# When Data Replace Norms: Platformisation of Knowledge Production

Completed Research

Lai Ma University College Dublin lai.ma@ucd.ie

# Abstract

Little attention has been paid to the research infrastructure that is tracing, tracking, monitoring and benchmarking individuals' and groups' performance and their implications for epistemic cultures and knowledge production. This paper discusses how the use of evaluative metrics and the dominence of data analytics can lead to platformisation of knowledge production by examining the normative view of science and epistemic cultures and the current development of research infrastructure such as vertical integration of research products. This paper argues that the dominance of commercial platformisation can decimate the negotiation powers of those who produce and review scientific outputs because researchers are acculturated to chase after funding, metrics, and data-driven economic and societal impacts. It is the objective of this paper to open up critical examination of the platformisation of knowledge production.

#### Keywords

Data brokers, epistemic cultures, evaluative metrics, knowledge production, platform capitalism, surveillance capitalism.

# Introduction

In the past few years, we are increasingly aware of the data businesses of big technology companies, where their platforms make profits by gathering and creating large amounts of data. In a recent book, Zuboff (2019) uses the term 'surveillance capitalism' to describe the monitoring apparatuses embedded in our everyday lives for predicting behaviour and for personalising and tailoring products and services for our perusal, while at the same time informing a plethora of companies, including insurance and mortgage providers, about our lifestyles by which social credits can be calculated and recorded.

While these studies mainly focus on major social media platforms including issues such as privacy, antitrust, political interference (e.g., Cambridge Analytica), little attention has been paid to the information systems used in universities and research institutions that are tracing, tracking, monitoring and benchmarking individuals' and groups' performance and the implications for epistemic cultures and knowledge production. These platforms are similar to Google and Facebook in that they collect and analyse data and then package them for sale. Whilst many digital platforms make use of user data to sell ads, as well as packaged data to third-party companies, the data products of research infrastructure are mainly sold back to universities and research institutions. It has been reported in the documentary, *Paywall: The Business Scholarship*, that an academic publisher, who also owns Scopus and Scival, can make about 35-40% profit margin—that is, higher than that of Google and Walmart.

Research infrastructures such as Web of Science (Clarivate Analytics) and Scopus were initially developed as indexing services, that is, they were/are databases for retrieving the most relevant sources by collecting and curating networks of citations. This is arguably one of the most important mechanisms in the history of search (e.g., Google's PageRank algorithm). The indexing services now also provides ranking of journals and individuals (e.g., WoS's Highly Cited Researchers) and data packages for university rankings and national performance-based funding systems such as the UK REF. Researchers in the field of bibliometrics, informetrics and scientometrics also use these services as data sources for their examination and analysis of the networks and progress of science, for example, by visualising the development and trends of research fields and the collaboration between individuals, disciplines and so on.

Although these platforms do not sell 'ads', they are essentially selling the data packages right back to the researchers and research institutions who are the sources of the data. By tracking publications and citations, they also create a hypercompetitive environment where researchers and scholars are pressured to publish more and faster (Ma, 2020, 2021). The more data researchers and scholars produce, the more data they can analyse and repackage as products for universities, governmental departments, and ranking companies. One should also note that many companies are data brokers and publishers at the same time. For example, Elsevier who owns data analytics platforms including Scopus and SciVal, is also a publisher who has an approximately 25% of market share of all journal publications (Fyfe et al., 2017) disseminated via the ScienceDirect platform. At the same time, they also run a reference manager product, Mendeley, which observes and collect readers' activity, makes recommendations (primarily Elsevier publications in the author's experience) and constitute a source of altmetrics data.

The business models of some data analytics companies and publishers, especially the so-called 'Big Deals', are based on the 'values' placed upon the number of publications and citations, and the perceived prestige of journals and publishers, in research evaluation and assessments including the review of grants and tenure and promotion applications and the allocation of block grants. Whilst metrics as 'value-tokens' (Biagioli, 2018) can be interpreted differently in different contexts, they undoubtedly have led to a market for comparing, classifying and ranking a wide range of entities (e.g., individuals, journals, etc.) in the research systems.

