

# **The Closed-Loop Academic Publication Data Conundrum**

UMD Libraries Innovative Practice Forum, June 8th, 10-11am

Joseph A. Koivisto and Jordan S. Sly, University of Maryland Libraries

## **Introduction:**

For this presentation we are looking at some of the potential implications of the increasingly closed-loop model of academic publication, student learning platforms, personnel management, and other elements of quantification being developed and sold by an increasingly small number of academic solutions companies. While there are a number of interpretations of these mergers, acquisitions, and ventures, from our perspectives as librarians from two ends of the research apparatus, we see many problems stemming from the increasingly tight connections between the scholarship that is produced, the perceived value of these publications in the market, and impacts on future research funding, student development, faculty positions, and more. We are uneasy, as we will discuss, by the ways in which this constricting pipeline may endanger our shared values in libraries and in higher education more broadly.

## **Outline**

As a quick outline, we will illustrate some of our animating questions that drove us towards this work, then highlight a select review of literature that has informed our ongoing project development as well as an explanation of one of the major examples that informs this study.

## **Animating Questions**

Animating question 1

Our first question focuses on the potential attenuation and narrowing of research areas based on the citation matrix and high-impact model for publisher reputation, vendor pricing, and library acquisitions and the lasting impact this may have on the research enterprise and student outcomes.

#### Animating Question 2

Similarly, our second question looks at the notions of the data focus on publications and how this is being packaged and sold to universities. Importantly, this is not just in the form of research databases, but also in the form of metrics counting for promotions and other human statistical purposes within the university. The question is threefold: 1, does the privileging of the citation information favor specific disciplines? Does this impact the research direction of the university? And whose hand is on the wheel with this privilege? Are there customers outside of the university apparatus who benefit from this direction?

#### Animating Question 3

What does this potential steering mean for future students and faculty?

#### Animating Question 4

Crucially, to what extent do the acquisitions and acceptance of these models clash with our professional ethics? Are we maintaining our goals as a profession or are we inadvertently helping to steer the direction of research to align with the needs of industries that may work against our collective morals?

#### Animating Question 5

Finally, to what extent have we, as in those in higher education broadly, but also academic libraries more specifically, done this to ourselves by volunteering our work into this quantified system and allowing, to paraphrase Theodore Porter, metrics and assessment to make decisions without seeming to decide –to trust in the purity and objectivity of numbers despite not always seeing the whole of the snake as it eats and potentially devours itself.

## Why it matters

For this sketch of literature we will provide a brief summary of some of the studies in this area as it helps to define our unique perspectives on this topic as we continue to develop our project given our roles within the libraries and our individual professional perspectives.

### Jordan's Start

While there are overlaps with the external critiques of modern higher education, internally within industry journals and conference cocktail hours, the bogey man has increasingly been the hyper-neoliberalisation, commercialisation, and quantification of the university. As universities seek to find efficiencies to maximize entrepreneurial outputs and profit, education enterprise companies are more than willing to provide all the tools necessary to research, manage, publish, hire, purchase, and develop people, technologies, and educational content.

To many, the quantification of the core mission of the university: to educate, is, effectively, anodyne as the rankings and the prestige that follows allows the selectivity both in the hiring of professors and the selection of students to create an elite university setting. As Pierre Bourdieu showed in his 1984 ethnographic study *Homo Academicus*, however, this is not an evenly distributed sentiment as the hierarchies of academic power are not distributed to all areas of the university equally. Central to Bourdieu's work is the notion of power, prestige, and

privilege within the universities he studied. The book itself is highly contextualized by its time and location, but Bourdieu's insight about the constricting influence of academic power is relevant. Bourdieu notes that the power of certain academics –by discipline, by impact, by status– has something of a stultifying effect on the development of research as the burden of this prestige hinders alternate work (Bourdieu, 1984). Extrapolating this to today we can think about the use of tools like Web of Science (Clarivate) which allows researchers to bother only with the most highly cited literature relevant to their field, which of course is a self-repeating cycle that privileges certain researchers, journals, etc. more than others.

What is important here is the monetization of this prestige-power cycle of the *Homo Academicus* both internally to the universities and to the external enterprise vendors. Bourdieu discusses the insular power dynamics of the university and the academic apparatus, but bringing this study up-to-date would require the examination of a more global cycle of information.

