

an outline of the bibliometric indicator used for performance-based funding of research institutions in norway

jesper w. schneider

Royal School of Library and Information Science, Fredrik Bajers Vej 7K, Aalborg, 9220, Denmark.

E-mail: jws@db.dk

doi:10.1057/eps.2009.19

Abstract

This article outlines and discusses the bibliometric indicator used for performance-based funding of research institutions in Norway. It is argued that the indicator is novel and innovative as compared to the indicators used in other funding models. It compares institutions based on *all* their publication-based research activities across all disciplines. Specific incentives are given to researchers to focus their publication behaviour on the most 'prestigious' publication channels within the different fields. Such aims necessitate a documentation system based on high-quality data, and require differentiated publication counts as the basic measure. Experience until now suggests that the indicator works as intended.

Keywords bibliometric indicators; performance-based funding; research institutions; publication counts

This article describes the bibliometric indicator used for performance-based funding of research institutions in Norway, commonly known as the 'Norwegian model' (Sivertsen, 2006). We reflect upon the model's incentives, advantages and potential problems in relation to the social sciences and other models with similar purposes. Finally, we present the latest experience with the Norwegian model. The general debate

with regard to the appropriateness of performance-based funding of research is not discussed in this paper. The focus here is on the mechanisms embodied in the bibliometric indicator and how it compares with other indicators used for similar purposes.

Government interest in performance-based funding of research institutions has been growing in recent years. Several European countries have implemented, or

are currently implementing, such funding models, whereas other countries are considering which model to choose. It is noticeable that the current interest is especially focused on metric solutions, such as bibliometric indicators, as opposed to panel-based peer review models like that of the Research Assessment Exercise (RAE) in the United Kingdom. In fact, the RAE is to be replaced after the 2008 evaluation by the metric-oriented Research Excellence Framework (www.hefce.ac.uk/research/ref/).

The Norwegian model is especially interesting in this context. It was commissioned by the Norwegian Ministry of Education and Research in 2002 and developed by the Norwegian Association of Higher Education Institutions in 2003–2004.¹ The purpose of the model is annually to redistribute basic research funding among institutions in the higher education sector (six universities and forty other institutions) based on a bibliometric indicator that counts scholarly publications. Publications have been counted since 2004. The Norwegian model was first used in connection with budget allocations for 2006 (which were based on publication counts for 2005). The funding model redistributes approximately 2 per cent of the annual budget for basic research in Norway, approximately €3.4 billion (Sivertsen, 2008).

At first, this may seem somewhat unremarkable, but the indicator is both novel and innovative. The aim is annually to count *all* scholarly publications within *all* research fields. Institutions are thus compared based on *all* their publication-based research activities. Such an aim requires publication counting as the basic indicator, and thus complete, valid and reliable publication data are necessary. In order to counter unintended publication behaviour and to encourage publication in 'prestigious' channels, counting in the model is differentiated. It is assumed that such a model is able to measure

'The aim is annually to count all scholarly publications within all research fields.'

'performance' beyond mere productivity. This is an innovative solution, which makes the Norwegian model an appropriate alternative to traditional citation-based indicators.

The following section outlines the background to the Norwegian funding model and discusses some issues important to the indicator. We then outline the model's basic components and discuss their rationale; the section after discusses the model in relation to the social sciences, and finally we describe recent experience with the model and briefly describe how the ideas underlying the Norwegian model seem to have inspired funding models elsewhere.

BACKGROUND

Research evaluation models are often characterised as either 'weak' or 'strong' (e.g., Whitley, 2007). Weak models are usually informal, private and primarily used for formative research policies and management issues. Strong models are formal, public and summative in the sense of having direct links to the allocation of resources. Research evaluation is also categorised as either *ex ante* or *ex post* (Kogan, 1989). *Ex ante* evaluation is applied to research proposals, and is conducted prior to the execution of research. *Ex post* evaluation comes once research has been completed, and assesses the output and the impact. Funding of basic research is usually done through annual block grants, with project funding being allocated thorough *ex ante* evaluation by the research councils or funding agencies, such as US federal funding, or *ex post* evaluation, in which

funding is either based on reviews of output and merit, or more mechanically, in which funding is directly connected to a quantitative formula. This paper focuses on the latter, as the Norwegian model is an *ex post* evaluation model directly connected to a funding formula based on a bibliometric indicator. But what do bibliometric indicators measure?

Performance-based funding of research presupposes distinct measures of performance; however these are not easily agreed upon. The principal methods are peer review, often in the form of panel rankings, and bibliometric indicators (Moed, 2005). We do not consider peer review in this paper.² The intention in applying performance-based bibliometric indicators has traditionally been either to identify 'high quality' research or to find out which research is 'better' (Moed, 2005). Not surprisingly, application of quantitative indicators has generated a debate about what is meant by 'research quality', and how various indicators are related to it.

