European Commission

MLE on Performance-based Research Funding Systems (PRFS)

PRFS Design: Policies and Ambitions

Thematic Report No 1

[Month of the report– 20XX]

Research and Innovation

PRFS Design – MLE on Performance-based Research Funding Systems

European Commission
Directorate-General for Research and Innovation
Directorate A – Policy Development and Coordination
Unit A4 – Analysis and monitoring of national research and innovation policies
Contact Marta Truco Calbet
E-mail marta.truco-calbet@*ec.europa.eu*
RTD-PUBLICATIONS@ec.europa.eu
European Commission
B-1049 Brussels

Manuscript completed in March 2017.

MLE on Performance-based Research Funding Systems (PRFS)

PFRS Design: Policies and ambitions

Thematic Report No 1

Erik Arnold

Directorate-General for Research and Innovation

2017 EN

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

In Europe, the general way to fund state universities in the post-War period was through a single block grant – what the OECD called the General University Fund. As university systems grew, funders tended to make an explicit distinction between funding for education and funding for research. In each case, this is 'institutional funding,' in the sense that the money is passed from the responsible ministry to the institution in one block. The next step in many countries has been to make some of the institutional funding dependent upon past performance. Such systems are known as performance-based research funding systems (PRFS).

In addition, universities generally get 'external' funding, in the sense of money for research or education, directly from industry, from research councils and other funding organisations within the state or from abroad, for example from the EU Framework Programme.

This report discusses institutional funding for research and does not deal with institutional funding for higher education. While in a minority of cases, Performance-based Research Funding Systems (PRFS) govern part of the institutional funding of research institutes as well as universities, in this paper we only address universities.

A PRFS has two components. First, it has an assessment process, which judges research output based on its scientific quality and increasingly also other criteria. The results of the assessment feed into the second component, a funding formula. This is an algorithm for allocating the performance-based institutional funding for research among the universities. Common to these formulae is that they move money away from the universities that have obtained low ratings in the assessment process and towards those universities that have done better.

2 THE SYSTEMIC CONTEXT OF PRFS

It has become normal in Europe to have a 'dual support' system for research funding. Under such a system, money for university research comes partly as 'institutional funding', which the universities can spend as they see fit, and partly from external, competitively won project-based funding, which in principle has to be spent on the subject matter of the projects. In most European countries, university autonomy means that no matter what formulae, processes of negotiation or incentives are built into institutional funding, once that money is in the hands of the universities they themselves decide how to spend it.[1]

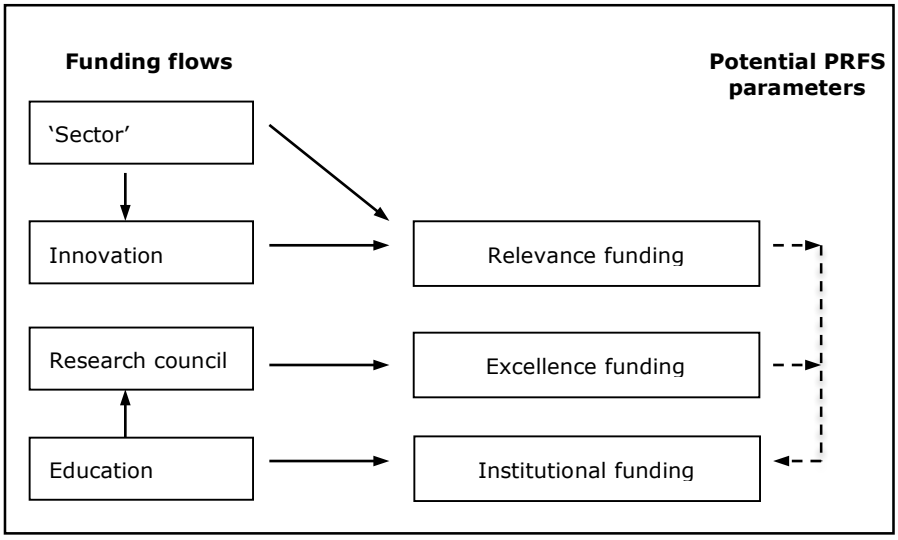
Throughout the twentieth century, in most countries research councils initially dominated the external research-funding stream, responding to investigator-initiated ('bottom up') proposals. This imposed quality control through peer review of project proposals but did not involve overtly directing research activities towards particular themes.[2] As a consequence of the OECD's work to promote 'science policy' in the form of a linkage between national (especially industrial) and scientific priorities, a new set of institutions ('innovation agencies') developed in many countries from the late-1960s that programmatically funded 'relevant' research. The innovation agencies thus generated 'focusing devices' (Rosenberg, 1976) (Arnold, Good, & Segerpalm, 2008) in the form of projects and programmes, implementing science policy through research-funding incentives.

Government university funding at national level today tends to comprise the blocks shown in Figure 1. An education ministry typically provides both institutional funding and money for 'excellence' research, the latter through a research council. In principle, institutional funding is infrastructural in nature: it provides a basis for strategy and planning and for capacity to do research in the universities. Innovation agencies are normally funded by an industry ministry. They are normally complemented by the research funding activities of ministries with other mission responsibilities, such as transport, environment, health and defence. Whether these ministries buy research directly or via their own funding agency, they – like the innovation agencies – provide incentives to do research that is relevant to specific societal roles or missions. (They also often use captive government labs or contracts with industry to do research.) Business provides a further funding stream for university research, usually offering resources for industrially relevant work. The external funders thus provide incentives for particular kinds of research. The balance among the various flows would be expected to influence the overall shape of the university research system.

1 This is a truth with modification. University systems are in different stages of progress towards full autonomy. UK universities' charters have always meant that they are supposed to be outside the control of the government. In other cases, some of the budget may be 'hypothecated', in the sense that the university has to spend it on particular things, such as academics' salaries. But the trend towards full autonomy is clear and strong

2 The allocation of budget among thematic or disciplinary research councils, of course, did generate some degree of thematic steering but these research councils tended not to programme their resources, largely responding to 'bottom-up' proposals and prioritising among them on the sole criterion of quality.

Figure 1 Stylised national university research funding system



Many PRFS now use the amount of external research funding from various external sources as quality indicators, so they can be used to magnify the effects of external funding – typically in the direction of 'excellence' or 'relevance' or internationalisation via participation in the EU Framework Programme of research and technological development.

While the distinction between institutional and project-based funding in a dual support system is conceptually tidy, it has to tackle some messy realities. For example, the traditional UK view was that institutional funding for research should finance the internal infrastructure of people and facilities needed for the university to play 'host' to externally funded projects and to some extent to do internally-funded research. The university was to provide a 'well-found laboratory'. The Research Councils would provide additional money to pay the variable costs of competitively-awarded research projects such as research assistance, PhD, post-doctoral or other temporary research staff, equipment and consumables but expected the institutional funding to cover the fixed costs, including the buildings and the salaries of permanently-employed people, such as professors. Hence, the Research Councils did not pay 'overheads' to the universities but only the marginal costs of research. Non-academic funders such as industry were expected to pay the marginal costs, plus an 'overhead' contribution to fixed costs.

From 2005, a principle of 'full economic costing' of research was introduced across UK higher education and this principle is also spreading internationally. This required universities to calculate the total costs of any project they won competitively. The principle was retained that the Research Councils should get a discount (typically 20%) because the state had already provided institutional funding for research while others who had not made a contribution had to pay the full amount. It therefore becomes important that the distribution of institutional funding for research among universities is not too dissimilar from the distribution of Research Council income, otherwise at least some universities will have to find ways to cross-subsidise Research Council work, for example from the teaching budget.

Historically, the trend has been for the ratio of external, competitive project funding to institutional funding to rise (Lepori, et al., 2007), suggesting increased competitive pressure for research resources and increased shaping of the research agenda and the research-performing institutions themselves by external forces. It also places increased pressure on universities' institutional research funding, which needs to support a growing volume of externally-funded research.

PRFS are adopted in order to pursue policy objectives. These in turn are related to the national context in terms of the problems that the PRFS is expected to address, the structure and performance of the research ecosystem as well as the range of other policy instruments used. Clearly, the mechanisms of the PRFS used need to be consistent with the wider set of incentives and to avoid – either on its own or in combination with other policy instruments – creating perverse incentives. For example, there is currently concern in Sweden that the focus of the performance-based funding system leads the educational dimension of university performance to be neglected. In the UK, the higher education funding councils propose to introduce a 'Teaching Excellence Framework' to counterbalance the negative effects of the REF and its predecessors on the status of the universities and their focus on teaching.

Another systemic issue is the need to ensure universities have sufficient institutional funding to be able to pay the non-funded costs associated with winning external funding, since this almost never covers its entire cost. This is particularly acute in systems that have a high ratio of external to institutional research funding, such as the UK. In the UK, there is a strong correlation between performance in the REF and research council grant funding, so the system is to a degree self-correcting (though this also gives rise to questions about whether it is necessary to have two parallel systems that more or less produce the same result). From the early 1990s, HEFCE introduced a second stream of quasi-institutional funding to compensate universities for the fact that neither industry nor research foundations tend to be willing to pay the full costs of research.

3 WHY DO POLICYMAKERS ADOPT PRFS?

Earlier studies show that four main categories of policy objectives lie behind the use of research assessment and PRFS (OECD, 2010) (Mahieu & Arnold, 2015), namely:

- to enhance the quality of research and the country's research competitiveness
- to steer behaviour in order to tackle specific failures in the research system
- to strengthen accountability
- to provide strategic information for research strategy at institutional and/or national level

Table 1 shows the policy motivations involved in more detail for a set of eleven countries.

National research assessment is not always linked to funding distribution. Countries that aim predominantly to allocate resources based on past performance use a PRFS. Others (such as Australia and the Netherlands) focus on informing research policies and institutional strategies but base the allocation of institutional funding on performance agreements between the universities and the responsible agency or ministry. An example of such an evaluation system is the Standard Evaluation Protocol in the Netherlands.

