



ARTICLE

Received 31 Aug 2016 | Accepted 3 Jul 2017 | Published 15 Aug 2017

DOI: 10.1057/palcomms.2017.78

OPEN

# Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective

Gunnar Sivertsen<sup>1</sup>

**ABSTRACT** Inspired by *The Metric Tide* report (2015) on the role of metrics in research assessment and management, and Lord Nicholas Stern's report *Building on Success and Learning from Experience* (2016), which deals with criticisms of REF2014 and gives advice for a redesign of REF2021, this article discusses the possible implications for other countries. It also contributes to the discussion of the future of the REF by taking an international perspective. The article offers a framework for understanding differences in the motivations and designs of performance-based research funding systems (PRFS) across countries. It also shows that a basis for mutual learning among countries is more needed than a formulation of best practice, thereby both contributing to and correcting the international outlook in *The Metric Tide* report and its supplementary Literature Review.

<sup>1</sup> Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway Correspondence: (e-mail: [gunnar.sivertsen@nifu.no](mailto:gunnar.sivertsen@nifu.no))

## Introduction

In seven major research assessment exercises, beginning in 1986 and concluding with the 2014 Research Excellence Framework (REF), the UK has used the *peer review of individuals and their outputs* to determine institutional funding. Many other countries have followed suit and introduced performance-based research funding systems (PRFS) for their universities. Most of these countries have instead chosen *indicators of institutional performances* as the method instead of peer review of individual performances.

The two alternatives, indicators of institutional performance versus peer review of individual performances, are discussed in *The Metric Tide* report (Wilsdon *et al.*, 2015), an independent review on the use of metrics in research evaluation. The review convincingly concludes that within the REF, it is currently not feasible to assess research quality using quantitative indicators alone. Peer review is needed. The review also warns that the use of indicators may lead to strategic behaviour and gaming. One of the main recommendations is that:

Metrics should support, not supplant, expert judgement. Peer review is not perfect, but it is the least worst form of academic governance we have, and should remain the primary basis for assessing research papers, proposals and individuals, and for national assessment exercises like the REF.

This recommendation could be interpreted as a formulation of best practice also for other countries, particularly since it is aligned with the first of the ten principles of the *Leiden Manifesto for Research Metrics* (Hicks *et al.*, 2015): “Quantitative evaluation should support qualitative, expert assessment.” The implication would then be that most other countries ought to change their PRFS. The trend, however, seems to go in another direction. The adoption of the UK model in Italy in 2003 has led to a semi-metric solution that differs considerably from the REF (Geuna and Piolatto, 2016). Sweden recently designed a UK-inspired model, but decided not to implement it, thereby keeping its indicator-based model, as we shall see below. It seems that PRFS need to be examined in their national contexts to understand their motivations and design.

An alternative to discussing *best practice* is to provide the basis for *mutual learning* among countries. This can be done by gathering information about the different motivations, designs and effects of PRFS from different countries. The *Metric Tide* report contributes to this purpose by including an international outlook with descriptions of PRFS in selected countries. It also has a supplementary report (Wouters *et al.*, 2015) with an extensive review of the literature on the use and effects of bibliometrics in research evaluation. Here, the conclusion is that the “literature does not currently support the idea of replacing peer review by bibliometrics in the REF”. However, as I will demonstrate in this article, I do not find a basis for this conclusion in the report’s treatment of the probably best documented indicator-based PRFS in the international literature, the so-called “Norwegian model”, which is applied in Denmark, Finland and Norway, partly also in Belgium (Flanders). The report with the literature review seems to be biased by the need to support the main recommendation on the REF in *The Metric Tide* report.

