



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/120144/>

Version: Accepted Version

Book Section:

Wilsdon, J.R. (2017) Responsible Metrics. In: Strike, T., (ed.) Higher Education Strategy and Planning: A Professional Guide. Routledge, Abingdon, pp. 247-253. ISBN: 9781138635265.

This is an Accepted Manuscript of a book chapter published by Routledge in Strike, T. (Ed.). (2018). Higher Education Strategy and Planning. London: Routledge on 21/06/2017, available online: <http://www.routledge.com/> 9781351797078

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

*Pre-proof version of chapter by James Wilsdon in Tony Strike (ed.) (2018)
Higher Education Strategy and Planning: A Professional Guide. Routledge.*

JAMES WILSDON

14. RESPONSIBLE METRICS

In the past decade, there has been an explosion in the range and reach of metrics and league tables to benchmark institutional performance, research qualities and impacts, teaching and learning outcomes. Yet some of the most precious qualities and contributions of higher education resist simple quantification. This chapter explores how to develop measurement and management systems that are both effective and supportive of responsibility, diversity and integrity. Building on *The Metric Tide* review, it also considers recent developments in the UK's approach to metrics for research and teaching, which are the focus of ongoing debate.

BETWEEN BIG DATA AND QUANTOPHRENIA

In his 2003 bestseller 'Moneyball', Michael Lewis describes how the fortunes of the Oakland Athletics baseball team were transformed by the rigorous use of predictive data and modelling to identify and invest in undervalued talent. These approaches soon spread through baseball and into other sports, and are now widely used in the financial sector, recruitment industry and elsewhere, to inform hiring and promotion decisions.

A recent study by researchers at the MIT Sloan School of Management argues that universities are ripe for their own 'Moneyball' moment (Bertsimas et al., 2015; Brynjolfsson and Silberholz, 2016). As the authors note:

Ironically, one of the places where predictive analytics hasn't yet made substantial inroads is in the place of its birth: the halls of academia. Tenure decisions for the scholars of computer science, economics, and statistics – the very pioneers of quantitative metrics and predictive analytics – are often insulated from these tools. (Brynjolfsson and Silberholz, 2016)

By analysing a set of metrics for publications, citations and co-authorship at an early stage in a researcher's career, and including these in hiring and promotion decisions, the MIT team suggests that it is possible to predict future performance with greater accuracy and reliability than through subjective judgements alone.

Given the role that citations, H-indices, journal impact factors, grant income and other conventional metrics already play in research management and decision-making (both explicitly and implicitly), some would no doubt welcome predictive analytics as a logical next step. Applications of 'big data' and 'broad data' within higher education institutions (HEIs) are still at a relatively early stage, given their longer-term possibilities (ICSU, 2016). Over the next decade, it is easy to envisage

JAMES WILSDON

increasingly granular data on research qualities and impacts being combined with more sophisticated metrics for teaching and learning to give HEI managers, planners and policymakers access to an unprecedented wealth of real-time data and analytics on which to base a wide range of choices: from individual hires and course selection, through to funding priorities and faculty structures.

Yet hand-in-hand with the expanding possibilities of metrics, debates have intensified about the pitfalls of an over-reliance on such measures. Some see metrics as one element of a more managerial, audit-driven culture in universities (Graeber, 2015; Collini, 2016, Martin, 2016). More specific concerns tend to focus on three issues. First, a growing ‘quantophrenia’ in higher education: a narrowing of managerial attention onto things that can be measured, at the expense of those that cannot (Sorokin, 1956; Burnett, 2016). Second, a reduction in diversity, as an emphasis on particular metrics or league table performance (itself weighted towards a few key indicators) drives HEIs to adopt similar strategic priorities, and individual researchers to focus on lower-risk, incremental work aimed at higher-impact journals (Hicks et al., 2015). Third, a distortion of incentives, which is in turn exacerbating problems of research quality, integrity and reproducibility (Benedictus and Miedema, 2016; Sarewitz, 2016).

EXPERIMENTS IN RESPONSIBLE METRICS

In response to such concerns, there have been a number of high profile efforts to reform how metrics are used in higher education and research. These include:

- *The 2013 San Francisco Declaration on Research Assessment (DORA)*, which calls for research to be assessed on its own merits and for an end to the use of journal impact factors in funding, hiring and promotion decisions. As of January 2017, DORA has over 800 organisational and 12,500 individual signatories, including a handful of UK universities;
- *The Leiden Manifesto*, which was published in 2015 by a group of leading scientometricians, and sets out ten principles for the use of quantitative indicators in research evaluation (Hicks et al., 2015);
- *Science in Transition*, a movement established in 2013 by researchers in the Netherlands, with the aim of tackling systemic problems in research and university culture, which “has become a self-referential system where quality is measured mostly in bibliometric parameters and where societal relevance is undervalued” (Dijstelbloem et al., 2013);
- *The Meta-Research Innovation Center (METRICS)* at Stanford University, which was launched in 2014 with a focus on transforming research practices and tackling problems of research integrity and reproducibility (Ioannidis et al., 2015).