Countries that use PRFS generally do so in an effort to increase the quality of research. Most also aim to trigger other behaviours, in line with policy priorities or a perceived need for change in the national research system. Objectives include: fostering critical mass; enhancing research-industry collaboration and the knowledge transfer; identifying or directing funding toward areas of research strength and emerging areas of research excellence; and strengthening the international competitiveness of research. (NZ Ministry of Education, 2012). Some also seek more accountability-related objectives, to stimulate efficiency in research activity and to demonstrate that investment in research is effective and delivers public benefits (Abramo, D'Angelo, & di Costa, 2011).

Table 1 Main PRFS policy objectives in eleven countries

	Quality of research	Systemic factors	Accountability	Strategic intelligence
Australia	Identify excellence across the full spectrum of research performance	Enable comparisons of research, nationally and internationally, for all discipline areas	Give government, industry, business and the wider community assurance of the excellence of research conducted in higher education institutions	Provide a national stocktake of discipline-level areas of research to strengthen institutions Identify emerging research areas and opportunities for development
Belgium /Flanders (BOF) (IOF)	Stimulate scientific performance and the quality of research (BOF)	Create an incentive for technology transfer (IOF)	Make quality visible	
Czech Republic	Reward research quality		Make quality visible to the national and international public	
Denmark	Improve quality, increase productivity and enhance efficiency in HE research	Sharpen international profile and international competitiveness of HEI		
Finland	Strengthen research quality	Sharpen international profile and international competitiveness of HEI	Demonstrate to the public that research funding is spent optimally	Assist the institutions to fulfil strategic goals and priorities
Italy (VQR, 2011)	Provide an objective and rigorous assessment of research	Define a national ranking per scientific area and 'structure' typology		Allow for a comparison of the national research quality with the quality in the major industrialised countries

	Quality of research	Systemic factors	Accountability	Strategic intelligence
Netherlands (2015)	Reveal and confirm the quality of research		Reveal and confirm the relevance of the research to society	Improve quality and relevance of research where necessary
New Zealand (2013)	Increase the quality of basic and applied research Support world-leading research-led teaching and learning		Provide robust public information to stakeholders about research performance within and across tertiary education organisations	Assist New Zealand's tertiary education organisations to maintain and lift their competitive research rankings relative to their int. peers
Norway	Strengthen research capacity	Enhance co-operation among research actors and knowledge transfer		
Sweden	Assess the quality of research	Stimulate HEI's to find a profile where they have a competitive advantage, which will help a clearer division of roles between HEI's and increased specialisation		
UK (REF, 2014)	Reward research excellence		Produce evidence of the benefits of public investment in research	Provide benchmarking information and establish reputational yardsticks, for use within the higher education sector and for public information

Sources: OECD, 2010; Mahieu & Arnold, 2015

While rewarding of performance is an intrinsic characteristic of all PRFS, only in a few countries does this selective distribution of funding deliberately aim also at a concentration of resources. This is the case of the United Kingdom where greater selectivity in funding allocation was initially an explicit goal of the Research Selectivity Exercise. It constituted a government response to limited resources and the increasing costs of research. The goal was to maintain research excellence but in fewer places (OECD, 2010). Sweden has also been looking to concentrate resources through a revision of the national PRFS, trying to foster more peaks of excellence in a system whose average level of quality is already quite high. However, in 2016, the government decided that it would not approve the revised system. In countries such as Italy, with a regionally centred policy for universities, concentration is undesirable: these countries need to use the PRFS to develop the quality of work within each university.

Other countries, instead, aim to strengthen research capacity in the weaker parts of the system rather than focusing money on 'excellent' researchers or institutions. The original purpose of the performance-based allocation system in Norway, for example, was to enhance the quality of research by motivating institutions to increase their research activities and by distributing resources according to research results. The main winners have been the provincial institutions (Aagaard, Bloch, & Schneider, 2015). Belgium (Flanders) established its performance-based funding mechanism with a clear intent to distribute research funding on a wide basis.

The strong research performance of The Netherlands and Switzerland in bibliometric terms, underlines that at least some goals of PRFS can be reached by other means. The Netherlands has a transparent national system for assessment of research at individual universities, which appears to exert a strong social influence over performance. The visibility and transparency of research output via the Croatian national RIS is thought to be a key factor in increasing the productivity of the research community in terms of numbers of papers produced. The reasons for the strong Swiss performance are less clear, but seem likely to include the governance

and culture of the leading universities, a willingness to import academic labour on a large scale and proximity to high-performing reach-based industry in a number of branches of industry.

Running a PRFS can be a very expensive enterprise. Technopolis has estimated that, including all the time the universities spent on preparing their submission strategies, impact statements and returns to the REF in 2014, the total cost of the exercise was £246m (Farla & Simmonds, 2015). Many countries have justified their decision to use metrics because of the perceived expense of a peer-review based system.

4 WHAT DO WE KNOW ABOUT THE EFFECTS OF PRFS?

There is little evaluation literature relating to PRFS. What there is suggests positive effects of the amount and quality of research output. However, these benefits are not costless. The 2010 OECD report on performance based research funding provides the following list of dimensions of research on which PRFS appear to have a negative influence:

- Interdisciplinarity
- 'Blue skies', 'high risk' and 'transformative research'
- Research on the periphery, or non-conventional research
- Applied research
- Researcher autonomy
- Collaboration among researchers (OECD, 2010)

4.1 … on the performance of research systems

Internationally, studies tend to associate PRFS with increased production and quality (measured as citations) (Moed, 2005) (Butler, 2003) (Jiminez-Contreras, Anegon, & Lopez-Cozar, 2003) (Sivertsen, 2010) (Adams & Gurney, 2014) (Smart, 2013). UK and Australian experience shows that rule changes to incentivise production over quality or vice versa lead to corresponding behaviour changes (Butler, 2010) (Moed, 2005).

Wang and Hicks (2013) compared trends in publication and Higher Education Expenditure on R&D (HERD) in a handful of countries, searching for structural shifts in publication output. They found one in the UK, associated with the second RSE (consistent with Martin and Whitley's (2010) assertion that the universities did not take the first exercise seriously) but no further structural shifts in the UK thereafter. They also found a discontinuity in Australia but date it to a point before the introduction of the Australian PRFS, when the university system was expanded and the universities were required to report their publications as part of a process of increasing monitoring and quality control. Other discontinuities link to changes in funding or system size, except for one in Germany that appears to be associated with a 1993 law allowing universities to make academics' pay dependent upon performance. In the Czech Republic, Vanecek (2013) links a discontinuous increase in publication volume to the introduction of a national evaluation system in 2004 and a context of increasing research funding, rather than to the start of the PRFS in 2008. It seems that PRFS can contribute to structural changes as part of larger processes of output-orientation rather than necessarily being the unique cause of such changes.

The JRC has recently produced two overviews of PRFS in EU member states (Jonkers & Zacharewicz, 2015) (Jonkers & Zacharewicz, 2016). The authors observe that almost all the countries considered have been improving their research performance measures in bibliometric terms. Thus, while all those that have adopted PRFS have improved their performance, those that have not adopted such systems have also improved, so there is not a clear relationship between PRFS use and performance. Other factors such as increasing investment in higher education research seem likely to have a positive influence but there is also no simple relationship between these and performance. It is possible that PRFS contribute to improved performance, therefore but it has not been possible so far to identify their net effect.

An evaluation of the Research Assessment Exercise (now REF) in the UK focused on the institutional views of the universities, which generally felt that the exercise had improved productivity and quality but had few negative effects (PREST, 2000). It was not able to measure effects in quantitative terms. However, the citation impact of UK research has

increased continually since the end of the 1980s, after the second research assessment exercise (Adams & Gurney, 2014).

Members of the MLE exercise pointed out a number of systemic effects from their own experience:

- In Italy, where some signs of increased convergence in the country in terms of the North/South divide were visible
- In Norway, where effects on publications were visible especially for the applied research universities
- In Portugal, where the evaluation system (i.e. the units for evaluation) has created a dynamic in how the HE system organises itself
- In Estonia, where PRFS are used as a threshold for institutional funding, no matter the legal form of the organisation (so: eligibility criterion)

Other effects mentioned were:

- In Portugal, an effect on education was created thanks to the introduction of PhD programs as one of the PRFS indicators
- In Norway, pushing colleges to publish was positive for young researchers
- In Italy, effects on quantity were marginal and they were hoping for effects on quality; there were no negative effects on young researchers. The effect on female researchers is not yet clear

4.2 … on research

Interdisciplinarity is viewed as important in research and research funding. The perception is that work at the boundaries between disciplines can be the birthplace of new disciplines and sub-disciplines. Work of industrial and social relevance may often be interdisciplinary because there are seldom mono-disciplinary answers to industrial or social problems. Research councils therefore tend to promote interdisciplinarity and to implement special procedures or structures into their peer review processes in order to do so. Innovation agencies tend to promote interdisciplinary work while applied industrial research institutes internationally tended to form polytechnic structures in recent decades as a direct response to their perception that their customers need multi- and inter-disciplinary solutions (Arnold, Barker, & Slipersæter, 2010). The MIT Media Lab even advocates 'antidisciplinary research'.

The general literature on interdisciplinarity suggests that it may be disadvantaged in peer review processes but there is little analysis specific to PRFS. An effect has been identified in the UK where cautious university managers are less likely to submit interdisciplinary than disciplinary work to the REF, reducing the incentives for researchers to do interdisciplinary work (Elsevier, 2015). There are examples of PRFS-driven focus on high-status journals marginalising interdisciplinary or heterodox forms of work (Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012) (Lee F. , 2007).

Transformative research is thought to involve higher risk and longer timescales than more conventional research. Transformative research tends to struggle in peer review, be that for grant funding or for journal publication (Wessely, 1998) (Horrobin, 1996) (Roy, 1985) (Lakatos & Musgrave, 1970) (National Science Board, 2007) (Arnold, et al., 2013) (Luukkonen, Stampfer, & Strassnig, 2015). This is often due to risk-averseness (Chubin & Hackett, 1990) (Wagner & Alexander, 2013), (Hävrynen, 2007) (Luukkonen, 2012). So the likely effects of PRFS on discouraging transformative research tend to lie in the way researchers and managers value and manage risk, rather than in the PRFS mechanisms themselves.