The harmful effects of the use of evaluative metrics have been discussed and documented in recent decades (e.g., Biagioli & Lippman, 2020; Wilsdon et al., 2005). The roles of publishers and data analytics, however, have seldom be questioned notwithstanding they play a substantial role in the system of knowledge production. Are there conflict of interests and/or anti-trust issues when some companies have products for the entire research lifecycle (see Chen, Posada & Chan, 2019)? Are researchers and universities both the sources of data and the consumers of these companies? Are there connections between the data/metric-driven research culture and the erosion of public trust in science and experts? And would the purpose of research and scholarship be swayed by commercial interests that exploit our desire to be the first, to be cited, and to be competitive?

In what follows, this paper will articulate the importance of norms and epistemic cultures in knowledge production and how the use of evaluative metrics and the dominence of data analytics can lead to platformisation of knowledge production. It is the objective of this paper to open up critical examination of the practices of the many platforms embedded in the research infrastructure.

#### **Normal Science**

In Thomas Kuhn's *The Structure of Scientific Revolutions* (1996, originally published 1962), the term 'normal science' is used to describe the everyday 'puzzle-solving' activities of scientists, while a revolution or paradigm shift is 'driven by the failure of the existing paradigm to solve certain important anomalies' (Bird, 2018). The scenarios of normal science or paradigm shifts show that *norms* play an important role as to what constitute or counts as scientific knowledge, which methods and methodologies are appropriate for understanding and interpreting data and findings, and when to progress towards a new paradigm. In other words, knowledge production is understood as a normative process whereby scientists interact and negotiate its aims, objectives and methodologies over time and space.

Debunking the understanding of science as 'objective knowledge', Knorr Cetina (1999) uses the term 'epistemic cultures' to reveal 'the fragmentation of contemporary science' and illustrates the many 'different architectures of empirical approaches, specific constructions of the referent, particular ontologies of instruments, and different social machines' (p. 3) in the construction of scientific knowledge. Epistemic cultures can be understood as the norms and practices co-constructed by the scientific communities. In other words, there is not a fixed form of 'science' as it is continuously changing and evolving through the interactions between people, instruments, and social conditions.

Whilst Kuhn (1996), Knorr Cetina (1999) and others have shown that norms and practices are central in the process of knowledge production, the sociologist of knowledge, Robert Merton (1973, originally published 1942), has discussed the ethos of science—universalism, communality, disinterestedness, and organised skepticism. In an ideal situation where everyone follows these norms, scientific research is aimed at producing common and public good, for scientists are expected to share their findings and collaborate with others without holding any personal or self-interests. Scientists are also expected to be reasonably critical and skeptical to uphold the quality of research and achieve the best understanding for humanity and society. The ethos of science represent the aspirations of science and scientists, presuming the conduct of science and scientists are not tainted by the desire for money, fame, or power. Hence, these aspirations can only be upheld by the norms and practices co-created and maintained by the scientific community.

How and why would scientists not act in accordance with the ethos of science? At the time of his writing, Merton (1973, originally published 1968) has discussed some implications of the reward system in science, including issues pertaining to competition and collaboration, as well as the Matthew Effect. For Merton, the reward system can have significant influences on the action and behaviour of scientists when rewards-driven practices run counter to the ethos of science. Nevertheless, the value of science and dynamics of knowledge production are perceived to be largely in the hands of scientists who interact and negotiate the norms of acceptable conduct and practices.

