

Commentary

A New Evaluation Culture Is Inevitable

Jack B. Spaapen

Royal Netherlands Academy of Arts and Sciences, PO Box 19121, NL-1000 GC Amsterdam, The Netherlands;
E-Mail: jack.spaapen@knav.nl

Published: 23 April 2015

Comment on:

Wolf BM, Häring AM, Heß J. Strategies towards Evaluation beyond Scientific Impact. Pathways not only for Agricultural Research. *Organic Farming*. 2015;1(1)3–18. DOI: 10.12924/of2015.01010003. Available from: <http://www.librelloph.com/organicfarming/article/view/of-1.1.3>.

Changes in the production of research (more collaborative, more inter- and transdisciplinary, more oriented towards societal demand) are influencing the ways in which research is evaluated. Traditional methods of evaluation primarily focussing on the production of scientific articles have long since given way to more comprehensive methods in which researchers' other activities are assessed too. Beyond these developments, evaluation also involves research endeavours concerning collaboration with other stakeholders in society, such as industry, NGO's, consumer groups, or governmental organisations.

However, this transformation does not happen without difficulties because there is no broad consensus about how to evaluate research in a more comprehensive way. When reviewing research in the broader perspective of its merits for societal questions, there are at least two kind of questions that arise. The first one is whether we should emulate the kind of indicators used in the evaluation of scientific quality or develop a different kind of methodological approach, for example a more qualitative one. The second type of question is whether we will be able to find data that is robust enough to perform the evaluation in responsible and justifiable ways. Both questions are important for the policy support necessary to develop reliable and acceptable evaluation procedures.

However, perhaps more important is the overarching

question of the function of evaluation itself in the newly emerging context. Is it an instrument primarily used for purposes of accountability, or is it an instrument for mutual learning and improving the research effort? Improving in this context does not mean striving for a higher position in one of the international rankings, but being more effective in reaching the scientific and societal goals intended. Further, to make this question even more demanding, societal goals are not undisputed; on the contrary, these goals are often the subject of fierce debates between, for example, policy makers and NGO's, or industry and consumer organisations. Agricultural research is therefore an excellent example, because it shows that it is not a matter of simply finding indicators for applied agricultural research, but that research in this sector is connected to much broader discussions (and controversies) in society about how to produce food in a sustainable way.

It is therefore both timely and necessary that Wolf et al. [1] take a closer look at the strategies necessary to change the mindset of those who are responsible in universities and at other levels of the scientific system for the development of alternative evaluation systems. To focus on the encouragement of connections between parts of the research and innovation system that already have a stake in the transition of science for its own sake to science for society may lead to innovative new networks in which the broader perspective is taken seriously. The debates referred to above will be part of such new networks, and of the development of different evaluation systems. As Wolf et al. show, there are several promising developments in this respect. Unfortunately, as is also made clear, these are still confronted by incentive systems that favour the old style of evaluation and the old method of producing research: mono-disciplinary, with a focus on publication in international journals.

This means that there is a major problem in building a new evaluation culture that is more fitting to the new ar-

rangements, in which much scientific research currently takes place. A key problem is the continued gap between the advanced understanding of this changing relationship between science and society—as developed by scholars of science and technology—and the policy context. Many policy makers and universities' governing boards still tend to rely heavily on more traditional ranking systems that are relatively easy to work with and work well in the institution's marketing strategy.

Therefore, the possible strategies mentioned by Wolf et al.—valuable as they are—should be extended to include a strategy to change the basic attitude of decision makers. They too should understand that a broader approach is both necessary and useful. This is especially valid for the research efforts addressed in this article, inter- and trans-disciplinary research that is the product of collaboration between different fields and expertise coming from science and society. If it is indeed the case that there is joint agenda setting, and co-production, it only makes sense to alter the evaluation process in a direction that does justice to these new arrangements and the kind of questions that are relevant in that context.

To a large extent, this is a question of ownership. In traditional academic research, there was one main funder, the government. Under such circumstances, evaluation becomes primarily an instrument for accountability.

The main questions then were whether tax payers' money was spent in a responsible way, and whether the government/researchers were doing the best they could (were they as good as possible?). However, when other funders and stakeholders become part of the equation, the prime goal of evaluation shifts from accountability to communication between partners—regarding goals and research design—and to mutual learning. In this situation, the ownership of the evaluation shifts from one principal funder to a joint responsibility shared among the most relevant stakeholders. Through a joint effort of these stakeholders one might be able to convince policy makers to allow for broader, more comprehensive methods of evaluation. It would help, then, if the availability and accessibility of data was at the best possible level, i.e. as robust and representative as possible for the activities and interactions that take place in the network. Among other things, this would mean that peer review has to be extended to reviews based on broader expertise.

References

- [1] Wolf BM, Häring AM, Heß J. Strategies towards Evaluation beyond Scientific Impact. Pathways not only for Agricultural Research. *Organic Farming*. 2015;1(1):3–18.