The Future of Indicators for Research Assessment and Open Science. Doing away with quantitative metrics.[1]

Dr.Dr.phil Rene von Schomberg[2]

Senior Research Fellow-RWTH Aachen University

We live in an evaluation culture. Both individuals and institutions are continuously evaluated on their performance: childrens' performance are already measured, at an early age, often well before they are able to express themselves properly and one can have the impression that teenagers at secondary schools spent more time in 'learning' for and doing all kinds of tests and evaluations, than the time they spent on actually learning and discovering new things and capabilities. This evaluation culture is also well established in the realm of science: involving institutions such as universities and individual researchers alike.

In the case of scientific researchers, evaluations are based on their productivity and quality of work. The productivity of researchers is conventionally measured by bibliometric indicators such as the number of publications whereas the quality of their work is captured by citation metrics and the venue of their publications. One can criticize the use of metrics by pointing out that the 'amount of citations' may indicate that the publication raised a particular interest, but this interest was constituted by the fact that the topic discussed in the paper just happened to be 'sexy' in the field of study or that the article was simply provocatively 'wrong' and 'required' an appropriate denial or correction of colleagues or, worse, simply caused an outrage. Equally, Einstein's original paper on the special theory of relativity received relatively little citations for such a monumental paper. The metrics employed 'measures' a particular interest raised in a particular scientific community rather than its quality. The same argument holds for 'productivity' metrics. The number of papers a researcher produces does not necessarily relate to the relative contribution the researcher has brought to the field of his or her specialisation.

The 'quality' of scientific journals is 'measured' by the contentious 'journal impact factor.' In all these cases we do not measure what we want to know, e.g., the quality, originality, and the relative contribution of the work to the body of science, e.g., its 'excellence.' We actually employ quantifiable factors as a proxy either for quality, excellence or productivity. To make it worse: at the same time, there are no consensual definitions for quality or excellence. In fact, we do not even have a definition of 'science', let alone 'good' science[3]. We leave it up

---

[1] Appers in German in: Julia Mörtel, Alfred Nordmann, & Oliver Schlaudt (eds) *Indikatoren in Entscheidungsprozessen.* Wiesbaden: Springer VS (forthcoming).

[2] The article was written in my capacity of Senior Research Fellow at RWTH Aachen university and that therefore thankfully *Sponsored by the German Federal Ministry of Education and Research (BMBF), through the Käte Hamburger Kolleg: Cultures of Research at the RWTH Aachen*"

[3] This situation leads to 'arbritary' reasons to receive governmental funding for research. A citizen, even a citizen with a Ph.D, cannot receive funding for a research proposal from national funding agencies. The assumption is that they do not do science with reference to the fact that they dont work at a research

to the peers of scientific community to evaluate what is understood by quality, excellence or good science, and these peers may well refer back to available metrics while doing their evaluation work. This makes the evaluation circular and self-referential, which is in itself defended by the scientific community by the fact that universities should be 'autonomous' and 'independent' in their (self) evaluations while vigorously rejecting any external or societal evaluation of their scientific work as this would otherwise endanger their (undefined) excellence of their work.

What is more, the metrics has a focus on researchers and their institutions (for example, university rankings) but not on the system a such: how does our collective system of knowledge production perform? Does the scientific system deliver sufficiently socially desirable outputs? We do not evaluate the system of science and instead seem to engage researchers and institutions in an irrational beauty contest among researchers and institutions: Do I or does our institutions do better than others? At the same time, we do not shy away to make the system subject of gamification: in order to maximise our publications, we break original papers up in several separate ones and in order to maximise citations, we incentivise colleagues to cite our work.

In the field of scientometrics, there is recently a reluctant acknowledgement that there is in fact no perfect metrics (Wilsdon, 2017). All quantifiable metrics can become subject of gamification, and one seems to agree that in order to arrive at a judgement of 'quality', a qualitative judgement is indispensable. Peer review is then still the main candidate to produce that qualitative judgement. But the field of scientometrics falls short of asking the question what type of science we actually want to foster. Do we want a science that primarily delivers new scientific insight? Or do we want a science that delivers on socially desirable outcomes? Or do want to have science that can engage in a variety of missions? Also research funders, those who actually would have an interest in digging into to this question have been caught for long in 'serving' the scientific community in providing the funds for the type of research they call for themselves. Unfortunately, we needed a global public health crisis, e.g., Covid-19, to make a start with necessary adjustments in the way, we organise, reward and fund and how we incentivise researchers to do the 'right' thing.

The event of Covid-19 has changed the modus operandi of science: the shift towards open research and scholarship

Open research and scholarship gradually evolved from two global trends: open access to research outputs and open source. The former refers to online, peer-reviewed scholarly outputs, which are free to read, with limited or no copyright and licensing restrictions, while open source refers to software co-created without any proprietary restriction, and which can be accessed and used.

