Scientific Publishing

In pursuit of open science, open access is not enough

Preventing monopolies in knowledge infrastructure is the next battleground for publishers and research institutions

By Claudio Aspesi* and Amy Brand†,‡

A
ter decades of debate on the feasibility of open access (OA) to scientific publications, we may be nearing a tipping point. A number of recent developments, such as Plan S, suggest that OA upon publication could become the default in the sciences within the next several years. Despite uncertainty about the long-term sustainability of OA models, many publishers who had been reluctant to abandon the subscription business model are showing openness to OA (1). Although more OA can mean more immediate, global access to scholarship, there remains a need for practical, sustainable models, for careful analysis of the consequences of business model choices, and for “caution in responding to passionate calls for a ‘default to open’” (2). Of particular concern for the academic community, as subscription revenues decline in the transition to OA and some publishers prioritize other sources of revenue, is the growing ownership of data analytics, hosting, and portal services by large scholarly publishers. This may enhance publishers’ ability to lock in institutional customers through combined offerings that condition open access to journals upon purchase of other services. Even if such “bundled” arrangements have a near-term benefit of increasing openly licensed scholarship, they may run counter to long-term interests of the academic community by reducing competition and the diversity of service offerings. The healthy functioning of the academic community, including fair terms and conditions from commercial partners, requires that the global marketplace for data analytics and knowledge infrastructure be kept open to real competition.

*University of Turin, Turin, Italy. †Academy of Medical Sciences and Humanities, Berlin, Germany. ‡Charité – University Medicine Berlin, Berlin, Germany.

INSIGHTS

POLICY FORUM
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 prefunded 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 (2). Ownership of the data may remain with the universities, though it is unclear whether they maintain perpetual access rights to analyses based on that data (6).

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 nonprofit 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 article processing charges, 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 academic community and the spread of data analytics are not without concern to researchers, who are among those most affected 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 largely 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. 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 artificial intelligence (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

BUNDELED METRICIZATION

The bundling of journals (often using undisclosed pricing models) has historically worked well for commercial journal publishers, often 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 or 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 and 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

---

1Independent Consultant, Zurich, Switzerland. 2The MIT Press, Cambridge, MA, USA. 3Knowledge Futures Group, Cambridge, MA, USA. Email: amybrand@mit.edu

---
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. Whereas basic service tiers might be free to researchers, premium tiers and institutional contracts could command high prices.

At the same time, access to the data and information exchanged by participants would provide the operator with valuable 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 highly concentrated control of the market. Minimal or nonexistent 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 multipurpose 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. In 2019, Elsevier also launched a new PracticeUpdate community focused on advanced melanoma (7) to supplement the other medical communities it has hosted over the past several years. Looking beyond the life sciences, the company signed a “content integration” agreement during the same year with the Society of Petroleum Engineers (8). 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 admittedly “worst case” scenarios, but we suggest that certain preventive measures could help ensure a robust and diversified ecosystem for data analytics and academic infrastructure and are worth pursuing even if the worst case does not come to fruition. 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 (9). Among the steps proposed: ensure that appropriate institutional policies and personnel are in place to manage research data and faculty productivity analysis; diversify the infrastructure ecosystem by investing in community-owned solutions and stronger cross-institution partnerships; and 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 copublish 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.

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 most 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 Press). Taylor & Francis’s acquisition of F1000 Research is a recent example of market consolidation in the OA publishing platform space, but several open-source hosting and workflow solutions have begun to emerge (10), 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, 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.

