



European
Commission

MLE on Performance-based Research Funding Systems (PRFS)

Peer Review in PRFS

Thematic Report No 3



July 2017

Research and
Innovation

Peer Review in Performance-based Research Funding Systems (PRFS) – MLE on Performance-based Research Funding Systems (PRFS)

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 July 2017.

This document has been prepared for the European Commission however it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

© European Union, 2017.

Reuse is authorised provided the source is acknowledged. The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p. 39).

MLE on Performance-based Research Funding Systems (PRFS)

Peer Review in PRFS

Thematic Report No 3

Dorothea Sturn

Table of Contents

1	INTRODUCTION	3
2	WHAT IS PEER REVIEW?	4
3	STRENGTHS AND WEAKNESSES OF PEER REVIEW	6
2.1	Limitations related to the traditional functions	6
2.2	New domains, new criteria, new challenges.....	8
3	PEER REVIEW IN PRFS AND RESEARCH ASSESSMENTS	10
3.1	Strengths and weaknesses of peer review – the MLE participants’ perspective	10
3.2	Criteria, dimension and outcome	12
3.3	Selection of peers and process management	13
3.4	Granularity, scope and costs	13
4	COPING WITH CHALLENGES	15
4.1	The Panels’ architecture	15
4.2	Social dynamic and management	16
4.3	Informed peer review: the triangulation of methods.....	17
4.4	Costs of Peer Review	17
5	CONCLUSIONS	19
6	ANNEX A: REFERENCES	20

List of Tables

Table 1:	Peer review in the MLE participating countries.....	3
Table 2:	Most important arguments in favour of Peer review	11
Table 3:	Most important arguments against Peer Review	11
Table 4:	Assessment Criteria.....	12
Table 5:	Costs per researcher	14
Table 6:	Panel members’ qualifications.....	15

List of Figures

Figure 1:	Uses of Peer review.....	4
Figure 2:	Measures and Mechanisms addressing Limitations of Peer review	7
Figure 3:	Extended Peer Review	9
Figure 4:	Different Panel solutions	14
Figure 5:	Panel architecture	16

1 INTRODUCTION

This paper is part of a sequence of papers written in the context of the Mutual Learning Exercise (MLE) in 2016-17 on Performance-based Research Funding Systems (PRFS) for institutional funding. The focus of this report is on the use of **peer review** in the context of such research funding systems. It aims to support discussion and mutual learning among the participating countries on different aspects, issues and questions related to peer review-based assessment systems.

Only few countries – both in the context of the MLE and in general – use peer review as an evaluation tool in the context of their PRFS. The scope of this paper is therefore widened to national research assessment systems that are **not directly linked to funding** distribution but nevertheless form an important part of the national assessment culture (eg the Standard Evaluation Protocol in the Netherlands).

The use of peer review in the PRFS of the countries participating in the MLE is as follows (Table 1):

- Some countries using metrics-based PRFS (like Norway) developed a specific “dual system” of funding and assessment tools. The PRFS systems are not viewed as research evaluation, but are instead complemented by other evaluation systems of a more formative character based on peer review
- Other countries such as Estonia, Moldova and Croatia use peer review based research evaluations to identify research institutions’ eligibility to apply for public research funding
- Both in Austria and Sweden, peer review is only used at the level of single universities or research organisations
- Spain is a borderline case as the Sexenio system has many similarities with a PRFS, and in fact, some authors (e.g. Hicks 2012) classify it as such. However, funding is distributed to individuals who apply for it rather than to institutions; it is therefore based on an individual rather than a nation-wide assessment

Table 1: Peer review in the MLE participating countries

Peer Review in PRFS as well as in systematic and comprehensive Research Assessment (Total No of answers: 12)			
Integrated in the PRFS	Outside PRFS		
Informed peer review or mixed models	Subject-specific evaluations and other assessments:	Accreditation (eligible for public funding):	At local (university’s) level only:
Italy Portugal Slovenia Turkey	Czech Republic Norway Spain	Croatia Estonia Moldova	Austria Sweden



A questionnaire sent out to all participating countries concerning the specific use of and experiences with peer review served as an important source of information for this paper.

2 WHAT IS PEER REVIEW?

As a traditional evaluation tool, peer review plays an important role in scholarly research and communication. Most often, the beginning of peer review is attributed to the 1752 Royal Society of London's development of a "Committee on Papers" to oversee the review of text for publication in the journal *Philosophical Transactions*. (see e.g. Spier (2002), Fitzpatrick (2009), Bornmann (2011)). Peer review as a means to support the decision about the suitability of articles proposed for publication is still probably its dominant function. It was not until the mid-20th century that new developments occurred, ie the use of reviewers outside the closed "Societies" and the use of peer review by Research Councils as a central mechanism to allocate research funding.

Today, peer review is widely used in nearly all research assessment exercises such as programme evaluations, field studies, national and institutional investigations and in some countries it plays a role in the PRFS. It often supports also the selection and promotion of scholars and the formulation of research programmes and strategies. Further extensions of the concept include the development of new methods like peer Panels or Expert Panels, the combination with other assessment methods such as bibliometrics (see "informed peer review" in Section **Error! Reference source not found.**, below), and the introduction of new assessment criteria such as relevance or impact. Figure 1, below, gives an overview.

Figure 1: Uses of Peer review

	Written Reviews	Peer Panels	Expert Panels	Other methods
 Scholarly Publishing	Dominant	Editorial Boards		
 Proposal Evaluation	Dominant	Dominant	Little use	
 Programme Evaluation		Little use	Little use	Dominant
 System Evaluation (national, regional, institutional)	Little use	Panels (different stages) combined with other methods		
PRFS	Little use	Panels (different stages) combined with other methods and indicators (bibliometrics)		

Peer review refers to the self-organisation of the academic world and is rooted in Merton's Theory of the "Normative Structure of Science" (Merton 1942), as it embodies all the so-called Mertonian norms of science:

- Communalism (common ownership of scientific discoveries, where scientists give up intellectual property rights in exchange for recognition and esteem),
- Universalism (claims to truth are evaluated in terms of universal or impersonal criteria and not based on race, class, gender, religion, or nationality),
- Disinterestedness (scientists are rewarded for acting in apparently selfless ways),
- Organized Scepticism (all ideas must be tested and subjected to rigorous, structured community scrutiny).

A definition of peer review taken from Gibbons and Georghiou shows clearly how peer review refers to an autonomous academic world with specific fields and reflects the impossibility for outsiders to judge the quality of research in a particular discipline.

