European Commission

Mutual Learning Exercise

Performance-Based Funding of University Research

Horizon 2020 Policy Support Facility

Research and Innovation

Performance-Based Funding of University Research

European Commission
Directorate-General for Research and Innovation
Directorate A — Policy Development and Coordination
Unit A.4 — Analysis and Monitoring of National Research and Innovation Policies

Contact (sector H2020 PSF):

Marta Truco Calbet, Responsible Policy Officer, marta.truco-calbet@ec.europa.eu
Román ARJONA, Chief Economist and Head of Unit A4 - Roman.ARJONA-GRACIA@ec.europa.eu
Stéphane VANKALCK, PSF Head of Sector, Unit A4 - Stéphane.VANKALCK@ec.europa.eu
Diana SENCZYSZYN, PSF Team Leader, Unit A4 - Diana.SENCZYSZYN@ec.europa.eu

RTD-PUBLICATIONS@ec.europa.eu
European Commission
B-1049 Brussels

Manuscript completed in December 2017

Luxembourg: Publications Office of the European Union, 2018

PDF ISBN 978-92-79-80413-7 doi: 10.2777/644014 KI-AX-18-003-EN-N

EUROPEAN COMMISSION

Mutual Learning Exercise

Performance-Based Funding of University Research

Koenraad Debackere (Chair)

Erik Arnold (Rapporteur)

Gunnar Sivertsen

Jack Spaapen

Dorothea Sturn

Directorate-General for Research and Innovation

2018 EN

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Full name
AM	Armenia
ASCR	Technology Centre of the Academy of Sciences of the Czech Republic
AT	Austria
AU	Austria
BA	Bachelor of Art
BMWFW	Federal Ministry of Science, Research and the Economy
BOF	Special Research Fund (Belgium, Flanders)
CNEAI	National Commission for the Evaluation of Research Activity
COST	European Cooperation in Science and Technology
CRAC	Careers Research & Advisory Centre
CRIS	Current Research Information System
CROSBI	Croatian Scientific Bibliography
CWTS	Centre for Science and Technology Studies (Leiden University)
CZ	Czech Republic
CZK	Czech koruna (currency)
DE	Germany
DG	Directorate-General
DG RTD	Directorate-General for Research and Innovation, European Commission
DK	Denmark
DOI	Digital Object Identifier
DORA	Declaration On Research Assessment
EC	European Commission
EE	Estonia
EN	English
ENID	European Network of Indicator Designers
ENRESSH	European Network for Research Evaluation in the Social Sciences and Humanities
ERA	European Research Area

Abbreviation	Full name
ERAC	European Research Area and innovation Committee
ERC	European Research Council
ERIH	European Reference Index for the Humanities
ES	Spain
ETIS	Estonian Research Information System
EU	European Union
EUR	Euro
FAO	Food and Agriculture Organisation
FFO	Ordinary Financing Fund
FI	Finland
FL	Flanders
FP	EU Framework Programme for Research and Innovation
FR	France
FTE	Full-Time Equivalent
FWC	Framework Contract
FWF	Austrian Science Foundation
GS	Google Scholar
GSM	Global System for Mobile communications
HE	Higher Education
HEFCE	Higher Education Funding Council for England
HEI	Higher Education Institute
HEPI	Higher Education Policy Institute
HERD	Higher Education Expenditure on R&D
HR	Croatia
HV	Switzerland
IOF	Industrial Research Fund
IP	Intellectual Property
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
ISI	Institute for Scientific Information

Abbreviation	Full name
ISSN	International Standard Serial Number
IT	Information Technology
JIF	Journal Impact Factor
JRC	Joint Research Centre
LT	Lithuania
MA	Master's of Art
MD	Moldova
MIT	Massachusetts Institute of Technology
MLE	Mutual Learning Exercise
MRC	Medical Research Council
NA	Not Available
NATO	North Atlantic Treaty Organisation
NGO	Non-Governmental Organisation
NIFU	Nordic Institute for Studies in Innovation
NL	Netherlands
NO	Norway
NOK	Norwegian Krone
NSF	National Science Fund
NSTS	National Science and Technology System
NZ	New Zealand
OECD	Organisation for Economic Co-operation and Development
OSIRIS	Oslo Institute for Research on the Impact of Science
PBRF	Performance-Based Research Funding
PEFS	Process Engineering Flow Scheme
PL	Poland
PRFS	Performance-based Research Funding Systems
PRO	Peer Review Organisation
PSF	H2020 Policy Support Facility
PT	Portugal
QR	Quality-related Research

Abbreviation	Full name
RAE	Research Assessment Exercise
RCN	Research Council of Norway
RDI	Research, Development and Innovation
REF	Research Excellence Framework
REPP	Research Embedment and Performance Profile
RFF	Regional Health Agency
RIS	Research Information System
RU	Russia
RUM	Research Uptake Management
SE	Sweden
SEP	Standard Evaluation Protocol
SHARE	Survey of Health Ageing and Retirement in Europe
SI	Slovenia
SIAMPI	Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society
SNIP	Source Normalised Impact per Paper
SP	Spain
SSH	Social Sciences and Humanities
STEM	Science, Technology, Engineering and Mathematics
STI	Science, Technology and Industry
STOA	Science and Technology Option Assessment (European Parliament)
TIPS	The Research and Innovation Policy Monitoring Programme
TR	Turkey
TV	Television
UK	United Kingdom
UNESCO	United Nations Educational Scientific and Cultural Organisation
US	United States
VABB-SHW	Flemish Academic Bibliography for the Social Sciences and Humanities
VQR	Evaluation of Research Quality (Italy)

Abbreviation	Full name
VTR	Valutazione Triennale della Ricerca
WHO	World Health Organisation
WOS	Web of Science
WWTF	Vienna Science and Technology Fund

SUMMARY

This Mutual Learning Exercise (MLE) on performance-based research funding systems (PRFS) was conducted under the Horizon 2020 Policy Support Facility run by the Directorate-General for Research and Innovation of the European Commission.