For a more current accounting of the concerns over quantification, Stefan Collini's 2017 work *Speaking of Universities* addresses the history and values of universities from the perspective of the internal industry concerns. Importantly, Collini, like Bourdieu, is looking at the cultures and structures of higher education and the threat of quantification on the scholarly apparatus. Collini picks up on Bourdieu's general notion of elite academic power, but further discusses the economic impact and incentive of universities to quantify and focus on the impact of particular disciplines over others. While the work ranges in its scope, it importantly discusses the ways in which the business incentives of the modern university finds purchase in the corporate sponsors on both ends of the cycle. Graduate job placement on the one end and research derived innovations on the other –sponsored and supported by interests outside of the university itself (Collini, 2017).

So what? This matters for the purpose of this research because the quantification furthers the internal stratification of academic prestige, increases the reliance on large enterprise companies who can leverage this prestige, and furthers the interests of stakeholders outside of the industry.

Companies like Clarivate are able to easily sell solutions to this perennial problem. Increasingly, too, as Collini notes, the competitiveness of a university depends on these precise metrics facilitated internally, yet levered externally (Collini, 2017). Administrators can turn to statistics and assessment metrics as a way of making decisions with a sense of rationality, despite the inbuilt biases of these numbers. Theodore Porter's 1995 work, *Trust in Numbers*, is a highly useful study for thinking about the ways in which these seemingly objective measurements belie the inequities demonstrated by Bourdieu and Collini and adds a perspective from outside of the university to avoid conspiratorial self-victimization (Porter, 1995). As Collini discusses, the values of a university reflect the revenue stream and as Porter discusses, the metrics working towards a positive reaffirmation of this focus. In other words, the quantification of university values allows an abstraction that works to justify a particular set of needs of the university apparatus through the guise of routine assessment and rankings and enforces a potentially stultifying hegemony and standardization that continues the funnel of academic power and prestige discussed by Bourdieu.

Open Access scholars like Peter Suber have long discussed these issues and the issues of the monopolistic scholarly publishing industry. Much of this work, which can be found across many publications, is summarized in his 2012 *Open Access*. Suber and many others address this publication condrium of prestige, citation tracking, and impact and note that this is one of the more difficult hurdles to clear in the way towards a more equitable publishing landscape

—although increasingly many in universities are understanding the perverse nature of these deals to a much greater degree (Suber, 2012). Professors in previously OA adverse areas in the humanities, too, as Martin Paul Eve discusses, have also embraced the Open Humanities despite the ingrained prestige publishing model (Eve, 2014). It is perhaps part of this increasing awareness that larger publishers are seeking new models, such as Elsevier seeking a more direct-to-consumer model connected to its Mendeley citation management platform, the merger of ProQuest and Clarivate which diversifies the reach of university enterprise needs, platforms like Ebsco’s Panorama which provides active use metrics and impact data, and companies like Proctoro controlling the educational experience of students—to name a few.

These companies understand the shifting nature of the academic landscape and the emphasis for an increased use of dynamic metrics. As Espeland and Sauder (2008) note, however, this push towards the quantification of impact is not merely an academic issue and is reflective of our society writ large. They emphasize the importance of what this says about us from a sociological perspective as we—both in universities and outside—are more focused on the ‘correctness’ of the data as opposed to questioning its construction as we continue to enforce what they refer to as an “ethics of numbers,” which contains the values of our institutions (Espeland and Sauder, 2008).

Working from a similar notion to Bourdieu, but from a frankly more scientific and rigorous methodological perspective, Callaham, Wears, and Weber (2002) found that the impact factor and prestige of the journal outweighed the value of the scholar themselves within a sample set of papers that they tracked (Callaham, Wears, and Weber, 2002). This is important because, as Collini and others have shown, the rise in metrics and quantification has benefitted the publishing industry in what might be an outsized way. The favoring of the journal as a method of