The normative view of science has contributed to the understanding of the process(es) of knowledge production in that knowledge is not simply 'out there' to be found but an interplay between facts and norms. Normal science, epistemic cultures and the ethos of science acknowledge norms and practices in the constitution of scientific knowledge (see also, Latour and Woolgar, 1986; Latour, 1987). Yet, at the time of this writing, the research systems and academia are often described as 'hypercompetitive market' where the market forces are driven by university rankings, allocation of grants, citation-based metrics, and so on. Scientists pursue 'hot topics' because they cannot get funding otherwise, not to mention funded research tend to create more publications and citations that are beneficial for career advancement. Scientific norms and practices are not only co-constructed by scientists, but have also been imposed by governments, commercial publishers and data analytic companies. One can argue that the invention of citation index and uses of evaluative metrics, including citation-based metrics and altmetrics, have introduced the competitiveness of the so-called academic marketplace, entwined in the web of commercial interests of publishers, data brokers, and news companies who own and maintain the research infrastructure.

# Metrics, Data Brokers and Marketing Experts

Evaluative metrics are embedded in a largely commercial research infrastructure—researcher profiles (e.g., Google Scholar and ResearchGate), benchmarking and data analytics tools (e.g., SciVal and Dimensions), data sharing and non-traditional research output platforms (e.g., figshare), whilst Altmetrics traces and tracks number of views, downloads, and mentions on platforms including Twitter, Reddit, Mendeley, blogs, and so forth. Most of these platforms provide data products and sharing features, but few, if at all, of which is involved in the heavy lifting of producing and reviewing the contents (i.e., the creation of research outputs including data and articles). Rather, they provide the platforms and/or data products to generate incomes and profits. As Andrews (2020) argues, 'research-sharing infrastructures and open tools and services are engaging with all these senses of platforms and platformization, with academic social networking services being seen as 'reputational platforms' and mediating both connections between researchers and the sharing of research outputs, processes, and information' (p. 266). They also report that vertical integration and acquisition have been undertaken by the biggest academic publishers and service providers for building 'suites of products based on data sharing and acting as intermediaries between libraries, universities, researchers, and the public—and platforms rarely have open and transparent governance' (p. 266).

As Chen, Posada and Chan (2019) show in Figure 1, Elsevier owns a variety of products throughout the research, publishing, and evaluation processes. They argue that there are potential conflict of interest when the company can 'privilege their own contents as well as researchers that participate within their integrated systems' (para. 26) and that 'users' including universities, lecturers and researchers can grow dependent on big publishers, not to mention the increasingly monopolising power of the ownership of data seemingly required by research evaluation at university and individual levels.



#### The Academic Knowledge Research Production Process

Figure 1 Elsevier Presence Throughout the Research Lifecycle (Chen, Posada & Chen, 2019)

Recently, there are also pay-to-publish platforms that provide services to increase research impact by repackaging academic articles for wider audience. The author of this article has been contacted by *Research Outreach, Research Features*, and *Science Animated*. These publication channels may be a symptom that the academic marketplace has become more competitive in the so-called attention economy, when academic writings become metrics-driven, perpetuated by the data and content-sharing products by a few main players in the publication and scholarly communication landscape.

The proliferation of these platforms, including 'data brokers,' can be attributed to the fact that evaluative metrics have become valuable data for scientists to assess the research quality of articles, journals, individuals (e.g., potential collaborators) and for governmental agencies/departments and universities to monitor, track, and plan research strategies and resource allocation. Evaluative metrics have become 'valuable tokens' (Biagioli, 2018) when they appear on researchers' profiles, grant applications, strategic plans, and so forth notwithstanding their meanings can be ambiguous and fluid. When data become literally valuable, that is, in monetary terms, indexing and data analytic companies develop their business models to produce and sell data packages to governmental departments, universities, news agencies such as Times Higher Education.

Why are metrics accepted as valuable tokens? It is perhaps because evaluative metrics are 'time-saving devices' (Ma, 2021): (a) evaluative metrics perform an important function in information retrieval, that is, to retrieve the most 'relevant' results based on citation counts in databases including Scopus, WoS, and Google Scholar, although these metrics can also confine you in an echo chamber since the lesser-known articles and authors are often pushed toward the end of retrieved results; (b) evaluative metrics are used to represent research quality and impact and hence shorten the time needed for evaluating candidates and grant applications, despite the many contestations about their validity and objectivity in research evaluation, and (c) evaluative metrics induce competitiveness amongst researchers to produce more and faster; the speed and volume are desirable for the progress of science and innovation.

