

OPINION

Let's make science metrics more scientific

Stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane** — how else shall we know what to reward?

Measuring and assessing academic performance is now a fact of scientific life. Decisions ranging from tenure to the ranking and funding of universities depend on metrics. Yet current systems of measurement are inadequate. Widely used metrics, from the newly-fashionable Hirsch-index to the 50-year-old citation index, are of limited use¹. Their well-known flaws include favouring older researchers, capturing few aspects of scientists' jobs and lumping together verified and discredited science. Many funding agencies use these metrics to evaluate institutional performance, compounding the problems². Existing metrics do not capture the full range of activities that support and transmit scientific ideas, which can be as varied as mentoring, blogging or creating industrial prototypes.

The dangers of poor metrics are well known — and science should learn lessons from the experiences of other fields, such as business. The management literature is rich in sad examples of rewards tied to ill-conceived measures, resulting in perverse outcomes. When the Heinz company rewarded employees for divisional earnings increases, for instance, managers played the system by manipulating the timing of shipments and pre-payments³. Similarly, narrow or biased measures of scientific achievement can lead to narrow and biased science.

There is enormous potential to do better: to build a science of science measurement. Global demand for, and interest in, metrics should galvanise stakeholders — national funding agencies, scientific research organizations and publishing houses — to combine forces. They can set an agenda and foster research that establishes sound scientific metrics: grounded in theory, built with high-quality data and developed by a community with strong incentives to use them.

Scientists are often reticent to see themselves or their institutions labelled, categorized or ranked. Although happy to tag specimens as one species or another, many researchers do not like to see themselves as specimens under a microscope — they feel that their work is too complex to be evaluated in such simplistic terms. Some argue that science is unpredictable, and that any metric used to prioritize research money risks missing out on an

important discovery from left-field. It is true that good metrics are difficult to develop, but this is not a reason to abandon them. Rather it should be a spur to basing their development in sound science. If we do not press harder for better metrics, we risk making poor funding decisions or sidelining good scientists.

Clean data

Metrics are data driven, so developing a reliable, joined-up infrastructure is a necessary first step. Today, important, but fragmented, efforts such as the Thomson Reuters Web of Knowledge and the US National Bureau of Economic Research Patent Database have been created to track scientific outcomes such as publications, citations and patents. These are all useful, but they are labour intensive and rely on transient funding, some are proprietary and non-transparent, and many cannot talk to each other through compatible software. We need a concerted international effort to combine, augment and institutionalize these databases within a cohesive infrastructure.

The Brazilian experience with the Lattes Database (<http://lattes.cnpq.br/english>) is a powerful example of good practice. This provides high quality data on about 1.6 million researchers and about 4,000 institutions. Brazil's national funding agency recognized in the late 1990s that it needed a new approach to assessing the credentials of researchers. First, they developed a 'virtual community' of federal agencies and researchers to design and develop the Lattes infrastructure. Second, they created appropriate incentives for researchers and academic institutions to use it: the data are referred to by the federal agency when making funding decisions, and by universities in deciding tenure and promotion.

Third, they established a unique researcher identification system to ensure that people with similar names are credited correctly. The result is one of the cleanest researcher databases in existence.

On an international level, the issue of a unique researcher identification system is one that needs urgent attention. There are various efforts under way in the open source and publishing communities to create unique researcher identifiers using the same principles as the Digital Object

SUMMARY

- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

Identifier (DOI) protocol, which has become the international standard for identifying unique documents. The ORCID (Open Researcher and Contributor ID) project, for example, was launched in December 2009 by parties including Thomson Reuters and Nature Publishing Group. The engagement of international funding agencies would help to push this movement towards an international standard.