citation analysis demonstrates that control over the publication and availability of these journals is a way to ensure demand and profitability. From a university perspective, the striving for faculty to perform within these journals then becomes a metrics cycle of outlay, talent development, and research material procurement –again, what these educational enterprise solutions are seeking. Carpenter, Cone, and Sarli (2014) highlight these metrics to enhance the promotions and tenure processes and to aid in obtaining research funds and grants by emphasizing the role of bibliometrics as a pragmatic way of demonstrating researcher value above the noise of the academic publishing morass. As a result, however, they too found that the journal prestige and impact was more important than the reputation of the researcher themselves in the overall citation index score (Carpenter, Cone, and Sarli, 2014). Klein and Chiang (2004) found that this citation emphasis on specific journals deemed high impact through SSCI (Social Science Citation Index) aggregation additionally showed some evidence of citational bias; that by following the high impact citations scholars were unintentionally promulgating a strain of academic ideology that favored specific disciplinary interpretations and not others (specifically conservative economic models that favored market solutions in notably conservative or centrist journals. Klein and Chiang, 2004). Again, this demonstrates some of the danger in the winnowing out of research avenues in favor of prestige citation metrics and the appearance of academic favoritism. Critically, to the author's point, this all happens in the opaque black-box of proprietary information.

Additionally, as Posada and Chen (2018) discuss, the move to open access by many universities has caused the big publishers to seek additional sources of revenue and consolidate control and market share through mergers and acquisitions specifically looking at Elsevier (RELX), Wiley, Taylor and Francis and the ways that these consolidations shored up the

academic publishing apparatus and developing tools to ensure that customers are driven towards these products. Crucially, they also found that university strategies and research funding followed the metrics of high impact research provided and supported by these companies (Posada and Chen, 2018). As Grossmann and Brembs (2021) discuss, the cost creep incurred by libraries reflects this pivot. Increasingly, the enterprise companies spend huge amounts on lobbying, technology, and direct-to-consumer initiatives, and additional contracts (government research, private research orgs, etc.) as well as researching and implementing differentiated costing models to ensure profit through off-set models that shift payments from the back-end to the front-end of the research cycle (Grossmann and Brembs, 2021). This all makes sense, of course, from the perspectives of these big publishers. This is not an evil practice, but it is one that detracts from our stated values and goals as research institutions.

#### SLIDE 12

Where do we find ourselves – as libraries and librarians – in this evolving environment of scholarly datafication and the production/consumption cycle of data-driven impact? Consider Jeff Pooley’s recent article “Surveillance Publishing”, out just this year in the *Journal of Electronic Publishing* (2022). In light of Proquest’s acquisition by Clarivate, Pooley presents surveillance publishing as a descriptor for companies that derive revenue by peddling predictive solutions to research questions based on aggregation and analysis of researcher behaviors. Be it searching for and accessing publications, research creation and publication, or citation, surveillance publishers are incentivized to bring insights to market that have been derived through user behavior tracking, distant machine-driven reading of the scholarly corpus, and impact metrics, all of which are aggregated and churned through the alchemical cauldron of



trade-secret-protected analytical processes and algorithms. Or as Pooley puts it, “minting money from behavioral by-products.”

#### SLIDE 13

This methodology brings with it the numerous attendant issues of ceding control of scholarly communications and material evaluation to blackbox systems. Inherent bias and algorithmic racism/sexism/&c. enter into a self-amplifying feedback cycle in which corporate-sourced assessments infect every corner of scholarly practice: hiring, promotion & tenure, grants, citation, and – ultimately – publication.

#### SLIDE 14

We then risk establishing a scholarly environment in which up-and-coming researchers are incentivized to perform in ways that align with the corporate algorithmic interpretive lens. All the while, we help to perpetuate this cycle by pumping the system full of more and more of its vital lifesource: cash money.

#### SLIDE 15

We, the academy, have become the product *and* the consumer, all in one. Where are libraries in this?

#### SLIDE 16

With the acquisition of ProQuest, Clarivate not only acquired a considerable cache of content, but they also got the corporate keys to library infrastructure through Ex Libris, one of the major producers of integrated library software (including Aleph, our current ILS) (Schonfeld, 2021). Now under the Clarivate corporate tent, library systems represent yet another data point to be fed into the scholarly insights machine. Additionally, and perhaps more alarmingly, libraries –

at least through their software – now become an additional module of monolithic corporate offerings for higher education software solutions.