In non-Anglophone countries, there is a tendency to push research into Anglophone publication channels as these are the ones best covered by the bibliometric databases (Scopus and Web of Science), often at the expense of nationally or locally relevant research and cementing a peripheral status for non-Anglophone countries (Stöckelova, 2012) (Meriläinen, Tienari, Thomas, & Davies, 2008). Lee and Harley (1998) show that the RAE has marginalised heterodox forms of economics.

There is no clear evidence about whether the potential for PRFS to discourage applied research (OECD, 2010) is actually realised. To the extent that it is often reported partly outside the indexed journals, this risk may be more significant in metrics- than peer review-based systems.

PRFS are sometimes argued to discourage collaboration. There are certainly some success stories. Both for Norway (Bloch & Schneider, 2016) and Morocco (Bouabid, 2014) there is evidence that the introduction of metrics-led research assessment has led to higher levels of collaboration, specifically international co-publication in the case of Morocco.

The OECD notes that the principal choice for PRFS designers in the area of co-authorship is between giving each institution or author full credit for outputs to which they contribute, or only to give fractional credit (OECD, 2010). The former is intuitively more conducive to collaboration.

> *Concerns about the effect of fractional counts on collaborative activities led Australian governments to reject this methodology in favour of whole counts. Norwegians appear less concerned and believe their use of fractional counts has not resulted in a decline in collaborative activities (Schneider, 2009, p. 372). Schneider believes that "'invisible colleges' and social networks within research specialties have eventually ensured collaboration". It is believed that the dependence of research on collaboration will counteract any adverse behaviour that might result from the funding model. (OECD, 2010)*

Studies in New Zealand (Edgar & Geare, 2010) and the UK (Henkel, 1999) point to the tension between the tendency in PRFS to encourage competition among individual researchers and the benefits of collegiality in research. This tends to confirm the idea that any effects on collaboration are determined by the detail of how particular PRFS are designed.

4.3 … on research careers

A clear effect of PRFS is an 'output imperative' (Henkel, 1999). In the UK, researchers' understanding is that successful publication rather than conducting research, is the objective of the job (Bence & Oppenheim, 2005). Moreover, as the outputs need be of a certain scientific quality, they need to appear in very particular publication channels, especially in high-impact factor journals, which have increased significantly as a proportion of overall research outputs both in the UK and elsewhere (OECD, 2010). The importance of scientific publication in PRFS appears to be a disincentive to popularisation (Elton, 2000).

The influence of PRFS on behaviour appears to be strongly mediated by the way researcher careers are managed and, correspondingly, the extent to which researchers conform to the demands of university Human Resource (HR) management. Thus, almost half of department heads in Norway make use of the Norwegian publication indicator (used in the Norwegian PRFS) for recruitment and promotion purposes, and almost 90% use it for monitoring departmental activity (Aagaard, Bloch, & Schneider, 2015). Institutional incentive systems linked to the national assessment exercise are also used in Italy (Abramo, D'Angelo, & Di Costa, 2011). In the Spanish system, performance in the *sexenios* influences the award of tenure and eligibility to become part of the panel that grants tenure as well as competitive grant funding and has been shown to drive individual performance (Jiminez-Contreras, Anegon, & Lopez-Cozar, 2003). A change in German law to allow performance to drive salary, had a similar effect. Evaluations consistently identify changes in research management

as an immediate effect of PRFS (Butler, 2003; Martin & Whitley, 2010) and these in turn affect staff appraisal and career development. Other studies demonstrate the link between individuals' REF performance and the extent to which they continue to be allowed to do research (Sikes, 2006) and at the upper end of the seniority ladder that considerations based on the REF have resulted in forced early retirements (Bence & Oppenheim, 2005).

5 DESIGN OPTIONS FOR PRFS

The methodologies adopted for research assessment differ, reflecting different government priorities. Policy objectives drive the approach to assessment, the number and type of indicators selected, and the relative weight placed on each indicator or assessment criterion in the construction of the final score for the allocation of the institutional funding.

Table 2 lists the design options visible in international practice with PRFS. These are discussed in more detail below. Many of these elements are interlinked. In particular, the overarching model chosen for the assessment (i.e. peer review, bibliometrics or a combination of both), drives the granularity and the periodicity of the PRFS.

Table 2 Key design parameters for the assessment component in PRFS

Key design parameter	Variations
Model used for the assessment of research quality	• Peer review-based • Informed peer review • Mix of peer review & bibliometrics • Metrics-based
Scope of research activity included	• Research • Innovation • Societal relevance
Type of indicators	• Output indicators • External funding indicators • Systemic indicators • Outcome/impact indicators
Assessment criteria in peer review-based systems	• Quality of outputs • Relevance of research activities • Institutional environment • Esteem measures
Granularity	• Units of analysis (grouping of scientific disciplines) • Inclusion of individual staff (inclusive/exclusive)
Periodicity	• Annual • Longer time frames

Source: Arnold, et al., forthcoming 2017

5.1 Overall 'model' used for the assessment of research quality

Assessment models may be based on peer review, metrics or a combination of the two. One possibility is to use peer review 'informed' by metrics, such as bibliometrics or indicators of innovation outputs. Another is to use metrics and peer review for different parts of the assessment process. For example, Lithuanian practice is to use bibliometric indicators of the volume and overall quality of research publications but to supplement this via peer review of papers selected by the research-performing organisations to represent the highest quality peaks within the overall set of research outputs.

5.1.1 Peer review versus metrics

The choice between metrics or peer review is contentious. On the one hand, metrics-based systems typically encounter criticism from the research community on the grounds that metrics provide imperfect measures of quality. The community tends to prefer peer review, thanks to its flexibility and its ability to assess a wider range of research outputs and research-related activities. PRFS that use peer review are more comprehensive and appear to have greater credibility and buy-in, meaning that PRFS provide status as well as funding incentives (NZ Ministry of Education, 2012). The recent proposal for a new PRFS in Sweden involved a shift from metrics to peer review.

Cost, however, is a major factor. PRFS that rely entirely on metrics are generally considered to be less expensive to administer and less compliance-heavy than systems that use peer reviews, which are seen as cost-intensive and time consuming. Both peer review and metrics suffer well-known weaknesses, which we do not discuss here. However, policymakers generally fail to adopt many of the more sophisticated indicators that bibliometricians can

provide. Use of journal impact factors is widespread, despite the growing understanding that these are inappropriate as indicators of the quality or impact of individual articles.

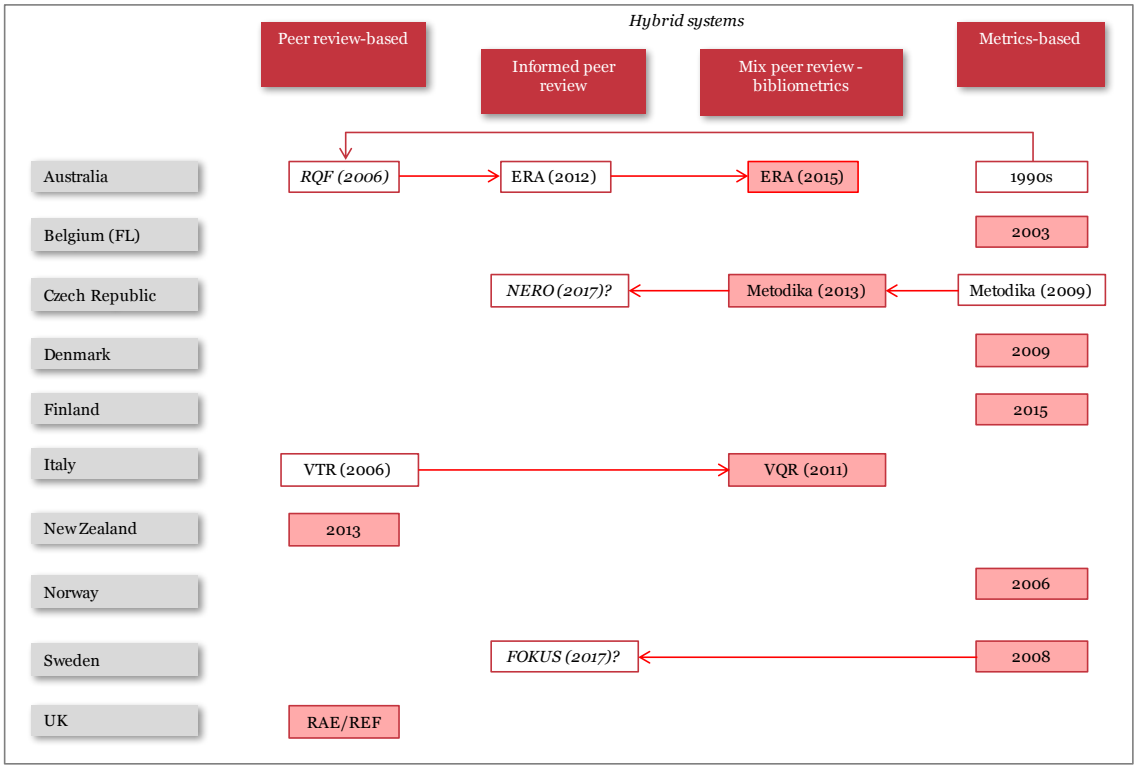
The development of altmetrics has so far left PRFS untouched. There may be interesting opportunities to use some of the approaches being considered in this field (Cronin & Sugimoto, 2014), provided it is possible to arrive at stable definitions and clear theory about what the new indicators actually mean.