Similarly, if all funding agencies used a universal template for reporting scientific achievements, it could improve data quality and reduce the burden on investigators. In January 2010, the Research Business Models Subcommittee of the US National Science and Technology Council recommended the Research Performance Progress Report (RPPR) to standardize the reporting of research progress. Before this, each US science agency required different reports, which burdened principal investigators and rendered a national overview of science investments impossible. The RPPR guidance helps by clearly defining what agencies see as research achievements, asking researchers to list everything from publications produced to websites created and workshops delivered. The standardized approach greatly simplifies such data collection in the United States. An international template may be the logical next step.

Importantly, data collected for use in metrics must be open to the scientific community, so that metric calculations can be reproduced. This also allows the data to be efficiently repurposed. One example is the STAR METRICS (Science and Technology in America's Reinvestment — Measuring the Effects of Research on Innovation, Competitiveness and Science) project, led by the National Institutes of Health and the National Science Foundation under the auspices of the White House Office of Science and Technology Policy. This project aims to match data from institutional administrative records with those on outcomes such

"If we do not press harder for better metrics, we risk making poor funding decisions or sidelining good scientists."

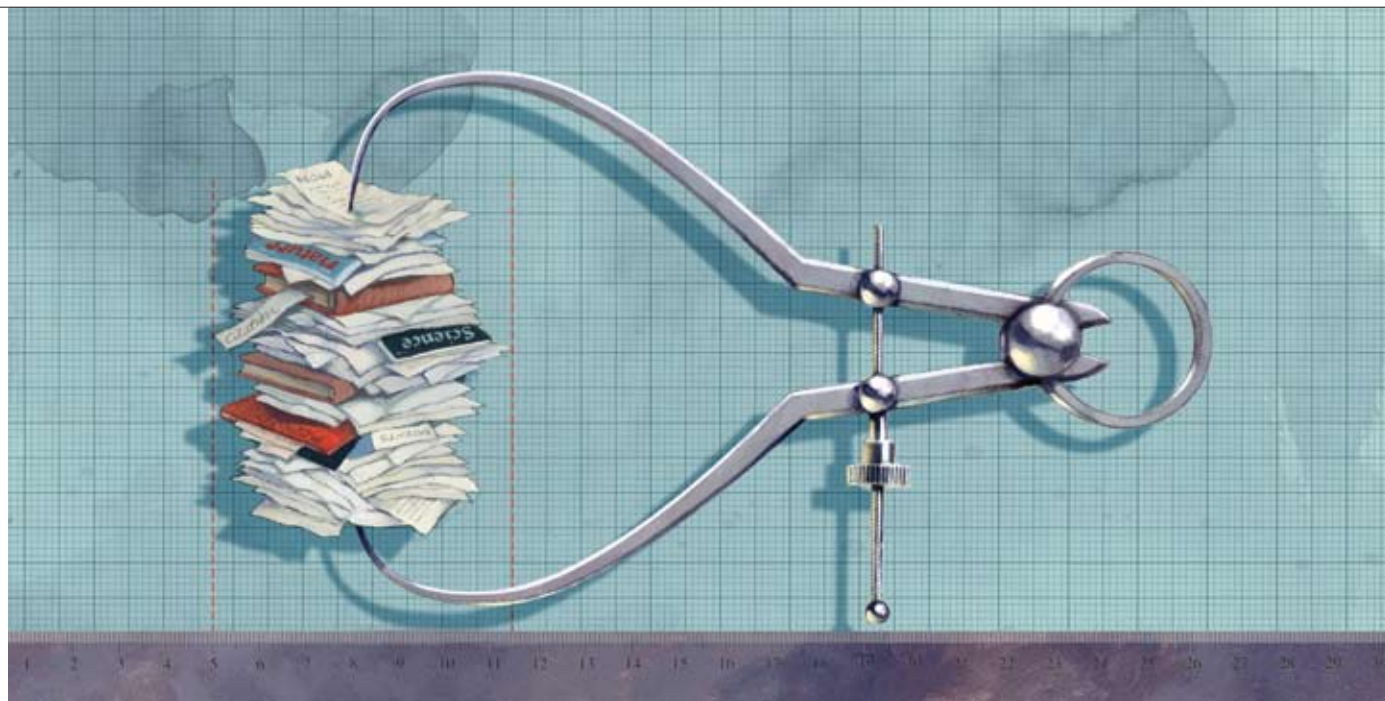


ILLUSTRATION BY DAVID PARKINS

as patents, publications and citations, to compile accomplishments achieved by federally-funded investigators. A pilot project completed at six universities last year showed that this automation could substantially cut investigators' time on such tasks.

