Science as big business

Wim Blockmans*[1], Lars Engwall†[2] and Denis Weaire‡[3]
*Institute for History, Leiden University, The Netherlands, †Department of Business Studies, Uppsala University, Sweden, and ‡Institute of Physics, Trinity College Dublin, Ireland

The expansion of the system

Over the last decades, new forms of evaluation and the intensity of their application have gradually taken such proportions that the whole character of the scholarly world is now infected by a kind of scientometric mania. Modern information technologies, the strong expansion in the number of researchers and their number of publications have generated new requirements for the management of evaluations and new tools to accomplish it [1]. Moreover, easier transportation facilities have considerably increased scholars' mobility on a global scale, which has accelerated and intensified their contacts. The budgets involved in the increasingly expensive material and personnel of laboratories have required closer scrutiny of the results of the huge investments, especially where most of them were made with public money, as prevails in Europe. This is indeed demonstrated by Nicola Gulley, who in Chapter 8 quotes an average annual growth of 3% in researchers and articles since the late 1940s in the STM (science, technology and medicine) fields. The accelerating tempo of research output is further illustrated by the 200-fold increase from 1985 to 2012 in the number of papers submitted to *Angewandte Chemie*, mainly due to the massive contribution of non-German researchers. The evaluation process, including desk assessment, thousands of referee reports and a redress procedure, leads to a very high refusal rate, which can be seen as the price that needs to be paid for top quality. More generally, Gulley quotes a 44% increase in the published papers in her field over the years 2000–2010.

Garfield's innovation

The huge expansion of the research area had implications for the organization of the critical assessment of the results, as well as for the overall management of the profession and the significant financial flows it absorbs. Responding to the first challenge, a radically new approach was presented in 1955 by Dr Eugene K. Garfield in his seminal article 'Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas' [2]. He claimed that "the Citation Index bypasses some of the limitations of classical subject indexing" ([3], p. 650), and, according to Yancey [4], he envisioned information tools that allow

[1]Email: wimblockmans7@gmail.com
[2]Email: lars.engwall@fek.uu.se
[3]Email: dweaire@tcd.ie

researchers to expedite their research process, evaluate the impact of their work, spot scientific trends and trace the history of modern scientific thoughts. Garfield realized that, since scientific indexes were discipline-oriented, researchers weren't finding all of the information relevant to their work. Researching a scientific area solely by its subject or keywords limited findings by ignoring relevant papers from other disciplines. By indexing scholarly work by citation, he allowed researchers to track what other works a paper has referenced, and how many times others have cited a paper. And by counting citations, the 'impact factor' could be measured, *assigning an indicator of quality* [Authors' italics] to more influential works.

Garfield founded the Institute for Scientific Information in Philadelphia which, at the moment when it was acquired by the Thomson Corporation in 1992, employed 500 people. It collected cited references found in 8700 of the world's leading scholarly science and technical journals covering more than 100 disciplines and going back to 1900. However, he stated himself that "a citation index must meet the same economic test that all products in our society must meet: does the cost justify the benefits?" [3]. Although his method demonstrated the central role played by a few journals in the natural sciences, he acknowledged that the fields of the humanities and most of the social sciences operate in a different way, which made its application to these domains economically not feasible. After all, he worked with a commercial company.

Although initially Garfield had in mind to facilitate bibliographic searches beyond disciplinary boundaries, increasingly his citation index was used primarily as a measure of the impact of an article, a researcher, a group, an institute, a journal, a discipline or even the research performance of a whole country. And increasingly, the simple and assumedly objective figure was assigned the role of a proxy for *an indicator of quality*, which would facilitate all kinds of evaluations. As Giuseppe Longo and several other authors in this volume argue, numbers replaced value and, to paraphrase Richard R. Ernst [5], reading was replaced by counting. Popularity of a topic within a time span of just a few years is not automatically an indication of innovative and significant research. Or, as Jane Grimson formulates it in her chapter:

> "*traditional bibliometrics* [...] *define what constitutes research quality rather than providing objective measures of research quality.* [...] [They] *measure, however imperfectly, only some of the dimensions of research quality, and there is no substitute for a detailed study of the researcher's output.*"