University leaders should be poised to revisit the lessons of past collective efforts to learn from what did and did not work, 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. 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 toward 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, 8 April 2020; https://group.spring-ernature.com/gp/group/media/press-releases/springer-nature-plan-s/37877245.
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. Courant, R. P. McAfee, M.A. Williams, Proc. Natl. Acad. Sci. U.S.A. 111, 8425 (2014).
4. L. Hinchliffe, Transformative Agreements: A Primer. The Scholarly Kitchen (23 April 2019); https://scholarlykitchen.sspnet.org/2019/04/23/transformative-agreements/.
5. Leaked document on Elsevier negotiations sparks controversy. ScienceGuide (2019); www.sciencugde.nl/2019/11/leaked-document-on-elsevier-negotiations-sparks-controversy/.
6. B. Upton, Dutch universities agree 'new way of working' with Elsevier. Research Professional News. 12 December 2019; www.researchprofessionalnews.com/fr-news-europe-infrastructure-2019-12-dutch-universities-agree-new-way-of-working-with-elsevier/.
7. Elsevier Launches PracticeUpdate, a Free Information Portal for Physicians; www.elsevier.com/about/press-releases/clinical-solutions/elsevier-launches-practiceupdate, a free information portal for physicians.
8. Elsevier and Society of Petroleum Engineers Collaboration will Empower Geoscientists to Make More Informed and Confident Decisions; www.prnewswire.com/news-releases/elsevier-and-society-of-petroleum-engineers-collaboration-will-empower-geoscientists-to-make-more-informed-and-confident-decisions-2019/.
9. SPARC Roadmap for Action (2019); https://sparcopen.org/our-work/roadmap-for-action/.
10. J. W. Maxwell et al., Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms, posted July 2015; https://mindthegap.pubpub.org/.

ACKNOWLEDGMENTS
We thank several anonymous reviewers. C.A. is a paid consultant to SPARC and was the lead author of (9).

10.1126/science.aba3763

SCIENCE AND DECISION-MAKING: COVID-19
Harnessing multiple models for outbreak management

Expert elicitation methods and a structured decision-making framework will help account for risk and uncertainty

By Katriona Shea1, Michael C. Runge2, David Pannell1, William J. M. Probert4, Shou-Li Li5, Michael Tildesley6, Matthew Ferrari1

The coronavirus disease 2019 (COVID-19) pandemic has triggered efforts by multiple modeling groups to forecast disease trajectory, assess interventions, and improve understanding of the pathogen. Such models can often differ substantially in their projections and recommendations, reflecting different policy assumptions and objectives, as well as scientific, logistical, and other uncertainty about biological and management processes (1). Disparate predictions during any outbreak can hinder intervention planning and response by policy-makers (2, 3), who may instead choose to rely on single trusted sources of advice, or on consensus where it appears. Thus, valuable insights and information from other models may be overlooked, limiting the opportunity for decision-makers to account for risk and uncertainty and resulting in more lives lost or resources used than necessary. We advocate a more systematic approach, by merging two well-established research fields. The first element involves formal expert elicitation methods applied to multiple models to deliberately generate, retain, and synthesize valuable individual model ideas and share important insights during group discussions, while minimizing various cognitive biases. The second element uses a decision-theoretic framework to capture and account for within- and between-model uncertainty as we evaluate actions in a timely manner to achieve management objectives.

EXPERT ELICITATION AND JUDGMENT

Formal methods for elicitation of information from individuals were developed to harness the collective knowledge of many minds while avoiding the failings of individual experts (e.g., overconfidence) and the problems that arise in group interactions, such as agreeing with field “leaders” (dominance effects), focusing on suggestions raised early in the process to the detriment of other ideas (starting-point bias, groupthink, anchoring), the dominating effects of “loud voices,” and overly rapid adoption of early ideas that might, on more careful consideration, be incorrect (4, 5). In these formal methods, idea generation and idea evaluation are deliberately separated, allowing a fuller range of possibilities to be explored and a wide range of uncertainties to be assessed. As one example, in the IDEA protocol for expert elicitation (6), once experts are clear about the questions, they individually provide initial best estimates and ranges, receive feedback on how their estimates compare with others, discuss the results, and then provide a final individual estimate. Some protocols, including IDEA, are designed to work remotely— an essential requirement in the present COVID-19 context.

To harness both the creativity of individuals and the insights of groups, variations on the Delphi method (developed by the RAND Corporation in the 1950s and included within the IDEA protocol) and the Nominal Group Technique (7) involve both independent and interactive stages in an iterative elicitation process (8, 9). The expert judgment literature shows that a failure to manage the elicitation process well can lead to generation of biased information and overconfidence (4, 5). Expert judgment approaches have been used for elicitation from individual experts in a wide range of relevant settings, such as development of clinical guidelines (8), and in conservation and ecology (9).

MULTIPLE MODELS

There are a number of existing approaches for dealing with multiple models in weather and climate research (10), fisheries (11), and disease forecasting (2, 12). Such forecasting efforts generally focus on statisti-