The same main recommendation, to continue with panel evaluation and peer review in the REF, has later been firmly supported by lord Nicholas Stern’s independent review *Building on Success and Learning from Experience* (Stern, 2016), which deals with criticisms of REF2014 and gives advice for a redesign of REF2021. The Stern review presents, however, several

suggestions for a redesign of the exercise that, in my reading, point in the direction of organizational-level rather than individual-level evaluations. It also expresses five other purposes than institutional funding for performing research evaluation. These other purposes also point in the direction of organizational-level evaluation. Hence, one might see the Stern review as opening for a “third alternative” beyond the question of peer review versus metrics. If so, one main difference between evaluation-based and indicator-based PRFS becomes smaller. Both types will be focusing on the performance of organizations, not on individual performances. Another and perhaps even more important difference might emerge instead. Organizational level evaluations have the potential of being formative. They can yield advice and contribute to strategic development. Indicators can only look backwards.

This article aims to contribute to the basis for mutual learning between countries regarding motivations, designs and effects of national PRFS. It starts by presenting a framework for understanding possible differences between countries. Such differences can be related to the relative weight given to the two main purposes of a PRFS, funding allocation and research assessment. I then continue with a critical discussion of the treatment of the Norwegian model in the *Metric Tide* report and its literature review. I demonstrate that the effects of a PRFS depends not only on its type, but also on its design. The last part of the article discusses the possible implications of the Stern report. With its further development, the REF might become very interesting from the point of view of mutual learning among countries.

## A framework for understanding country differences in the design of PRFS

Most European countries have introduced performance based research funding systems (PRFS) for institutional funding. An increasing trend is evident when comparing three overviews of the situation at different times (Geuna and Martin, 2003; Hicks, 2012; Jonkers and Zacharewicz, 2016). Countries can be divided in four categories regarding their use of bibliometrics in PRFS:

- A. The purpose of funding allocation is combined with the purpose of research evaluation. The evaluation is organized at intervals of several years and based on expert panels applying peer review. Bibliometrics may be used to inform the panels. Examples of countries in this category are: Italy, Lithuania, Portugal and United Kingdom.
- B. The funding allocation is based on a set of indicators that represent research activities. Bibliometrics is part of the set of indicators. The indicators are used annually and directly in the funding formula. Examples of countries in this category are: Croatia, the Czech Republic, Poland and Sweden.
- C. As in category B, but the set of indicators represent several aspects of the universities’ main aims and activities, not only research. Bibliometrics is part of the set of indicators. Examples of countries in this category are: Flanders (Belgium), Denmark, Estonia, Finland, Norway and Slovakia.
- D. As in category C, but bibliometrics is not part of the set of indicators. Examples of countries in this category are Austria and the Netherlands.

Regarding the use of bibliometrics, a main distinction can be made between informing peer review with bibliometrics (category A) and direct use of bibliometrics in the formula (categories B and C). As discussed in the introduction, the first alternative is recommended by *The Metric Tide* report (Wilsdon *et al.*, 2015). The latter alternative is more widespread. The choice may depend

on which of the two main purposes of a PRFS is given most attention in the national context.

The two main purposes of a PRFS, research evaluation and funding allocation, can be difficult to distinguish. Hicks (2012) defines PRFS as related to both purposes; they are “national systems of research output evaluation used to distribute research funding to universities”. Within a theoretical framework for regarding the two purposes as *inseparable*, “metrics (mis)uses” are no more than possible (mal)practices in the general domain of research evaluation systems (de Rijcke *et al.*, 2016). I will return below to a discussion of this theoretical perspective. Here, we simply observe that in different national contexts, one of the two purposes can be more relevant than the other for understanding the design of the PRFS. One example from each of the four categories presented above will demonstrate this.

United Kingdom has the best known PRFS in category A. It started in 1986 with peer review of individual performances as the chosen method for funding allocation, which initially was the main purpose of the exercises. Growing constraints on public funding and the prevailing political ideology resulted in policies aimed at greater accountability and selectivity (Geuna and Martin, 2003). Gradually, the method has become the more important purpose. The national research assessment exercise is now inextricably bound up with UK research culture and policy. This is the specific national context in which both The Metric Tide report and the Stern review, as discussed in the introduction, convincingly conclude that peer review is needed. The reason is that the PRFS is viewed as a research assessment system more than as a funding allocation mechanism.