The countries that participated were Armenia, Austria, Croatia, the Czech Republic, Estonia, Italy, Moldova, Norway, Portugal, Slovenia, Spain, Sweden and Turkey. The MLE was supported by a panel of experts: Koenraad Debackere (Chair), Erik Arnold (rapporteur), Gunnar Sivertsen, Jack Spaapen, and Dorothea Sturn. The MLE was overseen by Marta Truco from Unit 4 'Analysis and monitoring of national research and innovation policies', DG Research & Innovation, European Commission. The contributions of Bea Mahieu (project manager and quality reviewer) are acknowledged.

The work of the panel of experts was based on written and oral contributions from the participating states and field visits to some of them as well as wider literature and experience of PRFS. Regular meetings both in Brussels and in participating states provided considerable scope for discussion and exchange of ideas. Although not participating in the MLE, the Higher Education Funding Council of England kindly invited the members of the MLE to visit it in London and shared its experience of the UK system with the team.

Important lessons and recommendations emerged from this MLE. They are given *in extenso* in Chapter 7 and are summarised at a high level here.

Lessons learnt

PRFS are a useful addition to the set of instruments and policies governments can use to improve the quality and performance of university research systems.

In general, universities obtain income partly through 'institutional funding' and partly through 'external funding' for individual projects or programmes. Institutional funding may be provided as an unconditional 'block grant' or as 'performance-based funding' on the basis of past performance. **Performance-based research funding systems (PRFS) are two-part systems that assess the 'goodness' of university research on various dimensions** and use the results in a funding formula to allocate part or all of the institutional funding for research provided to universities. Most PRFS focus on the quality of research and

its scientific impact. In more recent systems, a growing number of other criteria are also being used, especially in relation to societal impact.

A PRFS is one of several mechanisms available to policymakers to encourage the improvement of quality, relevance, impact and other desiderata in university research. There is wide divergence among countries in their mix of funding mechanisms and no clear theory about what might constitute an 'optimal' balance. As with some other policy options, the introduction of a PRFS can administer a shock to the research system, changing culture and behaviour. Striking a balance between shock and continuity is another challenge for designers of PRFS. The PRFS and the other parts of the policy mix need to address policy goals in a complementary manner.

In introducing a PRFS, it is important to consider its systemic implications in the sense of its influence on the overall pattern of funding and incentives and whether the incentive system is then balanced so that universities are encouraged to fulfil well all their three missions of education, research and exchange of knowledge with wider society and not improve one at the cost of the others.

PRFS design is important not only in itself but also because there is a tendency for the design to be institutionalised into the way universities are funded and manage themselves, creating lock-in. Key parameters for PRFS designers include: the assessment model; the scope of research included; types of indictors used; assessment criteria; the granularity of the analysis; and the frequency (periodicity) with which the PRFS can be run. Most PRFS focus on the quality of research and its scientific impact.

There is continued disagreement about **whether peer review or metrics approaches are 'best'**. However, the greater the proportion of universities' research income governed by a PRFS, the more robust its methods need to be in order to withstand scrutiny by the beneficiaries. Small countries face particular design issues. The costs of running a PRFS appears to be driven by scale, the model used, and a range of contextual factors.

PRFS outcomes are influenced by the characteristics and quality of the information upon which the research assessment is based. Self-assessments are necessary for collecting qualitative or PRFS-specific information. It is preferable to obtain other data from quality-assured sources. A quality-assured current research information system (CRIS) is a very useful tool to support PRFS, especially in countries using 'small' languages.

Bibliometric tools are now widely available and can be used for PRFS. However, **generating and using robust bibliometric indicators requires skill** and considerable understanding of the limitations of the relevant databases and the characteristics of different fields of research. Hence, professional bibliometricians should be involved in the design and implementation of any PRFS that uses bibliometric information. Stakeholders should also be consulted.

Qualitative aspects of assessment can only be analysed by peer or expert panels. The composition of different panels as well as their appropriate staffing is one of the biggest challenges in the design of **peer review-based research assessment** approaches. Appropriate panel structures and measures to prevent conflicts of interest need to be put in place. The right choice of information provided to peer review panels depends upon the purposes of the PRFS and the criteria used. A well-balanced mix of different measures (e.g. peer review and bibliometrics) may offer additional insight and enhance the outcome of the research assessment.

Peer-based PRFS often make use of **self-assessment**s. These make it possible to introduce a formative element into the PRFS. Using a **cost-effective mix of panel and remote reviews** demands an aggregation mechanism. To ensure consistency some kind of calibration between different disciplines, interest groups and different panels is needed.