A simple undifferentiated count of a unit's number of publications is usually viewed as a measure of the productivity of research, in other words, 'quantity' rather than 'quality' (Moravcsik, 1973). Indicators based on publication counts have not received the same attention as citation-based indicators in discussions of validity because their connection to 'research quality' has been quite remote. The peer review process that scholarly publications undergo may be interpreted as a sign of 'quality'. But to many, a publication constitutes nothing more than an 'offer' to the scientific community. It is the subsequent reception of that offer that certifies the actual 'impact' of a publication (e.g., Moed, 2005; Glänzel, 2008).

The most important instance in which research funds were linked, at least in part, to productivity measures, undifferentiated by any measure of 'quality', is the

'The most important instance in which research funds were linked, at least in part, to productivity measures, undifferentiated by any measure of "quality", is the Australian funding model from the 1990s.'

Australian funding model from the 1990s (Butler, 2004). A striking effect of the model was a dramatic increase in the number of research publications from Australia. The largest proportion of this increase came, however, in lower-impact journals (Butler, 2004). The latter effect was clearly unintentional and suggested that a general increase in research output may have come with a lower overall impact (Butler, 2004). Undoubtedly, publication behaviour among a considerable number of researchers in Australia underwent a change that reflected an overriding emphasis upon quantity. This is what Weingart (2005: 125) calls a 'reactive measure'. Bibliometric indicators linked to funding models inadvertently become reactive measures when researchers alter their behaviour in ways unintended by those applying the indicator (Weingart, 2005). Notice that behavioural changes are indeed anticipated and intended politically when bibliometric indicators are linked to research funding. The problem arises when behavioural change has negative effects for the funding model as a whole, inducing 'game playing' by researchers without necessarily improving performance. In Australia, publication counts were conceived as a means of distributing research funds on the basis of the 'quality' of research. But undifferentiated publication counts are not measures of 'quality'. Institutions

were rewarded for publication activity only – to which they and their researchers obviously adapted.

The Norwegian funding model also focuses upon productivity. Contrary to the Australian model, the Norwegian model is based on differentiated counting of publications. Differentiated counts mean that some publication activities are favoured and therefore given greater weights. The Norwegian model encourages institutions and their researchers to publish through the most 'prestigious' publication channels within different fields of research. 'Prestige' is linked to such characteristics of the publication channels as 'having a tough peer review process', 'publication competition', 'visibility to the widest relevant audience', 'general reputation of the channel', etc. (Sivertsen, 2008). Notice that the model does not determine 'quality' or predict the 'impact' of individual publications. The model gives incentives to publish through a selected number of publication channels, somehow deemed 'qualitatively' better than the majority of channels within the specific field of research. At the same time, these incentives should counter excessive publication activity through less 'prestigious' channels, as happened in Australia. The notion that some publication channels are 'more prestigious' than others may seem controversial. Nevertheless, in many fields of research, different publication channels do have noticeably more or less status among researchers. This is made explicit in the Norwegian model. It is important to emphasise that publication through a 'less prestigious' channel is not the same as 'poor research'. It is merely a division between 'normal' and 'selective' publication patterns within the fields. The 'selective' patterns are weighted higher because they are deemed more competitive and attractive. It is assumed that this incentive can motivate researchers to publish more and *better* research, while

at the same time countering negative influences in publication patterns.

Some critics of the Norwegian model have argued that a proper quantitative performance measure should be based on citations (e.g., Sandström and Sandström, 2007). What the critics argue is that citations measure 'impact', which is often considered a proxy for 'research quality'. Obviously the use of bibliometric indicators presuppose that they are able to measure aspects of 'research quality' (e.g., Gläser and Laudel, 2007). However, faith in the possibility of measuring the scientific quality of individual publications through citations alone is misplaced (Van Raan, 1996; 2000; Moed, 2005; Glänzel, 2008). Within the bibliometric community, there is consensus that its complexity is such that 'research quality' can only be judged by peers (e.g., Van Raan, 1996). Citations, in general, and impact factors, in particular, are and remain primarily **the indicators of the reception of scientific information (Glänzel, 2008).** The observation that citations indicate reception, use and, therefore, usefulness, as well as impact, is the basic argument for using them as proxies for 'quality' at levels above individual publications. **Bibliometric indicators may therefore be more or less related to 'quality', and perhaps measure certain aspects of it, but they cannot exhaustively represent 'research quality'** (Martin and Irvine, 1983, Van Raan, 1996). The question is to what extent do publication-based indicators reflect 'research quality'? In the Norwegian model incentives are given to continue to focus on, or to shift the focus of, publication efforts towards the 'most prestigious' journals within the different fields of research. It is reasonable to suggest that such an incentive is related in some way to 'quality' as it rewards more than simple productivity. This is one of the reasons why the model is both novel and innovative compared to the existing *ex post* funding models. But it is

unlikely that publication incentives reflect the same aspects of 'quality' that citations are supposed to reflect.