5.1.2 Current international practice

PRFS are dynamic. A historical analysis of PRFS in scope to this study indicates a continuing search for improvement (Figure 2). Factors leading to change include pressure from the research community, developments in evaluation methodologies and concepts and a search for an improved cost-benefit ratio. In Italy and Sweden, a major driver was an envisaged increase in the proportion of institutional funding to be allocated based on the assessment results and a concomitant desire for more reliable assessment methods.

Currently, the UK and New Zealand are the only two countries relying close-to-uniquely on peer review. The UK REF allowed the use of informed peer review in cases where panels desired so by accessing citation counts and contextual analysis to help clarify citation behaviour and patterns in the relevant field. Use of journal impact factors and other bibliometric indicators not supplied through the REF administration was forbidden. One panel (Computer Science and Informatics) had planned to use Google Scholar data as a way to capture more of the conference activity that is central to the way that field communicates but was defeated by its inability to harvest the information needed from Google.

Figure 2 Trends in the models for research performance assessment



Source: Arnold, et al., forthcoming 2017

Notes: The Australian Research Quality Framework (RQF) that was proposed in 2006 was an RAE-like system, based on peer review panels but also including end-user assessments of impact on the economy and society. The system was criticised for being costly and non-transparent and when a new government took over in 2007, it was never implemented. Both in Sweden and the Czech Republic, the PRFS system is currently being reviewed

Belgium, Denmark, Finland and Norway use bibliometrics for the assessment of research quality in the PRFS. All of them use the "Norwegian model" for the bibliometric indicators, extending the data coverage of the international, commercial bibliometric databases (Web of Science and Scopus) by means of country-specific publication databases. It should be noted that in the Scandinavian countries, the PRFS is complemented with system-based or discipline-based national evaluations using informed peer review that have a more formative character.

In Belgium/Flanders, the BOF bibliometric indicators started in 2003. Those indicators were gradually refined, complementing the use of WoS with an endogenous Flemish Academic Bibliography. In 2004 BOF was complemented by IOF (the Industrial Research Fund) aimed at stimulating technology transfer. The IOF allocation rule also includes academic patents, academic spin-offs, competitive EU-funding (FPs) obtained, income from industrial contracts, income from clinical trials and licenses.

An increasing number of systems combine the two approaches. The 'mixed peer review-bibliometrics' model uses both bibliometrics and peer review. In Italy for example, the latest VQR (2011) used an informed peer review process for the greater part of the funding, based on outputs submitted by the research organisations as well as a self-evaluation but complements this with indicators to allocate the balance. The results of the peer review and bibliometric exercises are used separately to allocate units of assessment into broad quality bands and the combination of these bands with the volume of output then drives the funding provided. In Australia the ERA 2015 uses a broad range of assessment tools, including bibliometric and non-bibliometric indicators, as well as peer review. Crucially, these are not all used equally across all disciplines. Citation analysis is used more extensively in the sciences and peer review is used more extensively in social sciences, humanities and computing.

The 'informed peer review' model uses bibliometrics to inform the peer reviewers - to varying degrees and at the peers' discretion. This exploits the ability of indicators to represent large sets of data while exploiting the ability of peers to make more qualified judgments about excellence, coherence and other qualitative aspects that cannot be achieved through indicators alone.

5.2 Scope of research activity included

Over time, there is a clear trend to increase the scope of research assessment in the context of PRFS. While early PRFS focused only on scholarly outputs, the current trend is to encompass also aspects of innovation and the universities' 'third mission' of knowledge exchange with society (Sörlin, 2007). Increasingly, evaluations aim at assessing research performance also in terms of its impacts on research, innovation and society at large.

Most PRFS, no matter whether the assessment is peer-review or metrics-based, use indicators beyond academic outputs to measure performance. In practice, however, PRFS have not attempted a heavily metrics-based treatment of wider societal, cultural or economic impacts of research. Impact presents a significant challenge to research assessors, primarily because there often is a long time delay between publication and any social impact. Impacts are therefore predominantly assessed indirectly, i.e. by using proxy indicators such as the universities' capacity to gain external research funding (see Section 7.2.3, below). The UK REF 2014 was the first major attempt in a PRFS to demonstrate research impact in a systematic way across all disciplines (by means of narratives). Other national evaluation systems that are not linked to institutional funding have made similar attempts. An example is the Dutch Standard Evaluation Protocol where the self-evaluation component of the exercise requires the universities to provide narratives on the societal relevance of their research activities and outputs, albeit not at the rigorous and detailed level of the REF 2014. A similar approach is being attempted also in some field evaluations of the social sciences and humanities at the national level in Norway.

5.3 Indicators used

Indicators used in the context of PRFS can be grouped into three categories: indicators directly assessing research outputs; external funding indicators; and systemic indicators (Table 3). The two last categories act as proxy indicators for research quality, impact on innovation or societal relevance, and in the case of PhD recruitment/awards, (also) the size of the organisation.

The degree of emphasis given to each of these indicator classes depends on the function of the evaluation and its policy objectives.

- The 'direct' assessment of research productivity and research quality or excellence is a topic of importance in all PRFS;
- Several PRFS also directly assess the productivity, quality and value of non-academic outputs and innovation-related ones such as patents;
- Most systems use universities' ability to obtain competitive external project funding as a proxy for research quality or relevance;
- Several countries also use systemic indicators to assess the universities' broader research capacities and activities.

Table 3 Indicators used in PRFS in selected countries

		Belgium / FL (2009)	Czech Republic (2013)	Denmark (2009)	Finland (2015)	Italy (VQR, 2011)	New Zealand (2003)	Norway (2006)	Sweden (2008)	UK (REF 2014)
Output indicators	Academic outputs	√	√	√	√	√	√	√	√	√
	Non-academic outputs		√			√			√	√
	Innovation-related outputs (IPR)	√	√			√				√
External funding indicators	Competitive funding / national			√	√	√	√	√	√	√
	Competitive funding / international	√		√	√	√	√	√	√	√
	Contract research funding	√		√		√	√		√	√
	Non-competitive funding			√		√	√			√
Systemic indicators	Esteem (conferences, editorships, rewards etc.)					√				√
	Collaborations / national					√		√		√
	Collaborations / international					√				√
	International mobility				√	√				√
	Collaboration research-industry					√				√
	PhD recruitment/awarding	√		√		√	√	√		√
Outcomes/impact indicators	Academic impacts (citations)	√	√			√			√	
	Socio-economic outcomes/impacts (e.g. spin-offs)	√				√				√

Source: Arnold, et al., forthcoming 2017

In a few countries, citations are considered to provide proxies for impact on the research base. The use of indicators to assess socio-economic impacts is rare, perhaps because these are easily 'gamed' (Good, Vermeulen, Tiefenthaler, & Arnold, 2015).

In Table 3, above, the UK REF and the Italian VQR stand out for the breadth of the indicators they use, covering all indicator categories and focusing more than other PRFS on the use of systemic indicators. The UK REF pays considerable attention to differences among disciplines. For example, REF sub-panels are allowed to identify the types of academic and non-academic outputs they considered to be relevant to them. Some panels listed sixteen types of academic output beyond the traditional scientific publication categories that could be submitted, ranging from technical reports to textbooks. They also accepted the submission of nineteen types of non-academic outputs, including digital artefacts (such as software, archives, films etc.), seven types of physical artefacts (e.g. new materials or prototypes), and three types of temporary artefacts (exhibitions, performances, and 'additional' outputs). There is considerable international activity in trying to develop technometric indicators to address non-academic outputs.

5.4 Assessment criteria in peer review-based systems

Metrics-led research assessments have to accept the definitions of 'goodness' implicit in the mechanisms and systems that construct the metrics, such as the system of scientific journals, the conventions they use for citation and the indexing practices of the companies that maintain the commercial citations' databases. Peer review requires an explicit transition from expert observation to numerical grading or ranking. The criteria used in the small number of countries employing peer review are quite similar (Table 4).

Table 4 Criteria used in peer review based assessment frameworks

	Australia – ERA 2015	Netherlands – SEP 2015	Italy – VQR 2011	UK REF 2014
Outputs	Volume and activity; publishing profile; peer review; citations; research income	Research quality	Originality & innovation	Originality, significance and rigour
Relevance/ impact	Applied measures (IPR & research commercialisation)	Relevance to society	Relevance for the advancement of knowledge & social benefits	Reach and significance
			Technology transfer activities and (potential) socio-economic fallouts	
Environment		Viability		Vitality and sustainability
Esteem	Esteem measures (at eligible researcher level)		Internationalisation and/or international standing	

Source: Arnold, et al., forthcoming 2017

5.5 Granularity

Unit of analysis

The unit of analysis in an evaluation can be the individual researcher, a research group (field defined), the faculty, the department, or the institution. It is a fundamental component in the design of research assessment systems. The selection of the most appropriate unit of evaluation depends on the purpose of the research assessment: in assessments that do not drive funding, the deciding factor is the most suitable level for the collection of the information that is required for policy making or governance; in PRFS it is driven by the level at which the funding is allocated.

In practice, the tension between complexity and practicality means that while research groups are theoretically the ideal unit of evaluation, departments or universities are usually the focus of PRFS (OECD, 2010). This not only reflects the reality of research production but allows the treatment of non-disciplinary activities such as the development of Key Emerging Technologies.

In general, the methods used in assessment tend to correlate with the choice of focus: peer review is used for departmental or research group assessment, while metrics-based systems are used for university-level assessment. Metrics and peer review have different qualities.

Metrics-only systems can collect data about outputs at the level of the individual researcher. To our knowledge, nobody allocates institutional funding at this individual level. Normally, outputs are aggregated to the *organisational level* and used to determine the institutional funding for the research organisation as a whole. It is possible to aggregate results also to the level of individual groups or faculties (and some research organisations appear to run shadow systems in order to do this). For the external funders directly to allocate institutional funding to intra-organisational entities would challenge the principle of autonomy, so this is not done. Allocation practices within universities vary and are generally not transparent to outsiders. A rare insight is provided by a study of internal allocation patterns in Swedish universities, which reveals diversity of behaviour, not all of which is consistent with attaining the goals of the PRFS (Fridholm & Melin, 2012).