*"Peer review is the name given to the judgement of scientific merit by other scientists working in, or close to the field in question. Peer review is premised upon the assumption that a judgement about certain aspects of science, for example its quality, is an **expert decision capable of being made only by those who are sufficiently knowledgeable** about the cognitive development of the field, its research agenda, and the practitioners within it."* (Gibbons and Georghiou, 1987, my emphasis)

3 STRENGTHS AND WEAKNESSES OF PEER REVIEW

Peer review is perceived as the **gold standard** of assessing academic achievements, scholarly publishing and communication. It gives rigour and legitimacy to new ideas, improves the trustworthiness and clarity of academic work, and determines whether research can be viewed as scientifically valid. The most cited strengths of peer review include:

- Peers have a deep understanding of the research in question, the state of the art, the methods applied and the specific way of thinking in the field
- Peer review engages trust and is accepted by the community (especially compared to other evaluation techniques)
- Peer review improves research as it offers valuable feedback and comments
- Peer review creates scope for debate and facilitates formative evaluation elements (especially in the context of Review Panels)

Nevertheless, such enthusiastic statements are rare and most literature is sceptical. Peer review seems to be an unavoidable evil rather than a blessing. The Royal Society states:

*"Peer review is to the running of the scientific enterprise what democracy is to the running of the country. It may not be the most efficient system but it is the **least susceptible** to corruption."* (Royal Society 1995, my emphasis)

Twenty years later, this general appraisal still maintains, for example in the discussions on the appropriate evaluation tools for national assessments such as the UK REF:

*"Peer review, despite its flaws and limitations, continues to command widespread support across disciplines. Metrics should support, not supplant, expert judgement. Peer review is not perfect, but it is the **least worst** form of academic governance we have, and should remain the primary basis for assessing research papers, proposals and individuals, and for national assessment exercises like the REF".* (Wilsdon, J. et al 2015, my emphasis)

Furthermore, well-publicised fraud scandals involving famous researchers in physics and human cloning shook the academic world's confidence in the quality assurance system of journals like *Science* or *Nature*, and in peer review in general. This is only the tip of the iceberg, though. Potential limitations and weaknesses are widely discussed and partly proven by empirical studies and experiments.

In the context of this paper, with its focus on the use of peer review in PRFS and Research Assessment, we need to distinguish between failures and mistakes related to the traditional functions of peer review (ensuring scientific excellence and the autonomy of academia) and the problems that emerge due to the inability of peer review to cope with wider socio-economic criteria and strategic and governance questions.

2.1 Limitations related to the traditional functions

Academics used to be critical on their own standards and procedures even without any comments from outside. One of the harshest piece came out of the community:

"Peer review is slow, expensive, profligate of academic time, highly subjective, prone to bias, easily abused, poor at detecting gross defects, and almost useless for detecting fraud" (R. Smith, Editor of BMJ, 1999)

A multitude of studies, experiments and scholarly communication (in blogs, in journals, in platforms) have focused on the potential limitations of peer review:

- Many authors regard the traditional peer review mechanism as problematic as it causes **delays** to the publication and communication of novel research
- In addition, some criticise the closed and non-transparent system for **protecting the status quo** and suppressing research that is considered heterodox, radical or innovative

- Peer review disfavours **interdisciplinary** research (as well as multi- and transdisciplinary research)
- Other studies have shown peer review to be **poor at detecting errors** and misconduct
- **Biased** by conflicts of interest, and systematically biased against authors' reputations, locations, and even gender
- Especially in small fields, **personal stake** and vested interest could become critical, as there may be only a handful of experts worldwide who have sufficient experience with the methods, pitfalls and latest developments
- Furthermore, in a world of fast growing publication activities and massive increase in peer review effort required by funders and evaluators, the system is under considerable strain as peer review **overuses the scientific community**.

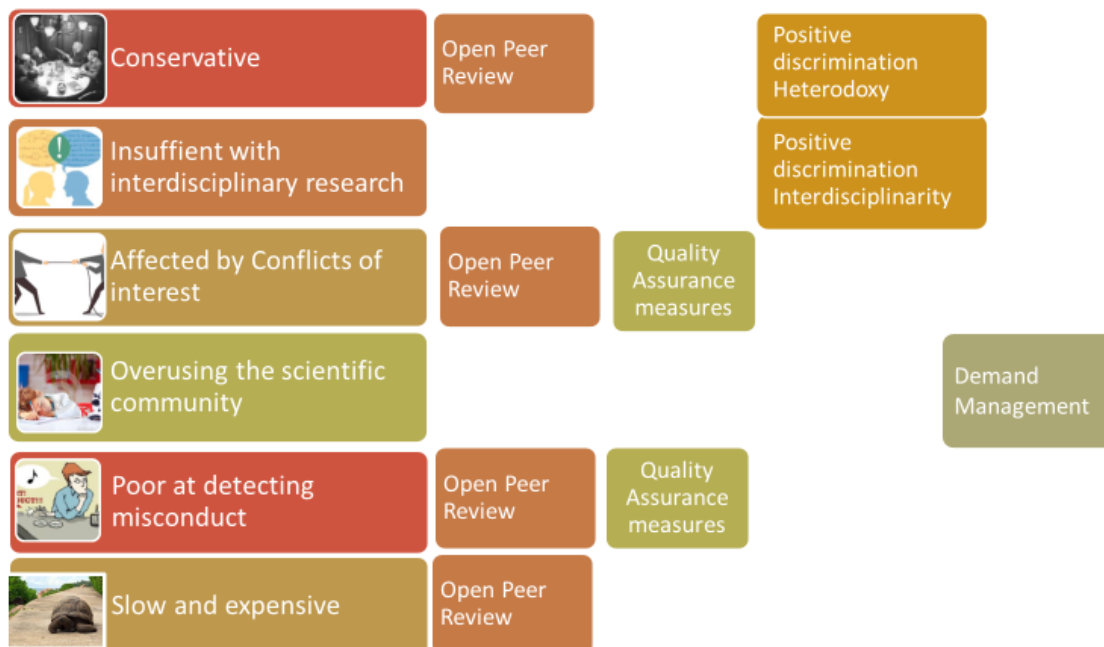
For arguments and evidence in detail see e.g. Cole et al 1981, Wessely 1998, Hodgson 1997, Wenneras et al 1997, Hopewell et al 2009, an overview is offered by Bornmann 2011. In recent years, the problems connected with peer review were accelerated by an increasing demand for competitive grants in combination with limited budgets and therefore declining success rates (see e.g. Fang 2011).

However, in the domains of both scholarly publication and research funding, many journals and councils introduced policies and practices that aimed at overcoming these limitations.

The picture in Figure 2, below, is far from comprehensive but it shows the most important measures and practices applied by funders and journals.

