SPARC

Landscape Analysis

The Changing Academic Publishing Industry – Implications for Academic Institutions

March 28 2019

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 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

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 COMPANIES

Elsevier

The Basics

Elsevier is the name usually attributed to RELX Group's Scientific, Technical, and Medical business. RELX Group (known until 2015 as Reed Elsevier) is a London-based provider of information, data, and analytical services for corporate, professional, and academic customers. Elsevier is RELX's largest area of activity, accounting for 33.7% of its revenues but 40.5% of its profits. It is also the most profitable business within RELX's portfolio, although Risk & Business Analytics is a close second.

How well does Elsevier compete in journal publishing?

Elsevier is not the largest publisher of STM journals in terms of titles (Springer Nature Group publishes about 3,000 titles), but it has the largest journal revenues: Springer Nature Group had 2017 revenues of €1.64/$1.9 (at an exchange rate of 1.16) million, but about 30% of that was books, leaving journal revenues at about $1.333 million, well below Elsevier's estimated $1.8 million.

There is no undisputed source for market shares in the journal business. STM estimates that the journal market is worth about $10 billion. Assuming that 55% of Elsevier's 2017 revenues came from journals, that translates into £1.363 million which – at the average exchange rate for the year – equals $1.756 million (a 17.5% market share). This number appears directionally correct, as the approximately 2,500 Elsevier journals account for about 9% of all journals published, and revenues per articles at Elsevier should be much higher than for the average of the industry[1].

[1]Revenues per article at Elsevier should be in the region of £3,100/$4,100 per article (we derive this number by taking 55% of the reported Elsevier revenues, which derive from journals, and dividing them by the 430,000 articles which, in 2018, Elsevier claimed it publishes every year). By comparison, the then CEO of Springer guided informally a few years ago in a meeting with investors that Springer (before the Nature Publishing Group merger) was earning about €2,000/$2,300 per article. This number will have risen thanks to the merger with Nature; at the same time, Elsevier has been lowering revenues per article. As a result, the gap is likely closing.

Elsevier, however, attracts a disproportionate amount of attention within the academic community because of its profitability and its business practices. Elsevier operates at a 37% reported operating profit margin compared to Springer Nature which operates at a 23% margin. Elsevier has often pursued tone-deaf business practices, which have been viewed as a land grab by the academic community. For example, Elsevier has refused in many countries to sign offsetting deals for publishing Open Access articles in subscription journals and has also engaged in relentless lobbying activity in both the U.S. and Europe to extend copyright in novel directions. Elsevier now routinely bypasses librarians and attempts to strike deals with other offices within large academic institutions. In response to the concerns of its critics, Elsevier would argue that it enjoys record revenues and profitability for the following reasons: it manages itself well, researchers want to be published in their journals, it funds activities for Open Science and for integrity in research more than any other publisher, and it sells data analytics services because the academic and research communities need them.

Both sides of this argument miss an essential point. The adversarial relationship between Elsevier and academic librarians (and some researchers) is an anachronism and a throwback to management practices that have long disappeared in virtually every other industry. The existence of SPARC, the periodic Elsevier boycotts from some groups of academics, and the announcements from European consortia (and, more recently, from the University of California) willing to lose access to Elsevier journals rather than agree to requested price increases all point to a flaw in how Elsevier (and some competitors) are run.

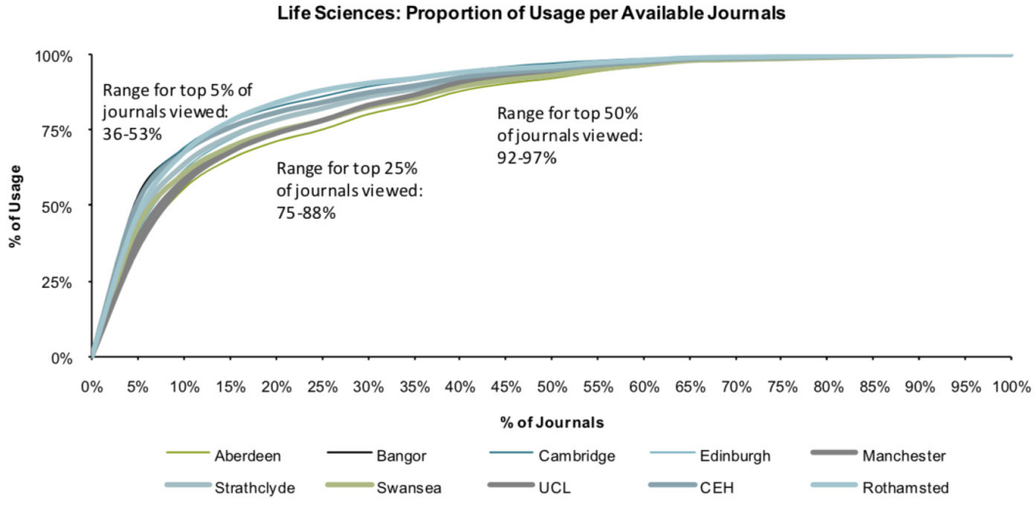
Historically, Elsevier has pursued price increases in the region of 5% annually, and justified this request with the parallel 5% growth in articles it publishes. This argument is flawed for two reasons: it assumes that there are neither productivity gains nor scale economies anywhere in the business. The first part of the argument is surprising – if productivity at Elsevier does not rise, then management is not doing its job properly – and the second part of the argument is wrong – administration or IT staff does not rise just because more articles are published. In light of the financial pressure on academic library budgets, refusing to pass along some of the savings is strictly a commercial and business decision, but it does invite retaliation from customers when possible. We have seen that Elsevier is effectively reinvesting some of these gains in new products and services which offer no benefit to libraries, as they are targeted to other users within universities (and even to completely different

customers). In other words, librarians are asked to bear the financial burden of investments which will benefit some other category of customers, as well as the shareholders and management of Elsevier.

Another issue with the request for an annual increase based on a larger number of articles published is the quality of those additional articles. Just because it is additional research does not make it equal in quality. In general, a large amount of published research may end up of limited or no use to subscribers. Data released by ten UK universities on readership in chemistry and life sciences journals show that 50% of journals account for 5 to 10% of all article readership (Exhibits 1 and 2). This pattern applies to Elsevier journals as well: when New Mexico State University decided in 2010 not to renew one of its collection contracts with Elsevier, it published the data relating access to readership for the journals included in the collection (Exhibit 3). In the absence of obligations for leading journals to add articles at a pace that is at least in line with the overall growth rate of articles published, there is a significant risk that the average "quality" of the articles will decline. This, in turn, is a contentious issue for librarians asked to fund a growing number of journals and articles with no guarantee that their quality will be adequate.

Exhibit 1

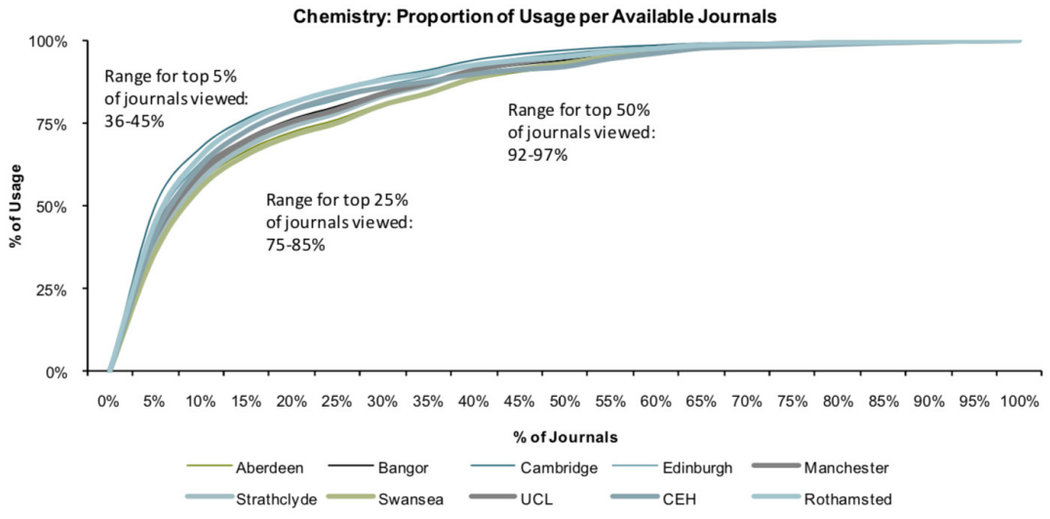
Half of the journals in Life Sciences account for 5 to 7% of articles readership among ten UK universities…



Source: Research Information Network, analysis

Exhibit 2

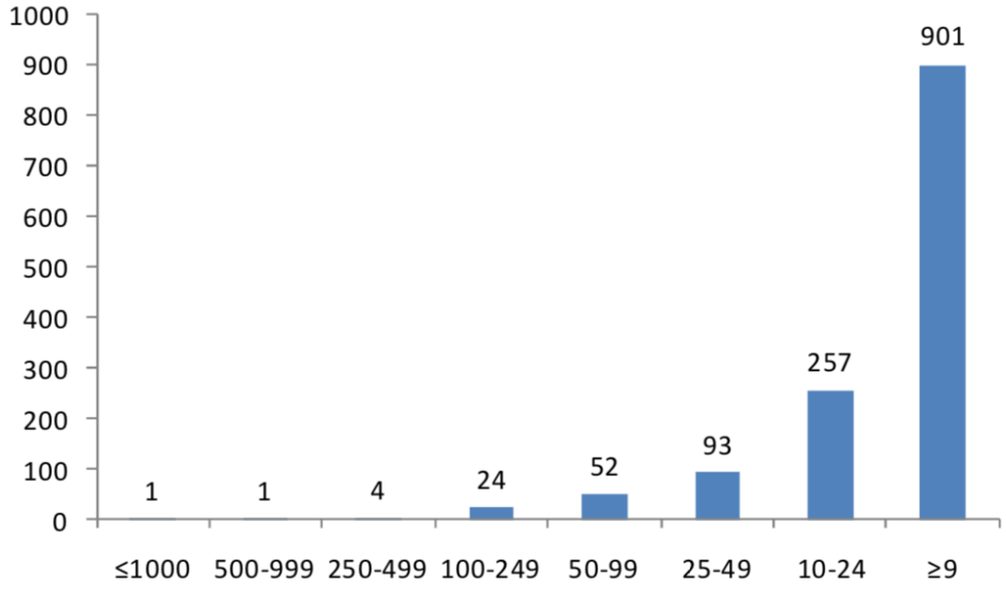
…and a similar ratio applies to chemistry journals.