Peer review systems cannot easily achieve this same flexibility, because they need to be implemented using discipline- or domain-based panels. It is considerably less complex for such a panel to assess at the research group level than at the individual level, although the latter is possible and is done in New Zealand[3]. Assessment at the *departmental or discipline level* is most common under peer-review systems (NZ Ministry of Education, 2012).

The inclusion of individual staff

There are two approaches to the way individual researchers' work is included in the assessment exercise:

- The assessment may be comprehensive, including all researchers at an institution. Systems that take this approach typically specify clear inclusion criteria, including most often that a researcher works for a minimum proportion of their time at the university. This approach allows a relatively representative overview of the outputs, quality and/or impact of research within the unit of analysis. This approach is taken in Australia, Finland, Italy and New Zealand
- In the UK RAE/REF, the universities are expected to identify a smaller selection of their 'best' researchers who will then submit their work. This reduces the burden on the evaluators, as the overall amount of submitted work is smaller. This approach does not give a representative view of all research activity that has occurred in an evaluated unit but instead indicates the maximum standard of which the unit is capable, in the view of those who prepare the submissions

There are weaknesses to both approaches: comprehensive inclusion of staff may, for instance, obscure the presence of a select few outstanding researchers in a unit whose level is otherwise ordinary, whilst selection of the best examples may obscure that relatively poor quality research is also being done.

In addition, normalising publication output by staff poses various measurement challenges (e.g. Glänzel et al., Journal of Informetrics, 2016 (10): 658-660).

[3] It is also done in specialised systems that assess individual researchers as part of a promotion or award process (as in current Spanish and German systems)

Even though the assessment of individuals is not an objective in PRFS, the inclusion of results at the individual level has effects on career prospects and the R&D system as such. There are similar effects where universities can reproduce the assessment results at the individual level, for example by running their own bibliometric analyses. Universities can also consult a national research information system such as CRIStin in Norway, which records individual outputs and assigns them to various quality categories based on the publication channel used.

Handling the differences among the scientific disciplines

Differences among fields present a major challenge for research assessment. Two models emerge in international practice:

- Some bibliometric-based systems try to overcome field differences in publication patterns by introducing a system of weights, field normalisations or field-independent indicators. These balance the differences in publication behaviour and subsequent patterns by presenting scholarly publications with complete data from research information systems, in order to compensate for the differences in the coverage of scientific fields in the commercial data sources, Web of Science and Scopus. Putting different fields into competition within metrics-based approaches means that the designer of the assessment has to produce a bibliometrics-based technique for comparing across fields without, in reality, having a deep understanding of what these inter-field differences are
- Peer review-based systems like the UK RAE/REF solve the problem by using discipline-based panels and units of assessment. This generally means that different fields are not put into direct competition with each other at the level of assessment, though the funding formulae may do so. Mechanisms are put in place which aim to ensure that peer review panels all use assessment scales that have the same meanings, irrespective of discipline

A topic closely linked to the choice of method is the ability to suitably assess and reward interdisciplinary research. Other assessment systems – like proposal assessment systems in research councils – struggle with interdisciplinary research. Peer panels may have to cooperate to address it. Indicator-based approaches do not have a good frame of reference for interdisciplinarity, precisely because the statistical frames of reference they use are defined within disciplines.

5.6 Periodicity

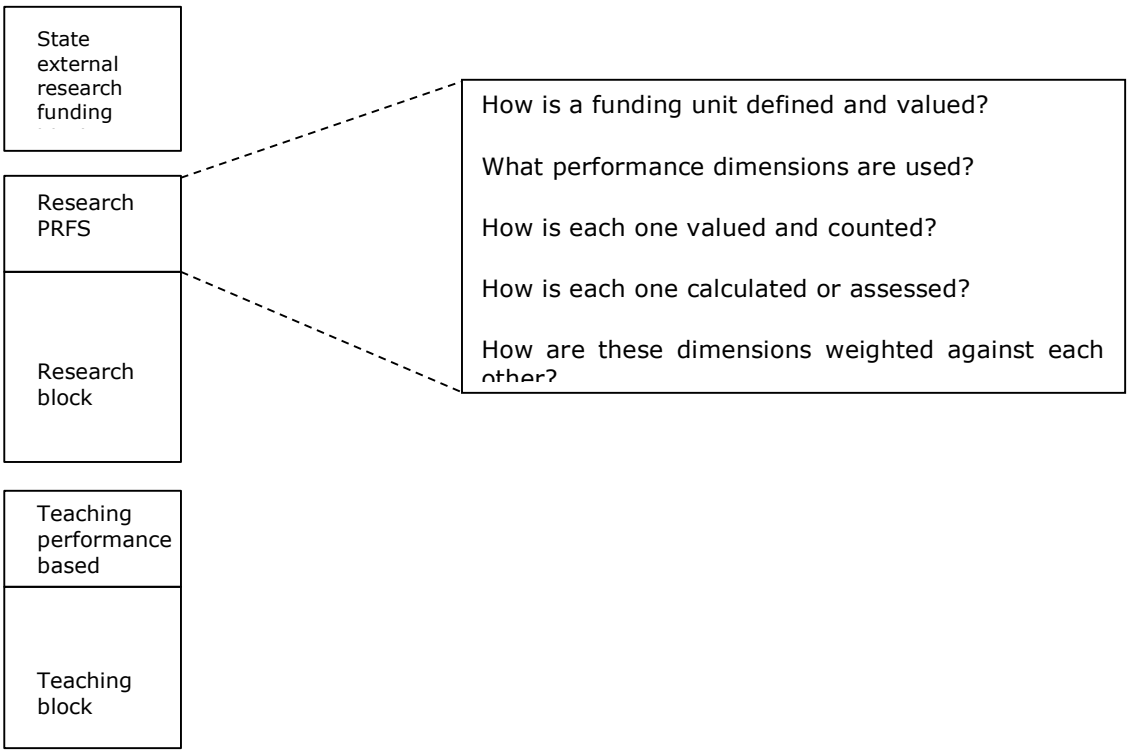
The frequency with which research activities are assessed tends to be influenced by system design and cost considerations. Collecting quantitative information is generally easier than peer-review evaluation, so information measured by metrics is collected more frequently – in contrast to peer review-based systems which are more resource- and time-intensive and therefore can only be conducted occasionally.

As a consequence, the information upon which PRFS are based is more up-to-date in metrics-based systems. The shorter time lag between evaluation and funding allows these systems to be more responsive to changes in policy objectives and in the research system at large. In countries with PRFS that have a more extended interval such as the UK REF, policymakers can only periodically understand the relative performance of the research organisations (in practice, the universities run internal systems using bibliometrics and sometimes peer review in order to obtain this information, and to devise their tactics for submitting work to the REF).

5.7 Funding formulae

Like any other policy intervention, the way and extent to which a PRFS is effective depends both upon its internal design and upon its systemic context. Figure 3 shows the main 'hard' elements that are likely to be of relevance ('soft' factors will also be influential, such as the culture in the universities, their governance, the amount of academic salaries and the way academics are promoted).

Figure 3 Funding system components



Describing a PRFS properly involves quantifying all the individual elements shown in the vertical column. That exposes its potential importance, relative to other incentives.

The right-hand box attempts to capture the generic building-blocks of the PRFS itself. Normally, the PRFS distributes a specific amount of money (budget) allocated by the government. To do this, the PRFS needs to calculate a basic unit or 'funding unit' into which the money is divided. The money will be allocated by multiplying the value of this funding unit by various performance indicators, which may be based on metrics or on scores allocated by peer reviewers. Describing the funding formula therefore also involves identifying one or more performance dimensions and deciding how to count performance along each of them. Individual dimensions may be composites: for example, an overall publication indicator may break down into a sub-system for assessing scientific publications, another way to handle other literature such as reports and a third way to count the extent to which researchers also produce popular science publications. Each of these sub-dimensions needs to be defined and explained. The calculation in which they are combined together into the higher-level indicator needs to be explained. Finally, the calculation that combines the performance elements together and translates them into funding should be described.

The level of detail countries were able to provide about the funding formulae they use was insufficient for a comparison to be made at the early stage of the MLE. The team will return to this and obtain more information so that comparisons can be made within the mutual learning exercise.

6 PRFS IN THE COUNTRIES PARTICIPATING IN THE MLE

6.1 Overview of background information provided by participating countries

Participating countries have supplied information about their PRFS. This is summarised in Appendix A and discussed here.

Armenia, Croatia and the Czech Republic use PRFS that cover Public Research Organisations (PROs) as well as universities. Norway maintains a separate system for the PRO sector, using similar indicators to the university system, thus preventing a reallocation of resources between the university and PRO systems.

Seven out of ten countries mentioned the need for increased research quality as a reason to introduce PRFS. Croatia and the Czech Republic focused more on transparency, though transparency and accountability were important in many places. Estonia introduced institutional funding for research only in 2005, using PRFS to allocate it from the start. Norway introduced PRFS as one component in a wide-ranging set of reforms intended to improve the quality of both higher education and research.

In many cases, structural reform was one of the aims for the PRFS. The type of reforms desired range from a significant reduction in fragmentation (often in countries with Academy systems), to a preference for existing universities to specialise more and build scale in high-quality areas. This was aligned with an aim for universities to develop themselves by implementing clearer strategies. Increasing university autonomy also underpinned the need to provide external incentives, rather than instructions.

Armenia, Estonia, Norway and Sweden all have a national tradition of field evaluations, alongside the PRFS. One of the ambitions of a recent project to develop new PRFS in the Czech Republic was to be able to combine PRFS' assessments at research group level in order to be able to tackle both institutional and field evaluations. However, the proposed new system has not been implemented. The Czech Academy of Sciences evaluates its' institutes every five years using international peers but there is no equivalent system for the universities or for individual fields of research. The Austrian system requires universities to do internal evaluations while the Swedish government has in recent years provided start-up funding to encourage Swedish universities to do the same.