The Budapest Open Access Initiative of 2002 established open access for the first time as an internationally desirable publishing practice. Although open access became primarily associated with a particular publishing or scientific dissemination practice, the Budapest Open Access Initiative already sought to induce a broader practice of open scholarship that includes the general reuse of many types of research products, not just

institution. A notable exception is the European Framework Programme for Research which can fund any 'legal' entity.

publications or data. However, it is only more recently that various forms of open scholarship have evolved to the concept of a transformed scientific practice, shifting the focus of researchers' activity from publishing as fast as possible to sharing knowledge as early as possible.

Michael Nielsen's (2013) book, *Reinventing Discovery: The New Era of Networked Science*, is arguably the first and most comprehensive coverage of Open Scholarship accessible to a broad readership. Michael Nielsen advocated open science as 'the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process'.

The ecosystem underpinning open scholarship is evolving very rapidly. Social network platforms for researchers, such as Research Gate or Academia already attract many millions of users and are being used to begin and validate more research projects in a brain-sourced way. Open scholarship is extended to open forms of using code, scientific discovery and analysis, open research assessment and review, as well as outreach.

Open scholarship is predominantly the result of a bottom-up process driven by a growing number of researchers, who increasingly employ social media and a variety of digital means for their research to initiate globally coordinated research projects and share results at an early stage in the research process, for example, through electronic notebooks that enable real-time global sharing of raw research data among collaborating partners and facilitate access to results for potential collaborators.

An early and well-known example of open scholarship from the pre-Internet stage is the Human Genome Project that started in 1990. The data on the human genome was widely shared among the scientific community in the course of the project, while a moratorium on publishing was kept in order to encourage optimal collaboration. Owing to this openness, they were able to decode the human genome in less than 15 years. Open scholarship has shifted the prime focus of researchers away from publishing towards knowledge sharing.

The ongoing changes are progressively transforming scientific practices, and innovative tools to facilitate communication, collaboration and data analysis are appearing. Researchers increasingly work together to create knowledge. Online tools create a shared space where creative conversation can be scaled up. As a result, the problem-solving process can be faster and the range of problems that can be solved can be expanded (Nielsen 2013).

Fuelled by the new opportunities for knowledge and data sharing, more open practices have emerged to address pressing issues at an early stage. For example, five months into the largest Ebola outbreak in history, an international group of researchers sequenced three viral genomes, sampled from patients in Guinea. The data was made public the same month. This open scientific practice was extended into making experimental vaccine available within a brief period and proved vital in combating relatively smaller outbreaks in 2018. The National Institutes of Health in the USA now require grantees to make large-scale genomic data public by the time of publication at the latest. The World Health Organization (WHO 2015) seeks a paradigm shift in the approach to information sharing in public health emergencies, from being limited by embargoes set for publication timelines, to open sharing using modern fit-for-purpose pre-publication platforms. Researchers, journals, and funders will need to engage fully for this paradigm shift to occur. The WHO acknowledged that patents on natural genome sequences could be inhibitory for further research and product development and wants research entities to exercise discretion in patenting and licensing genome-related inventions in order not to inhibit product development and to ensure appropriate benefit sharing. The organization also wants scientific publishers not to penalize, but to encourage or mandate public sharing of relevant data. Zikka was the next major emerging public health issue that was tackled with effective initiatives based on open scholarship.

The way in which the WHO dealt with these public health issues and the associated commendable initiatives of researchers who felt a moral obligation to engage in open scholarship practices with a view to the urgency of the matter, deserve to become a model for scientific practice. However, this practice is currently more an exception than the rule.

Despite this history of emerging public health issues, we had to wait until Covid-19, which was a public health crisis also hitting the 'rich' nations of the world, and only by then, the shift to open science became more than just a moral initiative of individual researchers to share data at the cost of giving up on original publications. Now it became fostered by public authorities, including the European Commission which made Open Science a policy priority since 2015. (for an overview of Open Science policies, see *Burgelman et al*, 2019).We have witnessed a necessary change in the modus operandi of doing science: Open research and scholarship by which researchers share data and knowledge with all relevant knowledge actors as early as possible in the research process[4] made it possible to deliver swiftly on vaccines. Without open science, the market introduction of these vaccines would have taken, under the usual circumstances of competitive research and intellectual property right constraints, minimally a decade.