Garfield himself saw the limitations of his metrics and tried to address them by ever more complicated calculations. Thus he checked the citation frequency of Nobel Prize winners, but it remained too labour-intensive to isolate the references to publications *before* the award for all the laureates. He observed that a "high level of citation will fail to detect, for example, publications from small, specialty fields", also in the domains of physics, chemistry, and physiology and medicine ([6], p. 183). He defined a 'Citation Classic' as a paper receiving more than 300 citations, but admitted that "only the rarest paper is cited more than 300 times within a few years of publication". And, even if "most citations to a paper [in these three scientific domains] are received in the first decade after publication", the phenomenon of delayed recognition occurs. He discussed, for example, the case of A.M. Cormack who was awarded the Nobel Prize for Medicine in 1979 for articles published in the specialized *Journal of Applied Physics* in 1963 and 1964. These had been virtually ignored until G.N. Hounsfield recognized in them the mathematical foundations for the CAT scanner and scored more than 800 citations with his paper. Cormack's invention had been obliterated by incorporation in another paper. And Garfield concluded: "Only peer judgement can determine which of these authors is uniquely qualified for the Nobel or some other illustrious award. Indeed, citation frequency by itself is not adequately indicative of outstanding and influential publications" ([6], p. 186).

The impact of Garfield

The citation index was nevertheless welcomed by administrators and scholars as a timesaving tool for all kinds of evaluations, which had become an increasingly heavy burden for senior scholars. Applicants to the European Union FP7 (Seventh Framework Programme) have been encouraged to mention the Thomson Reuters impact factor of their publications, and many research councils and universities apply this instrument in their selection procedures. In Chapters 5 and 9, Michel Gevers and Jan Reedijk respectively produce clear examples of the ways by which some commercial publishers seek to steer authors towards reinforcement of the 'impact' of their own journals.

As recounted by Milena Žic Fuchs in Chapter 11, a reaction came from the American Society for Cell Biology, which formulated on the 16 December 2012 DORA, a Declaration on Research Assessment. It has been subscribed by hundreds of scientific institutions and organizations, including the Academia Europaea. Its first recommendation reads as follows [7]:

> "*Do not use journal-based metrics, such as journal impact factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion or funding decisions.*"

This very welcome warning concerns just one of the various indicators which have been applied without the required prudence for some time. Self-evidently, candidates will adapt their publishing behaviour and their applications in function of the prevailing criteria.

By now, a variety of indicators have been developed, but all of them have their limitations, as most authors in the present volume demonstrate. Anthony van Raan, one of the leading initiators of bibliometrics in Europe, whose calculations have been based on the Web of Science and Scopus databases, acknowledges in Chapter 3 that bibliometric performance analyses have to be used as information in combination with peer review. His institute has been developing ever more sophisticated methods such as mapping on the basis of keywords (see his Figure 3 and the book cover), and new indicators, the latest being called the 'crown indicator'. But at the same time, he lists no fewer than 11 weaknesses of the existing tools.

Jane Grimson enumerates the biases by disciplines, dealing with interdisciplinarity, language, region, format of publication, and non-attributable multi-authorship. The trickiest shortcoming remains the short time span of the counting of citations, which makes it very unlikely that truly groundbreaking insights will be fully recognized in the citation indices. "Truly original work takes more than 2 years to be appreciated" [8]. Some book publications, including critical editions of sources in the humanities, remain standard references for a century and more [9].

Problems outside the 'hard' sciences

Despite such objections, these apparently objective and simple instruments were especially convenient for science administrators, who tended to prescribe and use them indiscriminately for all disciplines. By doing so, they may have systematically favoured those disciplines for which the bibliometrics had been designed, and harmed others. Best off were disciplines of which the object is global or even universal, so that all findings are equally relevant for specialists throughout the whole world. Taking the advantage of economies of scale, such disciplines have organized themselves globally, which made the production of widely disseminated journals attractive to international publishers. In turn, it became rewarding and prestigious to publish in such journals, to be invited as a reviewer and to serve as a member of the editorial board. That attraction made it even possible for science journals to require a fee for the publication of an article, which they sell to providers, libraries and users, whereas they did not contribute to the research costs. This self-reinforcing mechanism led to standardization of publications and the establishing of a ranking within large disciplines. North American and British journals became globally dominant. In the wake of this development, we have obtained institutional rankings, which, as pointed out by Michel Gevers, Giovanni Abramo and Linda Wedlin, have a number of shortcomings.