The majority of the PRFS considered are metric or use only a small amount of peer review and are conducted annually. Armenia and the Czech Republic use small amounts of peer review. Almost all the PRFS' designs have been subject to consultation with the research-performing organisations. Italy and Portugal use peer review-focused systems at intervals of several years and consult with the academic community ahead of each exercise.

Not all the countries had looked at international practice before designing their PRFS. The Nordic countries and the UK were most frequently mentioned by those countries that had done so. Italy had looked only at the UK but its PRFS was designed at a point where there was little other experience that could be observed. PRFS were usually designed by the responsible ministry or a committee established by it. Croatia had looked to the Institute of Economics in Zagreb for support in designing its funding formula. The Czech Republic's designs have been generated by committees appointed by the Ministry of Education, Youth and Sport or the RDI Council but did use a consortium of Technopolis, NIFU and the Technology Centre of the Academy of Sciences to design an alternative – which has not been adopted, although it has influenced the latest redesign of the system.

Most PRFS use the university as the unit of analysis. The metrics systems aggregate data about individual researchers. The Italian system feeds ratings back to the individual researchers but does not publish results below the level of departments.

Peer review systems use small numbers of criteria, focusing on scientific quality and impact but in Armenia and Portugal also considering research capacity in the institutions being assessed. Many of the metrics systems use data from the Web of Science (WoS) and Scopus but several countries also maintain a national research information system (RIS), intended to provide a complete listing of research outputs. The Norwegian system differentiates between two levels of journal quality, since publishing in more prestigious journals generates more points in the PRFS. Austria, Croatia and Portugal use self-reported data from the universities.

Peer review systems have to handle field differences by setting up a number of specialised field panels. Practices differ among countries using bibliometric indicators - some use field-normalised values; others do not. Norway has recently moved from using un-normalised values to normalising.

Countries using peer review see advantages in doing so. Those using metrics similarly regard their own way of doing things as successful. In either case, countries reported increases in quality (except in Norway, where the system increased production but not quality). Three of the countries are working actively to develop impact indicators, both qualitative and quantitative.

The majority of PRFS did not contain a prospective element. In Armenia, the system assesses past and proposed (future) work at the same time, so in that sense there is a prospective element. The Italian system looks at present and future research capacity. Austria's performance contracts are inherently prospective; Norway is starting to experiment with combining performance contracts with a PRFS (as is done in Austria).

The questionnaire did not produce detailed responses about how funding formulae work in the participating countries.

Half the countries said the PRFS' analyses were sources of strategic intelligence for the national authorities, as well as for participating institutions.

Benefits of PRFS were seen as being improved quality, restructuring and concentration of resources in the research sector, more accountability and internationalisation. In Estonia the results were seen as hard to dispute, while the Czech experience is that the academic community distrusts any form of research assessment, whatever method is used, and will happily 'game' any indicator system offered to it.

Peer review-based exercises were seen as costly but the growth of the national RIS was helpful, both because it simplifies assessment and because it reduces the compliance burden for researchers.

6.2 Details on practices and effects in the countries

6.2.1 Czech Republic

In the Czech Republic, the policy objectives were to increase both the research quality and productivity, to strategically steer research organisations in their research and to reach sustainability in the research system.

The PRFS that was launched at the end of the 2000s was uniquely focused on productivity. It counted research outputs registered in the national research information system and created perverse effects, especially in the field of applied research (where quality of the submitted outputs is harder to assess and the outputs harder to define). The current PRFS take a mixed approach: bibliometrics next to peer review and patent analysis. The system that is being developed is an informed peer review model. It covers all types of indicators and criteria. It is implemented annually.

The societal context of research is important for the Czech Republic as a central European country. There is also the question of assessment focusing on individuals or institutions and of where the money goes: to the institution or the research group directly.

6.2.2 Norway

The PRFS was introduced in 2001 and therefore has a long history. Policy objectives were to increase the quality and quantity of the research outcomes as the Norwegian system was underperforming in both these aspects.

The positive effects that were visible were related especially to productivity: the outputs doubled, even though publications account for only about 6% of the funding. However, the quality remained the same, i.e. there was an equal rise in quality and non-quality research outputs.

The publication indicators used are productivity indicators, based on counting the research outputs and information from Scopus. Scopus data are imported in CRISTin, through a semi-automated process, 30-40% of which has to be imported manually. Seeing the lack of effects on research quality, there is currently a consultation running among the research communities to decide upon the introduction of a citation indicator to stimulate the quality of the research.

Gaming effects were visible especially for the books as research output. Books are difficult to handle in the field of humanities. These can be textbooks, books for the wider public and scientific publications. The distinction is blurry and in the first years of the PRFS there was an over-reporting in the community. Over the years, the community self-disciplined; an important factor was that the registered research outputs were visible for all in the CRISTin [see the remark above on the importance of transparency as a factor allowing for social control]. They are talking about a specific citation indicator for books but it is very controversial.

The Norwegian system uses publication channels as an indicator of quality. There are about 20,000 journals included. The 'ranking' of publication channels is decided upon by the research community (a national committee) and revised every year. It is interesting to note that the same journals are categorised differently in Norway versus Finland versus Denmark.

The Ministry tried to include an indicator for 'outreach to the citizens' (e.g. radio/TV presence, webpages or articles in local newspapers) but stepped away from it as it was too difficult to implement and the collection and validation of the data was too resource-consuming.

6.2.3 Sweden

The HE system in Sweden is very diversified, with big and small universities, some general and some specialised, more or less tied to industry, as well as university colleges, so it is hard to fit them all into one system.

In Sweden the discussion on the PRFS and the model to be used is to be set against the context of the universities' autonomy. This was introduced in the 2000s with the intent to enhance the management of research within the universities, giving them direct responsibility.

The recent debate related to the institution that should be responsible for quality assurance (and therefore the assessment), i.e. the universities themselves or the national government. The argument of the universities was that they should bear the responsibility and that they already had systems of regular internal quality reviews in place. A national evaluation was therefore perceived as a waste of resources [see the topic of PRFS' 'cost-efficiency' above].

One outcome of this discussion was that in 2014, the new PRFS FOKUS, based on informed peer review, was not taken up when a new government came into power (the Research Council was given the task of developing new PRFS by the preceding government). The PRFS that is still in place has (only) two bibliometric indicators.

Effects were visible, especially in terms of research productivity, among the smaller institutes and colleges in particular, where the PRFS has been a reason for the university management to focus more on publications. Effects in terms of research quality were not visible; research excellence was already a point of attention for the larger universities, independently from the PRFS.

Also, the government started to demand the inclusion of societal impact (co-creation, collaboration and interaction with society) but there is very little knowledge about how to do that. There is data available only at a high aggregated level. In Sweden, "it is recognised that long-term results can't really be measured".

6.2.4 Portugal

The PRFS was introduced in 1996 It is based on international peer review, and researchers could choose to introduce bibliometric indicators in their application since2007. The evaluation has not fundamentally changed until recently, when the 2013 evaluation exercise introduced a number of changes, including a stronger use of bibliometrics. This is highly discussed in the community.

In terms of scope, since the beginning the PRFS evaluation covers all dimensions, but it grew stronger influenced by the criteria used by the peer reviewers, which changed slightly over time.

The periodicity is influenced by the cost of the exercise but also by the dynamics generated through the system of 'research units'.

In terms of granularity, the unit for evaluation is the 'research unit', ie research groups determined by the researchers themselves. In the last PRFS there were 322 research units. Research units can be defined internal to the organisation, across the departments (so detached from the organisational structure defined for the educational needs).

- The important factor is that these research units are dynamic, ie
- They change configuration over time, following the developments in research
- They grow because of the critical mass or because it is understood that additional expertise is needed
- They merge to allow for interdisciplinary research

There are three tiers in the funding system, block grant, performance-based funding and competitive funding. The retrospective as well as prospective dimensions of the PRFS force the research units to think strategically.

6.2.5 Croatia

The country adopts PRFS similar to the Norwegian system even though they're still working on the national research information system. In contrast to Norway, they use citations as impact indicators.

There are 8 universities and 25 research institutes. The PRFS has existed since 2013 and assess the universities and public research organisations (PROs). The unit is the institutions and the overall objective is to raise responsibility in the institutions for their own research strategies.

It uses bibliometrics and metrics, for which the criteria are:

- Research outputs
- External funding (national, international – number of funded projects and amount of funding as a share of the total funding)
- In- and out-going mobility
- Funding by industry or regions (excluding Structural Funds)
- Outreach activities to citizens

The PRFS in 2015/16 governed € 12.5m.

The model was developed by the Ministry and is based on consensus: it was discussed with the research community for about eighteen months. The visible effect is a rise in quantity; the effects on quality are not yet clear. The institutions took the exercise very seriously and have started making their own systems for internal funding allocation – down to the level of a single researcher.

The Ministry has collected information on these internal funding allocation models (about two and a half months ago) and are now analysing this data. The 'outreach to citizens' criterion, which intends to assess the societal impact and the 'popularisation' of science, is hard to measure and they are now considering whether to keep it in or replace it.

Another issue is the instability in the national government: there were three different governments last year and each one wants to change something.

6.2.6 Slovenia

The institutional funding of research groups was designed in 1999. Initially it covered only research groups within the organisations. It was based on peer review, using mostly national peers.

Now, international peers are used, and the units for evaluation are internal research groups (at the department level or lower) or cross-institutional ones.

The objective is to improve the quantity and quality of research, the impacts and the research capacity ('viability'). Productivity increase is not an issue any more, so it was dropped as criterion.

- Viability is assessed only retrospectively; it looks at the composition of the group (mixed ages) and the involvement of the PhDs and their mobility after gaining their degree). They have started now including also gender.
- Impact was measured in terms of 'relative impacts', but abandoned recently.

The evaluation is run every 6 years and it has a retrospective and prospective element.