Source: Research Information Network, analysis

Exhibit 3

The same distribution applies to Elsevier journals, as NMSU reported that 901 journals (out of 1333) were accessed once a month or less in the nine months before contract expiration.



Source: NMSU, analysis

Elsevier and the future of scholarly publishing – what strategy?

With this background in mind, it is no surprise that Elsevier is thinking about what the future may have in store. In spite of the fact that scale economies in the business would favor acquiring other publishers, in the past few years, Elsevier has not pursued significant acquisitions in academic journals. The company has invested, instead, to acquire platforms and tools which would broaden its range of products and allow the company to attract new customers beyond its traditional base of research libraries. In fact, in a 2015 investor presentation, Elsevier explicitly indicated its intent to increasingly serve university administrations, funding bodies, and governments with tools aimed at estimating and improving the productivity of research and optimizing funding decisions.

These investments have not been huge, but have been much larger than would make financial sense. Press reports indicate that Mendeley and SSRN cost between £100 and 125 million ($130 and 165 million) and Bepress may have cost between $100 million and $130 million. Mendeley and SSRN have business models which are unlikely to ever become truly profitable as standalone entities, while Bepress was profitable at the time of the acquisition (but the price paid still looks very expensive).

What is Elsevier's strategy?

Elsevier's answer to this question has probably evolved over time. As we previously mentioned, it is not unusual for a company to enter a business or acquire assets with one strategy in mind and then broaden its scope as it learns more about the opportunities offered by the acquired business and as the industry landscape evolves. Several interviews have surfaced four themes which we believe are central to the future of Elsevier.

1. **Protect the core journal business.** Elsevier must realize that Open Access (OA) is a significant threat to its economics, given that revenues per article today are in the region of £3,100/$4,100. Elsevier appears to have lowered its revenues/article over time (the total number of articles Elsevier publishes has grown by about 5% CAGR between 2006 and 2014, and reached about 430,000 in 2017), faster than its journal revenue growth of about 4% annually. Even so, Elsevier faces a potentially significant decline in revenues in the case that a full transition to OA takes place, since several surveys and reports indicate that the industry average APC is in the region of $2,250 to $2,400. Elsevier could cut some costs to compensate for

lower revenues, but we estimate that, with current estimated costs per article in the region of $2,650, a 10-15% cut would only save about $250 to $400, well below the decline in revenues. Incidentally, research Bernstein conducted years ago with the collaboration of a scholarly journal publisher surfaced that a full transition to OA would lead to savings of 12.5% in its cost base, so a 10 to 15% cost reduction appears eminently reasonable. In this scenario, however, the operating profits of Elsevier would be wiped out almost completely, as both revenues and costs would converge around the $2,250 to $2,400 mark. The best way for Elsevier to protect its profitability to would be to gain a massive market share at the high end of the market that could and would pay for expensive APCs. This, in turn, would be easier if Elsevier could win larger and larger numbers of high quality submissions from authors. In this scenario, Elsevier might be able to publish more high impact articles thanks to early insights into the quality of research (for example, through access to the desktops of researchers through Mendeley).

2. **Improve how Elsevier runs the journal business.** Elsevier already holds one of the largest repositories of data on citations and readership through its various databases (from Scopus and Science Direct to Mendeley, SSRN, and Bepress). Simply put, the more data points Elsevier develops, the better it positions itself to gain insights into developments that will affect its competitive position in the future, often with an early advantage which may be measured in years. For example, Elsevier could identify, through the analysis of research and publication patterns and the quality and reach of their collaboration networks, which researchers are likely to grow into future leaders in their respective fields and offer them editorial board positions on their journals ahead of other publishers. Elsevier could also identify which segments of various disciplines are likely to evolve into the next growth area for research by looking (for example) at project participation patterns, size and quality of teams, and funding bodies' decisions, targeting these segments with new, dedicated journals ahead of other publishers. Similarly, Elsevier could isolate in advance new trends in interdisciplinary studies, allowing it to establish publication forums where none exist today and even driving funding decisions which lead to accelerated growth for these types of research. Each of these levers could enhance the competitive position of Elsevier, particularly because the company has access to far more data on research patterns than any other publisher (except for Clarivate, which does not have access to content itself).

In addition to these strategic opportunities, there are more tactical ones, which may allow the company to run its operations better and at lower costs. For example, Elsevier could identify the editors best suited to work with specific authors to accelerate manuscript editing, or which peers are likely to respond faster and more constructively to requests for peer review on the basis of specific characteristics of articles. The possibilities are many and the opportunity for smaller publishers to replicate these advantages is almost nil because of Elsevier's size and reach in the marketplace.

3. **Sell insights to universities, funding bodies and governments.** The same insights that would allow Elsevier to offer editorial positions to rising star researchers or to launch new journals ahead of the competition are valuable to universities. Elsevier and Clarivate, as well as smaller competitors like Academic Analytics, already sell tools aimed at assessing the productivity of specific research. Some of these tools are relatively primitive, but it is reasonable to assume they will become better over time (interviews suggest that Clarivate is ahead of Elsevier in terms of quality, but Elsevier is working to fill the gap by offering to test their tools for free with leading research universities). In addition to driving resource allocation, these tools can also affect other core processes of universities. For example, the quality of published research (using the impact factor of journals as a proxy) has played a major role in hiring, promotion, and tenure decisions. Being able to better assess and predict the likely trajectory of Ph.D. candidates, research associates, and Assistant and Associate Professors would be of great value to the leadership of universities. Similarly, being able to target early on which emerging disciplines or subjects are most likely to grow in relevance (to the point of attracting significant incremental funding and students) would allow universities to improve the return on investment in new areas.

   Moreover, once these tools are deployed, the customers may find it difficult to discontinue usage for several reasons, including equity considerations (using the same assessment tools over time is encouraged in people-related processes, and lack of transparency on how algorithms work would make it difficult to substitute them). In general, the experience of the corporate sector is that users tend to rely on specific third-party data and find it difficult to abandon once it is embedded in their core processes. This means that these services can generate recurring revenues with strong pricing power, and Elsevier is uniquely positioned to offer

these services given its access to content and underlying data.

4. **Sell insights to industry or the investment community.** This opportunity may be the most speculative, but it is also, by far, the most valuable. Elsevier, like any other STM publisher, sits on a massive amount of intellectual insights and – increasingly – data. To provide some general estimates, NASDAQ believes that about 30% of its market capitalization (currently about $3 trillion out of a total capitalization of $10 trillion) is derived from academic research. It would make sense for Elsevier to try to capitalize on its early access to academic research (for example, by partnering with industry to exploit insights, or by establishing joint ventures with Venture Capital firms to improve the odds of VC investing).

   Elsevier seems well aware of the value of text and data mining, as it effectively attempts to prohibit text and data mining of articles outside of its own tools. It is possible that Elsevier only thinks of monetizing this value through the licensing of software that allows third parties to data mine articles and data repositories. One step in this direction is Elsevier's recent launch of its Entellect platform, a product targeted at life sciences companies that combines access to proprietary company data, Elsevier databases, and academic literature to mine text and data and streamline the R&D process. On the other hand, Elsevier may realize that much more value can be realized by mining the text and data by itself and then selling insights to interested third parties. Our interviews with industry participants indicate that Elsevier's is unlikely to take such a bold step, but it is entirely possible that management will see the opportunity at some point.

What could derail Elsevier?

We see three major threats to Elsevier's performance going forward and have laid them out below in what we estimate to be declining order of likelihood.

1. **Failure to properly execute their data strategy.** For all the competitive advantages which Elsevier has, their data strategy carries some risks as well. First and foremost, Elsevier is viewed with suspicion within large segments of the academic community. The launch of this SPARC project, for example, is the result of continuing ill will between Elsevier and the academic community. Elsevier has made several efforts to mend fences through a variety of approaches, ranging from donations and sponsorships of academic programs and events to the steady effort to communicate its support for research values through members of the

academic community. Despite these efforts, Elsevier has made little headway, in part because – in parallel with these pacification efforts – it continues to lobby for additional copyright protection, it staunchly defends rights which are counter-intuitive to academics (like refusing to allow text and data mining of articles with third party software), and it continues to demand price increases that are viewed as unreasonable and ignore the financial constraints of academic libraries.

In addition to these issues, which may lead universities to exercise special care to control or cap the usage of Elsevier data tools and services, Clarivate appears to have a significant lead in the quality of its competing tools. Clarivate also has the following advantages: a reputation for independence, no journal business to protect, and ample funding to maintain their current high-quality leadership. Lastly, Clarivate has the advantage of carrying an Intellectual Property business in its portfolio, which may make it the natural winner in selling services tied to IP exploitation.