Metrics have also become valuable because the need for accountability, which has often been stated as one of the reasons for collecting, tracking and monitoring research performance. The number of publications and citations are used to demonstrate the value-for-money (or return-for-investment) of public funds, and to provide evidence that research can and will lead to inventions and economic benefits. These platforms create an environment where accountability can seemingly be achieved by providing 'objective' data for monitoring, ranking and comparison. In turn, researchers and universities respond by producing more publications and attracting more citations, which create and provide more data for these platforms—data brokers—to be sold back to the research and universities, as well as university rankings and research assessment exercises.

Notwithstanding the validity and legitimacy of such functions have become questionable due to gaming behaviour, misconduct and manipulation (Biagioli & Lippman, 2020), the importance of evaluative metrics has been inflated and obfuscated by their uses in different settings despite that they cannot and should not be used as standards (Ma, 2020). The more evaluative metrics are perceived as important and essential for accountability, the more significance and power the date brokers hold in the system of knowledge production. The few major players also own a multitude of commercial products throughout the research process that enhances their ability to monitor research activities as well as to collect data that can be packaged and sold.

# Platformisation of Knowledge Production

Wouters (2014) has proposed the notion of the citation as infrastructure to explain how the use of citationbased metrics has become routinised, invisible, and are taken-for-granted. He maintains that infrastructures are 'multi-layered and complex and cannot be constructed top-down. Instead they evolve bottom-up' (p. 61). The idea that infrastructures are constructed and evolved 'bottom-up' assumes the structuration of academic norms and practices *without* interference of commercial interests (money) and/or strategic influences such as national priorities (power). However, the normative view of science and scientific practices can no longer account for citation rings and citation cartels, or misconduct and fraudulent research. How many scientists are taught and trained following the ethos of science? Or are they acculturated to chase after the number of publications and citations instead? If citation as infrastructure is not co-constructed bottom-up but rather designed and developed for maximising profits, then it is pertinent to understand the scholarly communication landscape in terms of platformisation.

There are, of course, many complexities when we try to untangle and understand the changing landscape of knowledge production and scholarly communication. Wars, for example, can have important impacts on what knowledge can be produced and exchanged (Richards, 1984). Similarly, the current evolving geopolitics can intensify secrecy and competition in knowledge domains such as artificial intelligence. Many have also commented on the impacts of the new public management and neoliberalism on knowledge production, in particular the instrumentalist view of research and scholarship in creating 'useful' knowledge for economic progress and growth, whilst changes in research policy including the implementation of Plan S in Europe can impede the production and dissemination of information in the so-called Global South. Nevertheless, the use of metrics and the proliferation of commercial platforms affect academic norms and practices and the relationships amongst scientists, scholars, universities, policymakers, and the public. Increasingly, what constitute or counts as scientific knowledge, which methods and methodologies are appropriate for understanding and interpreting data and findings, and when to progress towards a new paradigm are steered by the use of evaluative metrics in the larger system of knowledge production (Ma, 2020). Today, the consideration of normative science (Kuhn, 1996) and epistemic cultures (Knorr Cetina, 1999) must also take into account the major players of knowledge production, in particular the so-called 'Big Deals' publishers as well as the data analytic companies. In fact, the power bestowed upon the data brokers can be considered as 'a coup from above' as Zuboff (2019) describes the dominance of digital platforms:

'The rise of instrumentarian power is intended as a bloodless coup, of course. Instead of violence directed at our bodies, the instrumentarian third modernity operates more like a taming. Its solution to the increasingly clamorous demands for effective life pivots on the gradual elimination of chaos, uncertainty, conflict, abnormality, and discord in favor of predictability, automatic regularity, transparency, confluence, persuasion, and pacification. We are expected to cede our authority, relax our concerns, quiet our voices, go with the flow, and submit to the technological visionaries whose wealth and power stand as assurance of their superior judgment. It is assumed that we will accede to a future of less personal control and more powerlessness, where new sources of inequality divide and subdue, where some of us are subjects and many are objects, some are stimulus and many are response.'