There is no uniform or comprehensive set of **societal impact** indicators available, despite the considerable research effort that continues to be devoted to the task. Nor is there any case of systematic integration of third stream indicators into a PRFS in the countries studied. Many researchers are not used to analysing or reporting impact, so the quality of their self-reporting is likely to be variable. PRFS therefore have to use more or less ad hoc measures, preferably combining quantitative and qualitative methods.

Different **funding formulae** provide incentives for different kinds of behaviour. Overhead and infrastructure costs vary among fields. Consideration should be given to weighting the formula in order to take this into account. This appears to be especially important where the PRFS governs a large proportion of institutional income, leaving universities little 'slack' in using the block grant to tackle variations in costs among disciplines. The opportunities to use **PRFS results as a source of strategic intelligence** for policy are often under-exploited and are not necessarily considered in PRFS design.

PRFS favour the adoption of more performance-orientated policies, providing incentive mechanisms for the governance of the University system. Possible **uses of PRFS** include: favouring better recruiting

mechanism at the university level; greater transparency in the local allocation of funding at the Department level; providing important information to support the organisation of graduate and PhD courses; allowing the introduction of new policy instruments aimed at funding "excellence" both at the Department and individual level. Since the trend towards including incentives for non-scholarly performance in PRFS is relatively new, it is too early to expect to see much evidence that these are affected though there is anecdotal evidence that incentives such as those encouraging PhD education are effective.

While **performance generally rises after the introduction of PRFS**, it is also rising in countries with no PRFS and changes in performance are sometimes triggered by other policy changes than the introduction of a PRFS. It seems that PRFS can contribute to structural changes as part of larger processes of output-orientation rather than always being the unique cause of such changes.

Literature about the **effects of PRFS** indicates that university managers often believe PRFS improve performance while other academics are inclined to point to actual and potential negative effects on the character of research and on research carers.

Key risks identified in the scientific literature and in researcher surveys in connection with PRFS include: discouraging interdisciplinary research; discouraging 'blue skies', 'high risk' and 'transformative' research; discouraging research in non-mainstream topics; discouragement of applied research; under-valuing research published in languages other than English; reducing researcher autonomy; discouraging collaboration; and discouraging popularisation of science and third stream activities. While these concerns are credible enough to warrant investigation, there is too little systematic evidence about them. Some of these concerns are being considered in PRFS designs. For example, current research information systems have been build that incorporate national-language literature in order to combat the English-language bias, count joint publications in ways that encourage research collaboration and reward universities' education and knowledge exchange missions as well as research.

There is some evidence that because university research managers are not well able to predict the results of a PRFS, they encourage low-risk research and use misleading proxies in research management. **Effects of PRFS on universities and research are mediated by universities' human resource and research management practices**. They are likely over time to change the composition of the faculty, building any positive or negative effects of the PRFS into to the staffing and culture of the university.

Recommendations

PRFS in the policy toolkit

- Depending upon the specific developmental needs of the university research system, states should consider adopting a PRFS or an appropriate alternative if the national university system's research performance is in need of improvement
- The scope and design of the PRFS should be consistent with other policy measures operating and with the national context
- In particular, the proportion of institutional funding governed by the PRFS should be given careful consideration, based on national policy needs and the likely interplay between the PRFS and other policy instruments
- Both the assessment criteria and the funding formula used in the PRFS should encourage behavioural and institutional changes that reflect the policy goals
- Care must be taken in introducing new policy goals that old ones are not impeded, for example by creating a focus on research performance that is not balanced by incentives for universities to pursue their two other missions of education and knowledge exchange with society
- Where PRFS risk trigger abrupt changes in funding, use damping mechanisms that limit these changes to levels that are sustainable

PRFS design

- Designers should anticipate that it may be hard to make more than incremental adjustments to the design once the PRFS has been fully implemented
- Design requires expertise in the design and implementation of assessment and funding systems and in technical aspects of the assessment process, such as the use of bibliometrics, peer review and impact indicators
- The choice and weighting of criteria or indicators is intended to affect researcher behaviour. The likely intended and unintended effects of the PRFS on behaviour should be anticipated as far as possible at the design stage
- The assessment process and the funding formula should be distinct and independent
- Wherever possible, the effects of a PRFS should be simulated or piloted before the system is put into full-scale practice

- PRFS should be simple to understand and easy to communicate to universities and researchers alike
- They should avoid using large numbers of criteria and indicators as these create conflicting incentives and researchers cannot satisfy them all
- PRFS designers should routinely consult both the DORA declaration and the Leiden Manifesto during the design process

PRFS model

- The choice of a peer review-based or a metrics-based system should be made based on the needs that the PRFS is intended to satisfy and the constraints under which it will operate
- Where a PRFS reviews a sub-set of research output, the biases created through selection must be understood
- A combination of the two approaches is often useful and should be considered
- The character and biases inherent in bibliographic and other databases as well as in other metrics should be understood and accounted for
- Self-assessment is a useful tool in research assessment, especially where a formative element is desired. Where self-assessment is used, it should involve transparent questions which have been clearly explained by the PRFS managers. Researchers not used to self-evaluation may need particular guidance. Self-assessment should not be used to collect data available from quality-assured sources elsewhere as this increases the burden on those completing self-assessments and introduces errors
- A national CRIS should be considered as a way to maintain a high quality of input data to the PRFS, especially in smaller countries that work in 'small' languages