2. **A sudden collapse of subscription publishing**. Historically, shifts in the scholarly communications industry have taken place at a glacial pace. Open access (OA) revenues were still less than $500 million in 2017, or about 5% of total industry revenues, 15 years after the Budapest Open Access Declaration was signed. Of course, revenues undercount the extent of the success of OA because they do not include articles funded in OA through other programs, or the large number of articles made available in OA through repositories rather than journals – the articles available in OA in any form are now closer to 30% annually and rising. However, revenue is what has traditionally mattered to subscription publishers.

   Nonetheless, a collapse cannot be ruled out completely. Should governments and private funding bodies in the U.S., Japan, and Western Europe agree to abandon embargo periods before mandatory deposit in open repositories as well as disqualify hybrid journals from fulfilling OA mandates (and prove willing to enforce their mandates by publicly excluding violators from future grants), publishers may see submissions to all but a handful of leading journals dry up quickly. The industry may then convert to a pure Gold OA model to secure revenues, but – as mentioned earlier – the relatively high cost structure of Elsevier leaves it poorly equipped to compete in a Gold OA industry. If this were to transpire, Elsevier would have to go through a lengthy period of restructuring, and – in the meanwhile –

financial markets would likely show their displeasure by shrinking the valuation of RELX Group, both because Elsevier earnings would collapse and because it would be difficult to see a path to recovery.

Plan S may or may not lead to a full OA transition. The number (and spending) of the initial supporters is relatively modest, but more countries and funding bodies are expressing an interest in joining, at least in principle. Important issues remain open, however. Will a transition phase, allowing the publication of OA articles in hybrid journal, be allowed or not? For how long? What will trigger the sunset of such phase? Similarly, will caps to APCs be imposed? For how long? At what level? Just to be clear, an indefinitely long transition phase and/or the removal of any APC caps would decrease the impact of Plan S on subscription publishing; conversely, the outright banning of hybrid journals and the imposition of low caps on APCs would dramatically change the publishing landscape.

In addition, piracy might force a change in the business model. While librarians are careful to point out that using Sci-Hub is not legal and that there are better (and legal) alternatives, awareness of Sci-Hub continues to grow. Elsevier has been at the forefront of legal action to squash Sci-Hub, but efforts, so far, have not been effective. The experience of other media business – first and foremost the recorded music industry – suggest that it is close to impossible to litigate violators of copyright out of existence, since they can take refuge in jurisdictions outside the reach of Western courts, and, even when a company is finally taken down, a new one can quickly rise to take its place, starting a new cycle of litigation. The lesson from the music industry should not be lost on the management of Elsevier (and other publishers): what finally put peer-to-peer services out of business was not a recourse to legal action or technical mechanisms to foil searches, but the introduction of music streaming services which offered a convenient, reasonably priced service that satisfied consumers. In this sense, the scholarly communications industry is fortunate to have a model ready at hand to adopt – OA – rather than needing to look for one and test it. It took about 15 years between the revenue peak of 1999 and the widespread adoption of Spotify for the music industry to find a way out of its decline, and it had the good fortune of finding two companies (Spotify and Apple) willing to incur losses for several years to launch and grow the service.

We pointed out earlier that Elsevier's best defense in a fully OA world would be to win a disproportionate amount of high-quality submissions through predictive analytics. The other tool available to mitigate the impact of a shift to OA would be to add more value to articles than what other publishers can provide. Elsevier could try to facilitate search and access, streamline and accelerate editing and peer review, add personal recommendation and sharing tools, and justify higher APCs through demonstrable better usability of its journals and platforms. It is not by chance that Elsevier acquired Plum Analytics in February 2017, a leading provider of altmetrics, as well as Aries Systems (this deal was announced just at the beginning of August 2018). It would make sense for Elsevier to develop deep insights into how articles are used and demonstrate to researchers, funding bodies, and departmental heads the value of publishing with Elsevier (assuming, of course, that such value is demonstrable) to avoid being dragged in price competition with cheaper OA journals.

3. **Government intervention.** Even if the current hybrid system remains in place, it is questionable why Elsevier refuses to enter offsetting deals in most countries. It is difficult to believe that governments offering subsidies for Gold OA publishing (as it happens in Western Europe) intend them as a mechanism aimed at increasing publishers' revenues at the expense of taxpayers. It is a relatively minor issue compared to taxpayer subsidies of the energy or banking sectors, and it is unlikely to excite a populist backlash against Elsevier. On the other hand, because it is a somewhat arcane topic, it could be sufficient to find a handful of legislators or politicians in one country to adopt the cause and push for tough action. This could take many forms, such as excluding publishers who refuse to sign offset deals from receiving public funding for APCs or barring Elsevier from public contracts. Any governmental action would be litigated in court and would introduce further uncertainty and create a reputational risk which could affect the share price of Elsevier and force it to change course of action on OA.

   Similarly, Elsevier could be subject to anti-trust action. A complaint was lodged in October 2018 with the European Commission, alleging that RELX practices violate two articles of the Treaty of the Functioning of the European Union. Whether the European competition authorities will pursue the case remains unclear at this stage, but the expansion of Elsevier's activities into the provision of data analytics back to universities may raise further concerns.

Springer Nature Group

The Basics

The Springer Nature Group is the result of the 2015 merger of Springer Science+Business Media with the majority of the businesses owned by Holtzbrinck and operating as part of McMillan Science and Education. The company has two main shareholders: Holtzbrinck Group (a German family-owned publisher) which has 53% of the shares, and BC Partners (a European private equity firm) which has the balance (except for a small stake still owned by the Government of Singapore, which owned Springer before its sale to BC Partners). The Springer/Nature merger was quite expensive: BC Partners acquired Springer in 2013 for an estimated €3.1 billion and then agreed to give Holtzbrinck 53% of the equity of the combined entity. If by January 2015, when the Nature merger was announced, BC Partners valued its stake in Springer at least equally to its original investment, this means that Nature Publishing Group was valued at €3.5 billion, underlying the unique quality of its assets.

Springer Nature Group is the single largest publisher of scholarly journals, with about 3,000 titles. In terms of revenues, however, Springer is second to Elsevier with estimated journal revenues of €1.150/$1.333 billion (30% of Springer's reported 2017 revenues of €1.64/$1.9 billion, in fact, derive from the sale of books). Its operating profit margin of 22.8% is also well below Elsevier's 36.8%.

Historically, Nature Publishing Group, the publisher of Nature, has been perceived as the highest quality group of scholarly journals. Springer, on the other hand, was always viewed as having a good but not outstanding portfolio of journals. Springer was the first publisher to embrace OA through its acquisition of BioMed Central in 2008. At the time, the deal was viewed as an indication that OA would quickly gain traction among leading publishers. In the following years, however, Springer turned lukewarm on OA, and former CEO Derk Haank (who retired at the end of 2017) was publicly indicating by 2011 that he believed that the collection subscription (aka "big deal") is "the best invention since sliced bread". Regardless of whether one agrees with this statement, it is a good indication of the conservative and tough, business-oriented culture of Springer.

What is the strategy of Springer Nature?

Springer Nature under Derk Haank and its private equity owners has pursued a very traditional strategy, based on extracting synergies from industry consolidation. The merger with Nature Publishing Group gave Springer another large bucket of costs to cut on top of the relentless pressure on costs exercised by Haank over the years. It also absorbed a significant amount of management time, leaving little or no space for comprehensive initiatives on academic data and analytical services. In fact, its only recent noteworthy deal in research outside of the journal space has been lifting its stake in Research Square (a company that offers language and manuscript editing and translation services to authors of scientific content) to 60% in February 2018.

Springer's reluctance to enter the data analytics business may also be driven by the €3/$3.47 billion debt of the company. This debt equals about 5.5x EBITDA (as a comparison, RELX Group aims to keep its Net Debt/EBITDA ratio between 2.5x and 3x and was at 2.5x at the interim six months 2018 results), a ratio that is dangerously close to becoming unsustainable. Investors view a Net Debt/EBITDA ratio of 6x or higher as dangerous, particularly at a time when interest rates may start to rise. Management will have probably tried to lengthen maturities and convert debt to fixed rate after the failed IPO (which was meant to raise €1.2/$1.4 billion, earmarked for debt reduction). Nonetheless, it is difficult to imagine that, in the current environment, Springer could afford to shift its strategy to pursuing an Elsevier-like data analytics strategy, which would imply additional investments and losses for some time to come.

Why did the IPO fail?

Very simply, the IPO of Springer Nature failed because of a valuation issue. The price range offered to investors (€10.50 to €14.50/share, equal to $12.17 to $16.80/share) implied a valuation of 8.8 to 12.2 times 2019 forecasted EBITDA. This assumes that 2019 EBITDA will come in at about €600/$695 million (an almost 10% increase over the €551/$639 million EBITDA recorded in 2017).

The upper end of this range would have been expensive under almost all circumstances, and even the lower end requires significant confidence that the business will continue to generate adequate revenue and earnings growth for many years to come. In the short term, this is a reasonable assumption, as earnings of scholarly publishers tend to have very low volatility because most customers sign

multi-year contracts with pre-defined price inflation terms. On the other hand, there are so many things that could go wrong with this model in the future (in particular because of the rising potential of a shift to Open Access), leading investors to have little faith in the value of forecasted long term earnings.

Our interviews indicate that some investors expected the stock to fall at least another 10-15% after trading started following the IPO, effectively allowing them to buy it at a much more attractive valuation, in part because of the rising stream of negative news coming from both Springer and Elsevier on the negotiations of new subscriptions with leading European and North American customers.