However, as demonstrated in Chapter 13 by Christine Musselin and her co-authors, for a variety of reasons the science model does not fit all disciplines. This is further shown by Milena Žic Fuchs, who discusses data produced by Gunnar Sivertsen demonstrating that Scopus covers hardly more than 10% of the humanities publications and just one-third of those in the social sciences. For the Web of Science, the coverage reaches 16% and 22% respectively [10]. Similar results were obtained at an evaluation at Uppsala University in 2007, when the publications of university researchers were matched with Web of Science. It turned out that Web of Science covered 79.2% of the publications in the area of medicine and pharmacy, 57.0% in the area of science and technology and 8.3% in the area of the humanities and social sciences ([11], pp. 90–91).

Mathematicians, computer scientists, and technical and engineering scientists joined the large majority of the humanities and social scientists in stressing the adequacy of the particular publication methods typical of their disciplines, which do not relate well to the citations system. They address technical, societal and cultural problems for which the most appropriate audience is not necessarily to be found in a global forum and for which the most effective expression is not uniquely a short journal article in a highly formalized type of English, co-signed by numerous authors. Even more devastating is the observation that a blind eye is turned by all too many authors to the vast body of knowledge produced over time in various national languages. In culturally specific domains, this can lead to poorly informed publications in highly reputed international media.

As a consequence, it comes as no surprise that in a number of small European countries, beginning with Norway and followed by Denmark, Finland and Belgium (Flanders), governments have taken initiatives to construct databases of national publications which meet the international standards of peer-reviewed publication. Typically, they include books, collective volumes and articles in the national language. These systems are described and compared in Chapter 12 by Frederik Verleysen, Pol Ghesquière and Tim Engels who demonstrate that different procedures lead to considerable variation in the results. As with the European Reference Index for the humanities discussed by Milena Žic Fuchs, the inclusive databases offer the advantage of covering all the media of peer-reviewed scholarly publications, but lack fine tuning with regard to the status and circulation that journals and publishing houses really have, both within national borders and internationally. Such precise weighting of components is indeed important, since many of the databases have a direct financial impact. For instance, as the Flemish database is applied, it accounts for 6.28% of the budget distributed by the government to universities. It has been revised over the last years, and even now one may wonder if the proportion of 4 to 1 applied to the weight of a monograph (for example, a published PhD thesis) compared with a journal article, is fair or still detrimental to the book-publishing disciplines.

Problems in the assessments

All of these endeavours, contested by many serious researchers, are costly, time-consuming and bureaucratic. For instance, the 2008 Research Assessment Exercise is estimated to have cost the U.K. government approximately £60 million; £12 million to perform it and an additional £47 million for the universities to prepare the submissions [12]. The outcomes of this have been widely publicized and discussed publicly. Major newspapers published rankings and commented on shifts in comparison with the 2001 exercise. All of this enhanced a spirit of competition at all levels of the higher education system, which was not unequivocally healthy. Staff members are encouraged to aim for a high output of publications; other duties may be less valued and consequently neglected. Top journals have to keep their subscribers, so they have to cultivate reputations as guarantors of high-quality research. It is acknowledged, as also pointed out by Lars Engwall, that reviewers tend to give conservative assessments, which discourages risk-taking proposals. A recent analysis addressing three types of assessment, the subjective reviews after publication, the journal impact factor and the citations, observed the absence of correlation between the assessors. Experts showed to be prejudiced by the journal impact factor. But also the impact factor is ‘extremely error prone’ and the citations ‘inherently highly stochastic’ [13,14].

