In pursuit of open science, open access is not enough

OVERLINE

In the pursuit of open science, open access is not enough

Insert Deck Here

By **Claudio Aspesi[1] *and* Amy Brand[2]**

After decades of debate on the feasibility of open access (OA) to scientific publications, we may be approaching a tipping point. A number of recent developments including Plan S indicate that OA upon publication could become the default within the next several years, at least in the sciences. Despite uncertainty about the long-term sustainability of current OA models, many established publishers who had been reluctant to abandon the subscription business model are now showing their openness to OA. Most recently, Springer Nature announced its intention to transition the vast majority of its English language journals to OA, including *Nature* and the Nature Research journals *(1).* While more OA can mean more immediate, global access to scholarship, we are not the first to point to "the continuing need for much caution in responding to passionate calls for a 'default to open'" *(2).* In considering open access, the time has come to move beyond ideology, towards practical, sustainable models, and careful analysis of the consequences of our business model choices.

As subscription revenues decline in the transition to OA, some publishers are prioritizing other sources of revenue. In fact, the availability of more full-text OA further enables the development of new services. One potential downside for the academy is the growing ownership of data analytics, hosting, and portal services by large scholarly publishers, enhancing their ability to lock in institutional customers through combined offerings that condition open access to journals upon purchase of other services. Even if such "bundled" contractual arrangements have the near-term benefit of increasing openly licensed scholarship, they may run counter to the long-term interests of the academic community by reducing competition and the diversity of service offerings, which in turn paves the way for new *de facto* monopolies. The healthy functioning of the academy, including fair terms and conditions from commercial partners, requires that the market for data analytics and knowledge infrastructure be kept open to real competition, within a global marketplace.

BUNDLED METRICIZATION

The bundling of journals (often using undisclosed pricing models) has historically worked well for commercial journal publishers at the expense of academic libraries (e.g., conditioning access to high-demand scientific literature on the purchase of resources for which there is little/no demand)*(3).* Exemplifying a different form of bundling, many publishers are actively negotiating transformative "read-and-publish" agreements with libraries and consortia, in which payment terms for access to journals as well as for author fees for publishing in OA journals are bundled into a single contract *(4).* Such transformative deals may accelerate the transition to OA, but discounted article processing charges also have the potential to influence where researchers opt to publish their work, contravening basic principles of academic freedom.

As OA continues to gain ground, some publishers are seeking to protect their profitability by accelerating investment in research infrastructure and data analytics, and by bundling these and other offerings with journal access. For example, Elsevier announced a controversial framework agreement in late 2019 with several Dutch academic and funding bodies that ties a presumed zero increase in spending for content access with pre-funded open access for affiliated authors and a commitment to partner on the development of new research intelligence tools and services. Although the details are not public, the implication is that these universities are contributing institutional metadata for Elsevier product development in exchange for OA publication by their researchers. *(5).*

Bundling of access and analytics is worrisome in light of the ways in which scholarly publishers are positioning themselves, largely through the acquisition of existing commercial and non-profit technology companies, to compete with stand-alone vendors of analytics products to provide decision-support tools to university administrators. Consider that vendors of analytics-only services, unable to combine their services with price breaks on content access or APCs, could be disadvantaged, increasing the probability that data analytics will become a monopolistic or highly concentrated oligopolistic market. At the same time, shifting revenue growth among commercial providers from content to data analytics could generate deflationary pressure on the revenues of pure journal publishers, starving them of the capital needed to compete effectively and innovate.

Given how our established metrics already influence the academic ecosystem, the risks for universities and academic freedom could extend well beyond excessive spending and reduced competition. The use of data to inform institutional decision making is of course, in principle, a laudable goal, and inclusion of a range of commercial interests in the development of analytics can provide useful skills and insights to advance such efforts. That said, the metricization of the academy and the spread of data analytics is not without concern to researchers, who are among those most impacted by the use of quantitative measures but with the least control over whether and how algorithms are applied and their outputs interpreted. Indeed, rigid models for assessing productivity are often less helpful in the promotion review process than qualitative assessments of their work.

The impact factor, widely recognized within the academic community as problematic but nonetheless central in many academic appointment and promotion decisions, serves as a cautionary tale: an algorithm created to rank journal quality morphed into a universal academic metric simply because of its widespread availability. So too, the senior leadership of many academic institutions around the world is preoccupied with university rankings, regardless of their validity. The proliferation of algorithms for comparing productivity within standard academic disciplines across individuals, departments, institutions, and nations has the potential to exacerbate bias and exert an abundance of control over core decision processes, such as resource allocation and career advancement decisions.

[1]Independent Consultant and lead author of the SPARC *Roadmap for Action* report
[2]Director of the MIT Press and Co-Founder of the Knowledge Futures Group

Without a competitive market fostering alternative measures, leading academic institutions may be more prone to optimize for the same limited indicators of excellence and set the same research investment priorities.