The cases of public health emergencies provide an inspiring model for how global research collaborations can help address the societal challenges of our times. Such cases should not be an exception but the norm. However, to make open science the norm, as the dramatic cases of Ebola Zika and Covid-19 illustrate, open science policies that relate to the core of the work of researchers need to be implemented, addressing the necessary change of the rewards and incentive system for researchers. This implies as the case of emerging public health emergencies illustrate, that the importance of publishing in major scientific journals will be relativized in the context of a full operational open science, and other research outputs such as open data, open software, and so forth will become more important. Open research outputs will be available prior to publication rather than post-publication. Table 1 and Table 2 summarises the difference in, respectively, practising closed or open science and rewarding the two different modi of doing science.

Table 1 *Practising Closed Science versus Practising Open Science*

Practising Closed Science	***Practising Open Science***
Establishing individual prestige	Establishing a network, communicate and collaborate

[4] I defined open research and scholarship as ' sharing knowledge and data as early as possible with all knowledge actors' (von Schomberg, 2019, page 25). Open research and scholarship has thus two dimensions: openness to knowledge resources and openness towards collaboration with other knowledge actors.

Focus/motivation on scientific frontiers, new and sexy results	Focus- Motivation on societal impact: contribute to societal challenges and socially desirable outputs (SDGs etc.)
Endurance and good luck: 7 percent of Nature submissions are published	Share and publish early in open access or 'Facebooks' for scientists (Zenedo, Researchgate etc.)
Citations: please have patience!	Swifter and higher citation rates
	Open science practice of open data sharing increasingly a necessity in science: data-intensive science (see HGP)

Table 2. *Rewarding the current system of Science vs Rewarding Open Research and Scholarship*

<table>
<tr><th colspan="2">Rewarding the current system of science</th><th colspan="2">Rewarding Open Research and Scholarship</th></tr>
<tr><td colspan="2">Rewarding individual competing scientists - gaining scientific prestige</td><td colspan="2">Rewarding collaboration and sharing to achieve societal impact (e.g., Covid-19)</td></tr>
<tr><td colspan="2">Publish as much and as fast as possible: (publish or perish!)</td><td colspan="2">Share knowledge/data as early as possible in open collaboration: collaborate or have no impact!</td></tr>
<tr><td colspan="2">Excellence as a self-referential criterion</td><td colspan="2">Relative contribution to research missions with a focus on a societal challenge: collaborate with open research agenda's or have no socially desirable impact!</td></tr>
<tr><td>Incentivises researchers to produce specific outputs (mainly publications)</td><td>Use of quantitative metrics to 'measure' quality and productivity</td><td>Incentivises researchers to conduct particular research behaviour: share knowledge/data, collaborate, transnational, transdisciplinary, with all knowledge actors</td><td>Relative contribution to a research mission-qualitative assessment of research behaviour</td></tr>
</table>

Towards a new metrics or no metrics at all?

The normative question which should precede (but actually hardly does) the employment of any kind of metrics is the issue of what type of science we want to foster and what type of outputs of science we actually find rewardable. I have made the case that 'open research and scholarship is actually the type of science we should want to foster if we want to make the 'system' of science productive. The individual researcher should only be assessed in his or her relative contribution to a scientific mission. When construction workers wish to build a house, the workers will not be rewarded in accordance with a 'metrics' such as the numbers of nails they hit into woodwork, but to the way they contributed to the 'mission' of completing the house in accordance with particular quality standards and the quality of individual assignments and the effectiveness of their collaboration. Why would do it differently for rewarding scientist while completing a research mission? Excellent science is also seldom a matter of individual intellectual superiority rewarded by a competitive funding system, but increasingly more a matter of excellent collaboration among a large number of scientists. The article in which empirical evidence was revealed for the existence of gravitational waves conjectured one hundred years ago by Einstein was written by one thousand authors.(for which a Nobel Prize was awarded)[5] The five-hundred-and-second co-author of that article was probably less involved in the overall design and organization of the research than the first few authors, but essential for the completion of the research. Research is increasingly networked and collaborative by necessity, as it is for construction workers to complete a complex building. Researchers who had a 'minor' assignment in one mission may have a more major contribution in another, yet both missions are relevant for the 'productivity' of the system of science, and we need both minor and major assignments in order to deliver on socially desirable outputs[6].