Bibliometrics

- Professional bibliometricians (independent of the commercial database vendors) should be involved in the design and implementation of any PRFS that uses bibliometric information
- Building advanced bibliometric expertise requires a significant investment in a research centre in the field. There are a handful of such centres in Europe already. Especially small countries with limited resources should exploit existing expertise abroad rather than necessarily themselves making a heavy national investment in one

Peer review

- Care should be taken in defining peer review panel membership not to predetermine outcomes by excluding relevant disciplines or schools of thought and to ensure that members have a track record of being 'team players'
- Where criteria other than scientific quality are to be considered, panels should be 'extended' beyond academic peers to include relevant representatives of other parts of society
- Procedures should be in place to prevent conflicts of interest. In small countries, all peers should be international. A good proportion of international peers is also desirable in larger systems, to maintain contact with world science
- A PRFS cannot be run by a single panel. There have to be sub-panels and/or panels informed by remote peer assessment. Judgements have to be calibrated and made consistent across different panels and reviewers

Societal impact

- Societal impacts of research arise partly because of the value of new knowledge and partly through a range of other contextual factors, not the least of which may be luck. PRFS experimenting with assessing societal impacts should consider whether it is better to reward outputs and outcomes that can reasonably be expected to be steps on the way to impact or whether they want to reward impact itself
- Impact metrics should not be used on their own in a PRFS. Their meaning is unclear and they are easily gamed
- At the current state of the art, human judgement is the only way to assess impact. This can be supported by metrics but these must be interpreted by people
- Extended peer review of impact narratives can be used, though defining how these are to be prepared and assessed in a consistent way is very labour-intensive. Assessments are subjective and it is not known whether they are reproducible
- Impact assessment should be attempted with caution and can in certain cases be included in the assessment module of a PRFS while being excluded from the funding formula so that the incentive to generate impact is prestige rather than money

Using the results of the assessment

- Funding formulae should reflect the policy priorities of the PRFS. They can be skewed towards high-performing institutions if the desire is to concentrate funding in fewer places. They can drive quality or quantity or both, they can be used to adjust the relative incentives for researchers to perform along different dimensions (such as scientific quality versus societal impact), and so on
- In systems where the PRFS steers a high proportion of institutional funding, it is especially important that funding formulae take account of the need to align with influences on the institutional cost of doing research such as variations in cost among disciplines and the need to pay the overhead costs of doing research for external funders such as research councils, charities and sometimes industry that can or will not pay the full economic costs
- Assessment results can provide useful strategic intelligence for policymaking, such as offering a picture of national research performance in different fields. It is not necessarily difficult to combine and recombine assessment results at different levels to provide such strategic intelligence. PRFS systems should do so, where they are able to make a unique contribution to policy development

Intended and unintended effects of PRFS

- Member States should consider evaluating their PRFS periodically, if possible aiming not only to describe the gross effect of the current range of policies on research performance but also the net effect of the PRFS itself and the extent to which it leads to behavioural and organisational changes that render the PRFS itself redundant
- They should also conduct specific studies in order to search for stronger evidence about potential negative consequences
- Member states should understand those potential risks and consider them when deciding on the mix of university funding instruments deployed and in the design and architecture of their PRFS.
- Internationally comparative studies and evaluations are also needed in order to start to disentangle the operation of PRFS from national contexts

1 INTRODUCTION

This document reports the results of a mutual learning exercise (MLE), in which representatives from Armenia, Austria, Croatia, the Czech Republic, Estonia, Italy, Moldova, Norway, Portugal, Slovenia, Spain, Sweden and Turkey shared their experience of planning and implementing performance-based research funding systems (PRFS). Although not a member of the MLE, the Higher Education Funding Council for England (HEFCE) kindly received the members of the exercise at its office in London and presented information about the UK system[1]. The MLE members made two country visits – to the UK and Italy – in addition to meeting several times in Brussels. The exercise was organised and funded under the H2020 Policy Support Facility managed by Directorate General for Research and Innovation of the European Commission during 2016-7 and was led by a panel of experts under the chairmanship of Koenraad Debackere.

1.1 Background of the H2020 Policy Support Facility

To support countries in reforming their research and innovation (R&I) systems, DG Research and Innovation has set up a 'Policy Support Facility' (PSF) under Horizon 2020, aimed at "improving the design, implementation and evaluation of R&I policies". The PSF provides best practice, leading expertise and guidance to Member States and Associated Countries (on a voluntary basis) through a broad range of services to address their specific needs.

The PSF is a response to the strong need for more services to support evidence-based policymaking expressed by Member States (at the European Research Area and Innovation Committee, ERAC).

The PSF offers three main services to Member States and Associated Countries.