- In most journals and councils there are detailed procedures at work to avoid **conflicts of interest**, e.g. check of co-publications and self-declarations of unbiasedness
- Journals implemented **control- and quality-assurance** mechanisms to improve the detection of misconduct and fraud
- Some journals launched innovative initiatives like "Registered Reports", to enhance transparency and reproducibility (see Chambers 2014)

Figure 2: Measures and Mechanisms addressing Limitations of Peer review



The by far most important development is "**Open Peer review**" which forms part of the ongoing evolution of an open research system, and transforms peer review into a more constructive and collaborative process. Open peer review is a relatively new phenomenon (initiated in 1999 by the BMJ) one aspect of which is that the authors' and referees' names are disclosed to each other. In the 2000s, many journals have experimented with different

kinds of open peer review: some made reviewers' reports (with their consent) available to readers as part of an online pre-publication history alongside each research paper (BMJ open), some hosted papers on an open server for moderated public comment as well as simultaneous standard peer review (Nature) (see e.g. Pöschl 2012)

To overcome the tendency to discriminate **interdisciplinary research**, funding agencies and journal editors increasingly consider 'interdisciplinary' as an explicit criterion for evaluation. Other possible measures include specific schemes and programmes for the encouragement of interdisciplinarity.¹ As peer review panels require a grouping of their members along specific areas of expertise, their disciplinary structure is in tension with research that transcends these structures. In some countries, interdisciplinarity is accounted for by cross-referring to additional panels: members from more than one panel will be asked to consider proposals that cut across the remit of more than one panel (e.g. UK Research councils, German DFG). An alternative to this practice is the introduction of interdisciplinary panels which concentrate exclusively on interdisciplinary (and transdisciplinary) proposals (e.g. Italy's VQR, Swizz SNF, ERC). These forms of positive discrimination might be valuable first steps to incentivise collaborations between disciplines

In some cases, excessive **demand for competitive grants is limited** through specific mechanisms. This can be done by defining a maximum number of projects per principal investigator (eg one or two projects in parallel), by reducing the frequency of demand, by allocating some money through larger blocks of funds (eg to support a team or a set of projects rather than an individual project) and/or for longer blocks of time (eg five years instead of the typical three, or a series of rolling grants) or by introducing 'demand management' with different forms of pre-selection processes (eg United Kingdom or Switzerland).

2.2 New domains, new criteria, new challenges

As mentioned in Chapter 2 **Error! Reference source not found.**, above, nowadays peer review is widely used in nearly all research assessment exercises. As science became increasingly important for innovation processes, socio-economic development as well as for political decisions, peer review started to be applied also to new domains (project development, in-process evaluation, ex-post evaluation, etc.). Therefore, peer review was adapted to new needs in different ways and with different perspectives/dimensions:

- **New Criteria:** wider socio-economic criteria like relevance, impact, and new forms of interaction with society were considered, apart from scientific excellence
- **New Methods:** new forms of peer review including strategic peer review, in-process peer review and merit review were developed. They share the feature of considering multiple criteria and/or relying on non-scientific peers, at least partly
- **New Players:** depending on the topic in question, non-academic peers were involved, such as representatives from industry, users, members from NGOs and other civil society representatives. Peer Panels turned into **Expert Panels**. These new players are integrated in existing panels or the procedures are divided into double review systems with two sets of criteria (scientific excellence and socioeconomic relevance).² Some authors go even further and propose an open dialogue between science, industry, policy and civic society.³

The introduction of **broad socio-economic impact** as a criterion for evaluation, mainly as a justification against competing demands for resources in times of austerity cuts, puts a lot of pressure on traditional peer review-based research assessment. Peer review tends to ignore wider social and economic effects because of its strongly scientific orientation; it therefore seems not to be an appropriate tool for impact evaluation (see e.g. Ruegg and Jordan, 2007).

¹ Some funders developed specific schemes for the support of "blue sky" or unorthodox research in order to overcome the conservatism inherent in peer review systems.

² For example, the US National Institutes of Health (NIH) has such a system for grant applications (Scarpa, 2006).

³ Compare concepts like "Responsible Research and Innovation" (European Commission 2016), "Quadruple Helix" or "Open Innovation" (e.g. Carayannis et al 2015).

There are two interrelated solutions to this problem; both are difficult to implement, though:

- Peers can be informed by analyses of socioeconomic needs and priorities. In most cases this is done by providing 'impact statements'
- Non-academic experts can be included

The topic of socio-economic impact will be the focus of forthcoming papers and discussions in the context of the Mutual Learning Exercise (MLE) Performance-Based Research Funding Systems (PRFS). In this paper, it suffices to note that traditional peer review is unable to cope with this challenge.

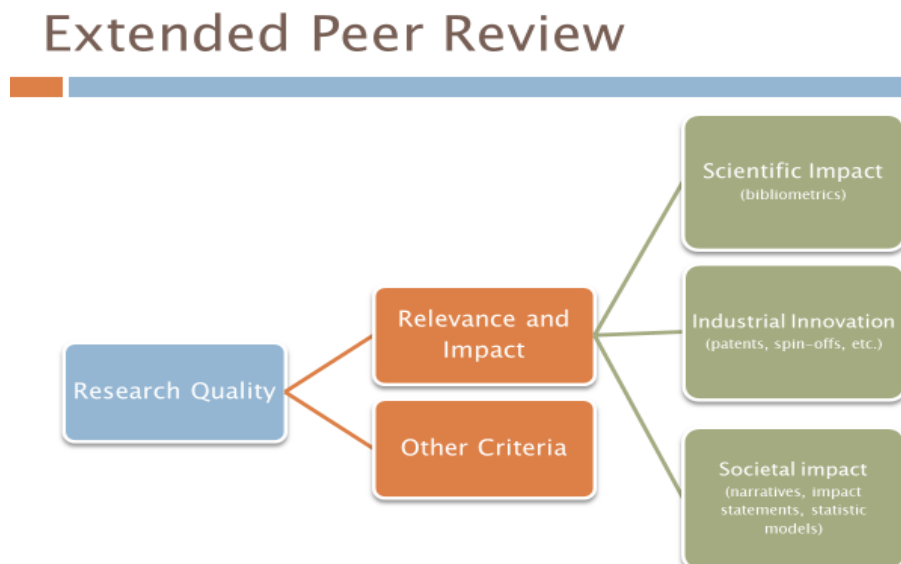
Nevertheless, nearly all participating countries include '**Impact outside academia**' as a criterion in their PRFS or national research assessment. In the context of the UK Research Excellence Framework (REF), the responsible council HEFCE states:

"The Research Excellence Framework was the first exercise to assess the impact of research outside of academia. Impact was defined as 'an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia'." (see <http://www.hefce.ac.uk/rsrch/REFimpact/>)

In a recent study, Derrick et al. investigate the divergence in opinions of evaluators involved in the REF prior to the assessment. They show the wide range of views about impact, and suggest that these views could conceptually be reflected in a range of different positions along the evaluation scale (Derrick et al 2016).