Obviously, the 'reception' of publications within a research field is an interesting metric for research evaluation. And perhaps our perception of citations is that it is closer to measuring 'quality'. Nonetheless, it is generally accepted that at least six modalities restrict the conditions under which citation-based indicators can be applied in order to produce valid and reliable results (Gläser and Laudel, 2007). These include first and foremost the aforementioned limitation that citations do not measure quality *per se*, but is rather 'an important aspect of quality' (Van Raan, 1996: 404). Second, bibliometric indicators must be applied to a large number of publications for the statistics to be reliable (e.g., Van Raan, 2000; Butler, 2001). Third, there is the issue of coverage; for valid conclusions about research performance to be drawn, the whole research output of evaluated units must be covered (Moed *et al*, 1985). Fourth, time frames are essential in citation analysis, as publications within different fields need different lengths of time to accumulate citations (Moed *et al*, 1985). Fifth, publication and reference patterns vary between fields of research. Bibliometric indicators are therefore field-specific and need to be normalised when aggregated and compared to each other (Moed, 2005). Furthermore, the delineation of fields affects the validity of bibliometric evaluations and therefore becomes a crucial task. Sixth, there are the critical technical issues with regard to access, coverage, quality and the limitations of bibliometric data (e.g., Glänzel and Schoepflin, 1994; Moed, 2005). Public access to bibliometric data suitable for research evaluation is severely restricted; such data are commercial to vendors like Thompson Reuters and Scopus. As a consequence, qualified research evaluation based on citation

analysis is commercialised and carried out by a few institutions worldwide that have access to a sufficient number of bibliometric data. Some of these constraints or modalities apply to all bibliometric indicators, whereas all of them apply to citation-based indicators.

As a result, citation-based indicators have several limitations that disqualify them for the purposes of the Norwegian funding model, as outlined above. Limitations include, for example, a lack of complete data reliable enough to enable comparison of institutions based on all their research activities across all fields, and the different time frames needed for citation counting. The latter is especially important in relation to funding models. In order for citation patterns to be reliable within a field of research, indicators based on citation counts require longer time spans compared to indicators based on publication counts (Moed, 2005). Accordingly, if citation-based indicators are used in a funding model, the performance measured is not that of the most recent research, but research of some age (often published 3–5 years prior to the year of evaluation) (e.g., Debackere and Glänzel, 2004). In some situations, and within some fields, this may be considered appropriate; in other situations, however, it may be seen as static and inflexible. An alternative is therefore a dynamic and flexible funding model, in which the most recent research determines the forthcoming year's or years' budget allocation(s). Such an approach disqualifies the use of citation counts due to the limited time span. Citation-based indicators are perhaps most useful for retrospective *ex post* evaluations.

A question remains whether *Google Scholar*, or other freely available citation databases, could be used for research evaluation in the future. Some studies have reported good correlations between traditional journal impact factors and citation data obtained through *Google*

Scholar (e.g., Harzing and van der Wal, 2008). Although there are certainly several things to commend about *Google Scholar*, such as its free access to data and its coverage of proceedings, books and international and non-English language journals, the fact remains, however, that the data quality is extremely poor and publication inclusion criteria are obscure. We therefore have no knowledge of coverage, and an extremely time-consuming task of data cleaning lies ahead of us if *Google Scholar* is to be used. Meho and Yang (2007) report a comparison of data cleaning between *Web of Science* and *Google Scholar* for the same citation analysis. It took 100 hours in *Web of Science* and a gruelling 3,000 hours in *Google Scholar*. More generally, *Google Scholar*, and other citation databases for that matter, are still restricted by the six intrinsic modalities mentioned above. In fact, the case of *Google Scholar* emphasises the need for complete and good quality data in order to produce valid and reliable indicators.

We now turn to the components of the Norwegian bibliometric indicator and discuss, in the following section, their rationales and purposes.

THE COMPONENTS OF THE MODEL

Bibliometric indicators tend to be aggregated constructs that often comprise complex formulae applied to data to which public access is restricted. Although their mechanisms are well understood, such complex indicators can be challenging for funding purposes due to their lack of transparency. It is difficult to perceive the extent to which individual publications, publication strategies or relative citation impacts affect the outcome of such funding models (see e.g., Debackere and Glänzel, 2004).

A priority in the development of the Norwegian model has been simplicity and

'A question remains whether Google Scholar, or other freely available citation databases, could be used for research evaluation in the future.'

transparency. It has been essential to construct an indicator that clearly delineates the different effects of publishing strategies within the model. The goal is threefold: (1) to give researchers and institutions incentives to publish in the most rewarding channels; (2) to create valid and reliable national publication data; and (3) to provide public insight into the model, its underlying data and reward mechanisms. The Norwegian model clearly specifies the actual number of points a specific publication contributes to the annual total number of points for an institution (see <http://dbh.nsd.uib.no/pub/>). Notice that these points are comparable across divergent fields from art history to astrophysics, as the model in Norway comprises all scholarly fields (with a few exceptions) and all scholarly publications within these fields.

The model has two interrelated components: (1) a transparent national research documentation system and (2) a simple bibliometric indicator. In the following section we will outline the two components and the rationale behind them.