Criteria for funds allocation:

- External funding from science funding body, evaluation score (science)
- External funding from business, abroad or 'thematic' ministries (relevance)

If the research unit scores above average for at least one of these indicators for each criterion, its funding can be increased; if it scores below the average for all criteria, funding is cut. The reduction of funding is marginal (1%), but it has strong effects thanks to the transparency: the evaluation results are published.

7 INITIAL POLICY RECOMMENDATIONS EMERGING FROM THE MLE

Participants also saw clear differences among the countries in terms of the relationship between national research governance bodies and the research-performing institutions (in some countries there is a tension between the universities' autonomy and the national government oversight). The national policy is also important in terms of the political rationalities of governments over time which are not forcefully directed towards addressing failures in the research system.

Participants indicated that the 'appropriate' share of institutional funding for research to be allocated through PRFS lies between 5% and 20%.

Other recommendations were:

- Do not put all your policy goals in the PRFS, use other instruments as well
- Remember the potential effects on institutions and individual researchers
- Design and count with care (e.g. how to handle papers with many authors)
- Transparency is key!
- There cannot be one single recommendation (objectives, context and design elements are different) beyond this: IF you do it, do it carefully
- Policy decisions for the system as a whole remain necessary to keep a balance
- Both metrics and peer review are useful. A mix of the two is best but remember that the effects of peer review are different from the effects of metrics
- Metrics-based systems are more appropriate for steering research behaviour (e.g. productivity) and can lead to a different research climate but can only look backwards
- Peer review is better equipped for looking forward and is better for the assessment of the institution as a whole - beyond the mere conduct of research; it is also more equipped for changing science policy and the overall science system
- Use a mix of retrospective and prospective measures where the latter can act as a buffer for the effects of the former ('positive discrimination')
- Field normalisation is important but do not overdo it
- Keep the system flexible

For the participants, however, the question remains: are we measuring too much? Are we making things too complicated?

8 ANNEX A: BIBLIOGRAPHY

Aagaard, K., Bloch, C., & Schneider, J. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation, 24*, 106-117.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: the effects of changing the rules of the game during the game. *Scientometrics, 88*(1), 229-238.

Abramo, G., D'Angelo, C., & di Costa, F. (2011). National research assessment exercises: a comparison of peer review and bibliometric rankings. *Scientometrics, 89*, 929-941.

Adams, J., & Gurney, K. (2014). *Funding selectivity, concentration and excellence - how good is the UK's research? .* HEPI.

Arnold, E., Barker, K., & Slipersæter, S. (2010). *Research Institutes in the ERA.* Brussels: European Commission.

Arnold, E., Good, B., & Segerpalm, H. (2008). *Effects of Research on Swedish Mobile Telephone Developments: The GSM Story, VA2008:04.* Stockholm: VINNOVA.

Arnold, E., Luukkonen, T., Boekholt, P., Nooijen, A., Jávorka, Z., & Zuijdam, F. (2013). *Evaluation of the Academy of Finland.* Helsinki: Ministry of Education and Culture.

Arnold, E., Simmonds, P., Farla, K., Kolarz, P., Mahieu, B., & Nielsen, K. (forthcoming 2017). *Review of the Research Excellence Framework: Evidence Report.* London: Department for Business, Energy nd Industrial Strategy.

Bence, & Oppenheim. (2005). The evolution of the UK's research Assessment Exercise: publications, performance and perceptions. *Journal of Educational Administration and History, 37*(2), 137-155.

Bence, V., & Oppenheim, C. (2005). The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions. *Journal of Educational Administration and History, 37*(2), 137-155.

Bloch, C., & Schneider, J. (2016). Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication Indicator at the individual level. *Research Evaluation, 25*(3).

Bouabid, H. (2014). Science and technology metrics for research policy evaluation: some insights from a Moroccan experience. *Scientometrics, 101*, 899-915.

Butler, L. (2003). Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts. *Research Policy, 32*(1), 143-155.

Butler, L. (2010). *Impacts of Performance-Based Funding Systems: A review of the concerns and evidence.* Paris: OECD.

Chubin, D. E., & Hackett, D. J. (1990). *Peerless Science: Peer review and US science policy.* Albany: SUNY Press.

Cronin, B., & Sugimoto, C. R. (2014). *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact.* Boston, Mass: MIT Press.

Edgar, F., & Geare, A. (2010). Characteristics of High- and Low-performing University Departments as Assessed by the New Zealand Performance Based Research Funding (PBRF) Exercise. *Australian Accounting Review, 20*(1), 55-63.

Elsevier. (2015). *A review of the UK's interdisciplinary research using a citation-based approach.* A report by Elsevier to the UK HE funding bodies and MRC.

Elton, L. (2000). The UK Research Assessment Exercise: Unintended Consequences. *Higher Education Quarterly, 54*(3), 274-283.

Farla, K., & Simmonds, P. (2015). *REF 2014 Accountability Review: Costs, benefits and burden.* Bristol: HEFCE.

Fridholm, T., & Melin, G. (2012). *"Med glädje, men inte med lätthet" –om högskolans fördelning av de direkta statsanslagen för forskning Distribution of governmental direct appropriations to research at Swedish universities .* Stockholm: Swedish Employers Federation.

Good, B., Vermeulen, N., Tiefenthaler, B., & Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation, 24*(2), 91-105.

Hävrynen, M. (2007). *Breakthrough Research: Funding for High-Risk Research at the Academy of Finland, 8/07.* Helsinki: Academy of Finland.

Henkel, M. (1999). The Modernisation of Research Evaluation: The Case of the UK. *Higher Education, 38*(1), 105-122.

Horrobin, D. F. (1996). Peer review of grant applications: a harbinger for mediocrity in clinical research? *The Lancet, 348*, 1293-1295.

Jiminez-Contreras, E., Anegon, F., & Lopez-Cozar, E. (2003). The evolution of research activity in Spain – The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy, 32*, 123-142.

Jonkers, K., & Zacharewicz, T. (2015). *Performance based funding: A comparative assessment of their use and nature in EU Member States .* Brussels: JRC Science for Policy Report.

Jonkers, K., & Zacharewicz, T. (2016). *Research Performance Based Funding Systems: a Comparative Assessment.* Brussels: JRC.

Lakatos, I., & Musgrave, A. (1970). *Criticism and the Growth of Knowledge.* Cambridge: Cambridge UP.

Lee, F. (2007). The Research Assessment Exercise, the state and the dominance of mainstream economics in British universities. *Cambridge Journal of Economics*(31), 309-325.

Lee, F., & Harley, S. (1998). Peer review, the Research Assessment Exercise and the demise of non-mainstream economics. *Capital & Class, 22*(3).

Lepori, B., van den Besselaar, P., Dinges, M., Potì, B., Reale, E., Slipersæter, S., . . . van der Meulen, B. (2007). Comparing the evolution of national research policies: what patterns of change? *Science and Public Policy, 34*(6), 372-388.

Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation, 21*(1), 48-60.

Luukkonen, T., Stampfer, M., & Strassnig, M. (2015). *Evaluation practices in the selection of ground-breaking research proposals.* Vienna: WWTF Working Paper: 13-08-2015.

Mahieu, B., & Arnold, E. (2015). *R&D Evaluation Methodology & Funding Principles - The R&D Evaluation Mthodology (Final Report 1).* Technopolis Group. Prague: Ministry of Education, Youth & Sports.

Martin, B., & Whitley, R. (2010). The UK Research Assessment Exercise: a Case of Regulatory Capture? In R. Whitley, J. Glaser, & L. Engwall, *Reconfiguring knowledge production : changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 51-80). Oxford University Press.

Meriläinen, S., Tienari, J., Thomas, R., & Davies, A. (2008). Hegemonic Academic Practices: Experiences of Publishing from the Periphery. *Organization, 15*(4), 584-597.

Moed. (2005). *Citation Analysis in Research Evaluation.* (Springer, Ed.)

National Science Board. (2007). *Enhancing Support of Transformative Research at the National Science Foundatio.* Washington DC: National Science Board.

NZ Ministry of Education. (2012). *An international comparison of performance-based research funding systems (PBRFS).* New Zealand.

OECD. (2010). *Performance-based funding for public research in tertiary education institutions: workshop proceedings.* OECD Publishing.

OECD. (2010). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings.* Paris: OECD Publishing.

PREST. (2000). *Impact of the Research Assessment Exercise and the Future of Quality Assurance in the Light of Changes in the Research Landscape.* Manchester: Manchester University: PREST.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy, 41*, 1262-1282.

Rosenberg, N. (1976). *Perspectives on Technology.* Cambridge: Cambridge University Press.

Roy, R. (1985). Funding Science: The Real Defects of Peer Review and an Alternative to It'. *Science, Technology and Human Values, 10*(3), 73-81.

Sörlin, S. (2007). Funding Diversity: Performance-based Funding Regimes as Drivers of Differentiation in Higher Education Systems. *Higher Education Policy, 20*, 413-440.

Sikes, P. (2006). Working in a 'new' university: in the shadow of the Research Assessment Exercise? *Studies in Higher Education, 31*(5), 555-568.

Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISI Newsletter, 6*(1), pp. 22-28.

Smart, W. (2013). *Analysis of the Impact of the PBRF. Interim Findings.* Wellington: Ministry of Education.

Stöckelova, T. (2012). Immutable Mobiles Derailed: STS, Geopolitics, and Research Assessment. *Science, Technology, & Human Values, 37*(2), 286-311.

Vanecek, J. (2013). he effect of performance-based research funding on output of R&D results in the Czech Republic. *Scientometrics, 10*, 1-25.

Wagner, C., & Alexander, J. (2013). Evaluating transformative research programmes: A case study of the NSF Small Grants for Exploratory Research programme. *Reseatch Evaluation, 22*(3), 187-197.

Wang, J., & Hicks, D. (2013). Detecting structural change in university research systems: A case study of British research policy. *Research Evaluation, 22*, 258-268.

Wessely, S. (1998). Peer review of grant applications: what do we know? *The Lancet, 352*, 501-505.