Error! Reference source not found. shows how peer review is extended to cope with requirements beyond the assessment of research quality and what additional information is needed to inform the peers and experts properly:

Figure 3: Extended Peer Review



3 PEER REVIEW IN PRFS AND RESEARCH ASSESSMENTS

In contrast to Chapter **Error! Reference source not found.** where we discussed peer review mainly in its classical domains of article and proposal evaluation, this Chapter will focus on peer review in PRFS. It also covers the use of peer review in other national research assessment systems that may not directly be linked to funding distribution but form an important part of the national assessment culture (see also Table 1 in the introduction to this paper, above).

Example: Peer review in UK's RAE and REF:

The UK was the first country to introduce research funding based on a peer review-based PRFS and in fact, since 1986 the UK experience with the Research Assessment Exercise (RAE) has informed policy initiatives in many other European countries and worldwide. After a series of extensive consultations and reviews, the Higher Education Funding Councils replaced the RAE with the new Research Excellence Framework (REF).

The REF was initially proposed as a metrics-based evaluation to replace the RAE after the completion of the 2008 exercise. It was intended to reduce the administrative burden on the academic community and better to demonstrate the economic and societal contribution of research (Martin and Whitley, 2010). However, after strong resistance from the academic community, HEFCE gave up on the switch to a bibliometrics based assessment. Disciplinary panels were given the possibility to use the results of centrally provided bibliometric analysis. A new element was introduced in the research assessment exercise, though, ie narratives on impacts reached, to be drafted by the assessed institutions.

Wilsdon et al. (2016) criticised the eclectic and non-systematic handling in the REF and argued for a responsible use of metrics as opposed to a full switch. Lord Stern's commission considered how the next REF exercise, to be launched in 2018, could be improved. The report didn't come as a revolutionary surprise, but includes many valuable proposals for an improved consistency and flexibility of the REF. Peer review should be continued:

"Panels should continue to assess on the basis of peer review. However, metrics should be provided to support panel members in their assessment, and panels should be transparent about their use."

(Lord Stern 2016)

The peer review-based assessment systems that are in place today vary in many aspects; however, they have some principal features in common. They all use **expert panel evaluation** as the central method, but show also other important characteristics that are common to comprehensive research assessments:

- Research is assessed at **different levels** (mostly panels and remote reviews), which demands some aggregation mechanism
- Peers are **informed by different sources**: nearly all countries integrate bibliometric indicators. In Italy, the use of bibliometrics differs between disciplines. Only Portugal relies close-to-uniquely on peer review; as in the UK, bibliometrics is little used
- **Impact outside academia** plays an increasingly important role. Except for Italy and Spain, all countries take some impact dimensions into account.

The sections below summarise the participating countries' answers to a tailored questionnaire on their use of peer review.

3.1 Strengths and weaknesses of peer review – the MLE participants' perspective

Following up on the discussion on the strengths and weaknesses of peer review in Section 3 above, **Error! Reference source not found.** and **Error! Reference source not found.**, below, provide a view on how the topic of peer review is discussed in the MLE participants' national context, based on their responses to a questionnaire.

All countries are well experienced with peer review procedures and struggle with some of the well-known weaknesses. Rules and regulations have been implemented to overcome the problem of conflict of interests and nearly all of them combine international expertise with local knowledge. Nevertheless, the answers reflect the existing ambivalence and the cost

issue is ubiquitous. A comment made by the Turkish participant provides a representative example:

"Peer review is a time consuming and costly process as it involves lengthy consultation with the academic community; however, it improves the outcomes of the assessment procedure."

Especially for smaller countries with limited research resources, the implementation of a lean assessment procedures is challenging.

Table 2: Most important arguments in favour of Peer review

No	Argument	Countries
10	Deep understanding (context, disciplinary cultures)	AT, HR, CZ, IT, MD, NO, PT, SI, SE, TR
5	Trust and acceptance	AT, IT, MD, NO, PT
3	Allows a diversity of opinions to be brought to the table	HR, ES, TR
2	Discursive method, creates scope for debates, facilitates formative elements	MD, NO
	Comments: Peer review is important for the improvement of the research system and adaption to European standards Peer review contributes to the legitimacy of the evaluation procedure, and of the decisions based on its results.	

Table 3: Most important arguments against Peer Review

No	Argument	Countries
2	Conservative ("old boys' networks")	MD, SE
4	Insufficient with interdisciplinary, transdisciplinary, translational, and problem-solving research	HR, MD, NO, SE
5	Conflicts of interest, schools of thought	IT, MD, PT, SI, SE
3	Overuse of the scientific community	AT, ES, SE
6	Time consuming and expensive	AT, CZ, IT, MD, PT, SE
	Comments: <ul style="list-style-type: none"> • Small countries: Insufficient pool of national experts • Contrast among different schools of thought can be particularly severe in some disciplines as, for instance, sociology. • There is always some concern over conflicts of interest, frequently overcome by using panels of foreign experts. • Overuse of scientific community: A bibliometric analysis could provide the basis for the funding decision in 50-60% applications. Only those with peculiarities or special difficulties in evaluation (e.g. multidisciplinary) should be submitted to detailed remote peer review and panel evaluation. • When it comes to PR in PRFS the main arguments have been that it would be inefficient if both the universities and the government use PR to assess research quality. Sweden sees that quality assurance is mainly a responsibility for universities and PR is therefore left for them to use. 	

3.2 Criteria, dimension and outcome

As already discussed in Section 2.2 above, peer review-based research assessment includes a large range of criteria; in some cases, the peers or experts were provided with additional information (mainly bibliometric indicators).

It is not surprising that in all participating countries, both **research quality and impact** are the principal **criteria for assessment (Error! Reference source not found., below)**. In the case of Croatia, impact - with a weight of 45% - is even ranked higher than research quality (35%). Italy and Spain are the two countries that concentrate on research quality only.

For all the other criteria, the patterns in the use of assessment criteria are less clear. Some countries include the promotion of young researchers, internationalisation and collaboration under different nomenclature; another important criterion is the ability to acquire external funding from competitive funds or from other sources outside academia (mainly industry). Other countries (eg Norway) follow a comprehensive approach and include the interplay of research and education. Many countries take aspects of management, efficiency and institutional strategy into account; for Turkey, management competence is a major criterion.

In cases where the assessments are mainly used for formative purposes, such as in Norway, the criteria are not weighted against each other; however, in Norway as in the Czech Republic, the weights differ for different types of organisations.