- **Peer reviews of national research and innovation (R&I) systems**, which are in-depth assessments of individual countries' R&I systems carried out by panels of experts and leading to concrete recommendations to the national authorities on reforms necessary to strengthen their R&I system

[1] The UK PRFS is implemented differently in each of the four nations of the UK (England, Wales, Scotland and Northern Ireland). We focus in this document on the English system, but for simplicity refer to it as the 'UK' system since it is by far the largest

- **Specific support to countries**, which can take form of 'pre-peer review' (providing a solid evidence-base and focus areas for a subsequent full peer review), 'post peer review' (providing concrete advice on how to adopt, adapt and implement peer review recommendations) and 'ad-hoc requests' (providing concrete recommendations on how to tackle specific R&I policy issues and implement reforms)
- **Mutual learning exercises** (MLEs), which are demand-oriented, expert-led learning exercises focused on specific R&I topics of interest to several participating countries, resulting in an exchange of good practice

1.2 Scope, objectives and outcomes of the Mutual Learning Exercise

ERAC expressed an initial interest in launching a Mutual Learning Exercise on PRFS via the Horizon 2020 Policy Support Facility in 2015. A call for interest was launched in June 2016 asking ERAC delegates wanting to participate briefly to describe the major challenge(s) they wished to address and their expectations of the MLE. Responses to the Scoping Paper were received from a first group of countries (Czech Republic, Austria Cyprus Estonia, Portugal, Italy and Norway).

Building on these responses, a **Scoping Workshop** was organised on 7 September 2016 to agree on the scope, objectives and potential outcomes of this MLE. A larger group of Member States and Associated Countries attended this meeting: the Czech Republic, Turkey, Cyprus, Estonia, Norway, Spain, Slovenia, Sweden, Portugal, Spain and Italy. Participants agreed on the following **Scope**.

- The MLE will be looking at national or regional funding systems based on methods to assess and measure research performance
- The scope will be limited to institutional funding of research in universities (albeit allowing for more granular analyses within universities) that is the result of (but not limited to) retrospective evaluation based on peer review and/or metrics. The MLE will pay particular attention to the inclusion of third stream activities in PRFS and to organisational practices. Also within the scope of the MLE are the methodologies or techniques that can be used for an economic analysis of the impact of PRFS systems as well as the issues pertaining to the supporting data infrastructure that is needed to manage PRFS systems

Within this scope, it should be understood that

- **The MLE is limited to institutional funding for research**. In some countries, this is called 'baseline funding'. Term may have different meanings in different contexts, the delegates agreed to develop a list of definitions of funding sources/systems (see Section 2.1.1) and agree on a common terminology. Competitive project funding is excluded from the remit of the group
- **The MLE is limited to universities as the unit of analysis**. Universities are to be the primary unit of analysis for the MLE. However, it was also agreed that the MLE should not exclude the possibility to address institutional funding issues at finer levels of granularity within institutions themselves
- **The MLE is primarily focused on, but not limited to, retrospective (ex post) evaluation**. Delegates agreed that although the focus of the MLE should primarily be based on funding for research in universities based on retrospective evaluation, prospective evaluation should not be excluded from the scope of the MLE. Some evaluations are both retrospective and prospective. In addition, performance contracts based on prospective analysis are in some countries complementary to performance-based funding systems.
- **The MLE will look at the major methodological options and performance dimensions on which to base a PRFS in order to support the policy ambitions**. After a thorough discussion, it was decided that the MLE would think broader than bibliometrics and peer review to include systems that combine both. It also addresses the inclusion of third stream activities as well as organisational and contextual issues such as infrastructure, staffing, systems as well as experimental dimensions (which will be incorporated within the other categories). **The MLE will look at data issues relevant to the functioning of PRFS systems**: type, quality, accessibility, management, maintenance, evolution of data sources and platforms. Data infrastructure and IT systems (including systems underpinning evaluation process) concerns will be addressed, to the extent that they are necessary enablers of PRFS
- The MLE will consider the economics of PRFS systems. Although there is little evidence available about whether PRFS systems have a positive impact, it was agreed that what is available would be addressed and that the economic and other costs and benefits of PRFS as far as possible be considered

Objectives

The objectives of this MLE are to consider policies and instruments/systems that incentivise those governance/decision/ behavioural processes within the institutions that will stimulate and achieve high(er) quality of research (output) and excellence.

Outcome

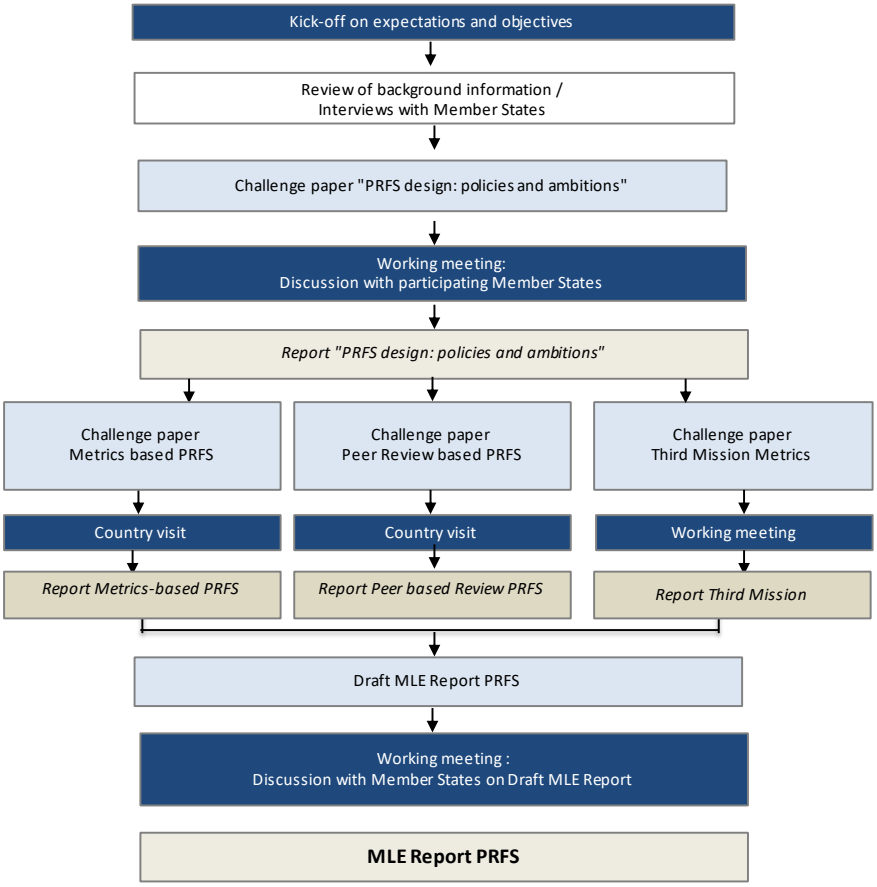
This report identifies good practices, includes a set of concrete operational recommendations, lessons learned and success factors – wherever possible based on robust evidence available to the participants about the impacts of the measures and the contextual factors that may explain them.