9 ANNEX B: SUMMARY OF INFORMATION FROM PARTICIPATING COUNTRIES

Problems and policy purposes

	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Increase scientific performance Improve quality and relevance Concentrate resources Increase competitiveness of RIs and HEIs Restructure PRO system Improve research efficiency Link public funding to performance Increase efficiency of state R&D spending	Need for more transparency in allocating institutional funding Need to increase research performance	Need for a transparent, 'hands off' funding system when universities are granted increased autonomy	Need to modernise the research assessment and funding system Desire for 'objective' assessment process; peers not regarded as trustworthy	PRFS introduced together with institutional research funding, 2005, to enable development of the research sector Until that point, institutions were unable to maintain their own research strategies	Increased quality Accountability	Increased quality and competitiveness	Introduced as part of a wider university quality reform in 2002 HE sector research productivity and quality were both seen as inadequate	Increased quality More efficient resource allocation	Encourage universities to give quality more priority

Aims

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Structural reform Modernise research infrastructure Reach international quality level Create commercialisation infrastructure Improve science-industry links	Ensure stability in institutional funding Improve research performance	Increase quality Transparent funding system Performance monitoring	Increase quality Incorporate a formative dimension Strategic intelligence for stakeholders Determine institutional funding Take account of differences among types of organisation	Enabling institutional development	Distribute institutional funding for research to the universities	Create incentives for institutions to focus on high-quality research activities	Reward institutions based on quality Stimulate high quality research Encourage development of institutional strategies	Competitive and flexible research system Foster strategic planning by institutions Internationalisation Quality	Create incentives for universities to focus research on areas where they could achieve high quality

Non-PRFS research assessment

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
PRO evaluations Field evaluations	Universities' own internal evaluation systems are audited every 3 years Annual intellectual capital reporting	None	Academy institutes are evaluated 5-yearly using informed peer review	Evaluations are undertaken at field- and programme-level	Universities are free to perform institutional evaluation, using data from the PRFS aggregated to department level	Some universities run internal assessment exercises using international peers	There have been 30+ field evaluations in the last 20 years	Universities are free to run internal assessment exercises	Large universities run internal research assessment exercises Research councils do field evaluations

Design

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Consultation									
National Academy of Sciences, RIs and HEIs	Ministries, universities, departments and individual researchers, interest groups such as university associations and student organisations	Universities and PROs	R&D & Innovation Council Research community	No	The academic community is consulted before each exercise	Discussion with relevant stakeholders	Reference group Association of Norwegian Higher Education Institutions	Research community is consulted ahead of each exercise	Original consultation led to only part of the proposed system being adopted Recent consultation led to the rejection of a peer review system
Other models inspected									
FR, DE, IT, AT, PL, EE, UK, RU, LT, US		'various' European countries	AT, NL, NO, SE, UK, Flanders		UK	UK, SE, NO, FI, DK, DE	SE, DK but these had little influence on research evaluation	None	NO, FI, DK, UK, AU
Designer									
State Committee of Science	Federal Ministry of Science, Research and the Economy (BMWFW)	Ministry of Science and Education (MoSE), Institute of Economics, Zagreb	RDI Council Ministry of Education, Youth and Sport Advised by Technopolis, NIFU & Technology Centre ASCR		ANVUR	Dedicated design committee	Overall system designed by the Ministry of Education and Research Publication indicator designed by the Association of Norwegian Higher Education Institutions	Ministry for Science, Technology and Higher Education	A new committee will design a system that also tackles impact

Assessment model

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Model									
Peer review and metrics – scholarly indicators	No. of active BA and MA students* No. of BA and MA graduations No. of PhD students employed by the university Weighted revenues from FWF and EU	Bibliometrics Competitive grants Researcher mobility Industry funding Popularisation See Table 5.	Bibliometrics Counts of innovation outputs, e.g. prototypes Peer review of a small number of submitted outputs	Publications Patents External research income PhD graduations Support to topics of national importance	SSH is pure peer review In other subjects, peers validate judgements on individual articles, based on citations and JIF	No system yet	Research outputs listed in the national RIS (CRIStin) External research income (RCN, EC, other) PhD graduations	Peer review, based on self-assessments and site visits Informed by bibliometric indicators	A bibliometric indicator An indicator of external funding
Units of analysis									
RIs HEI laboratories or departments	Universities	Individual researcher within field	Universities, institutes	Universities	Individual researchers	Universities	Individuals and institutions	'Research unit', is a group of people within an institution working in a field May be split into research groups	University
Unit of reporting									
RIs HEI laboratories or departments	Universities	Individual researcher within field	Universities, institutes	Universities	Individuals get their own rating Results are only published at department or institutional level	Universities	Universities	Research unit/group	University

*field-weighted

Assessment

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Assessment criteria for peer review element of PRFS									
Novelty Importance Feasibility Resources Performance indicators	N/A	N/A	?	N/A	Originality Methodological rigour Scientific impact	N/A	N/A	Productivity Scientific and societal relevance Research capacity Postgraduate training	N/A
Data sources used in assessment									
Publications, JIFs, conference contributions, patents, etc. National RIS, open access repositories, WoS, Scopus	University intellectual capital reports Ministry university and public research infra-structure databases	Annual reporting from Universities and PROs WoS, Scopus	National research information system Scopus WoS		WoS: 5-year impact factor and Article influence score Scopus: Impact per publication and Scimago journal ranking	National RIS (IBN) WoS/Scopus	CRIStin Institutional accounts National student data system	Self-assessments Scopus ORCID	Bibliometric indicator is based on WoS Funding indicator is from national statistics
How field differences are handled									
6 field committees Bibliometric indicators field-normalised Humanities panel decides how to treat non-indexed materials	Field weights	Eight fields with different weights Bibliometric indicators are not field-normalised	Bibliometric indicators are field-normalised In SSH books are reviewed by peer panels	Not considered	16 field panels	Field-normalised bibliometric indicators	Not handled before 2014. Now, publications in different fields attract varying numbers of points	Field panels	Bibliometric indicator is field-normalised

Experience with metrics

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Scholarly quality									
Indicators accepted		Bibliometrics Competitive grants	Bibliometric indicators contentious		Citations, journal impacts and esteem are used outside SSH		PRFS increased the quantity but not the quality of publication Considering a citation indicator to tackle this	Only used to inform peer review	Has increased universities' focus on research quality
Impact									
		Projects with industry	Counting outputs led to gaming to inflate their numbers		ANVUR is currently developing impact metrics		Income from business was introduced as an indicator a year ago. Too soon to judge results.		Vinnova has developed a qualitative impact indicator. In future this, the bibliometric and funding indicators will weigh 1/3 each
Prospective dimension?									
Included in assessment system	Performance contracts	No	No	No		Intention is to address gender equality, young researchers etc. to get a more even funding distribution	No experiments with performance contracts will start this year	PRFS considers present and future capacity	No

Funding

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Funding formula									
	10% PRFS, of which 87% is performance-based 90% performance agreements	Scientific productivity 60% Competitive funding and mobility 25% Industry cooperation 10% Popularisation of science 5% See **Error! Reference source not found.** for calculation	Bibliometric indicators are field-normalised Other outputs are weighted Each resulting point generates 'x' units of funding	Publications and patents 38% External research income 48% PhD graduations 9% Support to topics of national importance 5%	The PRFS covers 20% of institutional funding for research That 20% is distributed: 65% on general evaluation; 20% on results from researchers hired or promoted in the period; 15% on teaching quality		70% of institutional funding is a block grant and is not performance-based 24% is based on educational performance 6% is PRFS	Each unit gets an overall score (5-point scale) Each score has a different weight and is multiplied by FTEs to calculate funding	
Other uses of the assessment									
Strategic intelligence for the State Committee of Science	Strategic intelligence for BMWFW	Strategic intelligence for MoSE and institutions	No		Used in accredit-ation of PhD courses Universities say used for internal allocation		Aggregate data are published annually and used as strategic intelligence by policymakers		
Periodicity									
Annual	3 years	Annual	Annual	Annual	4 years		Annual	3-6 year intervals	Annual

Benefits and Costs

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Benefits									
Increased competition Concentration of resources Increased efficiency in research production	Transparency Increased performance		Initially performance improved	Results are difficult to dispute	Increased accountability	Expect quality improvement	Increased quantity of research output with no reduction in quality	Stimulated flexibility and re-organisation Promoted internationalis-ation and improved quality	Increased quality focus Some institutions have improved very significantly
Costs									
Difficult to balance peer review and metrics-based elements High costs of the exercise, especially when using foreign peer reviewers	The previous 11-indicator system was resisted, for fear of funding instability 4-indicator system produces little strategic intelligence		Since then, the innovation indicators have been gamed	Some difficulties in data cleaning Sometimes hard to define the boundaries of contract research	High compliance costs A national RIS is being developed for the next exercise		The main cost was developing the CRIStin RIS Compliance costs for researchers Some money moved from established to newer institutions	Expensive but seen as fundamental to the functioning of the system so the cost is rarely criticised	Most money is directed to the old universities Reduced attention to things that are not incentivised, e.g. PhD education

Getting in touch with the EU

IN PERSON

All over the European Union there are hundreds of Europe Direct Information Centres.
You can find the address of the centre nearest you at: http://europa.eu/contact

ON THE PHONE OR BY E-MAIL

Europe Direct is a service that answers your questions about the European Union.
You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: http://europa.eu/contact

Finding information about the EU

ONLINE

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU PUBLICATIONS

You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact)

EU LAW AND RELATED DOCUMENTS

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU

The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

This report provides an overview of the key concepts that guide and characterise the design of Performance-based Research Funding Systems (PRFS) in the international practice. While in a minority of cases, PRFS govern part of the institutional funding of research institutes as well as universities, the discussions and papers in this MLE only address universities.

This first paper in the context of the MLE sets the background for the thematically-focused papers that will follow. It covers the two core components of a PRFS: the assessment process, which judges research output based on its scientific quality and increasingly also other criteria, and the funding formula that are used to allocate funding to the universities, based on the assessment results