Research infrastructure, from publishing, indexing to data analytics, is largely owned by commercial entities. Whilst their services are presumably tailored for the needs of researchers and universities, they are also creating the needs for monitoring, ranking and benchmarking data, particularly when the various league tables and university rankings are based on these data packages. Before the invention of the citation index, publishing and indexing platforms were designed and maintained to facilitate the dissemination of knowledge and they were not built for individuals and universities to compete or compare with each other. When indexing services become data brokers, however, evaluative metrics become embedded in the search algorithms and throughout the research lifecycle. However, the built purposes of the research infrastructure are not necessarily to advance knowledge or facilitate information exchange and collaboration, but rather to increase research impacts and to compare one's work with others'. Akin to some social media platforms, these research infrastructures depend on user-generated contents (research articles) and they act merely as 'intermediaries' while prompting researchers to compete for more 'likes' (citations). The data become the authority about the good and bad of research and scholarship, whilst the structuration of academic norms and practices is gradually replaced by the steering effects of citations and metrics.

For many years and decades, libraries and librarians have taken pride in themselves as the gatekeepers of knowledge and information to ensure intellectual freedom in local communities and the society at large. It is their mission to provide all points of view in all forms of knowledge, scientific, societal or aesthetic. They remain largely neutral, or inclusive, whilst being responsive to the needs of their clientele. Yet, their role in the dissemination of scientific and scholarly information has changed. When Garfield (1972) suggests using citation analysis as a tool for journal evaluation, journal subscriptions were not bundled. Over the years, the big deal publishers have gained increasing powers in deciding not only what can be published, but also the information made available in the library catalogues. As the prices of subscription bundle increase and the library budget dwindles, learned societies and small publishers can become 'collateral damages' however making the positions of the big deal publishers even stronger. And the situation has somewhat worsened at the time of this writing, when many libraries are in the process of negotiating of transformative agreements including 'read and publish' and 'publish and read' options in the move to open access. Worse not because of the idea of open access, but because we are putting more powers in the hands of a few companies that can afford to provide the research infrastructure-digital platforms-and these companies have diversified their products in the research, publishing, and evaluation process. As data brokers gain more significance and power in the academic marketplace, the academic discourse that constitutes normal science and epistemic cultures can be hollowed out.

It is clear that the demand for data brokers and their products is largely driven by the dependence on quantitative analysis of research performance. The coverage of different databases (e.g., Scopus, WoS, Google Scholar, Mircosoft Academic) and their algorithms affect what we read and how we judge. The understanding of algorithmic governance in research infrastructure is no less important than in social media platforms. As Carson (2020) maintains,

'Machines, instruments, algorithms, and the like do not eliminate human subjectivity, they simply move it to another place: the person who designed or fabricated the machine, the individual operating the instrument, the specialist deciding which algorithm to use and what to include in the data to be analyzed, or the scientist making sense of the results.'

Discussions about the construction and development of research infrastructure have been limited to a small group of scientists, scholars and librarians for many decades, particularly those who are most enthusiastic about open access. Explicitly and/or implicitly, these discussions aim to uphold the ethos of science and they are the manifestation of the structuration of academic norms and practices. Yet, the dominance of commercial platformisation can decimate the negotiation powers of those who produce and review scientific outputs—data sets, programming codes, articles, and so forth—when researchers are acculturated to chase after funding, metrics, then data-driven economic and societal impacts. The new impact ranking,

for instance, have put another kind of data analysis in the spotlight (see Armitage, Lorenz & Mikki, 2020). If we do successfully turn the metrics tide and answer the calls for responsible use of metrics, there will be competition of impact, openness, collaboration, gender (see CWTS Leiden Ranking) and data and data brokers will continue to play a central role in knowledge production. Competition can be healthy for stimulating thoughts and innovations; commercial profits are vital for investing and developing research infrastructure, or simply starting a journal in an emerging area—that is, as long as data or data brokers do not replace or displace the discourse in constituting the norms and practices in knowledge production.