#### SLIDE 17

Through centralization and integration of libraries into a single suite of tools offering full coverage for university needs – enterprise management, library services, faculty performance tracking, tuition and fees management – claims of efficiency and seamless integration can be realized. But what for us would appear to be a seamless user experience is for Clarivate a seamless data collection activity.

#### SLIDE 18

Taylor (2021), in their piece “Mergers, Acquisitions, and my tinfoil hat”, opines that while it may seem like the hand wringing of skeptics and contrarians, there is simply too much money to be made. Library systems themselves do not represent a growth market and publishers have essentially pushed library collection budgets past the breaking point with usurious subscription fees. Now, with the inclusion of libraries in the corporate holdings of academic insight firms, there appears to be yet another piece of the carcass to be used.

#### SLIDE 19

Of course, this represents an ethical concern regarding patron privacy and extractive capitalist approaches towards user data. In light of the fact that we already know library vendors sell our data, it is also not a fanciful hypothetical (see LexisNexis, Thomson Reuters). Now, we are faced with a reality in which data extraction and capitalization are not only lucrative, but also convenient. The convenience is only furthered by the almost universal transition of library software to vendor-hosted software as a service (SaaS) models in which libraries use systems on servers that are beyond our control and scrutiny. Aside from vendor disclosures and assurances,

we won't know what data they are collecting and how. But we will know that the collection process will be one step less complicated as the data already lives on their machines.

With the high likelihood of library data entering the academic counting machine, we are faced with the probability that our data will help to drive the metrics-driven amplification cycle Jordan discussed. This is not a new prospect for libraries as bibliometrics and citation counts are frequently touted as objective levels of scholarly value. However, this new methodology is a sleeker more efficient machine that integrates our data without us even lifting a finger.

#### SLIDE 20

And while we have harped heavily on Clarivate, they do not stand alone in this new environment. Recently, Elsevier made moves to acquire Interfolio, a company that sells a variety of products that cover faculty searches and hires, promotion and tenure dossiers, faculty activity reporting, and more. Roger Schonfeld (2022) observes that this is in direct competition with Clarivate, evincing a tit-for-tat corporate arms race whose measures of escalation will be who has the bigger data sets to analyze. Again, there is too much money to be made for companies not to explore every possible revenue stream.

#### SLIDE 21

Schonfeld also notes resistance to these types of mergers, stating that this must lead us to consider the implications of buying in to on-stop-shop research infrastructure that bundles all of our data handling into a single pair of corporate hands.

#### SLIDE 22

In her 2021 book "A city is not a computer: Other urban intelligences", Shannon Mattern reminds us that "Procurement is political, – both in the police department and the library." As every aspect of higher education – from the library to faculty activity tracking to promotion and

tenure to research to impact assessment – becomes the target of techno-solutionist ideals of data connectedness and efficiency, we must consider the ethical and political implications of the technologies that we underwrite through our data and our dollars (and as a public institution, we are speaking about public dollars). Our dollars and data no longer exist within a silo of library-centric needs and uses, but rather help to feed the larger insight machine that, as Jordan so eloquently laid out, can have deleterious effects on scholarly behavior. As institutional data is conveniently and efficiently collected through vendor surveillance and fed back to the institution, at a hefty premium, we as librarians, through our procurement decisions, are implicated for our contributions. These metrics and insights shape behavior and demand, leading us to acquire journals because of impact, because they get cited the most, because they have the most impactful articles, because they court the most prestigious research, because their insights indicate what are the most impactful areas of scholarship, because we bought the most impactful journals, and so on and so on. As the trend towards vertical integration of ILS vendors, publishers, activity trackers, enterprise management solutions, etc. continues, the impulse to chase visibility within these vendor supplied models becomes less of a pursuit of efficiency and more of an echo chamber whose reach is comprehensive within the academic landscape, one that we are subsidizing with our data and our money.