Even if errors cannot be entirely excluded by peer review, that procedure offers far better guarantees than the mushrooming online open-access journals which have far less transparent structures and procedures than established printed journals. Online journals are not bound by limits in page space and are eager to publish in return for a fee. *Science* published in its 4 October 2013 special issue on 'Communication in Science: Pressures and Predators' the results of a fake research article submitted to 305 online open-access journals. It was rejected by 98, but accepted by 157 of them, mainly based in India, in return for a fee of $3100. The sheer volume of online publications these days makes it harder to distinguish between respectable and questionable journals [15].

Besides the perceived facilitation of review procedures, bibliometrics has been espoused by science managers seeking primarily quantitative, allegedly objective instruments for the distribution of resources and the control of the performance. In the context of the considerable growth of public spending for research, it is certainly justified to hold scientists accountable. However, as the progress of science is inherently not straightforward or predictable, the research community feels itself uncomfortable with over-simplified, 'auditable' measurements of immediate outcomes, especially as they impose a substantial extra workload and are considered to be speciously precise. The researchers' real target being discovery and innovation, they feel that the heavy burden of research evaluations is expensive and inadequate, leading to compliant attitudes, risk avoidance, cynicism and even unethical behaviour (see Chapter 10 and [16,17]). In general, bibliometric devices do not provide proper means to account for the societal mission that science also has to fulfil. The U.K. Research Evaluation, re-christened the REF (Research Excellence Framework), will include this function as a criterion, weighted by 20% in 2014 [18]. Which implies that all previous exercises have been biased in this respect?

> "*The REF will for the first time explicitly assess the impact of research beyond academia, as well as assessing the academic excellence of research. One purpose of the REF is to reward research departments in universities that engage with business, the public sector and civil society organizations, and carry new ideas through to beneficial outcomes* […] *Impact is defined as: any effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia.*"

New discussions will arise, and they have effectively started already, about the methods to be used to assess "the impact of research beyond academia" [19]. This appears from the conclusions of a prospective study published in November 2013 by the Royal Netherlands Academy of Sciences to advise the government about the 'public investment in knowledge and the value of research' ([20], pp. 11,13).

> "*It is essential to understand that the effect of research policy can only be usefully evaluated in the long term; that is where the effect is felt.*
>
> *There are plenty of empirical studies available that link public and private expenditure on R&D to GDP growth at the macro-economic level. These provide sufficient evidence that the economic value of research can be*

measured and analysed. It is understandable that the elasticities found in these studies vary; this is, to some extent, the result of spill-overs, serendipity, and variations in policy and in the broader innovation system.

Quantitative studies exploring the value of research do not make allowance for much of the broader economic and social value of research because that value is difficult to quantify, especially in the short term. The value of research is not limited to its present or future contribution to GDP, and only part of that value lies in its productivity effects. It is much more difficult to measure the overall economic and social value of research."

Of course, in calling into question recent developments, we have been dealing in generalizations. We should admit that there are honourable exceptions. In particular, the European Research Council, in its relentless pursuit of *excellence*, clearly a subjective concept, has so far resisted putting bibliometrics in the forefront of its evaluations. Rather, it relies primarily on the judgment of expert panels. For the Starting and Consolidators' Grants, panels meet face-to-face with the short-listed candidates, and then debate their merits across the table. It is a time-consuming process that some would call old-fashioned, but in a few years it has won admiration and respect.

Recommendations

To conclude, we venture to formulate some general recommendations which result from the papers in this volume.

1. Rich multidimensional assessment tools are needed to recognize and value the different contributions made by individuals, regardless of their discipline.
2. Evaluations, especially those on the individual level, should not be based on metrics only. They should be used only in combination with other factors, and taking into account their limitations.
3. It should be recognized that different factors are important to different people and depend on the research context, which calls into question any reliance on rigid weightings.
4. Criteria need to be adequate to the field, its size, standards and societal role.
5. Qualitative evaluations by experts, informed by quantitative indicators, remain the best possible method, addressing issues of scope, originality and value.
6. The journal impact factor should be dismissed, as is proposed in DORA.
7. In the case of multi-author papers, the contributions of individual authors are not distinguishable under current bibliometric systems; for the purposes of evaluation this is a very serious deficiency.
8. Metric indicators should be used for evaluation only to the extent that their inherent biases are well understood by those applying them.