THE NATIONAL RESEARCH DOCUMENTATION SYSTEM

A common predicament in all bibliometric analyses is the nature and coverage of the bibliographic data (e.g., Smith, 1981; Moed, 2005). This applies to citation databases, such as Thompson Reuters or Scopus, as well as national or institutional research databases that index scholarly

publications. Perhaps the most novel, and certainly the most innovative aspect, in our view, of the Norwegian model is the construction of a national research documentation system that supports the bibliometric indicator. Norwegian scholarly publications are not only all registered in the system, but the data describing them are also validated and standardised. As a result, Norway currently has one of the richest bibliographic databases used for bibliometric purposes. The system and its data are publicly transparent. It further secures complete, verifiable and structured metadata for all scholarly publications from all Norwegian research institutions.

Essential for the model and the registration of scholarly publications in the documentation system is the dynamic authority file of accepted scholarly publication channels. The authority file ensures that no 'non-scholarly' publications are entered into the system. Publication channels are defined as ISSN-titles (journals, e-journals and series) or publishers of ISBN-titles. Two criteria determine whether publication channels can be accepted for the model: (1) they must use external peer review, and (2) no more than two thirds of the authors that publish through a channel can be from the same institution. The reason this requirement is applied is that it is likely in such cases that authors with close professional relationships publish and review each others' works, and it is questionable whether such local, non-national peer-review processes function optimally. In addition, contrary to the intentions of the model, opportunities for publication through local channels may be linked to remuneration incentives, and thus stimulate high publication rates for that reason only.

Consequently, the scholarliness of publications is determined according to the status of the publication channel. Currently, some 18,000 publication channels

are accepted for the Norwegian model (for an overview, see <http://dbh.nsd.uib.no/kanaler/>).

Institutions are responsible for the quality of the data relating to their registered scholarly publications. Several measures are implemented to support this process. Bibliographic data from external sources, such as Thompson Reuters and the Norwegian National Library, are imported to the documentation system in order to facilitate registration and validation of publications. The primary validations and standardisations of the registered data include names of publication channels, document types, institutional affiliations of authors and institutional names. These procedures are crucial, not just for the present publication-based indicator, but for bibliometric analyses in general. They are an excellent basis for future citation analyses in areas where such measures would be appropriate. Authority files containing accepted publication channels, document types and institutional names ensure a predominantly automatic validation and standardisation process. However, manual procedures are necessary to some extent. According to Sivertsen (2008), the actual cost of running the model is low compared to the national gains achieved in data registration and data quality.

THE BIBLIOMETRIC INDICATOR

In order to make fields comparable, the simple bibliometric indicator delimits and weights publications in a way that balances field-specific publication patterns. The indicator's formula consists of two dimensions (see Table 1). The first dimension classifies publication types into three categories (papers in ISSN-titles, papers in ISBN-titles and ISBN-titles). The second dimension divides publication channels into two levels. Publication points are thus weighted according to publication type and level of publication channel.

Table 1: Publication points

	<i>Level 1</i>	<i>Level 2</i>
Scholarly articles in a journal, etc. (ISSN)	1	3
Scholarly articles in anthology etc. (ISBN)	0.7	1
Monographs (ISBN)	5	8

Level 1 corresponds to publication channels on 'a normal level' according to publication patterns within a specific field, whereas level 2 embraces the 'most selective' and 'prestigious' channels within the field. The latter is given extra weight. As stated above, the division of publication channels is made in order to give researchers incentives to focus their publication activity on a 'selected number of prestigious channels' within the research fields and to counter unintended publication behaviour. At any one time, publication channels on level 2 can account for a maximum of 20 per cent of the world's publications within a specific field. The intention behind this rule is to make the division of publication channels dynamic in response to changing publication behaviours among researchers.

The Norwegian national research councils in each field of research determine which channels are to belong to level 2. Evaluation and revision are carried out annually and on the basis of specific guidelines that consider primary field publication patterns.

Journal rankings of this sort are often controversial. Results are often contested and representatives of emergent fields especially are critical of lack of visibility or lower rankings that such processes accord to them. Nevertheless, the process of journal ranking in Norway has been relatively smooth, in contrast with, for example, a similar process recently carried out in Denmark. The main reason for the difference is likely to be found in the way the ranking process was approached in the two countries. In Norway, a top-

down approach was used. A detailed empirical study of all research fields identified three main groups of publication patterns. Notice that the groups do not completely correspond to the traditional divisions among scientific disciplines. As a result, the three groups were issued different guidelines for use in allocating publication channels to level 2. For example, fields belonging to group A have primarily international journal publication patterns, in which the most significant journals have high annual volumes, high rejection rates and cover a broad range of topics compared to other journals. Most journals in group A are included in Thompson Reuters, and calculations of field-specific citation indicators for journals are somewhat useful for creating a draft of publication channels for level 2. Consequently, provisional draft lists of journals were issued to the Norwegian research councils for each field belonging to group A. These lists were then used as a starting point for negotiation and nomination to level 2 based on the specific guidelines. The same procedure was followed for the other two groups of publication patterns (for further details see the link in note 1). Notice that the classification of publication patterns is only necessary for formulating different guidelines for use in nominating publication channels to level 2. The end result is a common list of publication channels at level 2, which is not divided by research fields. Researchers in all fields may use all channels included in level 2. This approach undoubtedly eliminated some of the