## Project Description

At the moment, we are still in the questioning and reading phase and trying to determine what a future project will look like that addresses these issues. There are many options, but many limitations when it comes to scope and output based on the nature of this project. On the one hand, it would be simple to put together a survey to see librarian opinion on this topic, but to

what end? One of the roadblocks comes in when we consider the bogeymen that these companies represent and the visceral reaction that happens when their names are mentioned in professional circles. That doesn't do us any good. On the other hand, a more complex and coded survey that seeks to gamify would be difficult to develop and possibly unwieldy in its conclusions. This approach would seek to think about the information landscape from the user and administrator perspectives to better understand their information needs. Understanding what gap the companies are looking to fill to better know how to combat the narrowing of the information landscape through this understanding. This focus on the users could help to create better education materials about the dangers of these companies that are less reactionary and more nuanced in the critique overall. This approach would also help in the messaging to university administrators who are seeing the problems from a wholly different perspective and these companies are providing easy and convenient solutions. Findings would not only help to frame the problem more effectively but would enable administrators to make evidence-based arguments for or against particular vendors or license terms that mortgage our data and labor in exchange for convenience and "efficiency". Yet, these solutions may be doing more lasting harm to the missions, values, and most importantly, futures of higher education than is immediately obvious. Other solutions for study-type would be an expanded literature and or systematic review that would provide both a wider perspective and a deeper statistical analysis to frame the impact of these deals, but I fear that due to the nature of the problem and the proprietary black box of many of these solutions, much of the data will be highly selective and the meta analysis somewhat short sighted without at least a dive into the corporate grey literature –but even then, the deals and numbers are likely to be highly individualized and fragmentary. Approaches such as this have traditionally been labor intensive and challenging – consider Matthew Reidsma's

systematic evaluation of discovery algorithms for which he had to custom develop computer applications to automated search and retrieval. What is clear from the literature is that many people are looking at these issues from many different perspectives from both within the library world and outside of it. Perhaps there would be value in simply expanding this presentation, which combines our two perspectives and the merging of a fairly disparate body of literature into a focused essay and critique to help connect the dots of those worried about the implications of this ouroboros of research

## Works Cited:

- Bourdieu, P. (1984). *Homo Academicus*. Stanford University Press.
- Callaham, M. (2002). Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals. *JAMA*, 287(21), 2847. <https://doi.org/10.1001/jama.287.21.2847>
- Carpenter, C. R., Cone, D. C., & Sarli, C. C. (2014). Using Publication Metrics to Highlight Academic Productivity and Research Impact. *Academic Emergency Medicine*, 21(10), 1160–1172. <https://doi.org/10.1111/acem.12482>
- Collini, S. (2017). *Speaking of Universities*. Verso.
- Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113(1), 1–40. <https://doi.org/10.1086/517897>
- Eve, M. P. (2014). *Open Access and the Humanities: Contexts, Controversies and the Future*. Cambridge University Press.

<https://www.cambridge.org/core/books/open-access-and-the-humanities/02BD7DB4A5172A864C432DBFD86E5FB4>

- Grossmann, A., & Brembs, B. (2021). Current Market Rates for Scholarly Publishing Services. *F1000Research*, 10(20).
- Klein, D. B., & Chiang, W. E. (2004). *The Social Science Citation Index: A Black Box—With an Ideological Bias?* 1(1), 33.
- Mattern, S. (2021). *A city is not a computer: Other urban intelligences*. Princeton: Princeton University Press.
- Pooley, J., (2022) “Surveillance Publishing”, *The Journal of Electronic Publishing* 25(1). doi: <https://doi.org/10.3998/jep.1874>
- Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.
- Posada, A., & Chen, G. (2018, June 15). Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers. *22nd International Conference on Electronic Publishing*. 22nd International Conference on Electronic Publishing. <https://doi.org/10.4000/proceedings.elpub.2018.30>
- Schonfeld, R. (2021). *Clarivate to acquire Proquest*. The Scholarly Kitchen. Retrieved from <https://scholarlykitchen.sspnet.org/2021/05/18/clarivate-to-acquire-proquest/>
- Schonfeld, R. (2022). *Elsevier acquiring Interfolio*. The Scholarly Kitchen. Retrieved from <https://scholarlykitchen.sspnet.org/2022/04/25/elsevier-acquire-interfolio/>
- Suber, P. (2012). *Open Access*. MIT Press.
- Taylor, J. (2021). *Mergers, acquisitions, and my tinfoil hat*. LibrarianShipwreck. Retrieved from <https://librarianshipwreck.wordpress.com/2021/08/16/>

[mergers-acquisitions-and- my-tinfoil-hat/](#)