PORTALS AND PLATFORMS

Longer term, we are also concerned about the potential rise of new discipline portals, or enhanced full-text databases. Organizing information within a particular subject domain into a searchable index is by no means new to scholarly publishing. Indeed, the first abstracting and indexing services date back to the early days of digitization. Whereas bibliographic databases typically contain only metadata, keywords, and abstracts, full-text databases contain complete documents, creating the potential for augmented discovery services through AI-powered mining and analysis of full-text.

From the researcher's perspective, full-featured subject portals may make good sense. Systematic collections of research data and publications, conference proceedings, discussion threads, relevant events, and perhaps even media coverage and job postings, could become natural destinations for scholars in many disciplines. One reason such robust disciplinary portals have thus far failed to become widespread is the cost, considering the expense involved in creating a platform that is truly comprehensive in coverage, reliably curated and cross-indexed, and kept up to date. For example, the tremendous resources of the Chan Zuckerberg Initiative (CZI) made it possible to create the AI-powered Meta platform that CZI acquired in 2017 and that indexes and links to more resources in the life sciences than any competitor.

A larger historic barrier to full-text portals has been the fragmentation of scientific content. As long as most articles sit behind paywalls, it is challenging (although not impossible) for any one publisher to secure access to a critical mass of content with the requisite legal rights. The comprehensive adoption of OA and less restrictive Creative Commons licenses changes this dynamic and makes it easier to imagine how a large publisher or funder, with the scale to invest in the technology, could layer onto full-text aggregations functions such as collaboration platforms, data hosting, literature and dataset search and linkage, open reviews, proceedings and discussion threads, faculty news, job searches, and perhaps other activities of learned societies. While basic service tiers might be free to researchers, premium tiers and institutional contracts could command significant prices.

At the same time, access to the data and information exchanged by participants would provide the operator with unique insights into both past and predicted future productivity of departments and individual faculty, potentially leading to new "information arbitrage" markets. With the rise in biological and medical research intelligence, the biomedical arena is likely to produce the first robust portals, and others would no doubt follow. However useful such portals may be, the potential benefits must be weighed against the potential costs of monopoly control. Minimal or non-existent competition is likely to result in less favorable terms for subscribing institutions, whether on price, user privacy, or overall service quality.

How likely is this outcome? Although fully open and not a multi-purpose portal, Meta is a good indication of what's technically possible when it comes to automated sourcing, analyzing, and connecting of published content. On the commercial side, one initiative that aims to aggregate multiple data sources is Elsevier's Entellect, *"a new cloud-based data platform designed to help life sciences companies overcome the challenges of modern R&D by enriching and harmonizing proprietary and external data and delivering it in an AI-ready environment*." In 2019, Elsevier also launched a new PracticeUpdate community focused on advanced melanoma *(6)* to supplement the other medical communities it has hosted over the last several years. Looking beyond the life sciences, the company signed a "content integration" agreement during the same year with the Society of Petroleum Engineers *(7)*. These are just some of the ways in which one organization can establish the building blocks of a subject portal strategy.

Developing portals across multiple disciplines would enable economies of scale in building and running the underlying software and in selling institutional subscriptions, potentially leading to even greater consolidation. It is hard to imagine more than a handful of enterprises being able to afford the upfront investment required to build and maintain these platforms. Once established, it would be difficult for new entrants to gain sufficient scale to compete, increasing the risk of monopoly control and pricing.

A COMMUNITY BEHOLDEN

We have highlighted some "worst case" scenarios to urge preventative measures that can help ensure a robust and diversified ecosystem for data analytics and academic infrastructure. If it doesn't invest in alternative solutions, the academic community may find itself beholden to a small number of vendors for managing communities, data flows, research assessment and learned society communications, all within digital silos that could hinder the growth of cross-disciplinary collaboration and discovery. In response to these concerns, the Scholarly Publishing and Academic Resources Coalition (SPARC) outlined a number of practical steps that university leaders should consider (8). Among the steps proposed are: (a) ensure that appropriate institutional policies and personnel are in place to manage research data and faculty productivity analysis; (b) diversify the infrastructure ecosystem by investing in community-owned solutions and stronger cross-institution partnerships; and (c) actively partner with research funders and learned societies in these efforts.

The relationship between academic institutions and learned societies is complex. Many faculty are members or leaders of learned societies, and some serve on the editorial boards of society journals. Learned societies have been among the *least* enthusiastic supporters of OA, stemming from concerns about the loss of the journal subscription revenues that subsidize their operations. Many societies co-publish with large publishers, and if the transition to OA results in a revenue decline, societies could choose to partner with subject portal providers as one way to replace lost revenues. In order to offer an alternative path to sustainability, institutional leaders would be wise to involve learned societies in the development of community-owned infrastructures and consortia. Compensating societies for applying their disciplinary expertise and convening power to these efforts could provide them with new sources of revenue.