potential conflicts usually experienced in journal rankings. The controversies and problems experienced with journal rankings in Denmark supports this. Denmark implemented an *ex post* funding model very similar to the Norwegian model. Surprisingly, the journal-ranking process in Denmark did not follow the Norwegian top-down approach with a thorough basis for negotiation and nomination to level 2. Instead, a less coordinated bottom-up approach was used. Members of the field-specific research committees were required both to identify relevant publication channels for their specific fields and subsequently to nominate candidates for level 2. Not surprisingly, this uncoordinated process created much confusion, frustration and fierce debate in the Danish research community.

Finally, the indicator uses fractional counting. Publications are attributed to institutions according to author affiliations. All institutions are given equal weight. If an author is affiliated with two or more institutions in a publication, the publication will be divided equally among the institutions. Consequently, a publication with one author is worth 1 point. A publication with n authors is worth $1/n$ point for each author. It was originally suggested that the maximum fraction should be $1/10$ point for publications with more than ten authors. This is not the practice in Norway at the moment. Consequently, an institution is granted publication points according to the following formula: Fraction of authors credited to the institution \times publication points for the individual publication (publication type and level of channel).

Fractionalised counting is a vital feature in the indicator. It is a provision that should prevent a focus on productivity only, as happened in Australia. If whole counting is applied, in which each participating institution gets full credit, one can easily imagine that the number of co-authored papers will increase

'The main argument levelled against fractional counting has been that such a scheme may diminish the incentives to collaboration between institutions, especially international collaboration.'

dramatically due to unintended reactive behaviour by authors and institutions.

The use of fractionalised counting in the indicator has nevertheless caused some debate. The main argument levelled against fractional counting has been that such a scheme may diminish the incentives to collaboration between institutions, especially international collaboration. However, the argument against such counting is partly based on a myth. Research co-operation is certainly a necessary and positive phenomenon in the era of 'big science', but the idea that collaboration is a recipe for guaranteed success is not true (Glänzel, 2008). It is generally accepted that the visibility and impact of collaborative research is on average moderately higher than that of non-collaborative research (Persson *et al*, 2004). But numerous counter-examples confirm that multi-authorship, and above all international collaboration, does not in itself guarantee increases in productivity, visibility or impact, and neither does it facilitate publication in high-impact journals (Glänzel, 2001; Glänzel and Schubert, 2001).

Hitherto, the position in Norway has been that 'invisible colleges' and social networks within research specialties have eventually ensured collaboration, where collaboration has been needed and wanted. Much research depends on

collaboration and it is assumed that the funding model will generally not undermine such behaviour. Until now there has been no empirical indication that the Norwegian model causes a drop in national or international collaboration.

No optimal counting scheme exists. Fractional counting serves an important function in the Norwegian model, as described above. As no adverse effects can be seen, there is no reason to give special incentives to collaboration.

The following section discusses the problems related to the use of bibliometric indicators for research evaluation in the social sciences, and outlines the advantages of the Norwegian model in that respect.

THE NORWEGIAN MODEL AND THE SOCIAL SCIENCES

Existing bibliometric approaches to research evaluation most often characterise the social sciences according to the data coverage of Thompson Reuters or more recently Scopus. They usually either compare institutions based on this coverage, or simply exclude social-science fields all together in the evaluation of institutions. The former approach has validity problems. The latter approach is clearly limited and also deficient, in as much as it restricts the research fields that come to characterise an institution's research activity and perhaps impact. Indeed, there has been considerable debate about the usefulness of bibliometric indicators as an evaluative tool for the social sciences (Katz, 1999: 1). There is consensus in the literature that application of bibliometric indicators to the social sciences (and humanities) is difficult primarily due to the limitations of publication coverage in the citation database (e.g., Hicks, 1999; 2004; Katz, 1999). Citation databases are usually required for bibliometric analyses whether they are publication- or citation-based

or both. The reason is that these databases contain country and institutional affiliations for *all* authors in contrast with most domain-specific databases. But bibliometric indicators are only appropriate when the database being analysed adequately covers the publications that are the principal carriers of knowledge within a field. The coverage is good in Thompson Reuters and Scopus for most natural- and life-science fields, but not nearly sufficient in the social sciences and humanities. The databases lack coverage of monographs, book chapters and conference proceedings, and they have a highly selective language and geographical coverage as well. This fact undoubtedly throws doubt on the validity of evaluations of performance in the social sciences and humanities relying on these databases. Hicks (2004: 474) has observed that '[w]hen challenged to evaluate scholarly work in the social sciences and humanities we are rudely forced to work outside this comfort zone in a frankly messy set of literature'.