Each of these initiatives influenced the main UK contribution to these debates: the Independent Review of the Role of Metrics in Research Assessment and Management, which I chaired on behalf of the Higher Education Funding Council for England (HEFCE). After fifteen months of evidence gathering, analysis and

consultation, this published its findings as *The Metric Tide* in July 2015 (Wilsdon et al., 2015).

The main motivation behind *The Metric Tide* was a desire by government to look afresh at whether metrics could play a greater role in the next cycle of the Research Excellence Framework (REF). But the review group interpreted its role more broadly, and made a series of targeted recommendations to university leaders, funders, publishers and researchers designed to ensure that indicators and underlying data infrastructure would support the diverse qualities and impacts of higher education and research. Over and above these detailed points, we proposed a framework for *responsible metrics*, built on five principles:

- *Robustness*: basing metrics on the best possible data in terms of accuracy and scope;
- *Humility*: recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment;
- *Transparency*: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results;
- *Diversity*: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system;
- *Reflexivity*: recognising and anticipating the systemic and potential effects of indicators, and updating them in response.

SURFING ON THE METRIC TIDE

When *The Metric Tide* was published, it provoked a lively debate in the UK and further afield. Despite the spread of opinion encountered over the course of the review, the steering group were encouraged by the degree of consensus in support of our detailed recommendations and the broader idea of responsible metrics.

However, in the UK, these issues were far from settled. Following the May 2015 general election, the government announced ambitious plans to reform the higher education and research system. These were set out in a November 2015 green paper (BIS, 2015), a May 2016 white paper (BIS, 2016) and a Bill, which at the time of writing is still making its passage through Parliament.

Proposed reforms include a reshaping of the regulatory architecture for HEIs and research funding, the replacement of HEFCE with a new Office for Students, and the introduction of a Teaching Excellence Framework (TEF) “to identify and incentivize the highest quality teaching” (BIS, 2015). From the start, the government presented metrics as a non-negotiable element of the TEF, albeit with scope for peer review and expert judgement alongside. This has provoked intense arguments about the use and limitations of particular indicators for teaching and learning, mirroring established debates on the research side of the system. For its

JAMES WILSDON

initial cycle, the TEF is set to proceed using flawed indicators. As one recent commentary describes the situation:

It is now accepted sector wisdom that the Teaching Excellence Framework is neither a measure of teaching nor a measure of excellence. The designers know that and don't want to keep hearing it said (Strike, 2016).

Debate is already shifting to which additional indicators can be introduced into the TEF cycles that follow, and whether these will address the sector's legitimate concerns. Work is ongoing, but the idea of responsible metrics remains a useful starting point from which to consider the place of metrics in the design of an evaluation system.

On the research side, the November 2015 green paper also reopened questions over metrics and the REF, and a comprehensive review of the REF, chaired by Lord Stern, was initiated in December 2015 and issued its findings in July 2016 (Stern, 2016). Despite pressure from some quarters to move to a metrics-based REF, the Stern Review concluded that peer review should remain the primary method of research assessment, supported by responsible uses of data. The approach that Stern outlines – maintaining the primacy of peer review, using carefully-selected metrics in the environment section of the REF, and improving data infrastructure and interoperability – is fully in line with the findings of *The Metric Tide*. A further technical consultation by HEFCE is now looking in detail at the choices and challenges involved in implementing Stern's recommendations, and the framework for REF 2021 will be finalised by the summer of 2017.

UK FORUM FOR RESPONSIBLE RESEARCH METRICS

One of the conclusions of *The Metric Tide*, and of related initiatives like DORA and the Leiden Manifesto, is the need for ongoing effort to shift institutional cultures, practices and incentive frameworks that sustain damaging or irresponsible use of metrics. To this end, a UK Forum for Responsible Research Metrics has been established to bring together research funders, HEIs and their representative bodies, publishers, data providers and others. Chaired by David Price, Vice-Provost of UCL, the Forum will carry forward activities on three fronts:

1. *Supporting the effective leadership, governance and management of research cultures within HEIs.* The Forum will encourage UK HEI leaders to develop a clear statement of principles on how research is managed and assessed, and the role of metrics within these processes. Particular attention will be paid to criteria and indicators used in academic appointments and promotions.
2. *Supporting the responsible use of metrics by key organisations in the funding system.* The Forum will work with funders to develop their own context-specific principles for the use of quantitative indicators in research assessment and management. There will be a particular focus in

RESPONSIBLE METRICS

the first 12-18 months on specific technical issues around indicator definition, selection and use in the environment section of the REF, and on wider guidance to assessment panels for REF 2021.

