

## **Conference “Beyond impact factor, h-Index and university rankings: Evaluate science in more meaningful ways”, Bern, 21 November 2018**

### **Background**

The Swiss Academy of Sciences (SCNAT) organised this international conference on behalf of the Swiss Academies of Arts and Sciences, as part of its “We Scientists Shape Science” initiative. Renowned experts highlighted the limitations of metrics in capturing scientific quality and the resulting pressure on the quality of scientific output and presented approaches that challenge conventional metrics. The implications for the Swiss science landscape long-term were the subject of a stakeholder roundtable and discussion with the audience.

Due to unforeseen circumstances, Gero Federkeil was unable to contribute, and Marcel Tanner was replaced by Michael Stauffacher (Board Member of the Swiss Academy of Humanities and Social Sciences, Chair at TdLab of ETH Zurich).

### **Summary**

**Ellen Hazelkorn** Two drivers account for the quantification of science: globalised knowledge and a changed “social contract”. In global science, national pre-eminence no longer suffices, instruments compare quality and performance beyond national boundaries. Rankings of scientific institutions has become inevitable, and has resulted in global prestige becoming a dominant driver of policy decisions, rather than quality, societal impact, greater equity or diversity. As a result, the priorities set by commercial rankings effectively undermine institutional autonomy and national sovereignty. The science-society contract has changed because more people are involved in science today (a shift “from the elite to the mass university”), with implications on what kind of research is funded, how it is measured and by whom. Renewing the science-society contract requires making it explicit by agreeing on the goals to be achieved by science, and on the most appropriate and meaningful indicators that can capture its contribution to the public. The scientific community must be involved in this debate through its leaders.

**Stephen Curry** h-Index and Journal Impact Factor are unhealthy for research and researchers. Neither reflects the relevance of scientific work adequately, both are pseudo-objective and do not correlate with the opinion of scientists on which publications in their field are important. An overreliance on the two figures has several negative effects: reduced productivity, positive bias in the literature, impact on reliability and public trust, devalues other important activities and stress on the individual. The h-Index needs to be balanced with human judgement and complemented with alternative skill factors. Two good examples were provided: i) the University Medical Centre Utrecht assesses scientists on the basis of their research, publications and grants, but also on their performance on managerial/academic duties, mentoring and teaching, possible clinical work, as well as on entrepreneurship and community outreach; ii) the Charité University Hospital Berlin also considers the contribution to open science. Changing metric systems is not simple and happens slowly. This effort requires concerted action across the globe, which is what the San Francisco Declaration on Research Assessment (DORA) intends to foster.

### **Discussion**

- More transparency and a wider discussion on business models of learned societies is needed because of large differences in the income they generate from publishing scientific journals.
- Incentives to do scientific research can be a mixture of intrinsic scientific curiosity and extrinsic factors, for example societal relevance.
- Open Access journals, e.g. PLOS ONE, can lead to conflicts of interest and pressure for publication.

**Martin Reinhart** presented the results from an interview study in Switzerland in the life sciences: More funding and more evaluation does not produce better science; scientists are well-informed and opinionated on issues of research evaluation and see the sole criterion of publications as inadequate metrics; current evaluation systems limit diversity, because they produce what they measure, therefore calling for more diverse and qualitative evaluation criteria.

**Stephanie Meirmans** The main conclusion from their comparative study of the Netherlands and Switzerland was that funding streams do not directly affect research integrity. Researchers in Switzerland are still in a relatively comfortable situation, because of high level funding rates, relatively little administration, flexibility in project plans, available funding for fundamental research and institutional core funding. It is nonetheless important to give more consideration to what “good” science is and to foster the framework in which it can thrive.

**Michael Ochsner** Current metrics measure scientific quality for maybe less than 50%. Their validity is limited because an indicator measures only what it purports to measure, but not the quality of the science itself. Metrics by themselves are thus not responsible, so the discussion should focus on how they can be used responsibly. Furthermore, the societal impact of the science is not to be equalled with scientific quality. Ochsner proposed 19 criteria related to the quality and performance of research, for the measurement of which indicators may be developed, even though not all criteria can actually be measured. Specifying these indicators requires bottom-up procedures as well as respect for disciplinary differences and different levels of assessment.

**James Wilsdon** As DORA and Leiden Manifesto advocate, responsible metrics need to include peer review. Science institutions and research funders are called upon to sign DORA and to adapt their internal evaluation policies accordingly. Evaluation frameworks need to be expanded by developing a better/richer set of criteria and indicators to be used in hiring, promotion and assessment. Pressure should be brought to bear on the private sector providers of metric and ranking systems to improve their assessment frameworks. Investment should be made into “meta research”, i.e. “research on research”.

**Sarah de Rijcke** In the area of tension of scientific excellence versus societal relevance, science evaluations with the right criteria can serve as an important bridge between policy and research by delivering science to society. The indicators for that purpose should be co-designed by both their producers and users, whilst applying different methodologies to different disciplines. The portfolio approach to assess research with a suite of broader and more advanced indicators was proposed: expertise (scholarly, educational, knowledge transfer, communication, organisational), output (publications, teaching, datasets, public media, web/social media, software/tools, infrastructure, grant proposals) and influence (on science, on teaching, on society, on economy).

## **Discussion**

- University rankings are relatively new, so they can in principle be adapted by broadening the range of indicators. Alternatively, and because ranking systems are a business, they will collapse if all universities were to opt out. Seeking meaningful and constructive dialogue on this with university presidents is important.
- The portfolio approach to assess science entails the challenge of appropriately weighing clear indicators and diffuse alternative indicators. To do justice to both, the evaluation context and goal are important, as well as the relevance of the research in question.

## **Panel with Matthias Egger, Andrea Schenker-Wicki, Nicole Schaad, Michael Stauffacher**

- Apart from exerting pressure on the indexing business, changing science assessments means bringing about a change in research culture and the attitudes of scientists. This process has begun and needs to be taken further, e.g. by the science academies, supported by good examples of universities that have changed their practice or developed useful tools and practices.
- Research organisations need professors with different skill sets yet equal levels of recognition.
- The scientific research system loses some of its best talents to the private sector, because they do not want to be part of this “bad” system.
- The publication record still forms the basis when universities hire staff, neglecting skills relevant for working with students (coaching, mentoring, etc.) and other “portfolio” criteria (outreach, teaching, management, vision) that vary according to the discipline.

- Funding agencies should broaden evaluation factors in grant selection processes.
- Recognising the limitations of peer review in attempting to distinguish very good from excellent science, an alternative evaluation approach could be to fund the top category on selection basis, and choose the remaining very good projects at random.

*Roger Pfister, Swiss Academies of Arts and Sciences & Swiss Academy of Sciences, 30 November 2018*