Going forward…

The board and management team of Springer Nature are now in a tight spot. It looks difficult to get anything close to the valuation they expect by just running the business as they have to date. Assuming investors will not accept a valuation higher than 7x EBITDA, EBITDA must rise by 20% (+€120/$139 million) to €720/$835 million to achieve the lower end of the valuation range, and by about 75% (+€450/$522 million) to €1.050/$1217 billion to achieve the higher end of the range. After years of cost cutting, there is probably limited scope to radically lower the cost structure to achieve this goal: the company has about €1/$1.16 billion in operating costs (before interest, taxes, depreciation, and amortization), so even cutting €120/$140 million requires taking out another 10% of expenditures (and over 40% to take out €450/$522 million).

Alternatively, management could try to lift the growth rate of the business. This option, however, looks difficult to execute for two reasons. As we discussed earlier, the company is saddled with significant debt (€3 billion) and this leaves little scope to invest in growth initiatives. In addition, some of the possible growth initiatives would add to the losses in the near term, further clouding the prospect of trying to IPO the business in 2019 or 2020. It is widely thought, for example, that Holtzbrinck did not include Digital Science into the assets of Springer Nature to avoid saddling it with the estimated €10 to 20 million ($11.6 to 23.2 million) in annual losses which Digital Science used to incur at the time (per our interviews – Holtzbrinck does not disclose numbers). Recently, however, Digital Science revenues are rising at a much faster pace (again, per our interviews), and this improvement in performance may lead to reconsideration over whether it should have a future role within Springer Nature (and, in fact, become a key element of the investment case in Springer Nature).

In the end, BC Partners, like all private equity (PE) firms, wants a liquidity event that will allow it to exit the investment. In many cases, PE firms do have the recourse of distributing to their investors the holdings in acquired assets, but this outcome is unpopular with investors and becomes a black mark on a PE firm. The size of Springer Nature may preclude a sale to another scholarly publisher on anti-trust concerns (although, in recent years, regulators have accepted mergers in sectors of the media industry which are even more concentrated, like recorded music). It seems unlikely that a traditional media company would want to step in and own a minority stake in a scholarly publisher. This leaves the Holtzbrinck family or an IPO as the only long term exits for BC Partners. The first course would require a significant effort by Holtzbrinck, who may have to part with other assets (or take a significant debt) to buy out BC Partners – both options appear unpalatable. An IPO remains, therefore, the most logical outcome, but valuation remains a sticky issue, as we have seen. It is also worth noting that the stock market is near its highs, raising the question of what can lift the appetite of investors for this asset. In summary, the timing for an IPO looks increasingly uncertain.

Implications for the academic community

Elsevier is the net winner from the IPO's failure. Had the IPO succeeded, Springer Nature would have had less debt, more flexibility to pursue growth through investments, and it could have also created an uncomfortable peer in the eyes of investors: a low valuation of Springer Nature would have raised questions on the valuation of RELX Group. In fact, the current RELX Group total market value (about £40.7/$53.8 billion, including its Net Debt of £6.2/$8.2 billion at the end of June 2018) implies a valuation of 14.5x the estimated 2018 EBITDA of £2.8/$3.7 billion. Since RELX Group derived about 40% of its 2017 operating profit from Elsevier, even a top of the range 12.2x valuation for Springer Nature would have raised a few questions about the RELX valuation – and a Springer Nature valuation at the bottom of the range at 8.8x much more so. Finally, a publicly listed Springer Nature would have provided a quarterly score card on the performance of Elsevier in terms of revenue growth and margins expansion – and any underperformance would have provided further problems for the management of Elsevier. Instead of facing all these issues, Elsevier stands alone among journal publishers in its strategy of transitioning towards serving the broad academic community on data analytics. Even if Springer Nature does, at some future point, take control of Digital Science, this move will be delayed.

A publicly-listed Springer Nature, with less debt, could have moved towards replicating Elsevier's strategy, offering a welcome alternative to dealing with Elsevier. This opportunity now looks more remote in time, if at all feasible, leaving Clarivate as the only likely competitor to Elsevier in data analytics going forward.

Wiley

The Basics

Wiley is the smaller of the three main journal publishers. It publishes about 2,300 journals, so – in terms of titles – it is very close in size to Elsevier. Journal revenues, however, are much lower: in the fiscal year that closed on the 30th April 2018, journal revenues totaled $901.5 million. This number is driven by the lower number of articles published: the company claims to publish only one third of the 500,000 articles submitted every year, translating into about 166,000, substantially fewer than the 430,000 articles which Elsevier publishes; in terms of revenues per article, both Wiley and Springer Nature now appear to earn more than Elsevier (Exhibit 5).

Exhibit 5

Revenues per article, 2017

Publisher	Journal revenues	Articles published	Revenues / article	Revenues / article ($)
Elsevier (£)	1,363,000,000	430,000	3,170	4,089
Springer Nature (€)	1,164,400,000	300,000	3,881	4,386
Wiley ($)	901,500,000	166,000	5,431	5,431

Source: Websites, annual reports, analysis

Research is the largest activity within Wiley, accounting for 52% of revenues (largely journals, with a small contribution from Atypon). Books account for 34%, and the balance derives from learning services (increasingly digital, but not exclusively). The company is publicly listed and led by an independent Board, but the Wiley family is still involved.

What is Wiley's strategy?

Wiley's acquisition in 2016 of Atypon for $120 million, at first glance, seems to

underline management's belief that the traditional subscription business model will survive. We have modelled the deal based on the data made available by Wiley, supplemented by interviews. Based on these data points, Wiley would have to believe that Atypon could grow its standalone cash flow (i.e. the cash generated by the business for its previous owners, with no synergies) by 2% in perpetuity to justify the price.

We doubt that Wiley's management is just taking unwarranted risks. Wiley could justify a very expensive valuation because of its synergies with Atypon. Our interviews suggest that Wiley never achieved a satisfactory digital distribution. Interscience was heavily overstaffed with as many as 400 FTE employees and may have cost Wiley as much as $60 million/year. If Wiley could save only two thirds of its Interscience costs and it lost all external revenues, the deal would still make sense financially. Of course, this conclusion hinges on several assumptions. The greatest sensitivity is to the actual costs savings, which in turn depend on the number of Interscience employees at the time of the deal. If that number is directionally correct, then all other sensitivities are relatively minor. If we are correct, Wiley has invested substantial resources to eradicate an operational problem. This makes a lot of sense financially and operationally, but it also means that management has had less time (and resources) to think, develop, and execute a data strategy.

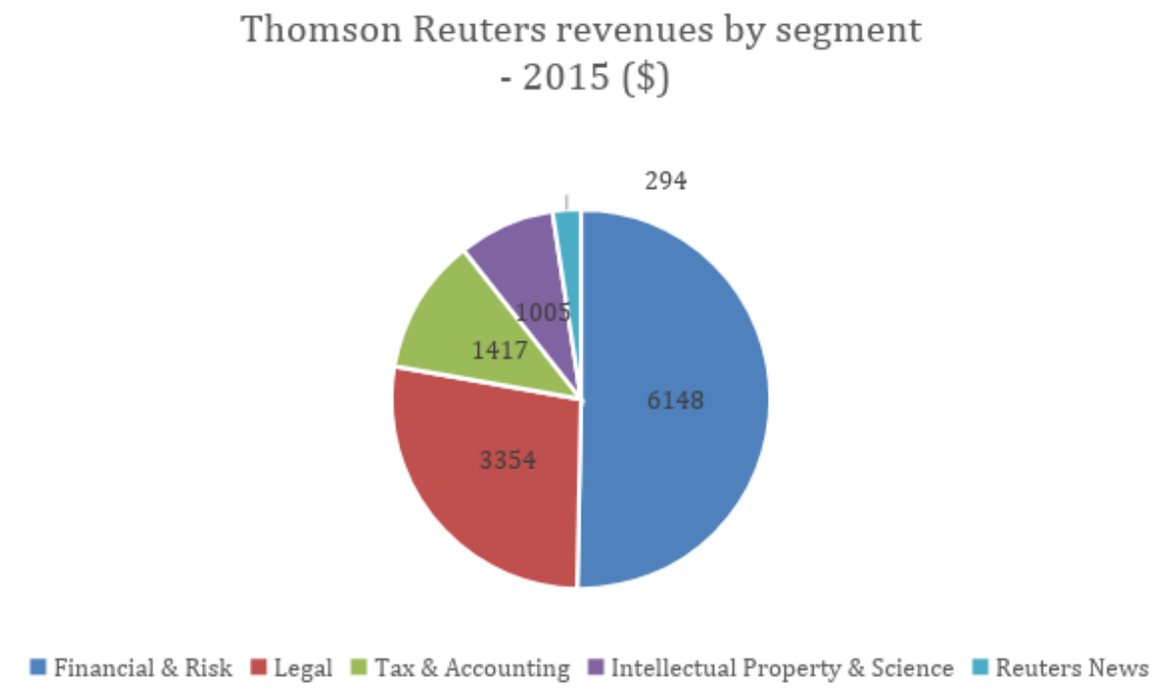
Clarivate

The Basics

Clarivate is the name the new private equity owners (Onex and Baring Asia) gave to the assets previously owned by Thomson Reuters when they acquired them in 2016 (but management had already decided to sell it by 2015). In our view, Thomson Reuters decided to sell the business for two reasons: it was small, relative to its portfolio at the time (Exhibit 6), and it was not providing the same reliable mid-single digit revenue growth it had enjoyed until a few years earlier (Exhibit 7). In addition, the management of Thomson Reuters believed that changes in the journal business and the rise of OA would add to uncertainty over the future business model, potentially putting pressure on valuations. At the same time, the business' high profitability meant that a sale would generate substantial amounts of cash.