The national research documentation system in Norway provides a 'comfort zone' of reliable and highly structured bibliographic data. Here we have complete annual research publication data for the social sciences and the humanities. The documentation system is not a citation database; instead, complete publication activity for all scientific fields can be measured, and publication behaviour and profiles can be monitored and influenced through the model. From the point of view of the social sciences (and humanities for that matter), the Norwegian model clearly has advantages compared to existing bibliometric approaches to research evaluation at the level of institutions. The most obvious advantage is the emergent visibility of the research activity of these major scientific fields. This enables comparison of institutions based on all their research activity, and likewise direct comparison among different

fields of research within and among such institutions.

The final section briefly presents some recent experience with the model in Norway, and describes how the Norwegian model seems to have inspired other funding models around the world.

BIBLIOMETRIC EXPERIENCE IN NORWAY AND INFLUENCE UPON OTHER COUNTRIES

Information about the effects of the funding model in Norway is gradually beginning to appear. According to Sivertsen (2008), there has been a substantial growth in publications from the higher education sector in Norway. Notably, the growth in publication output has affected both levels of publication channels. Validation studies on a subset of Norwegian publication data indexed in Thompson Reuters' *Web of Knowledge* show that the increased publication rates are distributed evenly between high impact journals and other journals (Sivertsen, 2008). In fact, Norway's relative citation rate has been stable at 25 per cent above the world's average, according to data from Thompson Reuters' *National Science Indicators* (Sivertsen, 2008). While there may be several explanations for this increase in production, and the stability in citation rates, it appears that the model seems to work as intended. Until now we have not seen the adverse effects experienced in Australia with their undifferentiated publication-based indicator. We should of course be careful when assessing the immediate effects of the Norwegian model. Bibliometric analyses may show that Norway's performance has increased in comparison to other countries. The problem is that attribution of this improvement to the Norwegian model may turn out to be a *post hoc ergo propter hoc* fallacy. Experience from the RAE in the United Kingdom suggests this. While

'Several comparative studies indicate that the absolute amount of money invested in university research is a much stronger predictor of research performance than any research evaluation model.'

the United Kingdom improved its relative performance, so did other countries without research evaluation models. Several comparative studies indicate that the absolute amount of money invested in university research is a much stronger predictor of research performance than any research evaluation model (e.g., Liefner, 2003). Hence, the RAE was not the only factor influencing improved performance in the United Kingdom. More generally, it is important to monitor performance-based funding models because they may over time lead to 'homogenisation' of research, discouraging experiments with new approaches, and rewarding 'safe' research, irrespective of its benefits to society. The resulting decrease in diversity may be harmful (Guená and Martin, 2003). There is also a danger with models that focus on a one-dimensional concept of research 'quality' and link the results directly to funding. In the Netherlands, by contrast, performance is assessed along four dimensions – 'scientific quality', 'scientific productivity', 'scientific relevance' and 'long-term viability' – and results are not directly linked to funding.

Nevertheless, based on the 4 years of experience with the model in Norway, it is reasonable to conclude that it is indeed possible to include *all* scholarly publications within *all* scientific fields in a bibliometric indicator. It is particularly

important that institutions with different research profiles can be compared based on all their research activity, and thus the different publication patterns among fields are comparable with the Norwegian model.

Current approaches to *ex post* funding models can be classified as to (1) whether they are based on peer review (panel evaluation), bibliometric indicators or a combination of these; (2) whether all or only some research fields are included; and finally (3) whether some or all publications (usually data from Thomson Reuters) are included. In a survey of fourteen countries, Guena and Martin (2003) identified three countries that at the time had implemented *ex post* evaluation models for allocating research funds: Britain, Australia and Finland. Only Australia had a model based on a bibliometric indicator directly linked to funding. Today, more countries have implemented, or are considering implementing, such directly linked *ex post* funding models. Norway and Belgium (Flanders) have implemented, and Denmark and Sweden are in the process of implementing, such models. Britain and Finland are changing their current models, but maintaining *ex post* funding. Only Australia has for the time being given up *ex post* funding of research.

Obviously, the model in Norway is based on a bibliometric indicator, and includes all fields and all publications. The RAE in the United Kingdom has been re-considering the standard panel evaluation, where in principle all fields are included and the institutions are allowed to submit a selected number of publications within each field for evaluation. As stated above, the RAE is to be replaced after the 2008 evaluation with a metric-oriented model. Denmark is currently implementing a model profoundly influenced by the Norwegian model. It seems that there will be only minor differences between the two models. One difference

might be the extra incentive to collaboration in the Danish model. Finland has not yet decided upon whether to use a panel- or bibliometric-based model. Flanders currently relies upon a model based on bibliometric indicators (both publication- and citation-based indicators) for selected fields using data from Thompson Reuters.³ However, the model is to be modified due to the above-mentioned problems, with measuring performance in the social sciences and humanities (Debackere and Glänzel, 2008). Most interestingly, the Norwegian model is to be the basis for the modification in Flanders. For the time being, Australia has turned down a performance-based funding model; however, an initiative has been implemented that will evaluate Australian research in general. This initiative is inspired by the Norwegian model, as a comprehensive list of publication channels is compiled and the channels weighted. Finally, Sweden has recently chosen a model comparable to the one that is to be modified in Flanders, based on bibliometric indicators for selected fields using data from Thompson Reuters. In fact, the indicator in Sweden is rather unorthodox as it tries to *estimate* and *predict* performance for all fields based on a limited set of data from Thompson Reuters (Sandström and Sandström, 2008).

