineties of the 20th century and the mid-teens of the 21st, when the STM industry grew its revenues by about 5% annually, available data suggests that the industry's subscription revenue growth has slowed down to perhaps 1-2%. OA revenues have been rising faster, lifting overall revenue growth to 3-4%, still below the growth rate of a decade ago, but introducing a new uncertainty over the future revenue trajectory of the industry. In 2018, STM (a publishers' trade body) estimated that growth in the next few years would average 4% annually, suggesting that contracts were still being signed at about that annual growth rate, but this growth is contingent on the outcome of many unknowns.

Will caps be introduced on APCs? Will hybrid journals remain viable? Will new contractual forms (offsets, R&P, P&R, etc.) lead to overall revenue growth? The paradox of the STM publishing industry is that it has very high visibility over its revenues in the next two to three years (because of multi-year contracts) and very little beyond this time frame.

It is not surprising that many publishers would seek additional opportunities for revenue growth, and leveraging data and data analytics is the most obvious "next step" for companies that already own vast amounts of research data, both in terms of content and usage.

THE PRODUCTS

The Challenges Ahead

The decline of print textbooks is forcing every publisher to move towards digital content. Once content is digitized and delivered, however, it opens vast possibilities to collect and analyze data. While data can be used for good, there are also manifold challenges that institutions have only begun to grapple with.

Digital products include not only textbooks, but also homework systems, assessment tools, adaptive content customized based on the learning profiles of students, standalone platforms, Learning Management Systems, lecture capture, etc. Anecdotal evidence suggests that these systems built and maintained by publishers capture massive amounts of data about student and faculty behavior that go beyond what is necessary for accomplishing their core objectives (i.e. improving student outcomes). Institutions, faculty and students should think about the accumulation and use of data collected and retained by schools and commercial vendors.

Student and Faculty Privacy

Digital tools collect and analyze data in a wide variety of ways, including to establish what is a student's learning profile, where and when students access content or complete homework, what resources are used to complete tasks, how long it takes to complete individual exercises, which digital library materials have they accessed, and so on. While publishers may justify collecting this information for the purposes of improving educational outcomes, there are also serious questions about the potential risks. This data, if hacked, re-sold, or surrendered to governments without judicial review, it can be used to classify students, screen them for employment or access to graduate education, infer their political views, and even map their network of friends, mentors, and followers. While there are federal and state regulations concerning student privacy, some (such as FERPA) have not been updated in decades and cannot be assumed to cover all possible uses.

While many students today have a choice between acquiring a textbook in print or digital form, the trends towards restricting access to digital products is unmistakable:

digital content lowers costs for publishers and enables the collection of data, and helps universities increase productivity and slow their cost inflation. For example, digital study guides supplementing digital textbooks allow colleges to reduce the number of teaching assistants required, particularly for large introductory classes. Are students better off if these gains in productivity are inadvertently purchased with vast amounts of their data?

Another area of concern is the content laid out in the Terms of Use of digital textbooks, which may often be signed by the student, rather than the institution. Since it is the decision of a faculty member – an employee of the institution – to assign a digital textbook, it is therefore a foregone conclusion that students must accept a publisher's terms in order to access their course materials. Generally speaking, it is standard for terms of use for digital products to include a clause allowing the provider to change the terms at any time without notice, possibly retroactively. Faced with increasing financial pressures and tempting opportunities to monetize data, could publishers resist?

The risks to student and faculty privacy are significant: they range from hacking to unmonitored re-sale of data to third parties. Could commercial vendors find themselves selling student data, even inadvertently, to the next Cambridge Analytica? Would commercial vendors resist government requests for data? Would universities resist requests for selective data from prospective employers, possibly dangling a greater number of hires from the institution if they could (for example) only know how students answered a specific set of questions or which students have desirable collaboration patterns? Would students read the fine print before sharing their data with an “app” offering to predict their dream job or lifetime earning potential?

Algorithms and Analytics

Also, the algorithms themselves used by publishers are notoriously not transparent, raising a spectrum of ethical questions. For example, how do adaptive learning algorithms conclude that an individual should be served one of several types of customized content? Are all student profiles considered and valued equally, or are systems effectively classifying students on the basis of perceived abilities and tendencies, handicapping some even before they complete a class? Numerous examples in recent news stories illustrate how algorithms can be influenced by the unconscious bias of the humans who design them, which can manifest in forms of

unintended discrimination. In a higher education context where algorithms are trusted with increasingly important decisions, the lack of transparency raises not only ethical concerns, but also potential legal exposure.