Moving down to category B, we find Sweden as an interesting example of a country that has decided between the two main alternatives. Sweden recently designed a UK-inspired model for resource allocation based on expert panels, FOKUS (Swedish Research Council, 2015), but decided not to implement it, mostly for reasons of cost, but also because the universities are concerned about their institutional autonomy and want to organize research evaluations themselves (Swedish Government, 2016). Sweden consequently continues with its solution since 2009: A small part of the resource allocation for research is based indicators of external funding and of productivity and citation impact within Web of Science. The understanding in Sweden is now that the purpose of research evaluation must be achieved by other means than the PRFS. The emerging alternative is that each university runs a research assessment exercise by itself and with the help of international panels of experts. As an example, Uppsala University is presently running a research evaluation named “Quality and Renewal” where the overall purpose is to “analyse preconditions and processes for high-quality research and its strategic renewal”.<sup>1</sup>

Norway in category C not only has a PRFS. It also has a UK-inspired research assessment exercise (Geuna and Martin, 2003). It is not used for funding allocation. The purpose is to provide recommendations on how to increase the quality and efficiency of research. Norway’s PRFS, on the other hand, is designed for other purposes that also typically may motivate these systems (Jonkers and Zacharewicz, 2016): increased transparency of the criteria for funding, enhancing the element of competition in the public funding system, and the need for accountability coupled to increased institutional autonomy. Norway is typical of category C by using performance indicators that also represent educational activities. Hence, the idea of replacing the indicators with panels performing research evaluation has never been discussed. Research evaluation takes place in a different procedure and resource allocation is the single purpose of the PRFS.

The Netherlands in category D and Norway are similar in the sense that there is a research assessment exercise at certain

intervals which does not influence resource allocation. The exercise in the Netherlands is self-organized by each of the universities and coordinated on the national level by a Standard Evaluation Protocol. With this autonomous self-evaluation system in place, there is an agreement with the government that performance indicators representing research should not be part of the PRFS. This decision places the country in category D.

These four examples demonstrate that PRFS need to be examined in their national contexts to understand their motivations and design. While research is mostly international, institutional funding is mostly national. Much of the institutional funding comes from tax payers and is determined by democratic decisions. Hence, country differences in the design of a PRFS and its motivations should be expected and respected.

While the literature on the possible adverse effects of PRFS seems to be increasing (Larivière and Costas, 2016), there is little information at the international level about how these systems are motivated and designed in their national contexts. The effects of a PRFS may also depend on the design, as I will demonstrate now.

### Research assessment detached from institutional funding: Norway as an example

*The Metric Tide* report characterizes Denmark as having a “metrics-based evaluation system” (p. 23), disregarding that the country regularly organizes research evaluations based on peer review for other purposes than funding. A more correct description would be that Denmark has a metrics-based institutional funding system. All the Scandinavian countries are observed as using indicators only in their approaches to research assessment (p. 28). I would instead say that they use indicators in their approach to institutional funding. Norway’s UK-inspired research assessment system is overlooked although it has been described in the international literature earlier on (Geuna and Martin, 2003; Arnold *et al.*, 2012) and the Research Council of Norway (RCN) continuously updates information about it in English.<sup>2</sup> The following short description of the Norwegian system is given not only to supplement the perspective of *The Metric Tide* report, in which only national institutional funding schemes are regarded as country practices in research evaluation, but also to demonstrate that there can be reasons for detaching a national research assessment system from funding implications.

The RCN is not only the main body responsible for “second stream” funding in Norway. As part of its role in providing strategic intelligence and advice for the Government, the statutes say that the RCN should “ensure the evaluation of Norwegian research activities”. Always with the help of international expertise, and with international panels of experts in all cases where direct judgment of research quality is concerned, the RCN regularly conducts subject-specific evaluations (often going across the higher education sector, the institute sector and the hospital sector) and evaluations of research institutions as organizations. It also asks external experts to evaluate its own programmes, activities and other funding instruments.

Subject-specific evaluations are organized with intervals of five to ten years and are meant to give recommendations for further improvement. Examples of recent evaluations are: Humanities research in Norway (recently finished), Basic and long-term research in technology (in 2014–15), Norwegian Climate Research (in 2012), Sports Sciences (in 2012), Basic Research in ICT (in 2012), Mathematical Sciences (in 2011), Earth Sciences (in 2011), and Biology, Clinical Medicine and Health Science (in 2011).