Table 4: Assessment Criteria

Criteria	HR	EE	IT	MD	NO	PT	SI	ES	TR
Quality/Excellence/Scientific impact	35%	x	100%	x	x	x	40%	100%	39%
Sustainability/Human Resources		x		x	x	x			9%
Impact economy/society	45%	x		x	x	x	40%		12%
Management/Efficiency/	25%				x	x	20%		40%
Interplay of research and education					x				

In all countries, peers receive a good deal of information, data and indicators from different sources:

- A **self-evaluation report**, often including qualitative evaluation. In the case of Croatia all quantitative information and even bibliometrics is integrated into the self-evaluation report. This can be an advantage as the evaluated unit must check the correctness of the data in advance
- Norway as well as Estonia use their **national CRIS** as a source for bibliometric indicators

All countries use some indicators and metrics to inform the peers, in most cases bibliometrics. In this context, one should note two special cases:

- Portugal's PRFS has many characteristics of a 'pure peer review model' as no bibliometric is used
- In Italy, the use of bibliometrics differs between disciplines: no use of bibliometrics is made in the humanities; whereas in the social sciences, the use of bibliometrics is restricted to economics and statistics.

No clear picture can be given to the question how the peers are required to express their final vote and how the scores are linked to the funding formula (in cases where the assessment serves as a basis for funding decisions).

3.3 Selection of peers and process management

In line with the broad range of criteria in use, the **requested qualification profile** for reviewers encompasses not only scientific/scholarly excellence but also industrial background, management skills and experience in assessment processes. Italy and Spain, which both concentrate on research quality only, are exceptions. The smaller the number of panel members, the more demanding the search for appropriate peers turn out to be.

Both Norway and Portugal rely on international peers only; the other countries mix national with international experts. Generally, great care is taken to ensure an appropriate representation in the panels. In Italy and Norway, the guidelines contain detailed instructions for ensuring an appropriate representation.

3.4 Granularity, scope and costs

All the participating countries limit the number of papers and other inputs into the peer assessments whereas bibliometrics covers all research (at least in Italy⁴, Norway, Spain, Slovenia and in the Czech Republic). Furthermore, there are **no restrictions concerning the researchers themselves**. This differs from the UK's practice so far, where only selected researchers can submit. Lord Stern's report recommends that "All research active staff should be returned in the REF" (Lord Stern 2016).

The **number and architecture of the panels** vary widely and range from 1 main panel in the Case of Croatia, Moldova and Slovenia up to 6 main panels and 24 sub-panels in the case of the Czech Republic.⁵ The latter is alike the UK REF which contains 4 main panels and 36 sub-panels.

These important differences are partly explained by both Croatia's⁶ and Estonia's focus on a single university whereas in Italy, Slovenia, Portugal, Turkey, Spain, Czech Republic and Norway, the assessments are carried out nationwide. Croatia and Estonia include on-site visits, most others do not.⁷

Norway as well as the Czech Republic apply a two-step approach with sub-panels, while the other countries follow a one-step approach, partly assisted by remote evaluations. Interestingly enough, the panels' architecture of RCN evaluations are in the process of being reorganised into a multi-step approach. Furthermore, physical meetings are supplemented by remote interventions and supported by SharePoint as a platform.

Abstracting from some country-specific particularities, **three different panel models** can be distinguished:

- Single Panel Model,
- Discipline-specific Panel Model,
- Multi-step Panel Model.

Most countries combine subject-specific international reviews that are carried out remotely with mainly national panel members. Only in Portugal and Norway, all panel members come from abroad.

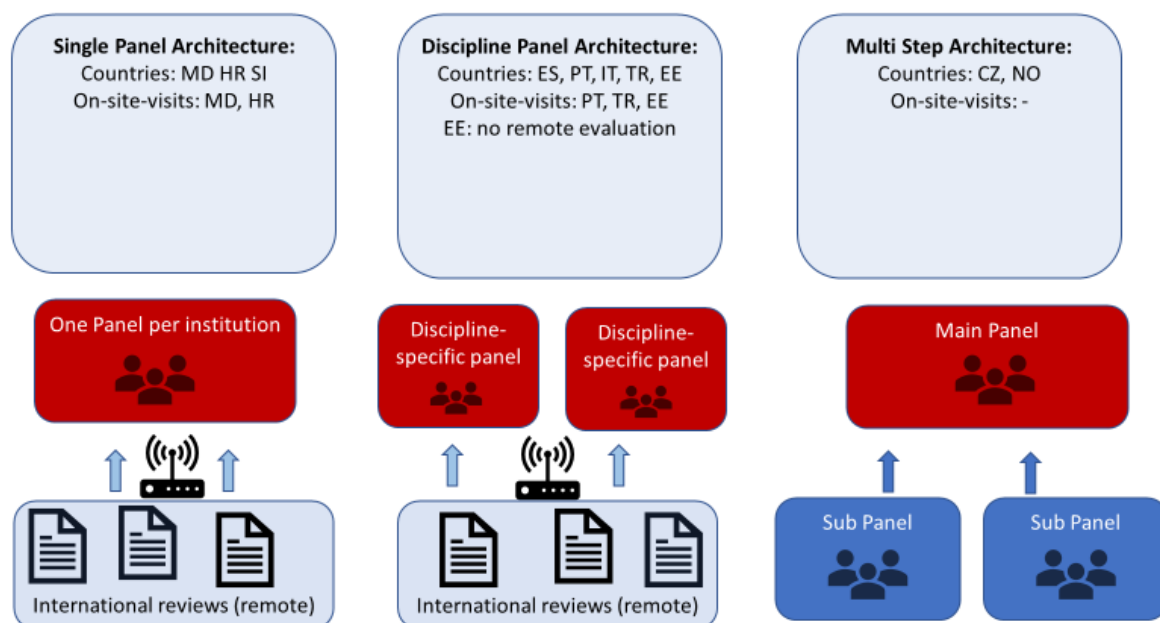
⁴ Note that in Italy no bibliometrics is used in the humanities and in the social sciences apart of economics and statistics.

⁵ The Spanish Sexenio system works even with 20 main panels and 51 sub-panels.

⁶ If universities and research organisations work on similar research topics they are often evaluated together.

⁷ The Czech Republic is "considering" on-site visits.

Figure 4: Different Panel solutions



Error! Reference source not found., below, shows the roughly calculated **costs** for the research assessments in Croatia, Italy, Norway, Slovenia, Portugal, Estonia and the Czech Republic. They range from as little as €5,000 up to €10m because of the considerable variations in assessment and reporting levels (single University versus whole country). In all cases only direct costs are reported; costs carried by the research community for the submission process are not included.

By lack of exact figures on the funds distributed as a result of the assessments, FTE researchers employed in the HE sector are included in order to compare the **costs per researcher** in the different countries (data source: Eurostat, all numbers refer to 2015). The very well investigated case of the UK (REF 2014, see Farla et al 2015) is added.

It doesn't come as a surprise that costs per researcher tend to be higher in smaller countries than in larger ones.⁸ Therefore, small countries with a limited research base may struggle with the implementation of a proper PRFS at reasonable costs (see Section 5.4 below).