This MLE thus aims to support participating states and others in designing, implementing and evaluating different Performance Based Research Funding Systems through

- Better understanding of various PRFS designs developed and/or deployed in the participating states
- Understanding how they are linked to policy ambitions and underpinned by the various performance dimensions
- Understanding and monitoring expected and realised impact as well as unintended consequences
- Coming up with an overview of relevant insights on PRFS through combination of sharing practice and expert input regarding
  - Policy ambitions driving PRFS and their evolution over time,
  - Dimensions of PRFS and their link to evolving policy ambitions, both intended and unintended,
  - Indicators and data underpinning the dimensions,
  - Methods and techniques for the assessment of economics in various PRF-systems.

To this end, the process shown in Figure 1 has been followed between September 2016 and September 2017.

Figure 1 Process and Activities for this Mutual Learning Exercise



1.3 University Research Funding Systems

In Western Europe, the general way to fund state universities in the post-War period was through a single block grant covering both teaching and research – what the OECD called the General University Fund – provided by a ministry for education and science. As university systems grew, ministries in some countries started to make an explicit distinction between funding for education and that for research. In each case, this was 'institutional funding' in the sense that the money was passed from the responsible ministry to the institution in one block. The principle of university autonomy meant and continues to mean that the universities themselves then decided how to spend that money. From the 1990s,

governments became increasingly explicit that universities should pursue a 'third mission' of sharing knowledge with the wider society and supporting innovation in addition to teaching and research, though this was rarely reflected in the way institutional funding was allocated. The next step in many countries has been to make some of the institutional funding dependent upon past performance via PRFS.

In addition to institutional funding, universities compete for 'external' funding, in the sense of money for research or education provided directly by industry, research councils, innovation agencies, foundations, charities and other funding organisations within the state or from abroad, for example from the EU Framework Programme. In order to focus our discussions of policy on the parts of the national system that the government can influence directly, we generally refer to external state funding and only treat other kinds of external funding where the analysis demands (for example, where the level of non-state funding may be a parameter in the PRFS).

This report discusses institutional funding for research and does not analyse institutional funding for higher education. While in a minority of countries (including Norway, the Czech Republic, Estonia and Latvia) PRFS govern part of the institutional funding of research institutes as well as universities, here we address only universities.

1.4 PRFS

The introduction of performance based research funding systems (PRFS) is one of the central mechanisms through which many EU Member States have tried to increase the effectiveness and performance of their public sector research systems in recent years, in line with ERA Priority 1. The efficiency of funding in terms of the capability to meet certain policy goals in a cost-effective way has become increasingly important.

Performance based funding refers to the component in the organisational level (institutional) funding system that is allocated on a *competitive basis*, as opposed to block funding and to core grants based on performance contracts/agreements. It aims to incentivise those governance/decision/behavioural processes within the research organisations, and in particular universities, able to stimulate and achieve high(er) performance in relation to a number of selected criteria (for instance by increasing the volume or quality of research output, prioritising certain fields of research, developing greater interaction with industry, fostering internationalisation or improving gender balance). It may also be a means to concentrate resources in the best-performing organisations.

Many countries have implemented some forms of PRFS over the past decade(s) and the share of institutional funding which is allocated competitively on the basis of performance assessments has increased in many countries.

The nature of the systems and the methodologies used in the allocation of funding vary considerably.