The institutional evaluations will consider research quality as well as relevance, and they will give more attention to organizational performance such as efficiency, flexibility and competence

development, cooperation, task distribution and internationalization, leadership, management, and finances. An example is the recent institutional evaluation of independent social science institutes (2016–2017). A recent example of an evaluation of funding instruments is the Midway evaluation of seven Centres for Research-based Innovation (2015).

Since the evaluations inform national strategies and priorities in funding, and since they also inform local leadership at the institutions, they are not without economic consequences. However, the evaluations are not translated into a formula for institutional funding. The possibility of doing so has been discussed from time to time, but the decision has so far been to avoid it. The reasons given have been the following:

- The evaluations mainly have a formative and advisory function.
- The results of the evaluations should be expressed qualitatively. Quantitative information should instead be used in the analysis that precedes the conclusions.
- Gaming should be avoided in the information given to panels.
- Direct funding should support institutional autonomy.

Along with other sources of information, bibliometrics is used to inform peer review whenever relevant in RCN's evaluations. Bibliometrics was even used extensively in the ongoing evaluation of humanities research in Norway. The data source is in this case not Web of Science or Scopus, but the nationally integrated Current Research Information System in Norway (CRISin). Owing to the bibliometric indicator used in institutional funding (see below), this system has comprehensive coverage all peer-reviewed scholarly publishing in the country, also in the formats and languages that are practiced in the humanities (Sivertsen, 2016).

### Indicator-based funding: the Norwegian model

Norway not only has a national research assessment system based on peer review. It also has an indicator-based PRFS in which one of the indicators is bibliometric. This indicator, known as the “Norwegian Model” (Schneider, 2009; Ahlgren *et al.*, 2012; Sivertsen, 2016), has so far been adopted at the national level by Denmark, Finland and Norway, partly also by Belgium (Flanders), as well as at the local level by several Swedish universities. In its section on Denmark, The Metric Tide report describes it as “primarily a distribution model”. The supplementary literature review for The Metric Tide report (Wouters *et al.*, 2015), however, describes it as an evaluation system and discusses its effects in its chapter on “Strategic behaviour and goal displacement”. The same literature review has since then appeared in updated version as a scholarly journal article on “Evaluation practices and effects of indicator use” (de Rijcke *et al.*, 2016) in which it is theoretically argued that any use of metrics in a funding regime can be considered as evaluative. I will shortly describe the Norwegian model before I return to the discussion of effects, both empirically and theoretically.

The Norwegian model has three components:

Component A. A complete representation in a national database of structured, verifiable and validated bibliographical records of the peer-reviewed scholarly literature in all areas of research;

Component B. A publication indicator with a system of weights that makes field-specific publishing traditions comparable across fields in the measurement of “publication points” at the level of institutions;

Component C. A performance-based funding model, which reallocates a small proportion of the annual direct

institutional funding according to the institutions' shares in the total of publication points.

In principle, component C is not necessary to establish components A and B. The experience is, however, that the funding models in C support the need for completeness and validation of the bibliographic data in component A. Since the largest commercial data sources, such as Scopus or Web of Science, so far lack the completeness needed for the model to function properly, the bibliographic data are delivered by the institutions themselves through Current Research Information Systems (CRIS).