Table 5: Costs per researcher

	Yearly direct costs in Mio. €	Researchers FTE in HEI, 2015	€ per researcher
Italy	1,10	74.892	15
Norway	0,50	13.746	36
UK	2,86	188.434	15
Slovenia	0,17	2.555	67
Portugal	0,30	27.766	11

⁸ Please note that this calculation is a very rough estimate.

4 COPING WITH CHALLENGES

4.1 The Panels' architecture

The **composition of different panels** as well as their **appropriate staffing** is one of the biggest challenges in the design of a peer review-based research assessment approach. There are a lot of trade-offs at stake, as shown in **Error! Reference source not found.**, below.

Furthermore, there should be an appropriate balance of gender and age in the peer review panels and the panel chairs should have leadership qualifications and management abilities. In the UK REF, the panels include several international and UK-based experts, as well as traditional academic evaluators, and research users (stakeholders) as evaluators. The user evaluators were predominantly from outside the academic sector and represented a variety of private, public or charitable sectors that either use university-generated research, or commission or collaborate with university-based researchers (see Derrick et al 2015).

Another important point is that to ensure consistency in the assessment of actors in different fields, some kind of **calibration** needs to be organised between the panels. In other words, there should be a common understanding of the assessment criteria, the standards and the application of the quality scores.

Table 6: Panel members' qualifications

To exploit the advantages of peer review, deep and subject-specific scientific/scholarly knowledge must be included	Broad multidisciplinary knowledge is needed, especially to supplement very specific peers in small disciplines
For the acceptance of the outcome, members should be persons respected by the scientific community	Panel members should have time and energy to concentrate on the evaluation
To guarantee independence and to avoid conflict of interest, international peers should be preferred	Local and regional background knowledge is important
Especially for the judgement of impacts, non-academic panel members are important (industry, users, NGOs)	Panel members should share a common language and understand each other

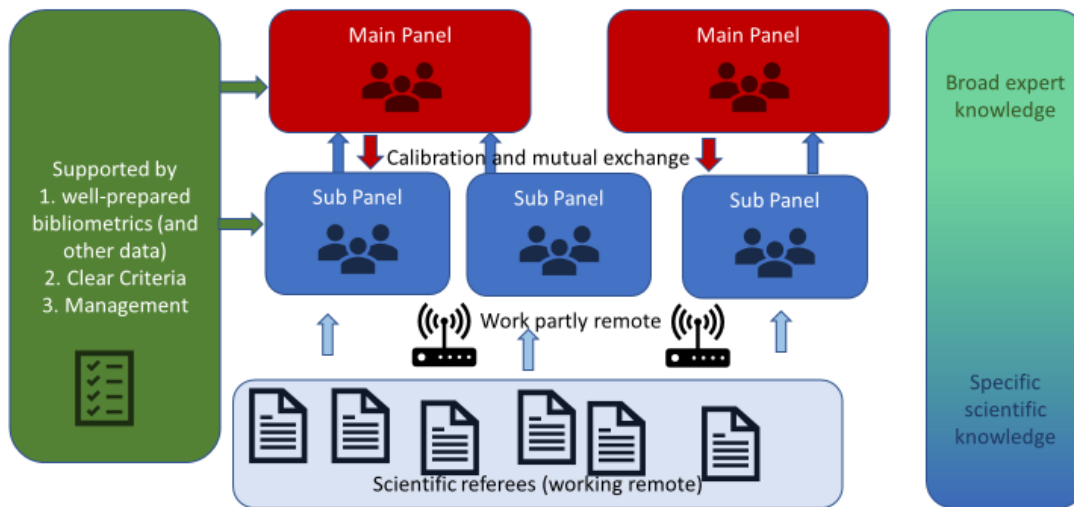
In contrast to this broad range of requirements, the need to control **costs and coordination tasks** implies that the number of persons, interest-groups, aspects and criteria involved should be limited. Costs are one of the most serious arguments against the use of peer review-based evaluations, so mechanisms are needed to **cope with the different requirements of a multi-criteria, multi-interest effort at an affordable price.**

A partial solution to this problem is the introduction of **sub-panels** as well as single academic peers only assessing research quality. With a well-composed division of labour between the different panel stages, even very subject-specific aspects can be considered without losing track of the main issues. Not only the UK, but many other countries as Norway or the Czech Republic are using sub-panels.

Another partial solution is to introduce **'remote'** panel evaluation and renounce to on-site visits (see Mahieu et al 2015). This reduces the costs of the evaluation as well as the burden on the reviewers in terms of time investment. In most cases, the interaction among the experts themselves is kept because of its importance in the decision-making process, so a (minimal) number of physical panel meetings are foreseen. Such an approach has been adopted, for example, in the UK RAE/REF exercises and in disciplinary evaluations in Norway.

Error! Reference source not found., below, shows a possible architecture for the panels.

Figure 5: Panel architecture



4.2 Social dynamic and management

Panel sessions are complex social situations where more or less convincing arguments are exchanged and different thinking traditions, role definitions and behavioural tendencies contribute to the review process. Olbrecht et al. 2007 show how evaluators quickly acquire an “**evaluation culture**” during panel discussions. This is also important as it is this committee culture that ultimately influences the review outcomes (Kerr et al. 1996; Langfeldt 2004), and the future evaluation behaviours of peer reviewers in similar situations.

Panel peer review reaches a common judgement through what Olbrecht and Bornmann (2010) described as “mutual social exchange”, where the final judgement is based on the common judgement of all evaluators. Often, **tacit negotiations and compromises** affect the decision and few panel members dominate the field. Langfeldt (2006) notes:

“Another aspect of the uncertainty in judging research quality is that reviewers often hold different views – including different assessments of the adequacy of scholarly approaches and methods and the scholarly value and relevance of research questions and topics. The outcome of peer review consequently depends on what kind of expertise is included in the review process – for example, which research fields or what kind of interdisciplinarity, and the inclusion of conservative and mainstream-oriented reviewers or more controversial and non-established directions.”

In a critical study of the evaluation of research grants application, the Swedish research Council argues that panel evaluation isn’t **gender neutral** and proposes some measures (Ahlqvist 2014). Further problems occur if panel members are poorly prepared, are unsure about their role or simply shy.

Many of these weaknesses can be neutralised by means of **procedural guarantees** and a good **process management**. The remedies can be summarised as follows:

- Use **unambiguous criteria** and define participants’ **roles** clearly,
- **Prepare** the meetings with care and structure them thoroughly,
- Offer **trainings** for both management stuff and panel members.

4.3 Informed peer review: the triangulation of methods

Peer review is still the **dominant method** in the context of panel assessments. And there are good reasons for this, as the assessment of quality requires a detailed understanding of the research in question. It is widely accepted – especially in the research community – that only peers possess the specific knowledge needed to reflect the distinctive features of the disciplinary cultures and traditions of thinking.