Even if monopolistic subject portals never come to pass, the possibility remains that a small number of companies will own most of the critical data assets, analytics, and platforms used by the scientific community. There have long been a limited number of academic journal hosting platforms, and in recent years the majority of these services have been acquired by publishers (e.g., Wiley's purchase of Atypon and SAGE Publishing's purchase of Global Village Publishing) or private equity firms (e.g., Accel-KKR's large stake in HighWire). Taylor & Francis's acquisition of F1000 Research is a recent example of market consolidation in the OA publishing platform space, but sev-

eral open source hosting and workflow solutions have begun to emerge *(9),* leading to welcome diversity in the technology choices of new OA publishers. Most of these solutions, though, lack solid plans for long-term growth and sustainability.

A first step to support competition and avoid monopolistic consolidation would be to engage in efforts to model consortial funding for and ownership of these and other non-commercial platforms. Universities should step up to invest in home-grown research infrastructures and cross-institution consortia, with the goals of establishing competition, sustaining best-in-breed open alternatives, and perhaps eventually providing a suite of services that can substitute for all-in-one commercial workflow solutions. Open, community-owned discovery and analytics services such as lens.org and meta.org, along with Stanford Libraries' home-grown analytics solution RIALTO, merit further attention from academic leadership, as do grassroots efforts to develop indicators of excellence focused on humanities and social sciences (e.g., HuMetricsHSS). So does the widespread use of altmetric indicators in journal publishing and the growing adoption of open standards like the CRediT taxonomy, which links standardized roles to author names in multi-authored publications, providing a qualitative indicator of researcher contribution.

Finally, university leaders should be poised to revisit the lessons of past collective efforts to learn from what did and did not work, in order to design effective and durable collaborations going forward. There is much to be learned, for example, from the long-standing success of the arXiv e-print repository in the fields of physics, mathematics, and computer science, fueled by a combination of grants, in-kind support, and institutional memberships.

The struggle for control over information and knowledge looms large everywhere we turn. When Berners-Lee created the World Wide Web, his intention was to enable researchers to share their work. Not only have our research communication tools and practices thus far fallen short of the decentralization that the Web made possible, but the evolution of the Web itself also reminds us that making vast amounts of linked data readily accessible to third parties can trigger a number of unintended consequences. The dominance of a limited number of social networks, shopping services, and search engines shows us how internet platforms based on data and analytics can tend towards monopoly. In the research information space, contracts are being negotiated establishing *de facto* terms and conditions for how data analytics services are being provided. Learned societies are being wooed. Research assessment metrics are being proposed. Building blocks for establishing discipline portals are being assembled. The time for the academic community to act in coordination is now.

REFERENCES AND NOTES

1. Press release, April 8, 2020, https://group.springernature.com/gp/group/media/ press-releases/springer-nature-plans/17877246
2. K. Smith, Three Questions with Don Waters. *Council on Library and Information Resources (CLIR) Newsletter* 131 (2019), https://www.clir.org/2019/10/clir-issues-131/.
3. T. C. Bergstrom, P. N. Courant, R. P. McAfee, M. A. Williams, *PNAS* July 1, 2014 111 (26) 9425-9430; first published June 16, 2014 https://doi.org/10.1073/pnas.1403006111.
4. L. Hinchliffe, Transformative Agreements: A Primer, *The Scholarly Kitchen* (April 23, 2019), https://scholarlykitchen.sspnet.org/2019/04/23/transformative-agreements/
5. Leaked document on Elsevier negotiations sparks controversy, *ScienceGuide https://www.scienceguide.nl/2019/11/leaked-document-on-elsevier-negotiations-sparks-controversy/* (2019) [no author; the easiest access to this source is via the URL].
6. Elsevier Launches PracticeUpdate, a Free Information Portal for Physicians, https://www.elsevier.com/about/press-releases/clinical-solutions/elsevier-launches-practiceupdate,-a-free-information-portal-for-physicians.
7. Elsevier and Society of Petroleum Engineers Collaboration will Empower Geoscientists to Make More Informed and Confident Decisions, https://www.prnewswire.com/news-releases/elsevier-and-society-of-petroleum-engineers-collaboration-will-empower-geoscientists-to-make-more-informed-and-confident-decisions300969190.html?mkt_tok=eyJpIjoiTVdaaVpEaGhZVE14WW1ZeCIsInQiOiJnWlp0NitJY083aGlBU2hMQkhVQzZDMktTNFwvVStFbXRPUFk3Mkg0U0Z0NnN4MEpzMUErSlZGZ0paQU1BUklkeGkyV0plQlA2ZURIQWlKQUp6VVM4QnVJcHVzczZNWTBhQUFxY0dJc1hiN3JZVTZHWHV2V1FJOFdaYVBoN3dhOEwifQ%3D%3D
8. SPARC Roadmap for Action, https://sparcopen.org/our-work/roadmap-for-action/ (2019) [the easiest access to this source is via the URL].
9. J. W Maxwell, E. Hanson, L. Desai, C. Tiampo… and E. Michelle, Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms, posted July 2019, https://mindthegap.pubpub.org/.

Acknowledgments: The authors would like to thank several anonymous reviewers and the excellent editorial team at *Science* for their invaluable input.

10.1126/science.aba3763