The Norwegian model has been extensively documented and evaluated, also in the international journal literature. An overview is given by Sivertsen (2016). Most of what is known about its effects comes from a relatively thorough evaluation that was initiated in 2012 by the Norwegian Association of Higher Education Institutions (representing the funded organizations) in collaboration with the Ministry of Education and Research (the funding organization). An independent Danish team of researchers studied its design, organization, effects and legitimacy (Aagaard *et al.*, 2014). As well as advising improvement and further development, the exercise provided the basis for four in-depth studies of internationally relevant questions (Aagaard, 2015; Aagaard *et al.*, 2015; Schneider *et al.*, 2015; Bloch and Schneider, 2016). Since 2014, the funded and funding organizations have collaborated on following up the evaluation to improve the model and its practices. One example is that the possible disincentive to collaboration (Bloch and Schneider, 2016) and the imbalance in the representation of research fields (Aagaard *et al.*, 2015) have been solved by a redesign of the indicator. The fact that the indicator is also used locally, in some contexts for purposes where it is not appropriate and can do harm (Aagaard, 2015), has been accepted and followed up, not only using the advice of establishing inter-institutional learning arenas for proper managerial use of the indicator (Aagaard *et al.*, 2014), but also by agreeing on national recommendations for good conduct on the local level.

The effects of the Norwegian model are also discussed in the above-mentioned article on evaluation practices and effects of indicator use (de Rijcke *et al.*, 2016), which presents the results of the literature review for *The Metric Tide* report within a theoretical framework in which any use of metrics can be considered as evaluative in itself (Dahler-Larsen, 2014) by having constitutive effects on research practices, including strategic behaviour and goal displacement, task reduction and potential biases against interdisciplinary research, just to mention some of the negative effects reported in this review. I agree with this theoretical position. I, therefore, accept that metric funding systems are studied as evaluative regimes and incentive systems. I only suggest that if proper research evaluation systems are in place in a country, which is the case of Norway, such systems should not be completely overlooked when studying institutional funding systems as evaluative. Different systems might influence, reinforce or balance each other with regard to effects.

The evaluation of the Norwegian model reported stability in the publication patterns (Aagaard *et al.*, 2014; Schneider *et al.*, 2015). Even so, Norway is given as an example along with Denmark, Flanders and Spain “that researchers' quality considerations may be displaced by incentives to produce higher quantities of publications when funding is explicitly linked to research output” (de Rijcke *et al.*, 2016). However, none of the six publications cited to support this statement prove that it is true. I contributed to one of them myself (Ossenblok *et al.*, 2012). It neither supports the statement or is concerned with this question. Instead, it tells another story: The effects of a PRFS depend just as

much on the *design* of the PRFS as on the *type* of PRFS. We observed somewhat different trends in the publication patterns in Flanders and Norway and could relate them to different designs. Quite different effects from those in Flanders and Norway have recently been reported from Poland (Kulczycki, 2017), which has the same type of PRFS with direct use of bibliometrics, but with a design quite different from those of the two other countries.

### From individual-level to organizational-level evaluation?

Lord Nicholas Stern's independent review *Building on Success and Learning from Experience* (2016) deals with criticisms of REF 2014 and recommendations for REF 2021. Hence, it is also about *design* of a PRFS.

The Stern review has been used as part of the basis for the Higher Education Funding Council for England's consultation with the universities on the design of the next REF, scheduled for 2021. I shall not try to predict the outcomes of the process. The following is merely some reflections on the Stern review as I read it.

The Stern review regards peer review as the cornerstone of research evaluation. Its relevance for decisions on institutional funding is not questioned. However, as mentioned in the introduction, the Stern review also highlights five additional goals of the REF: It informs strategic decision making (i) and local resource allocation (ii). It provides accountability and transparency (iii) as well as performance incentives (iv), and it contributes to the formation of the institution's reputation (v). It seems to me that all of these goals could be reached without evaluating the performance of individual researchers. An organizational-level evaluation with peer review as one of several tools could perhaps meet these goals even more efficiently and accurately.

The review's most important recommendations likewise point towards organizational-level evaluation. If implemented in full, societal impact and the research environment will be assessed at the institutional level. All research-active staff will be evaluated, not only a selection as before. Researchers' outputs will be assigned to the institutions that contributed to them and not be "portable" as before.

Stern also recommends the systematic use of metrics, particularly at the institutional level, along with the adoption of open and standardized data for the evaluation. Such data will need to be produced and exchanged at the institutional level. Finally, Stern recommends that the REF should be coordinated with the forthcoming Teaching Excellence Framework (TEF). One could ask: Will the TEF also take the evaluation down to the individual level and assess the performance of each student, disregarding the other contexts where such evaluation takes place? It is difficult to imagine the TEF as anything but an organizational-level evaluation.