CONCLUSION

Let us end with some general comments on bibliometric indicators and the apparently expanding 'ranking industry' (e. g., van Raan, 2005). It is necessary to warn against the unthinking use of indicators and rankings in research policy applications and evaluations. Indicators and rankings are not natural objects. They are cultural constructs that are theoretically informed by underlying assumptions (Leydesdorff, 2008). In general, such tools reduce the complexity of a

multidimensional problem to a simple number. It is therefore important that users of indicators and rankings, such as policymakers and administrators, be aware of their underlying assumptions and potential effects. Different indicators have strengths and limits, and no indicator alone can express the multidimensional complexities involved in research evaluation. The choice of indicators depends on priorities and aims. In Norway priority was given to transparency and evaluation of institutions based on *all* research fields and *all* their scholarly publications. This aim determined the initial choice of indicator. However, it does not exclude the later use of alternative indicators or peer review evaluations, in which this would be appropriate. Indeed, different indicators and peer reviews should be productively

combined rather than being made to compete with each other (Moed, 2005). But regardless of the indicator chosen, the documentation system created in Norway sets new standards for national research registration, and thus for the quality of the publication data needed for bibliometric analyses. All countries should strive for this.

Acknowledgements

The author thanks Gunnar Sivertsen from the Norwegian Institute for Studies in Innovation, Research and Education (NIFU STEP) for his kind support and his generous sharing of data and information in relation to the model and the experiences with it. The author also thanks the anonymous reviewers for their valuable comments.

Notes

1 A detailed description of the indicator, its background and development, is available in www.uhr.no/documents/Rapport_fra_UHR_prosjektet_4_11_engCJS_endelig_versjon_av_hele_oversettelsen.pdf. Most of the detailed description of the documentation system and the bibliometric indicator is based on the document.

2 For comprehensive reviews of subjectivity and reliability issues, see Cicchetti (1991) and Langfeldt (2001), and for comparisons with bibliometric indicators, see Moed (2005).

3 A different model is used in the remainder of Belgium.

References

- Butler, L. (2001) *Monitoring Australia's Scientific Research: Partial Indicators of Australia's Research Performance*, Canberra: Australian Academy of Sciences.
- Butler, L. (2004) 'What Happens When Funding is Linked to Publication Counts?', in H. F. Moed, W. Glänzel and U. Schmoch (eds.) *Handbook of Quantitative Science and Technology*, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 340–389.
- Cicchetti, D.V. (1991) 'The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation', *Behavioral and Brain Sciences* 14: 119–186.
- Debackere, K. and Glänzel, W. (2004) 'Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key', *Scientometrics* 59(2): 253–276.
- Debackere, K. and Glänzel, W. (2008) 'Evidence-based bibliometrics: A decade of bibliometrics-based science policy in Flanders', in J. Gorias and E. Schiebel (eds.) *Book of abstracts, 10th international science and technology indicators conference*, 17–20 September 2008, Austria: University of Vienna, pp. 123–125.
- Glänzel, W. (2001) 'National characteristics in international scientific co-authorship', *Scientometrics* 51(1): 69–115.
- Glänzel, W. (2008) 'Seven myths in bibliometrics. About facts and fiction in quantitative science studies', in H. Kretschmer and F. Havemann (eds.) *Proceedings of WIS 2008, Berlin, Fourth International Conference on Webometrics, Informetrics and Scientometrics and Ninth COLLNET Meeting Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)*: <http://www.collnet.de/Berlin-2008/GlanzelWIS2008smb.pdf>, (accessed 7 October 2008).