Exhibit 6

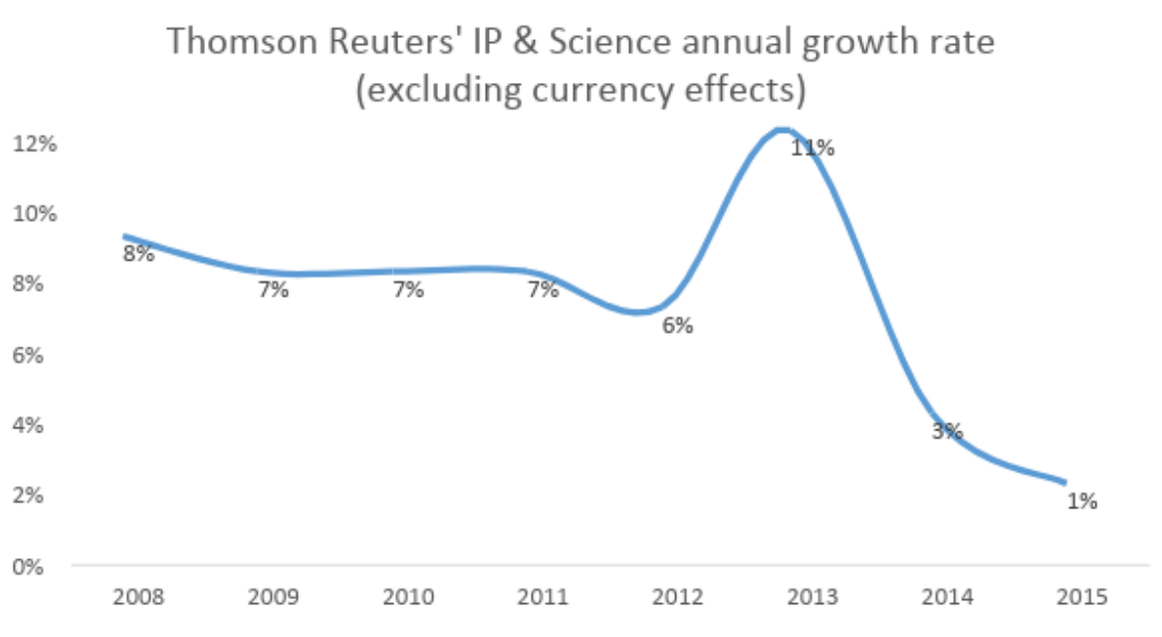
Science and IP accounted for only 8% of Thomson Reuters revenues…



Source: Annual report

Exhibit 7

…and, by 2016, its growth had slowed down considerably.



Source: Earnings presentations

The sale price ($3.55 billion) implied a valuation of about 11.5x 2017 EBITDA ($310 million). At the time of the sale, the latest reported full year EBITDA (for the Fiscal Year 2015) was $313 million (and in 2016, when the assets were mostly owned by Thomson Reuters, the EBITDA was $315 million). At the end of 2018, the company carried a substantial amount of net debt ($2.007 million, slightly higher than the $1.965 million it reported at the end of 2016), implying a net debt/EBITDA of 6.5x (6.2x at the end of 2016), higher than the 5.5x of Springer Nature. This means that management operated under significant constraints in terms of investments. However, in January 2019 Clarivate announced that the company would merge with Churchill Capital Corp, becoming a public company listed at the NY Stock Exchange. The company guided that the transaction implied an entity value of $4.2 billion, with valuation of 12.5x 2019's estimated EBITDA (implying a 2019 EBITDA of about $335 million). Clarivate also disclosed that its net debt had declined to $1.3 billion thanks to the closing in Q4 2018 of the sale of its Intellectual Property Management business (in part offset by the acquisition of TrademarkVision). This means that the Net Debt/EBITDA ratio of Clarivate has declined to about 3.9x – a ratio which is still high, but is more manageable and provides management with some room to invest going forward.

What is Clarivate's Strategy?

When Thomson Reuters owned the science business, it kept it separate for many years from its other assets in the space (healthcare and IP), pursuing a strategy narrowly aimed at serving libraries and academic institutions with leading tools which would be neutral in the eyes of its customers and the scholarly communications industry. After the disposal of most of the healthcare assets in 2012, it became apparent that there was an opportunity to consolidate science and intellectual property in one group, with the view of offering, over time, products which bridged data on basic research with data on commercial applications.

It is unclear to what extent the new owners of the business support this strategy. Despite the high level of debt, the company has made two acquisitions aimed at modernizing its research offerings: Kopernio, which is integrated into Web of Science, and Publons, which fits with ScholarOne, the manuscript management system. However, neither of these upend the company's strategy as a neutral data analytics supplier to researchers and librarians. For many smaller publishers that are unable to compete with the investment budget of Elsevier, Clarivate offers the ideal partner to enhance their services.

Integration within Clarivate between scientific and academic research and the other areas of activity is less prominent. In fact, the few interviews given by Annette Thomas, the CEO of the Scientific and Academic Research Group, are focused on the core business. Nonetheless, we think that integration between core data analytics and intellectual property remains a significant opportunity, as it can lead to serving both universities when seeking deeper ties (and funding) with industry, and industry when looking at identifying, understanding, and supporting relevant research in its early stages.

Other Publishers

In general, we view any publisher below the size of Wiley as unlikely to have the financial resources to mount a major challenge to the research data analytics businesses of Elsevier and Clarivate. There are, however, a couple of noteworthy comments:

1. Holtzbrinck must decide what to do with Digital Science. The business was not included in the merger of Springer and Nature Publishing Group (reportedly, because its losses would saddle the performance of the combined entity). Digital Science's original strategy was to build an open, neutral infrastructure, but that course is slightly changing. In part, the culture internal to Digital Science, with its collection of entrepreneurs pursuing their vision means that integrating the companies into a single, cohesive offering is very difficult. In part, Holtzbrinck probably does not wish to invest vast amounts of money to grow it to the point where it can mount a serious challenge to Elsevier or Clarivate. On the other hand, adding Digital Science to Springer Nature may allow them to tell a more credible story to investors, which may help selling an IPO which looks difficult to attempt again.

2. Some smaller, independent data analytics companies are also gaining traction. Academic Analytics is one of the most prominent, but other companies are also launching specific products and services targeting the academic research community.

THE PRODUCTS

Compiling a map of all the products and services that the three leading research data analytics vendors (Clarivate, Digital Science, Elsevier) market outside libraries is inherently a best effort exercise (Exhibit 8). Many of the companies or businesses within each company rely on a large trove of data and, effectively, repackage this data to serve as many different purposes as possible. In other words, only ingenuity limits what can be sold. The Elsevier and Clarivate offerings tend to rely heavily on data they first collect as part of their library services: Scopus and Web of Science represent significant foundations for many of the services and products offered respectively by Elsevier and Clarivate.

Exhibit 8

A map of the data tools and services offered by Clarivate, Digital Science, and Elsevier.

	Clarivate	Digital Science	Elsevier
Research Intelligence solutions for universities			
performance assessment	InCites	Dimensions	Pure, SciVal
researchers profiling	InCites	Symplectic	Pure
researchers networking			Pure
expertise discovery	InCites	Dimensions	Pure
productivity analysis	InCites		SciVal
identification of new trends			SciVal
Identification of partners	InCites	Figshare	SciVal
benchmarking	InCites	Dimensions	SciVal
funding			Funding Institutional
ad hoc consultation	Professional Services	Digital Science Consultancy	Analytical services
results promotion	WoS Author Connect		
Research Intelligence for funders	InCites, Essential Science Indicators	Dimensions	
ad hoc consultation	Professional Services	Digital Science Consultancy	Analytical services
applications peer review		Dimensions	Expert Lookup/ Reviewer finder
Research Intelligence for governments	InCites, Essential Science Indicators		
ad hoc consultation	Professional Services	Digital Science Consultancy	Analytical services
Research productivity tools	EndNote, EndNote Site Licence, Kopernio, Converis	Readcube, Altmetric, Overleaf, BioRAFT	Mendeley, SSRN, PlumX
Discipline-specific solutions			
chemistry			Reaxys
life sciences		Labguru, BioRAFT	

Source: lit search, interviews

The map we have compiled raises several issues. It is not the purpose of this document to suggest or recommend solutions to the many issues posed by the deployment of these tools, but we believe it is important to raise at least a few questions for the benefit of readers.

1. First, all three companies have targeted customers inside and outside the academic community. This raises an important question for universities: to what extent is it in their interest to share additional data with companies that will likely feed the data (even if in aggregate form, at least some of time) to funding bodies and governmental decision-makers?

2. Second, the use of data is obviously important when assessing productivity and impact, but no algorithm can capture the complexity of some of the judgmental decisions that university administrations and departmental heads are asked to make. Wealthy and well-funded institutions can probably afford to spend the time to evaluate every individual researcher. On the other hand, institutions affected by limitations on budgets will be tempted to give disproportionate weight to tools that can be advertised outside the institution or that affect the behavior of governments and funding bodies. Also, each institution is inherently different in how it wants to weigh quality, performance, and impact. Using a standardized algorithm (that may not even be transparent) could lead to decisions in conflict with the values (and the policies) of any one institution.

3. Third, what is the use of data within these companies once it is made available by research institutions? What degree of privacy is afforded to university and researcher data once it is uploaded to utilize the productivity tools made available? Is this data fed into other businesses within the same company? Is it made available or re-sold to third parties? What would be the response of a commercial vendor to a government request for disclosure or a subpoena?