#### REFERENCES

- Andrews, P. C. S. 2020. "The Platformization of Open," In M. P. Eve (Ed.), *Reassembling Scholarly Communications*, pp. 265-276. <u>https://doi.org/10.7551/mitpress/11885.003.0027</u>
- Armitage, C. S., Lorenz, M., & Mikki, S. 2020. "Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results?" *Quantitative Science Studies*, (1:3), pp. 1092–1108. https://doi.org/10.1162/qss\_a\_00071
- Biagioli, M. 2018. "Quality to impact, text to metadata, publication and evaluation in the age of metrics," *KNOW*, (2:2), pp. 249-275.
- Biagioli, M., & Lippman, A. (Eds.) 2020. *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge, Massachusetts: The MIT Press.
- Bird, A. 2018. Thomas Kuhn. In *Stanford Encyclopedia of Philosophy*. Retrieved February 16, 2021, from <a href="https://plato.stanford.edu/entries/thomas-kuhn/">https://plato.stanford.edu/entries/thomas-kuhn/</a>
- Burrows, R. 2012. Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, (60:2), pp. 355-372.
- Carson, J. 2020. "Quantification Affordances and Limits," *Scholarly Assessment Reports*, (2:1), 8. https://doi.org/10.29024/sar.24
- Chen, G., Posada, A., & Chan, L. 2019. "Vertical Integration in Academic Publishing," in *Connecting the Knowledge Commons From Projects to Sustainable Infrastructure*. OpenEdition Press. <u>https://doi.org/10.4000/books.oep.9068</u>
- Cronin, B. (2005). *The Hand of Science: Academic Writing and Its Rewards*. Lanham, Maryland: Scarecrow Press.
- Fyfe, Aileen, Coate, Kelly, Curry, Stephen, Lawson, Stuart, Moxham, Noah, & Røstvik, Camilla Mørk. 2017, May 25. Untangling Academic Publishing: A history of the relationship between commercial interests, academic prestige and the circulation of research. Zenodo. <u>http://doi.org/10.5281/zenodo.546100</u>
- Garfield, E. 1972. "Citation analysis as a tool in journal evaluation," *Science*, (178), pp. 471-479.
- Giddens, A. 1984. The Constitution of Society: Outline of the Theory of Structuration. Polity.
- Kornberger, M., Pflueger, D., & Mouritsen, J. 2017. "Evaluative infrastructures: Accounting for platform organization," *Accounting, Organizations and Society*, (60), pp. 79-95.
- Knorr Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, Massachusetts: Harvard University Press.
- Kuhn, T. S. 1996. The Structure of Scientific Revolutions. Chicago: The University of Chicago Press.
- Latour, B, & Woolgar, S. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, New Jersey: Princeton University Press.
- Latour, B. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Massachusetts: Harvard University Press.
- Ma, L. 2020. "The steering effects of citations and metrics," *Journal of Documentation*, (77:2), pp. 420-431. DOI 10.1108/JD-06-2020-0093
- Ma, L. 2021. "Metrics as time-saving devices," in F. Vostal (Ed.), *Inquiring into Academic Timescape* (pp. 123-133). Bingley, UK: Emerald.
- Merton, R. K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.
- Richards, P. S. 1984. "Aryan Librarianship": Academic and Research Libraries under Hitler. *The Journal of Library History*, (19:2), pp. 231-258.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B., 2015. *The Metric Tide: Report of the*

*Independent Review of the Role of Metrics in Research Assessment and Management.* London: Sage Publications Ltd. DOI: 10.13140/RG.2.1.4929.1363

- Wouters, P. 2014. "The citation: From Culture to infrastructure," in B. Cronin and C. Sugimoto (Eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 47-66). The MIT Press.
- Zuboff, S. 2019. The Age of Surveillance Capitalism: The Fight for a Human Future at the new Frontier of Power. London: Profile Books.