- Glänzel, W. and Schoepflin, U. (1994) 'Little scientometrics, big scientometrics ... and beyond', *Scientometrics* 30: 375–384.
- Glänzel, W. and Schubert, A. (2001) 'Double effort = double impact? A critical view of international co-authorship in chemistry', *Scientometrics* 50(2): 199–214.
- Gläser, J. and Laudel, G. (2007) 'The Social Construction Of Bibliometric Evaluations', in R. Whitley and J. Gläser (eds.) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems, Sociology of Sciences Yearbook* 26, Dordrecht, The Netherlands: Springer, pp. 101–123.
- Guená, A. and Martin, B.R. (2003) 'University research evaluation and funding: An international comparison', *Minerva* 41(4): 277–304.
- Harzing, A.-W.K. and van del Wal, R. (2008) 'Google scholar as a new source for citation analysis', *Ethics in Science and Environmental Politics* 8: 61–73.
- Hicks, D. (1999) 'The difficulty of achieving full coverage of international social science literature and the bibliometric consequences', *Scientometrics* 44(2): 193–215.
- Hicks, D. (2004) 'The Four Literatures of Social Science', in H. F. Moed, W. Glänzel and U. Schmoch (eds.) *Handbook of Quantitative Science and Technology*, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 473–496.
- Katz, J.S. (1999) 'Bibliometric Indicators and the Social Sciences', Report prepared for UK Economic and Social Research Council, www.sussex.ac.uk/Users/sylvank/pubs/ESRC.pdf, (accessed 7 October 2008).
- Kogan, M. (1989) 'The Evaluation of Higher Education: An Introductory Note', in M. Kogan (ed.) *Evaluating Higher Education*, London: Jessica Kingsley Publishers, pp. 11–25.
- Langfeldt, L. (2001) 'The decision-making constraints and processes of grant peer review, and their effects on the review outcome', *Social Studies of Science* 31: 820–841.
- Leydesdorff, L. (2008) 'Caveats for the use of citation indicators in research and journal evaluations', *Journal of the American Society for Information Science and Technology* 59(2): 278–287.
- Liefner, I. (2003) 'Funding resources allocation and performance in higher education systems', *Higher Education* 46: 469–489.
- Martin, B.R. and Irvine, J. (1983) 'Assessing basic research. Some partial indicators of scientific progress in radio astronomy', *Research Policy* 12: 61–90.
- Meho, L.I. and Yang, K. (2007) 'Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google scholar', *Journal of the American Society for Information Science and Technology* 58(13): 2105–2125.
- Moed, H.F. (2005) *Citation Analysis in Research Evaluation*, Dordrecht: Springer Verlag.
- Moed, H.F., Burger, W.J.M., Frankfort, J.G. and Van Raan, A.F.J. (1985) 'The use of bibliometric data for the measurement of university research performance', *Research Policy* 14: 131–149.
- Moravcsik, M.J. (1973) 'Measures of scientific growth', *Research Policy* 2: 266–275.
- Persson, O., Glänzel, W. and Danell, R. (2004) 'Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies', *Scientometrics* 60(3): 421–432.
- Sandström, U. and Sandström, E. (2007) 'A Metrics for academic science applied to Australian Universities', <http://eprints.rclis.org/archive/00011776/>, (accessed 7 October 2008).
- Sandström, U. and Sandström, E. (2008) 'Resurser för citeringar', Högskoleverkets rapportserie 2008:18 R, Utgiven av Högskoleverket 2008: <http://forskningspolitik.se/DataFile.asp?FileID=15>, (accessed 7 October 2008).
- Sivertsen, G. (2006) 'A bibliometric model for performance based budgeting of research institutions', in K. Debackere and W. Glänzel (eds.) *Book of abstracts, 9th international science and technology indicators conference*, 7–9 September 2006, Leuven, Belgium: Katholieke Universiteit, pp. 133–135.
- Sivertsen, G. (2008) 'Experiences with a bibliometric model for performance based funding of research institutions', in J. Gorias and E. Schiebel (eds.) *Book of abstracts, 10th international science and technology indicators conference*, 17–20 September 2008, Austria: University of Vienna, pp. 126–128.
- Smith, L. (1981) 'Citation analysis', *Library Trends* 30: 83–106.
- van Raan, A.J.F. (1996) 'Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises', *Scientometrics* 36: 397–420.
- van Raan, A.J.F. (2000) 'The Pandora's Box of Citation Analysis: Measuring Scientific Excellence – The Last Evil?', in B. Cronin and H. B. Atkins (eds.) *The Web of Knowledge*, Medford, NJ: Information Today Inc, pp. 301–319.
- van Raan, A.F.J. (2005) 'Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods', *Scientometrics* 62(1): 133–143.

- Weingart, P. (2005) 'Impact of bibliometrics upon the science system: Inadvertent consequences?' *Scientometrics* 62(1): 117–131.
- Whitley, R. (2007) 'Changing Governance of the Public Science: The Consequences of Establishing Research Evaluation Systems for Knowledge Production in Different Countries and Scientific Fields', in R. Whitley and J. Gläser (eds.) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems, Sociology of Sciences Yearbook 26*, Dordrecht, The Netherlands: Springer, pp. 3–27.

About the author

Jesper W. Schneider is an associate professor at the Royal School of Library and Information Science in Aalborg, Denmark. He holds a PhD degree in information science, and his specialties are bibliometrics and research evaluation. His main research interest is science mapping. Some principal publications in this area can be found at www.db.dk/jws. For several years, he has been working as a bibliometric consultant for the Danish Agency for Science Technology and Innovation advising them in relation to implementing a performance-based research funding model in Denmark, as well as doing *ad hoc* bibliometric analyses.