The possible move towards organizational-level evaluation could strengthen the REF's ability to provide universities with advice for further development, not only grades based on past performance. It would also allow for additional assessment without implications for institutional funding and thereby make it easier to evaluate interdisciplinary areas, such as—to take a recent example from my own country—climate research.

Organizational-level evaluation could also reduce expenses. REF 2014 included almost 200,000 individual outputs that had already been evaluated in other contexts. It cost £246 million according to the Stern review, up from £66 m in 2008. As mentioned above, the Stern review recommends to include all researchers, not only a selection as before. If the individual-level approach is retained, costs could raise further. The sheer size of

the REF itself now seems to lead to organizational-level evaluation.

There is a chance that the choice in the UK will not be between peer review and metrics, but between individual-level and organizational-level evaluation. With the latter, individual performances would be seen as already evaluated by internal institutional procedures and external procedures in journals and second-stream funding. Peer review would become one of several tools for assessing the performance of a university as a whole.

In addition, the REF could become an inspiring example for smaller countries that are concerned about costs and institutional autonomy when they design their performance-based funding schemes for universities.

### Establishing a basis for mutual learning

The Stern review and the consultations and preparations for the REF2021 show that PRFS are not stable, but constantly further developed and changing. Another example presented above is the evaluation of the PRFS in Norway and how it was followed up. Solutions in some countries are sometimes inspiring other countries. The REF has inspired the PRFS of Italy and the national research assessment systems in Norway and Portugal, while the Norwegian model has influenced the PRFS in Denmark and Finland. It seems important to develop further the basis for mutual learning.

One such step forward already occurred in an Organisation for Economic Co-operation and Development workshop, initiated by Norway, on performance-based funding for public research in tertiary education institutions in Paris in June 2010.<sup>3</sup> One of the outcomes of the workshop was Linda Butler's (2010) chapter on "Impacts of performance-based research funding systems: A review of the concerns and the evidence". This is the most important forerunner of the more academic review (de Rijcke *et al.*, 2016) discussed above. Unlike her successors, Butler could respond in her work to feedback from the participating countries. From reading the results of the evaluation of the PRFS in Norway, de Rijcke *et al.* (2016) note that the Norwegian model is used "despite explicit advice against adoption on lower levels" (Aagaard, 2015), but are not informed about how this problem has been dealt with after the evaluation. It seems that academic literature reviews are not able, as a method, to keep trace of continuously changing national PRFS and thereby to inform mutual learning.

I am presently participating as an expert in a *Mutual Learning Exercise on Performance Based Funding Systems*, which involves fourteen countries and is organized by the European Commission in 2016–17. In this exercise, representatives of governments are actively contributing, learning from each other and taking home advice and inspiration. Still, the process has so far shown that national contexts heavily influence the type and design of PRFS and the needs that these funding systems are expected to respond to. Differences should, therefore, be expected and respected. This does not mean that there is no need for a discussion and clarification across countries of responsible metrics in higher education and research. But rather than trying to formulate best practice statements from the perspective of one or two countries, I have presented a framework for understanding differences with the aim of facilitating mutual learning.

### Notes

- 1 <http://www.uu.se/en/about-uu/quality/evaluation/evaluation-of-research/>.
- 2 Updated information from the Research Council of Norway can also be retrieved directly from the "Evaluations" webpage at [www.forskningradet.no](http://www.forskningradet.no).
- 3 <http://www.oecd.org/sti/sci-tech/oecd-norwayworkshoponperformance-basedfundingforpublicresearchintertiaryeducationinstitutions.htm>.