4. Fourth, some of these businesses effectively text and data mine materials submitted by researchers. What is the licensing agreement underlying these relationships? Do other researchers enjoy similar access rights? One of the issues brewing in the controversy around the boundaries of IP is the limitation demanded by many publishers to the use of third-party software for the purpose of extracting insights. It is debatable whether these demands are strategically sound (we believe they are not, at least in the case of the leading publishers) and whether

they make any sense in the case of mining for insights on broad themes, such as the emergence of new disciplines. On the other hand, we find the limitation of text and data mining for conducting scientific research unacceptable, full stop. If text and data mining can be patented, publishers should do so. If they cannot obtain a patent, then demanding that scientific research is conducted exclusively by using their tools is unreasonable.

Once again, this is not (and is not meant to be) a comprehensive list of all the issues posed by the provision of data tools and services outside university libraries. It is also not meant to provide recommendations on possible solutions, which will be articulated at a later stage and in different forums. We expect the list of issues to grow substantially over time, as more people weigh in from their vantage points and based on their experience and concerns.

TEACHING AND STUDENT ACTIVITIES

Background - The State of the Textbook Market

In 2016, the publishing industry was finally forced to acknowledge that the poor performance of higher education textbook publishing was being negatively affected by trends that are structural, rather than cyclical, in nature.

The events of 2016 were startling: at the end of 2015, Pearson's management expected U.S. higher education revenues to decline by 1 or 2%, reflecting a similar decline in college enrollment and flat spending per student – while holding market shares stable. By the end of 2016, results were much worse: organic revenue decline in courseware was -10%. Wiley reported even worse results: for the fiscal year (FY) ending in April 2016, textbooks sales declined by -15%, and for the FY ending on the 30th April 2017 organic revenues declined by a further -13%. McGraw-Hill Education reported a -8.9% decline for domestic higher education revenues in the FY ending on the 31st December 2016. Finally, Cengage reported for the fiscal year ending on the 31st March 2017 a revenue decline of -15% for the domestic Learning business.

The publishers entered 2017 with significant misgivings. Even if Pearson blamed a one-time inventory adjustment amongst bookstores for the revenue decline of 2016, its guidance for the following year implied a further 6-7% decline in underlying demand for textbooks and courseware in 2017 (management indicated that revenues may come in between +1 and -7%, depending on stocking strategies of retailers). Against such dire scenarios, the -3% organic revenue decline of 2017 was almost a relief, and it led management to issue a slightly improved guidance for a further -6% decline in demand for courseware in 2018 (against a previously issued guidance of -6 to -7%). Once again, other publishers posted similar or even worse numbers: Wiley reported a -6% revenue decline for textbooks in the FY which ended on the 30th April 2018, McGraw-Hill Education Higher Education revenues declined by -2.8% and Cengage a -3.4% decline in domestic learning revenues for the FY ending on the 31st March 2018 (Exhibit 9).

Exhibit 9

US Higher Education courseware revenue decline for selected publishers.

	2016	2017
McGraw-Hill Education	-8.9%	-2.8%
Pearson	-10.0%	-3.0%
Wiley	-13.0%	-6.0%
Cengage	-15.0%	-3.4%

Source: annual reports, presentations

The continued decline of the higher education courseware business in the U.S. is driven by the interplay of three factors: student enrollment, pricing, and participation rates.

Factor 1: Student Enrollment

The first factor of student enrollment is cyclical in nature based on the economy (even if there are longer term trends in terms of participation). Pearson estimates that every 1% change (up or down) in the U.S. unemployment rate drives a 3% change (in the opposite direction) in the U.S. college enrollment rate. The robust labor market of 2016 and 2017, therefore, could be expected to drive down enrollment rates in college, which – in fact – declined by -1.4% in 2016 and by -1.0% in 2017 (Exhibit 10). Drops were not equal across all categories of institutions; however, for-profit and community colleges continue to show above average rates of decline, while both public and private non-profit four-year colleges show below average rates of decline (or some modest growth).

Exhibit 10

Changes in student enrollment in US colleges, 2015-2017

	2016	2017	2017
Four year, public	0.8%	0.2%	-0.2%
Four year, private nonprofit	-0.3%	-0.6%	-0.4%
Four year, for-profit	-13.7%	-14.5%	-7.1%
Two year, public	-2.9%	-2.6%	-1.7%
Overall	**-1.7%**	**-1.4%**	**-1.0%**

Source: National Student Clearinghouse Research Center

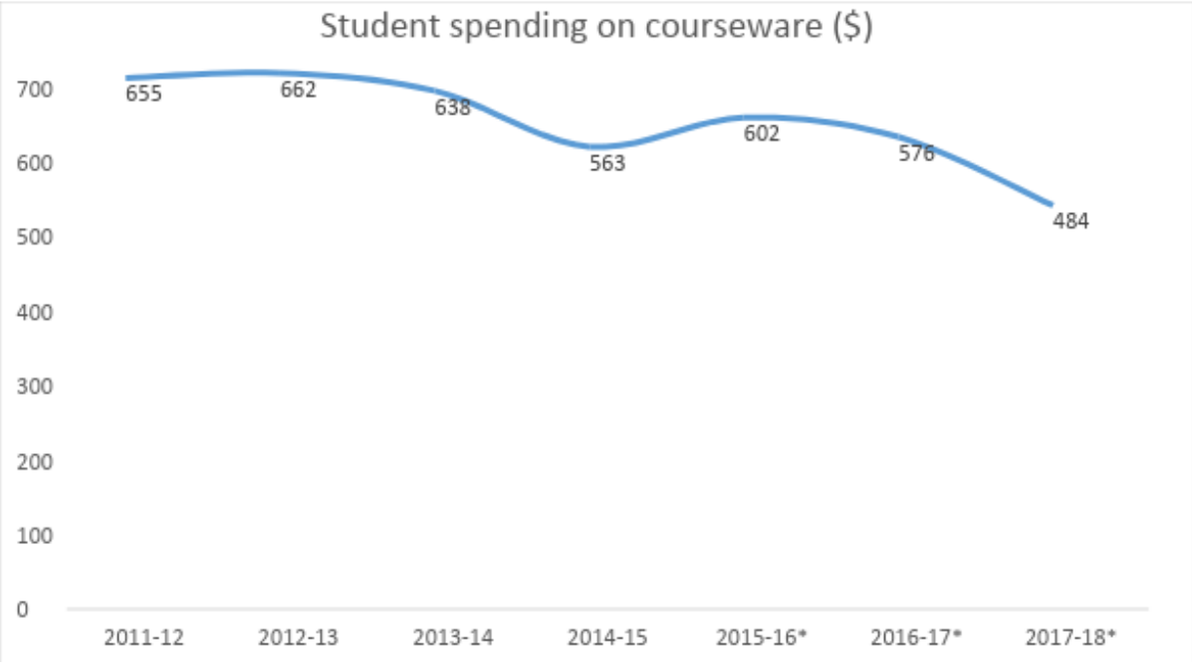
Factor 2: Pricing

Assessing changes in average pricing is rather difficult. Contrary to trade books, college textbooks do not have uniform suggested retail prices. College and digital bookstores therefore are free to set any mark-up above the wholesale price set by the publisher. In addition, in the transition from print to digital, publishers can lower their wholesale prices to reflect lower costs, such as the savings on printing, binding, shipping, and warehousing costs (which account for an estimated 7% of publisher revenues and 9% of costs). In fact, because of the potential to sell digital materials directly to students, publishers could stand to increase their revenues despite lower prices by appropriating at least part of the estimated average 25% mark-up practiced by bookstores. Also, "inclusive access" (a model where a time-limited subscription to digital textbooks is automatically charged to students) allows publishers to also increase overall revenues by increasing their sell-through rate and potentially convincing institutions to subscribe to a broader range of content.

Student spending habits may show some impact on price changes. The annual student survey published by the National Association of College Stores suggests that student spending on textbooks has declined over recent years, thanks at least in part to a variety of strategies to access the materials, including textbook rentals, alternate editions, and used books (Exhibit 11). However, other studies suggest that students opting out of purchasing textbooks likely also contributes to the decline in student spending, which calls into question whether students are in fact better off.

Exhibit 11

Student spending on courseware continues to decline according to some sources.



Source: National Association of College Stores

Product mix also influences realized prices. Follett explains a fairly broad range of options and prices: for a $100 print textbook being sold new in a college bookstore, the used version will command about $75, rental should cost anywhere between $50 and $65, digital editions will retail anywhere between $40 and $60 and custom editions (which are assembled exclusively for a specific campus) will retail at $30. Custom programs are arrangements in which the publishers produce materials that are specifically customized for a course by assembling chapters and articles from multiple sources. These materials are often offered at a steep discount, in large part because of the increased sell-through rate, since students are unable to seek used or rental copies off campus. This has a similar effect as inclusive access, and though different, sometimes the two approaches are combined. As the availability of lower cost options have proliferated over the last decade, it is easy to see how average spending could appear to be in decline.

Factor 3: Participation Rates

Student participation rates (the rate at which students buy new books) are also difficult to assess. The National Association of College Stores survey indicates that 63% of college students buy new books, 56% buy used books and 25% acquire digital materials. This data does not paint the full picture, since most students apply a mixed strategy based on the available options and whether they believe they may need the textbook beyond completion of the class. Most publishers mention seeing market research pointing to about 25 to 33% of students in an average U.S. college course buying new books, about 50% buying used books and the balance resorting to other strategies (borrowing, renting, photocopying, just not using any course material, etc.)

This factor is perhaps the most controversial driver of structural change in the market. Until a couple of years ago, publishers maintained that student segments are different, and that students who buy new books are unlikely to shift to renting, borrowing, or even buying used books. Since the sharp decline in 2016, however, publishers started to review their beliefs. Pearson's management, for example, has been conceding for about two years that renting is indeed affecting demand for new books.