3. *Improving the data infrastructure that supports research information management.* In light of ongoing reforms to the UK's research system – notably the establishment of UK Research and Innovation (UKRI) as a strategic umbrella body for public funding – there is scope to be ambitious about the design of a “next generation research data infrastructure”, which can ensure greater efficiency and interoperability of data collection, and its intelligent and responsible use to inform HEI strategy, research assessment, funding prioritisation and national policy (e.g. around industrial strategy). Organisations like Jisc are heavily engaged in this agenda, and the Forum will add weight and build support for such efforts.

Outputs from the Forum's deliberations are likely to include good practice guidance on the use of metrics in HEI management and academic recruitment practices, recommendations for publishers on responsible uses of metrics in promotional materials, and recommendations on gaps and opportunities for technical infrastructure development.

MEASURING PROGRESS

The Forum for Responsible Research Metrics is a step in the right direction. But despite all the initiatives outlined here, it too often remains the case that, in the words of Peter Lawrence, poorly designed evaluation criteria are “dominating minds, distorting behaviour and determining careers” (Lawrence, 2007). More work is needed to link debates across teaching and research, to develop better indicators, and to foster more sensitive management frameworks. Alliances can be forged beyond the higher education sector, by linking to wider streams of scholarship and advocacy around algorithmic accountability and the future of the workplace. And UK efforts need to be aligned and joined to parallel work across Europe, in the United States and further afield.

We now have the evidence we need to influence how the metric tide washes through higher education and research. Planners, strategists, managers and information professionals have a crucial role to play – alongside academics – in determining whether we sink or swim.

BIBLIOGRAPHY

- Benedictus, R. and Miedema, F. (2016), Fewer numbers, better science. *Nature*, Vol. 538: 453-454, 27 October 2016
- Bertsimas, D., Brynjolfsson, E., Reichman, S., Silberholz, J. (2015), Tenure Analytics: Models for Predicting Research Impact. *Operations Research* 2015 63:6, 1246-1261

JAMES WILSDON

- BIS (2015), Higher education: teaching excellence, social mobility and student choice. London: BIS, November 2015
- BIS (2016), Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice. London: BIS, May 2016
- Brynjolfsson, E. and Silberholz, J. (2016), 'Moneyball' for Professors? Frontiers blog. *MIT Sloan Management Review*, 14 December 2016
- Burnett, K. (2016), Universities are becoming like mechanical nightingales. *Times Higher Education*, 19 December 2016
- Collini, S. (2016), Who are the spongers now? *London Review of Books*, Vol. 38 No. 2, 21 January 2016
- Dijstelbloem, H., Huisman, F., Miedema, F., Mjinhardt, W. (2013), Why Science Does Not Work as It Should And What To Do about It. Science in Transition Position Paper, October 17, 2013
- Graeber, D. (2015), The Utopia of Rules: On Technology, Stupidity and the Secret Joys of Bureaucracy. London: Melville House
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. (2015), The Leiden Manifesto for Research Metrics. *Nature*, Vol. 520:429-431, 23 April 2015
- ICSU (2016), Open Data in a Big Data World. Paris: ICSU/IAP/ISSC/TWAS
- Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., Goodman S. N. (2015) Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLoS Biol* 13(10): e1002264. doi:10.1371/journal.pbio.1002264
- Lawrence, P.A. (2007), The mismeasurement of science. *Current Biology*, 17 (15): R583-R585.
- Lewis, M. (2003), Moneyball: The Art of Winning an Unfair Game. WW Norton: New York
- Martin, B. (2016), What is Happening to our Universities? SPRU Working Paper Series 2016-03. Brighton: SPRU
- Porter, T. M. (1995), Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. New York: Princeton University Press.
- Sarewitz, D. (2016), Saving Science. *The New Atlantis*, Spring/Summer 2016.
- Sorokin, P. (1956), Fads and foibles in modern sociology and related sciences. Chicago: Regnery
- Stern, N. (2016), Research Excellence Framework Review: Building on success and learning from experience. London: BIS, July 2016.
- Strike, T. (2016) The TEF is a statistical wonderland. *Wonke.com*, 5 December 2016. <http://wonkhe.com/blogs/the-tef-is-a-statistical-wonderland/> (accessed 10 January 2017)
- Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. London: HEFCE, July 2015

AFFILIATIONS

James Wilsdon
University of Sheffield