It is worth noting that publishers are not the only vendors of data analytics to universities. Many (if not most) colleges are using data analytics to varying degrees in the recruitment process, and there are many questions in terms of fairness raised using non-transparent algorithms. Are algorithms perpetuating, even involuntarily, biases based on ethnicity, geography, occupation, and likelihood that students or their families turn into donors, etc.?

The Future

It is very important to emphasize once again that this report is not intended to take an adversarial view about the deployment of data analytics in academic institutions. We acknowledge that the issues posed by data are here to stay. We would strongly recommend that academic institutions analyze separately the issues posed by metrics (“what is being measured”) from those posed by algorithms (“how is this being measured”). Of course, the two categories feed on each other. For example, it can become easy to measure performance by using what is made available by vendors and can be procured easily, instead of devoting resources to evaluate hard to quantify (or just to collect) elements of performance. But metrics and algorithms pose very different issues and should be addressed separately.

Academic institutions need to take control of metrics. It is their own responsibility – and theirs alone – to ensure, for example, that faculty are evaluated on the basis of multiple factors. These factors may include the impact factor of journals that published their research, but may also extend and weigh appropriately, for example, collaboration, collegiality, management of junior staff and team work. Of course, these other elements may be complex or expensive to gather and analyze, can be ambiguous and leave room to criticism. We are not advocating that academic institutions choose any specific metric over another – just that they deliberately address which should be used in the evaluation of faculty, rather than just using those that are easily available through commercial vendors.

PRELIMINARY RECOMMENDATIONS

The goal of this document is to describe the current landscape of publishers moving into core activities of universities. Actively formulating and implementing solutions to these problems is complex but critically important, and something SPARC intends to work closely with the community on. However, we think it is worthwhile to conclude this document with a recap of some of the options for action currently available to the academic community.

In general, we think that solutions must be tailored to individual colleges or universities, although some of them would likely require a significant number of institutions to work together. For any institutional context, the key drivers will be a mix of cultural elements (how centralized or decentralized are decisions? What is the appetite for establishing deep ties to commercial vendors?) and financial considerations (what are the spending priorities? What magnitude of savings does leadership hope to get from the deployment of data analytics? Would targeting significant revenues from IP represent an acceptable decision? What kind of companies are off-limits to partnerships because of their activities?) We aim to offer a menu of choices and illustrations of best practices, rather than prescribing a single course of action. That said, we think that there are two different sets of solutions.

Risk Mitigation

We think of the first set as risk mitigation solutions. These are actions aimed at protecting colleges and universities from the unintended consequences of deploying a rising number of data analytic tools and collecting larger and more intrusive amounts and categories of data.

These actions could include establishing detailed data policies and mechanisms for ensuring compliance. These policies could demand that institutions maintain the ownership of data, that terms and conditions of contracts with commercial vendors are not covered by non-disclosure agreements (i.e. open procurement), that

algorithms used by vendors are made fully transparent, that results from their usage are portable (i.e. that institutions maintain the right to historic output series, in order to facilitate switching to other vendors), that data is not re-sold to third parties (or at least is not resold without the explicit consent of the institution), and that data sets that are deemed particularly sensitive are not turned over to government agencies without first resorting to the appropriate court, etc.

We also think that risk mitigation requires appropriate organizational decisions. Some institutions may opt for coordination group or task forces, while others may decide that presidents or provosts may want to identify individuals tasked with both issuing data policies, monitoring execution, and helping individual offices negotiate with vendors and adjudicate possible conflicts of interest across different parts of the institution. Over time, we think that the latter solution will prevail in most complex institutions, but each one will have to determine its own transition pace.

Weighing Trade-Offs

A second set of solutions will require individual institutions to decide among the many trade-offs. For example, using data analytics to drive the admissions process is less expensive than hiring a large staff, and institutions may choose different levels of human intervention to balance their budgets and oversight priorities. Similarly, research institutions must balance an aggressive policy of commercial exploitation of intellectual property to supplement funding with faculty demands for independence and funding of basic research and disciplines with limited commercial upside. Some institutions will support community solutions to some of these issues, while others will be skeptical. Our goal is to offer a broad array of detailed solutions and to help, when needed, individual institutions think through the issues.

In conclusion, we believe there is still time for the academic community to act, and now is the time to do it. By taking stock of the situation, asking the right questions, and choosing the right course of action, the academic community can prevent itself from winding up in a position where it is obliged to follow a path out of its control and harmful to its future.