Doing closed science for Covid-19 would have produced some 'high quality' papers in high impact journals several years after Covid's emergence. We would have engaged only the scientific community to the extent they would have received funding under competitive conditions. This would have resulted in good scientific work by a few over an extended period of time and in the meantime, we would have lost some human lives…..Open Science save human lives! The lack of collaboration and the tense competition of researchers may make researchers 'productive' in terms of producing papers but makes the system as such unproductive. The employment of metrics which supports a closed science endangers the outcome of socially desirable outputs such as the production of a vaccine. Open Science not just makes the system of science more productive, but also more responsive to societal demands. The makes the issue of creating an 'altmetrics' which would also capture the 'impact' of the work of scientists superfluous and inappropriate. An altmetrics 'measuring' the impact of the work of individual scientists would equally support a system of science which is 'unproductive' in terms of the 'right' impact of science, in the same way as bibliometrics currently supports a science which is unproductive in delivering 'quality' science. The normative question which precedes the issue of 'measuring' impact is ofcourse the direction we wish to give to research (and innovation)[7]. Open Science addressing Covid 19, combines the 'quality and productivity' of the system of science (producing relevant new

---

[5] It is ironical that the prestigious Nobel Prize, continues to assume the ' brilliance' of an individual scientist, whereas virtual all recent Prizes where awarded to individuals who explicitly acknowledged that their work would not have been possible without substantial collaboration with other scientist in their field. More often these prizes simply have to be awarded to more than individual.

[6] Very often, research on very relevant socially desirable outputs are not considered 'sexy' enough to be published in scientific journals and land up in the vast growing 'grey' scientific literature (for example, research on improving sanitary systems in developing countries: very relevant, but not a candidate for publishing in a 'high impact' journal)

[7] To incentivise researchers to 'drive research and innovation towards socially desirable objectives' is the field of Responsible Research and Innovation, see extensively: Von Schomberg (2019)

insights and scientific discovery) with the effectiveness of delivering on a socially desirable output (e.g. a vaccine). Any use of metrics would have been inappropriate.

We thus not only do not need any metrics, but it would also always constitute a barrier for making the system of science productive in terms of producing scientific insight as in terms of delivering on socially desirable outcomes. But how then to evaluate 'open science?

We still need to incentivise researchers to do the 'right thing', and the incentives are at the level of incentivising the appropriate research behaviour (e.g., early data and knowledge sharing, and collaborating on the 'right' issues to deliver socially desirable outputs). Some may suggest that this again entails measuring and counting for example, how many open data sets a researcher produces or in how many open collaborations the researcher has been involved in. Obviously, this would fail in view of the fact that the sciences are truly diverse, and the 'amount' of data sharing has a different status in the field of 'history' than in the field of the life sciences. It has been suggested nonetheless to employ distinct indicators for various scientific endeavours (Wouters, 2020). But this type of thinking will not consider, the necessity to evaluate researchers to the extent they contributed to a 'research mission,' be it a frontier research mission (such as the one on gravitational waves) or an applied research mission (on a covid 19 vaccine). This is essentially a qualitative judgement. Such judgements could be arrived at in the framework of conventional peer reviews but also in extended peer reviews, with relevant stakeholders when the 'impact' of research missions have to be evaluated. Hence 'measuring,' e.g., the employment of a metrics, is never an option for whatever science. Possibly we could think of indicators, rather than a metrics, for the extent researchers engage in open science practices and thereby contribute to the system of science as such. For this purpose, the 'number' of collaborative networks and the amount of data sets that are early shared, are candidates. They could help us evaluate the global system of science in terms of their engagement with open science. However, the evaluation of individual researchers should always remain of qualitative nature. There are no sensible metrics for that exercise.

References

Burgelman J-C, Pascu C, Szkuta K, Von Schomberg R, Karalopoulos A, Repanas K and Schouppe M (2019), Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. Front. *Big Data* 2:43.

Nielsen, Michael (2013), *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press

Von Schomberg, R (2019), 'Why responsible innovation' in: R. Von Schomberg and J. Hankins (eds) *International Handbook on Responsible Innovation: A Global Resource.* Cheltenham: Edward Elgar, pp. 12–32.

World Health Organization (WHO) (2015). Developing Global Norms for Sharing Data and Results During Public Health Emergencies. Statement arising from a WHO Consultation held on 1–2 September 2015. Available online at: https:// www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/ en/

Wilsdon, J (2017), *Next-generation metrics, responsible metrics and evaluation for open science*, Publication office of the European Union

Wouters, P (2019) *Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship*, Publication office of the European Union

Further Reading

European Commission high-level advisory group report (2020) “Progress on open science: Towards a shared research knowledge system. Final report of the Open Science Policy Platform”. https://doi.org/10.2777/00139

UNESCO recommendation on open science: https://en.unesco.org/science-sustainable-future/open-science/recommendation

Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I (2015) “Bibliometrics: The Leiden Manifesto for research metrics.” Nature 520, 429 https://doi.org/10.1038/520429a

Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN (2018) “Assessing scientists for hiring, promotion, and tenure”. PLoS Biol 16(3): e2004089. https://doi.org/10.1371/journal.pbio.2004089

Wilsdon J et al. (2015) “The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management” https://doi.org/10.13140/RG.2.1.4929.1363