In general, there is a sense that publishers have limited knowledge of the drivers of student behavior, and that much of what determines how a student will behave is still unknown. In part, this limited knowledge is the legacy of a cultural bias among the publishers, who have always viewed college professors as their true customer, as they are the ultimate decision makers responsible for adopting textbooks. This attitude will change; for example, Pearson's Chief Strategy Officer comes from Kantar, WPP's market research unit and is likely to make a major effort to change the culture of the company and spend more resources to understand students and their needs and choices (in order to more effectively market to them).

The Rise of "Inclusive Access"

In 2017, for the first time Pearson's North American digital higher education revenues were higher than print revenues. For McGraw-Hill Education, digital learning solutions accounted for 62% of revenues in 2017. At Cengage, in the FY ended on the 31st March 2018, digital solutions accounted for 53% of learning revenues. While these numbers still leave most students in most courses using print materials once the used and physical rental markets are taken into consideration, it still means that the rising adoption of digital solutions in the years to come will generate vast amounts of data on students and the teaching performance of faculty.

Publishers are responding to these negative trends by actively supporting "inclusive access." In its simplest form, inclusive access is an arrangement where a digital subscription to required course materials is included when students enroll in a course by a price pre-negotiated with the publisher. The institution may recover the cost of these subscriptions by directly billing student accounts for each applicable course or by building the cost of course materials into tuition and fees (the latter is less common right now, but is likely to grow with ongoing state and federal lobbying efforts by the publishing industry). While an opt-out mechanism is typically made available for students who wish to purchase their course materials on their own (it is required by U.S. federal regulations under 34 C.F.R. Sec. 668.164(c)), processes are far from streamlined and students may miss the narrow window of opportunity to opt-out. Some inclusive access programs include a flat-rate subscription to a publisher's catalog of digital materials. Theoretically, students could lower their cost of acquiring books across courses by effectively replacing individual purchases with an annual subscription. However, much remains to be seen.

Consortia like Unizin are beginning to act as aggregators around the inclusive access model, negotiating lower rates across publishers so that students do not have to pay for individual subscriptions (akin to how Spotify or Apple Music aggregate content across major music companies and make it available to consumers through a single subscription, instead of forcing consumers to subscribe to the catalog of each recorded music company). From the publishers' perspective, the combination of lowering costs (because digital materials replace print books), adding users (because very few students opt out and because there are no books feeding the used book market), and the potential to reduce distribution costs by circumventing bookstores

means that the economics of these aggregated subscriptions can be equal or better.

On its face, some argue that inclusive access is a win-win situation for both publishers and students: equal or higher profits for publishers and lower, more predictable spending for students. However, things are not as simple as they appear, particularly for students. All the students who lower their spending by reselling books, renting them, sharing them, or by checking them out of libraries may not be better off (and the lowest spenders will be worse off), depending on actual prices for new and used books and the cost of inclusive access deals. There is also no mechanism preventing publishers from resuming their historical rate of price increases once the model becomes widespread. Most important for the purposes of the issues raised in this document, once students transition to digital materials it enables both their institutions and the commercial vendors to collect vast amounts of data on them: their physical location when they use them, their study habits, their learning profile, and granular knowledge on their performance. This poses significant privacy issues, and – potentially – legal liabilities which could become, at some point, very grave.

Inclusive access is beginning to draw legal challenges. In January 2019, a used bookseller filed a complaint against Trident Technical College (TTC) in South Carolina. The suit alleges that TTC made it hard for students to determine the price of inclusive access materials and for making opt out procedures difficult or unfeasible. The court documents also state that the inclusive access contract with Pearson required TTC to meet a quota in order to be guaranteed a discount, which illustrates the kind of Faustian bargain these deals can entail. The contract pits the financial interest of the institution against the financial interests of students (the institution loses money if too many students opt out) and also the academic freedom of faculty (the institution loses money if too few faculty assign Pearson materials). It also means that the institution cannot be sure of the ultimate price they will pay at the time they sign the contract. This type of deal is emblematic of the concerns that are raised in this document.

THE COMPANIES

Pearson

The Basics

Pearson is the remnant of what used to be a very large British conglomerate. The company started as a construction company in 1844, but by 1920 it had exited the construction business and started to diversify into finance (it owned a stake in Lazard) media and non-media assets (Chateaux Latour, the Madama Tussaud operations). Media assets included a chain of newspapers managed within the Westminster Press Group, the Financial Times and a 50% stake in The Economist, Penguin (a trade book publisher), various TV production businesses (which were later sold to RTL). All these assets have been divested totally or partially over time (like in the case of Penguin), leaving the education publishing assets at the core of what the company does.

Longman was the initial core of Pearson's education business. Over time, the company acquired, among others, Scott Foreman, the textbook business of Harper Collins, Prentice Hall, Addison Wesley, and the textbook business of Simon & Schuster. The acquisition of NES (National Education Systems) in 2006 effectively set Pearson on the path to transform itself into a digital learning company. Since 2011, as revenue growth started to falter, Pearson has been sizing down its educational portfolio as well.

At its core, the company aims at participating in education sectors that it expects to transition to full digital delivery of content and services, enabling itself to add value to content through data analytics. Thus, the company decided to exit the U.S. K-12 courseware sector, despite holding a leading market share. Management, in the end, decided that financial constraints, the opposition from segments of the educational community, and the highly-decentralized nature of U.S. school districts when it comes to technology decisions means that digital transformation of U.S. schools would either not happen or happen in ways that would make it difficult to build a profitable digital business for Pearson.

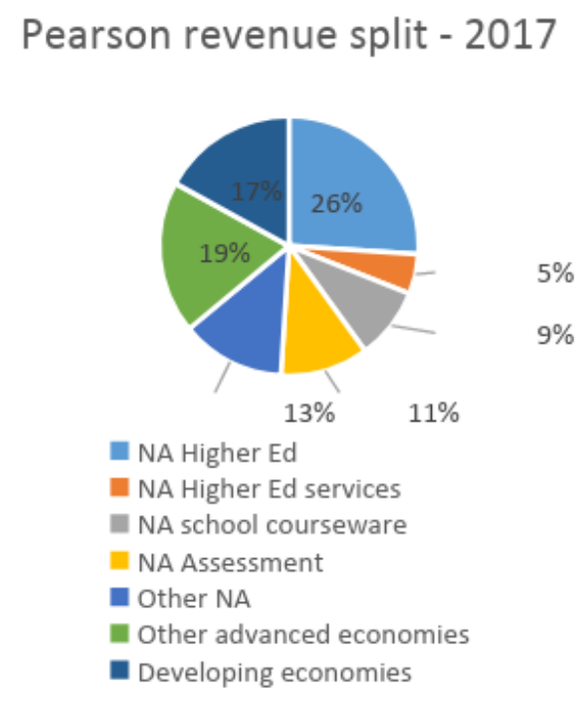
This leaves higher education at the core of Pearson's educational activities in the U.S.,

with another assessment business that is smaller and significantly less profitable than the higher education business. The company does not disclose operating margins by line of business, but we think that higher education should be a low twenties operating margin business, vs. low teens for assessment.

North American Higher Ed revenues account for 26% of Pearson's total revenues of £4.513/$5.972 billion and for 40% of North American revenues. In addition, Higher Ed services (which include activities like running the online activities of several colleges and universities, account for another 5% of total revenues and almost 8% of North American revenues (Exhibit 12).

Exhibit 12

North American Higher Ed accounts for 31% of all Pearson's revenues.



Source: Annual report, presentations

The company does not disclose the profitability of individual business lines. North America recorded an operating profit margin of 13.5% in 2017. Historically, we estimate that higher education operated at an operating profit margin in the low twenties, while school courseware would operate at mid to high teens and assessment at low teens. These margins have declined in recent years as demand

contracted across the board.

The Issues

Pearson has been the only publicly listed leading textbook publisher for several years. For many years, the main competitors (McGraw-Hill Educations and Cengage) struggled to perform adequately. In the aftermath of the financial crisis of 2008, at first Pearson enjoyed a grace period, as U.S. Higher Ed enrollment grew because of the weakening labor market, and K-12 textbooks budgets could still benefit from the lag between the downturn in the economy and pre-approved spending decisions. By 2011-12, however, as the labor market started to show signs of improvement, enrollment in U.S. colleges started to peak, and school districts pulled back spending. In 2012, Apollo Group (a private equity firm) acquired McGraw-Hill Education, while in 2014, Cengage reorganized after an organized bankruptcy. New management teams started to improve the competitive performance of both companies at the exact time when the market started to decelerate, aggravating the performance issues of Pearson.

Because of several profit warnings, the share price of Pearson has undergone a substantial decline (and a recovery in recent months). The stock peaked in March 2015 at £14.92; by January 2016, it had declined to £6.90 and, by September 2017, to £5.69. Since then, it has recovered to £9.17 thanks to more stable performance and the decision to dispose of several assets, as well as reduce costs aggressively.

Management has been very vocal, at least with the financial community, about the declining value of print-based textbooks in the long run, to the point that many investors interpreted the message as “the long-term value of the print textbook business is zero”. Whether management meant to convey this message is irrelevant; Pearson has educated the financial community that only digital courseware and data analytic services will have value in the future. Effectively, Pearson is now committed to driving the fastest possible adoption of digital courseware, and exiting business where this transition is unlikely or impossible.