The independent review of the Research Excellence Framework was guided by a Steering Group and included views provided by the community through over 300 responses to the Call for Evidence. The review states:

"Responses to the Call for Evidence reiterate the importance of peer review. They argue that, with the exception of some sub-disciplines, metrics capture only some dimensions of output quality. However, applying the 'gold standard' of peer review does depend on panels having a very broad range of expertise and sufficient time to analyse each output in detail. At best, peer review is not a perfect 'measure', and with the time pressures on some REF panels, maintaining consistency and quality of review is very challenging." (Lord Stern 2016)

The idea that the **combination of peer review with metrics**, ie quantitative indicators and especially bibliometric indicators may enhance the evaluation process lays behind the introduction of 'informed peer review', where peers get informed by metrics which allows comparisons and triangulation across methods.

Some authors argue that bibliometric indicators can make peer review **more transparent and offer additional insight** in cases of diverging panel members' opinions (eg Arnold 2012). Norway, with its "dual system" of funding and assessment tools points in the same direction:

"There is little discussion on the problems of peer review. We use bibliometrics as support for qualitative assessment. Metrics is used in order to make comparisons and aggregate assessments across disciplines." (Norway, quoted out of the questionnaire)

One very practical argument in favour of informed peer review would be that peers are much better informed by professional and properly designed indicators than by a quickly googled h-Index.

Critical voices remark that errors and distortion of the peer review cannot be balanced by the combination with bibliometrics as both are prone to the same distortions (mainly conservatism). Furthermore, the '**Matthew Effect**' will be strengthened by the interaction of quantitative and qualitative methods. (e.g. Fröhlich 2006)

Out of the countries participating in the MLE, Slovenia, Italy, Turkey, Portugal and Spain apply PRFS based on informed peer review. Others show combinations of peer review and bibliometrics in different forms.

4.4 Costs of Peer Review

Much of the debate about the value of performance-based funding allocation systems focuses on the disadvantages versus the benefits of their implementation; there is very little evidence on their absolute costs or their costs relative to other funding allocation systems.

This contrasts with the fact that **costs are one of the most important issues** when it comes to the decision of implementing a PRFS. Furthermore, costs are extremely difficult to calculate (see eg the estimation in Section 3.4). Internal and direct cost estimates for the research organisations are high, and if the opportunity costs of refereeing and panel membership are included into the estimation, the total costs increase dramatically.

In the case of the UK, some ambitious efforts were made to estimate direct and indirect costs of RAE and REF:

- An independent consultancy, PA Consulting Group, estimated the cost of RAE 2008 at £60m (PA Consulting Group, 2008).
- For REF 2014, the figure was £55m solely for the cost of participating HEIs (Manville 2015). According to an assessment made by Technopolis (Farla et al 2015), the total cost of REF 2014 was £246m; the HEIs accounted for most of the costs (around £212M for the submission process and around £19M for panellists' time).

Of course, all these costs may include double-counting, reflecting the challenge of distinguishing additional REF-related costs from 'business as usual' (i.e. the underlying cost of managing research quality). For a balanced view of the whole exercise, benefits should be taken into account as well: direct benefits such as the development opportunities for institutions receiving more funding as well as indirect benefits such as researchers' focus on quality and change of publishing behaviour. On the other hand, much academic time was spent on discussing, reflecting, and protesting. Even for a large country such as the UK, the RAE/REF turns out to be an expensive endeavour.

Geuna et al (2015) compared costs of peer-review based PBFS in Italy and the UK with other allocation systems. They conclude that the costs of peer-review systems are:

"significantly higher than both historical and metrics based allocation systems when the internal and direct costs of PRBRA [peer review based research assessment] are considered in relation to the funding allocated, the ratio is lower than in the case of project based allocation systems such as in the case of the Research Councils in the UK." (Geuna et al 2015)

All these comparisons are highly complex because the costs of metric-based allocations depend on the quality of the national information systems in place, whereas the costs of project-based allocations are related to the success rates. Very low success rates may result in arbitrary decisions at high costs.

Furthermore, the significant amount of fixed costs can discourage the implementation of peer review-based funding systems especially in small countries that experience a decline in total public funding.

In the future, developing smart and tailored PRFSs especially for smaller countries will remain a central challenge. **International co-operations and joint assessments** with other countries should be considered to share both costs and experience.

5 CONCLUSIONS

There is no clear evidence that systems with a peer review-based evaluation culture - whether or not linked to the funding decision – systematically perform better than others. The Swiss system, which is at the top of nearly all European rankings both in terms of innovation and scientific excellence, is solely based on the “Harnack-Principle”.

The experiences of the MLE participating countries differ widely due to differences in size, in research orientation, and in national evaluation cultures. Nevertheless, both the peer review systems applied by the countries and most of current literature allow for the following summary of **common issues and challenges**:

- Peer review is **indispensable** when the assessment of quality requires a detailed understanding of research in question
- **Costs** are one of the most important issues when it comes to the decision of implementing a PRFS. Internal and direct university cost estimates are high, and if the opportunity costs of refereeing and panel membership are included into the estimation, the total costs increase dramatically. The significant amount of **fixed costs** can discourage the implementation of peer review-based funding systems, especially in **small countries**
- Apart from scientific excellence, **wider socio-economic criteria like relevance, impact, and new forms of interaction with society become increasingly important**. Consequently, new forms of peer review considering multiple criteria and relying at least partly on on-scientific peers were developed
- To establish and maintain confidence in the system, procedures and rules should be developed to **avoid conflicts of interest**. Especially in small countries, international expertise should be integrated
- A well-balanced **mix of different measures** (eg peer review and bibliometrics) may offer additional insight and enhance the outcome of the research assessment. Both high quality of the **bibliometric indicators** and transparency in the use by peers or panels are important in this context
- Research is assessed at **different levels** (mostly panels and remote reviews) which demands some aggregation mechanism. The **composition of different panels** as well as their **appropriate staffing** seems to be one of the biggest challenges in the design of peer review-based research assessment approaches. In order to ensure consistency some kind of **calibration** between different disciplines, interest groups and different panels is needed, and social dynamics shouldn't be ignored

Finally, some concrete good practice and recommendations can be formulated on how benefits can be achieved by applying a properly designed peer review-based assessment:

- **High costs** of peer review can be mitigated by introducing:
 - Sub-panels and remote review,
 - International co-operations both in the design and in the application phase.
- **Other limitations** of peer review can be mitigated by:
 - Adequate management of processes and social situations,
 - Careful selection of panel members,
 - Properly designed bibliometrics,
 - Specific rules concerning conflict of interest and interdisciplinary research.