Funding agencies currently invest in fragmented bibliometrics projects that often duplicate the work of proprietary data sets. A concerted international strategy is needed to develop business models that both facilitate broader researcher access to the data produced by publishing houses, and compensate those publishers for the costs associated with collecting and documenting citation data.

Getting creative

As well as building an open and consistent data infrastructure, there is the added challenge of deciding what data to collect and how to use them. This is not trivial. Knowledge creation is a complex process, so perhaps alternative measures of creativity and productivity should be included in scientific metrics, such as the filing of patents, the creation of prototypes⁴ and even the production of YouTube videos. Many of these are more up-to-date measures of activity than citations. Knowledge transmission differs from field to field: physicists more commonly use preprint servers; computer scientists rely on working papers; others favour conference talks or books. Perhaps publications in these different media should be weighted differently in different fields.

People are starting to think about collecting alternative kinds of data. Systems such

as MESUR (Metrics from Scholarly Usage of Resources, www.mesur.org), a project funded by the Andrew W. Mellon Foundation and the National Science Foundation, record details such as how often articles are being searched and queried, and how long readers spend on them. New tools are available to capture and analyse 'messy' data on human interactions — for example, visual analytics intended to discover patterns, trends, and relationships between terrorist groups are now being applied to scientific groups (<http://nvac.pnl.gov/agenda.stm>).

There needs to be a greater focus on what these data mean, and how they can be best interpreted. This requires the input of social scientists, rather than just those more traditionally involved in data capture, such as computer scientists. Basic research is also needed into how measurement can change behaviour, to avoid the problems that Heinz and others have experienced with well-intended metrics that lead to undesirable outcomes. If metrics are to be used to best effect in funding and promotion decisions, economic theory is needed to examine how changes to incentives alter the way research is performed⁵.

How can we best bring all this theory and practice together? An international data platform supported by funding agencies could include a virtual 'collaboratory', where ideas and potential solutions can be posited and discussed. This would bring social scientists together with working natural scientists to develop metrics and test their validity through wikis, blogs and discussion groups, thus building a community

of practice. Such a discussion should be open to all ideas and theories and not restricted to traditional bibliometric approaches.

Some fifty years after the first quantitative attempts at citation indexing, it should be feasible to create more reliable, more transparent and more flexible metrics of scientific performance. The foundations have been laid. Most national funding agencies are supporting research in science measurement, vast amounts of new data are available on scientific interactions thanks to the Internet, and a community of people invested in the scientific development of metrics is emerging. Farsighted action can ensure that metrics goes beyond identifying 'star' researchers, nations or ideas, to capturing the essence of what it means to be a good scientist. ■

Julia Lane is the director of the Science of Science & Innovation Policy programme, National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA.
e-mail: jlane@nsf.gov

1. Campbell, P. *Ethics Sci. Environ. Polit.* **8**, 5–7 (2008).
2. Curtis, B. *Globalis. Soc. Edu.* **6**, 179–194 (2008).
3. Kerr, S. *Acad. Manage. J.* **18**, 769–783 (1975).
4. Thrash, T. M., Maruskin, L. A., Cassidy, S. E., Fryer, J. W., & Ryan, R. M. *J. Pers. Soc. Psychol.* (in the press).
5. Leydesdorff, L. in *Beyond Universal Pragmatics: Studies in the Philosophy of Communication* (ed. Grant, C. B.) 149–174 (Peter Lang, 2010).

The opinions expressed are those of the author and may not reflect the policies of the National Science Foundation.

Comment on this subject and view further reading online at go.nature.com/AXZ7d1.