Pearson has explicitly highlighted, in some of its investor presentations, how data analytics can affect outcomes for students. Very clearly, management believes that digital tools are not just aimed at the face value benefits of streamlining learning, increasing faculty productivity, and lowering costs. A combination of student profiling

through adaptive learning tools (which identify the best way to teach to different segments of students), and reporting (through the analysis of study and testing performance of students) is expected to transform schools with very large enrollment numbers, where it is less cost effective to monitor individual students. Pearson's strategy of focusing its textbooks on institutions with large numbers of students and a lower rate of student achievement naturally leads to developing and marketing these tools.

McGraw-Hill Education

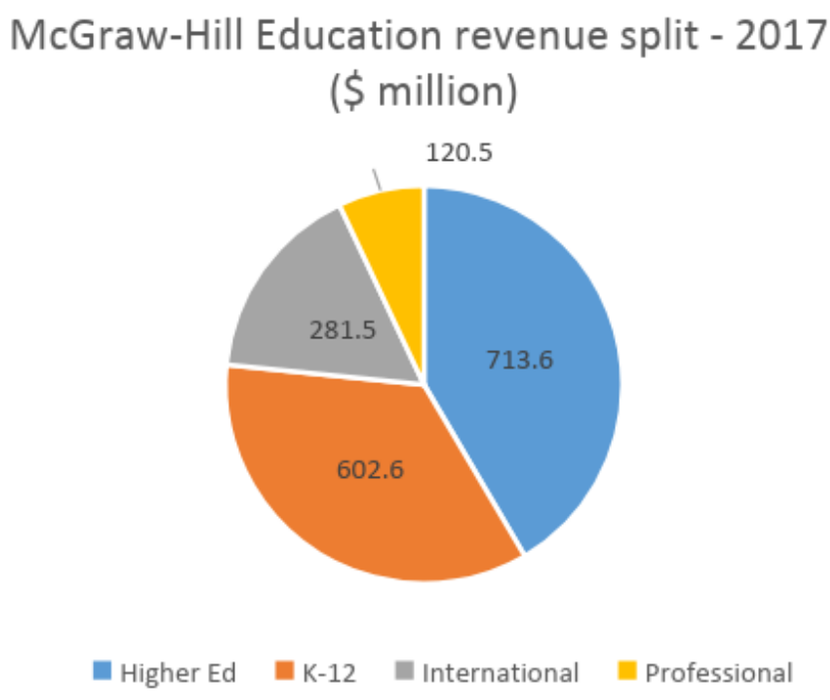
In March 2013, McGraw-Hill closed the sale of its education business to Apollo, a private equity company for $2.4 billion in cash. In the nine months before the sale was announced, the company had achieved a 15.8% operating margin, while EBITDA in 2013 was negative (making any multiple not meaningful).

McGraw-Hill Education operates across higher education, K-12, and professional markets, but is very dependent higher education, which accounts for 41.5% of its revenues (Exhibit 13). The company exited the assessment business in 2015, effectively going in the opposite direction of Pearson, which has decided to abandon K-12 courseware and stay in assessment.

Apollo has hoped for a long time to IPO the business to exit the investment. Net debt at the end of 2017 stood at $1.832 billion, implying a net debt/EBITDA ratio of 8.5x. An IPO that would raise fresh funds and pay down debt would be quite a relief to management, but so far has proven impossible to pull off. The company filed papers in 2015 and again in 2016, but market conditions have prevented Apollo from carrying out this step. The company has spent about $700 million in technology since Apollo acquired it, and this demand on its cash flow shows in the net debt, which peaked in 2016 at $1.927 billion, rising from $1.242 billion at the end of 2013 (the year of the acquisition).

Exhibit 13

Higher Ed accounts for 41.5% of McGraw-Hill Education revenues



Source: Annual reports

Cengage

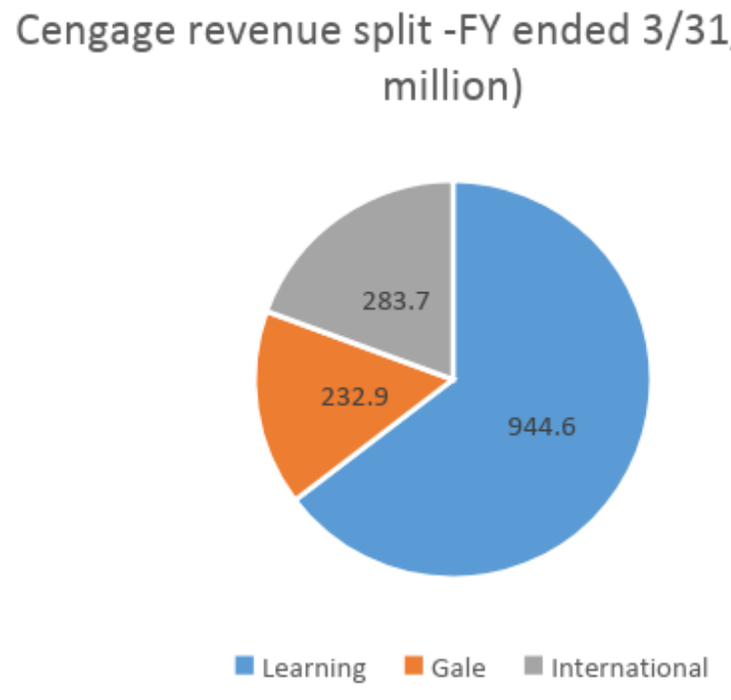
Cengage is the successor company to the old Thomson Learning (Thomson education business), which was sold to APAX (a London-based private equity company) and Omers (an Ontario pension fund) in 2007. The company is largely focused on the higher education market. Its learning segment accounts for almost 65% of revenues, and Gale, the primary sources business which sells mostly to colleges and universities, accounts for another 16% (Exhibit 14).

Cengage is a pioneer in inclusive access and Cengage Unlimited, its full catalog subscription, was the first full offering of this kind. Cengage was also the first major publisher to offer a product branded as open educational resources (OER), although the content is still purchased by students enrolled in a course. In recent years, Cengage has been the most visible of the major companies in slashing their prices and reframing their products as a solution to the high cost of textbooks (although

they continue to offer traditional products as well). Anecdotal evidence has revealed aggressive marketing tactics for Cengage Unlimited, including contacting student governments to request help lobbying the administration to purchase the full catalog subscription.

Exhibit 14

Higher education accounts for almost two thirds of Cengage's revenues (even excluding Gale's higher education revenues).



Source: Annual reports

Cengage struggled significantly in the years after the Apax/Omers acquisition. It tried to improve its performance by raising prices aggressively, but it could not avoid a reorganization and an organized bankruptcy in 2013, with the goal of shrinking the $5.8 billion of net debt it had at the time. The original acquisition cost Apax $7.75 billion, to which must be added the $750 million to acquire the college textbook business of Houghton Mifflin in 2008 and the textbook business of National Geographic in 2011 for an undisclosed amount. In the end, the debt burden proved too onerous, and the restructuring allowed the company to bring it back to acceptable levels. As of the 31st March 2018, the company had a net debt of $1,932.3 million. Since the adjusted EBITDA was $298.5 million, the net debt/equity ratio is a high, but

still manageable at 6.5x.

The long-term exit plan for the owners of Cengage is not clear. An IPO would be subject to the same investor sentiment which has prevented McGraw-Hill Education to complete its own IPO. An industry sale could run into anti-trust concerns, since the three largest publishers hold a substantial share of the market and any consolidation among them would raise concerns.

Other Publishers

It is worthwhile to note that there are other, smaller publishers in the Courseware business. Wiley, MacMillan, and SAGE all have higher education textbook businesses, often with one or more areas of specific focus (for example, Macmillan has an historic focus on the social sciences and humanities, and Wiley has always been very visible in business). OpenStax, a nonprofit open textbook publisher operating out of Rice University, has successfully launched a set of introductory for high enrollment courses. These textbooks are openly licensed, free online, and available in print at a low cost, and are also backed by a marketplace of optional supplemental materials.

Estimating market share among publishers has been difficult. Higher education publishers report their revenues to the AAP (Association of American Publishers), which aggregates and circulates them. The consistency of the reporting has been called into question, however, in part because there is no clarity about whether all of the publishers are consistently reporting their courseware revenues (some may include sales of teaching materials, while others may only report courseware shipments). Additionally, revenues are reported at the time of shipment (minus provisions for estimated returns), so publishers could potentially inflate revenues in any one quarter or year by shipping larger numbers of books during that time (or lowering their estimates for returns). Finally, as publishers scrambling to ink deals with institutions that trade higher sell-through rates in exchange for temporary discounts, revenues may not be the most reliable indicator of market share at this point of transition.

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.

Algorithms, on the other hand, do not need to be developed by each academic institution, just so long as they are transparent and can be analyzed and properly understood. "Black box" algorithms used in academic settings may contain any number of issues that are incompatible with the values of the institution. But, as long as the algorithms remain hidden, there is no way to know, and – as many anecdotical reports indicate – biases can easily be built into them, even inadvertently.

We do see some academic institutions actively opposing the use of data analytics. For example, the University of Ghent in Belgium announced in December 2018 that it would change how it evaluates its faculty. In the announcement, Rector Rik Van de Walle wrote – among other things:

> *"No more procedures and processes with always the same templates, metrics and criteria which lump everyone together" and "The model must provide a response to the complaint of many young professors that quantitative parameters are predominant in the evaluation process. The well-known and overwhelming 'publication pressure' is the most prominent exponent of this. Ghent University is deliberately choosing to step out of the rat race between individuals, departments and universities. We no longer wish to participate in the ranking of people".*

This is a courageous and daring course, and we acknowledge that most North American academic institutions may not be ready to abandon data and data analytics in order to inform decisions outright. Our goal is to ensure that institutions approach these decisions deliberately and in a manner consistent with their values.

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.