6 ANNEX A: REFERENCES

- Abramov, G., D'Angelo, C., di Costa, F. (2011). National research assessment exercises: a comparison of peer review and bibliometric rankings. *Scientometrics*, 89, 929-941.
- Ahlqvist, V., Andersson, J., Söderqvist, L. (2014). A gender neutral process? *Swedish Research Council*.
- 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 and Industrial Strategy.
- Bornmann, L. (2011). Scientific Peer Review. *Annual Review of Information Science and Technology*, Section III: Scholarly Communication 45 (1),197-245
- Bornmann, L. (2013). What is the societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society of Information Science and Technology* 64, 217-233.
- Carayannis, E., Buzás, N. (2016). Special Issue on Targeted Open Innovation. *Journal of the Knowledge Economy* 7.
- Chambers, C. (2014). A step change in scientific publishing. <https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/registered-reports-a-step-change-in-scientific-publishing>
- Cole S., Cole J., Simon G. (1981) Chance and consensus in peer review. *Science* 214, 881-886.
- Deiters, H. (1960): Wilhelm von Humboldt als Gründer der Universität Berlin, in „Forschen und Wirken“ Festschrift zur 150-Jahr-Feier der Humboldt-Universität zu Berlin 1810 – 1960. Bd 1 „Beiträge zur Wissenschaftlichen und Politischen Entwicklung der Universität“ Berlin: VEB Deutscher Verlag der Wissenschaften.
- Derrick, G., Samuel, G. (2016). The Evaluation Scale: Exploring Decisions About Societal Impact in Peer Review Panels. *Minerva* 54, 75-97.
- EERE Peer Review Guide (2004): Based on a Survey of Best Practices for In-Progress Peer Review. <https://energy.gov/sites/prod/files/2016/08/f33/EERE%20810%20-%20Peer%20Review%20Guidance.pdf>
- European Commission (2014): *Responsible Research and Innovation. Europe's ability to respond to societal challenges*. Brussels: European Commission.
- European Commission (2016): *Open Innovation, Open Science, Open to the World – A Vision for Europe*. Brussels: European Commission.
- Fang, H. (2011). Peer review and over-competitive research funding fostering mainstream opinion to monopoly. *Scientometrics* 87, 607-616
- Farla, K., Simmonds, P. (2015). *REF 2014 Accountability Review: Costs, benefits and burden*. <http://www.hefce.ac.uk/pubs/rereports/year/2015/refreviewcosts/>
- Fitzpatrick, K. (2009). *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. NYU Press: New York University.
- Fröhlich, G. (2006). "Informed Peer Review": Ausgleich der Fehler und Verzerrungen? Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung? *Hochschulrektorenkonferenz HRK*, 193-204.
- Geuna, A., Piolatto, M. (2015). Research Assessment in the UK and Italy: Costly and Difficult, But Probably Worth it (At Least For a While). SPRU working paper

<https://www.sussex.ac.uk/webteam/gateway/file.php?name=2015-27-swps-geuna-piolatto.pdf&site=25>

Gibbons, M. Georghiou L. (1987). *Evaluation of Research: A Selection of Current Practices*, Paris: OECD.

Greaves, S., Scott, J., Clarke, M., Miller, L., Hannay, T., Thomas, A., Campbell, P. (2006). Nature 's trial of open peer review. *Nature* 2006. doi:10.1038/nature05535. <http://www.nature.com/nature/peerreview/debate/nature05535.html>

HEFCE (2017). *REF Impact* <http://www.hefce.ac.uk/rsrch/REFimpact/>

Hodgson C. (1997). How reliable is peer review? A comparison of operating grant proposals simultaneously submitted to two similar peer review systems. *J Clin Epidemiol* 50, 1189-1195.

Hopewell S., Loudon K., Clarke M., Oxman A., Dickersin K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*

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

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

Kerr, N., Robert J., Kramer, G. (1996). Bias in judgement: Comparing individuals and groups. *Psychological Review* 103, 687-719.

Langfeldt, L. (2004). Expert panels evaluating research: Decision-making and sources of bias. *Research Evaluation* 13, 51-62.

Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and multidisciplinary assessments. *Research Evaluation* 15, 31-41.

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

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

Mandavilli A. (2011). Trial by Twitter. *Nature* 469, 286-287.

Martin, B. (2011). The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? *Research Evaluation* 20, 247-254.

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

Merton R. (1968). The Matthew Effect in Science. *Science* 159, 56-63.

Merton, R. (1973) [1942], The Normative Structure of Science, in Merton, R., *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago: University of Chicago Press.

Nature Editorial (2006). Peer review and fraud. Two assessments of the refereeing process highlight challenges for journals. *Nature* 444, 971-972

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

Olbrecht, M., Bornmann, L. (2010). Panel peer review of grant applications: What do we know from research in social psychology on judgement and decision making in groups? *Research Evaluation* 19, 293–304.

Olbrecht, M., Tibelius, K., D'Aloisio, G. (2007). Examining the value added by committee discussion in the review of applications for research awards. *Research Evaluation* 16, 79–91.

PA Consulting Group (2009). *RAE 2008 Accountability Review*. <http://www.hefce.ac.uk/pubs/rereports/year/2009/RAE,2008,Accountability,Review/>

Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Front. Comput. Neurosci.* <https://doi.org/10.3389/fncom.2012.00033>

Royal Society (1995). *Statement of the Royal Society on Peer Review 1995* <https://royalsociety.org/topics-policy/publications/1995/peer-review/>

Ruegg, R., Jordan, G. (2007). *Overview of Evaluation Methods for R&D Programs*. Prepared for U.S. Department of Energy Office of Energy Efficiency and Renewable Energy. https://www1.eere.energy.gov/analysis/pdfs/evaluation_methods_r_and_d.pdf

Scarpa, T. (2006). Research funding: peer review at NIH. *Science* 311, 41-44. DOI: 10.1126/science.1122796

Smith, R. (1999). A beginning that should lead to complete transparency, *BMJ* 318 doi: <https://doi.org/10.1136/bmj.318.7175.4>

Spier, R. (2002). The History of the Peer-Review Process. *Trends in Biotechnology* 20, 357-358.

Stern, Lord N. (2016). *Building on Success and Learning from Experience. An Independent Review of the Research Excellence Framework*. <https://www.gov.uk/government/publications/research-excellence-framework-review>

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

Wenneras C., Wold A. (1997). Nepotism and sexism in peer review. *Nature* 387, 341-343

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

Wilsdon, J. et al (2016). *The Metric Tide. Independent Review of the Role of Metrics in Research Assessment and Management*. DOI: 10.13140/RG.2.1.4929.1363 HEFCE. London.

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.



The focus of this report is on the use of peer review in the context of such research funding systems. It supports the discussion and mutual learning among the participating countries on different aspects, issues and questions related to peer review-based assessment systems. Few countries use peer review as an evaluation tool in the context of PRFS, so when discussing concepts and good practice, the scope of this paper is broadened to include also national research assessment systems that are not directly linked to funding distribution (e.g. the Standard Evaluation Protocol in the Netherlands).