- Some countries use a funding formula partially based on **quantitative assessment** of research outputs. These include Belgium (Flanders), Czech Republic, Denmark, Estonia, Finland, Norway, Poland, Sweden and Slovakia
- Another set of countries base their funding on evaluations of research output through **peer review**. This category can be separated into 'informed' peer review (ie the peers base their judgments also on metrics) and 'pure' peer review. This is the case for France, Italy, Portugal, Lithuania and the United Kingdom

Some other countries have implemented a limited PRFS. This is the case of Austria, Germany and Netherlands. Other countries such as Bulgaria, Cyprus, Greece, Hungary, Ireland, Malta, Romania, have not implemented a systematic performance based funding system yet – but in several cases, they are in the process of considering or introducing one. The small size of the country or resistance from the academic community can impede the introduction of PRFS. It is clear that the PRFS landscape in the various Member States is evolving continuously. For instance, Slovenia has evolved towards an informed peer review PRFS. While Latvia has introduced a "3-pillar HEI funding system" with some elements of a PRFS. Spain can be considered a borderline case as the Sexenio system has many similarities with a PRFS and some authors like Diana Hicks (2012) even classify it as such, although its function is to establish a quality threshold for higher status and pay at the individual level rather than being used to fund universities at the institutional level.

There is an ongoing debate regarding the benefits and disadvantages of each PRFS system and detractors as well as proponents of the principle of using PRFS. While a PRFS aims at increasing the performance of the public research system, it has also the potential to generate perverse effects. The costs involved in setting up different types of assessment are also a factor to consider.

As countries have implemented PRFS with different characteristics, there is a large potential to explore how each system works, what can be learned from each experience and how these learnings can be used to improve the various systems.

1.5 Structure of the Report

The structure of this report is as follows.

- PRFS are instruments used to implement policy. We first discuss how they fit into the wider mix of policy instruments used to influence university performance. We point out that there is a need for the criteria used in the PRFS assessment to be consistent with the objectives of policy. The funding formula is another key aspect of the link to policy because it involved deciding the relative importance of the various policy objectives. PRFS assessments can also be used to inform policy at both the national and the institutional level, so this chapter ends by discussing these additional uses
- Next, we consider the different types of information that can be used in doing the assessment, which may involve direct inspection of research outputs, the use of bibliometric or impact indicators, analysing self-assessments by the universities or qualitative ways of describing impact
- Third, we discuss the process of assessment, which may involve peer review, metrics approaches or a mixture. We also consider the scope of the assessment, its granularity (does it work at the level of the university, the department, research group or individual?) and the frequency with which the assessment is done
- Fourth, we look at the ways the research assessments are used – primarily to drive funding but also to provide strategic intelligence to policymakers and research performers. We look at evidence about the effects of the PRFS on universities, the kind of research that is promoted and researchers' careers
- Fifth, we examine some system-level questions: How does system design relate to policy goals? What does operating a PRFS cost? What issues arise in the case of small countries and those where people speak a minority language? What are the limits to the complexity of a PRFS?
- Finally, we draw out lessons for policymaking

The coverage of our analyses in this report varies a little, depending on the availability of information from the participating states. Since our objective is learning rather than benchmarking, we also introduce examples from other countries where that is instructive.

2 PRFS IN THE POLICY TOOLKIT

This Chapter begins by defining some terms and locating PRFS within the wider mix of policies available to increase the quality of university research, how PRFS fit into the system of funding incentives through which policymakers aim to steer university research. We stress that PRFS are policy instruments: they exist to implement policy objectives, usually going well beyond the simple end to distribute money. These policy objectives need to be reflected in the assessment criteria PRFS use.

We go on to set out a generic structure for PRFS and spell out the chief design options available to policymakers before discussing three other systemic issues of which designers need to take account: the cost of running the PRFS; limits on the complexity of the systems; and special issues arising in small countries, of which there are several in this MLE. Finally, we draw out some lessons from the analysis.

2.1 PRFS in the policy mix for improving university research performance

2.1.1 The elements of funding systems

The composition of university funding systems varies by country, as does much of the terminology. In this report, we use the following definitions.

- **Institutional funding** is money provided to a university at the central level. It is normally provided by a ministry of education or education and science. The principle of university autonomy dictates that once the money is handed over, the university itself decides how to spend it, even though the funding ministry may indicate its preferences for how the money is used. Institutional funding may be provided separately for teaching and research or for the two taken together
- Institutional funding may be allocated in two ways
  - As a **block grant**, essentially without conditions. The amount of the block grant will be influenced by history, the size of the institution, its particular costs and so on but is not driven by its previous performance and tends to be rather stable or to grow in line with the university cost base
  - As **performance-based funding**, dependent upon various measures or judgements of past performance. Performance-based funding can be used to reward teaching, research or third mission performance. Because national budgets tend to be limited, universities often find themselves in competition with each other for performance-based funding so this income can go

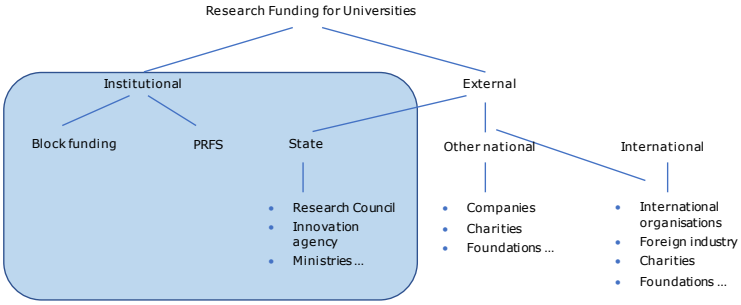
down as well as up between funding periods. Some countries have used a PRFS to distribute only the increase in institutional funding for research provided by government between one period and the next