## References

- Aagaard K, Bloch CW, Schneider JW, Henriksen D, Lauridsen PS and Ryan TK (2014) *Evaluation of the Norwegian Publication Indicator (in Danish)*. Norwegian Association of Higher Education Institutions: Oslo.
- Aagaard K (2015) How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*; **42** (5): 725–737.
- Aagaard K, Bloch CW and Schneider JW (2015) Impacts of performance-based research funding systems: The case of the Norwegian publication indicator. *Research Evaluation*; **24** (2): 106–117.
- Ahlgren P, Colliander C and Persson O (2012) Field normalized citation rates, field normalized journal impact and Norwegian weights for allocation of university research funds. *Scientometrics*; **92** (3): 767–780.
- Arnold E, Mahieu B and Carlberg M (2012) *Evaluation of the Research Council of Norway*. Background Report No 2-RCN. Organisation and Governance Technopolis: Brighton.
- Bloch C and Schneider JW (2016) Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication level at the individual level. *Research Evaluation*; **25** (4): 371–382.
- Butler L (2010) Impacts of Performance-based Research Funding Systems: A Review of the Concerns and the Evidence, in OECD, Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings. OECD Publishing: Paris.
- Dahler-Larsen P (2014) Constitutive effects of performance indicators. *Public Management Review*; **16** (7): 969–986.
- De Rijcke S, Wouters PF, Rushforth AD, Franssen TP and Hammarfelt B (2016) Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*; **25** (2): 161–169.
- Geuna A and Martin BR (2003) University research evaluation and funding: an international comparison. *Minerva*; **41** (4): 277–304.
- Geuna A and Piolatto M (2016) Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). *Research Policy*; **45** (1): 260–271.
- Hicks D (2012) Performance-based university research funding systems. *Research Policy*; **41** (2): 251–261.
- Hicks D, Wouters P, Waltman L, de Rijcke S and Rafols I (2015) Bibliometrics: The Leiden manifesto for research metrics. *Nature*; **520** (7548): 429–431.
- Jonkers K and Zacharewicz T (2016) Research Performance Based Funding Systems: a Comparative Assessment. doi:10.2791/659483.
- Kulczycki E (2017) Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland. *Research Evaluation*; **26** (1): 41–52.
- Larivière V and Costas R (2016) How many is too many? On the relationship between research productivity and impact. *PLoS ONE*; **11** (9): e0162709. doi:10.1371/journal.pone.0162709.
- Ossenblok T, Engels T and Sivertsen G (2012) The representation of the social sciences and humanities in the Web of Science—a comparison of publication patterns and incentive structures in Flanders and Norway (2005–9). *Research Evaluation*; **21** (4): 280–290.
- Schneider JW (2009) An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*; **8** (3): 364–378.
- Schneider JW, Aagaard K and Bloch CW (2015) What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation*; **25** (3): 244–256.
- Sivertsen G (2016) Publication-based funding: The Norwegian model. In: Ochsner M, Hug SE and Daniel HD (eds). *Research Assessment in the Humanities: Towards Criteria and Procedures*. Springer Open: Zürich, pp 79–90.
- Stern N (2016) Research Excellence Framework (REF) review: Building on success and learning from experience. Ref: IND/16/9. Department for Business, Energy & Industrial Strategy.
- Swedish Government. (2016) Regeringens Proposition 2016/17:50. Kunskap i samverkan—för samhällets utmaningar och stärkt konkurrenskraft.
- Swedish Research Council. (2015) Research Quality Evaluation in Sweden - Fokus: Report of a Government Commission regarding a Model for Resource Allocation to Universities and University Colleges Involving Peer Review of the Quality and Relevance of Research. Swedish Research Council: Stockholm.
- Wilsdon J *et al* (2015) The Metric Tide: The Independent Review of the Role of Metrics in Research Assessment and Management. doi:10.13140/RG.2.1.4929.1363.
- Wouters P *et al* (2015) The Metric Tide: Literature Review. HEFCE. doi:10.13140/RG.2.1.5066.3520.

## Data availability

Data sharing is not applicable to this paper as no datasets were analysed or generated.

## Additional information

**Competing interests:** The author declares no competing financial interests.

**Reprints and permission** information is available at [http://www.palgrave-journals.com/pal/authors/rights\\_and\\_permissions.html](http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html)

**How to cite this article:** Sivertsen G (2017) Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*. 3:17078 doi: 10.1057/palcomms.2017.78.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017