9. Expert reviewers should be adequately recognized and rewarded.
10. The academic community has been overtaken by events: it should make greater efforts to anticipate and influence future developments.

References

1. Drori, G.S., Meyer, J.W., Ramirez, F.O. and Schofer, E. (2003) *Science in the Modern World Polity: Institutionalization and Globalization*. Stanford University Press, Palo Alto
2. Garfield, E. (1955) Citation indexes for science: a new dimension in documentation through association of ideas. *Science* **122**, 108–111. Also available at: http://garfield.library.upenn.edu/papers/science_v122v3159p108y1955.html
3. Garfield, E. (1964) Science citation index. A new dimension in indexing. *Science* **144**, 649–654 [Reprinted in (1984) *Essays of an Information Scientist* **7**, pp. 525–535, quote on 526]
4. Yancey, R. (2005) Fifty years of citation indexing and analysis. http://wokinfo.com/essays/50-years-citation-indexing/ (Accessed 21 November 2013)
5. Ernst, R.R. (2010) The follies of citation indices and academic ranking lists. A brief commentary to 'Bibliometrics as weapons of mass citation'. *CHIMIA* **64**, 90
6. Garfield, E. (1986) Do Nobel prize winners write citation classics? *Current Comments* **23**, 3–8 (Reprinted in *Essays of an Information Scientist* **9**, 182–187)
7. San Francisco Declaration on Research Assessment. http://www.ascb.org/SFdeclaration.html
8. Lawrence, P.A. (2007) The mismeasurement of science. *Current Biology* **17**, R583–R585
9. Nederhof, A.J. (2006) Bibliometric monitoring of research performance in the social sciences and humanities: a review. *Scientometrics* **66**, 81–100
10. Sivertsen, G. (2011) Achieving complete and comparable data and output indicators for scholarly publishing in the humanities and social sciences. http://bibliometrie.univie.ac.at/fileadmin/user_upload/ub_bibliometrie/sivertsen.pdf (Accessed 22 November 2013)
11. KoF07. Quality and Renewal 2007. An Overall Evaluation of Research at Uppsala University, Uppsala. Uppsala University
12. RAE manager (2009) RAE Manager's report. http://www.rae.ac.uk/pubs/2009/manager/manager.pdf; PA Consulting Group (2008) RAE 2008 Accountability Review. http://www.hefce.ac.uk/media/hefce/content/pubs/2009/rd0809/rd08_09.pdf
13. Eyre-Walker, A. and Stoletzki, N. (2013) The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biology* **11**, e1001675
14. Eisen, J.A., MacCallum, C.J. and Neylon, C. (2013) Expert failure: re-evaluating research assessment. *PLoS Biology* **11**, e1001677
15. Bohannon, J. (2013) Who's afraid of peer review? In *Science*. 4 October 2013, http://www.sciencemag.org/content/342/6154/60.full
16. Power, M. (1997) *The Audit Society: Rituals of Verification*. Oxford University Press, Oxford
17. Sheikh, A. (2000) Publication ethics and the research assessment exercise: reflections on the troubled question of authorship. *Journal of Medical Ethics* **26**, 422–426
18. http://www.ref.ac.uk/panels/assessmentcriteriaandleveldefinitions/; http://www.ref.ac.uk/media/ref/content/researchusers/REF%20guide.pdf
19. Magrs, P. (2013) Re: your REF impact request. In The Times Education. 25 April 2013, http://www.timeshighereducation.co.uk/features/re-your-ref-impact-request/2003342.article; http://blogs.lse.ac.uk/impactofsocialsciences/2013/05/07/absence-of-impact-used-to-be-the-fashionable-thing-to-claim/
20. Public investment in knowledge and the value of research. https://www.knaw.nl/nl/actueel/publicaties/publieke-kennisinvesteringen-en-de-waarde-van-wetenschap