- In some countries, students pay **tuition fees** to their university. This constitutes a separate income stream. However, it is not considered in this report as it relates to teaching rather than research
- All other funding is **external funding**, which is normally won by the university on a project-by-project basis in competition with others. Each project involves a contract with the external funder, so the university is not normally able to spend the money on things other than those agreed with the funder
- In this report, our analyses of university income consider only institutional income and state external funding for research (we ignore any external funding for education), in order to focus on the research resources that are directly under the control of the nation-state
  - **State external funding for research** typically comes from research councils, innovation agencies and other government agencies and ministries. We include EU structural funds where they support the budgets of these organisations (or similar arrangements) because they are under the control of the national government as well as containing a small element of national funding
  - In this report, state external funding for research does not include research funding from industry, charities or foundations because these are not under the control of the state
  - There are usually additional **international sources of external funding for research** from the EU Framework Programme, multilateral organisations such as NATO, UNESCO or the World Bank, bilateral research funding agreements as well as industry abroad and sometimes state funding from other countries, for example via research contracts for the US National Institutes of Health or the US Department of Defence. None of these is considered in our analyses of state research funding, although some of them are used as indicators in national PRFS
- Some countries distribute part of institutional funding on the basis of **performance contracts** or agreements between the responsible ministry and individual universities. These agreements often use indicators of past performance to help judge whether the university is living up to expectations but they seem rarely to have direct consequences for the amount of institutional funding universities receive. Their use is often to support future change in universities more than to reward past performance

- A **performance-based research funding system** (PRFS) has two elements: an assessment process and a funding process, which uses the results of the assessment to drive a **funding formula** and allocate institutional research funding to institutions
- Some countries incorporate **PhD education** in their PRFS while others treat it as part of education. In this report, we have aimed consistently to treat performance-based funding relating to PhD education as part of the PRFS (and where necessary we have reallocated such funding from the education category to research, so that comparisons among countries are consistent)

The shaded area of Figure 2 shows the scope of the research funding we consider here.

Figure 2 Scope of funding considered in this report



2.1.2 The policy mix

PRFS are one of several elements in the policy mix for research funding. The main 'levers' available to policymakers for improving the quality and relevance of university research go beyond funding and include the following.

- Making more of the institutional funding for research that universities receive dependent upon past performance via a PRFS
- Providing state external funding for research, which is in some way quality assured, to supplement institutional funding. The general expectation is that this improves the quality of the research done in the universities by imposing quality requirements on the best work they do and thereby showing where the expected quality level should be for the whole system. This funding is quality assured through peer reviewing proposals and open competition for project funding. It is all

about promises for the future. Few funders make a systematic link between past performance and future funding. If that link is strongly made, the effect is that older, established researchers will dominate. Research councils generally run schemes specifically for young researchers (and less often specifically for women, who are also systematically disadvantaged) in order to combat this tendency

- The ratio between institutional and project-based funding provides another potential steering mechanism. One would expect that the bigger is the proportion of state external (quality-controlled) funding, the higher the quality of the research undertaken. However, the international data on funding and research performance in bibliometric terms do not support this idea
- Another driver of quality is internationalisation, not least international collaboration. This is clearly visible in countries with weaker research systems, where overall bibliometric performance tends to rise as the degree of international collaboration increases. Internationally collaborative research is systematically more highly cited than national work. Countries with strong bibliometric performance tend already to have a high degree of international co-publication
- We generally believe that governance of the universities is a driver of quality in research as well as in teaching, which is why Europe is going through a wave of university reforms, increasing autonomy, empowering the leadership and reducing the lock-in produced by electing university leaders
- The other 'lever' is overall investment in research in the higher education sector. There it is noticeable that some of the best performing countries have invested strongly in doing more university-based research in the last decade or two. But not everyone has done this – for example, the UK and The Netherlands both do pretty well in bibliometric terms despite the fact that the growth in HERD over the past couple of decades has been slow in both countries

Available data about these levers are rather confusing. There seem to be no simple relationships that can be demonstrated based on robust evidence. Policymaking therefore depends strongly on contextual knowledge and judgement rather than hard evidence. It is notable, however, that **changes** in the policies discussed are intended to change performance and often appear to do so. PRFS are often intended as ways to administer a shock to the research system, triggering behavioural and cultural change and hopefully embedding that in the routines and processes that underpin good performance. Unfortunately, there has been little evaluation of PRFS implementation. That little does support the idea that introducing a PRFS changes performance by changing individual and institutional behaviour. So far, the question of behavioural additionality – whether the new behaviour would persist if the PRFS were taken away –

has yet to be explored. This is an important gap. Because PRFS are expected to change culture and behaviour – at both the individual and the institutional level – it is reasonable to expect these changes to persist even without the incentive. There are also many other forces acting on the universities that enforce quality culture so it is not necessarily the case that the PRFS has to persist over time or that the level of the incentive it provides needs to be maintained. This may be an important issue in countries where the PRFS drives a large proportion of institutional funding and where, therefore, any negative effects it may have on the nature of research and research careers is likely to be strongest.

2.1.3 The funding system

Through 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 generate thematic 'focusing devices' (Rosenberg, 1976) (Arnold, Good, & Segerpalm, 2008) in the form of projects and programmes, implementing science policy through research-funding incentives.

While the distinction between institutional and project-based funding is conceptually tidy, reality can be more messy. 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 use institutional funding 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 'overhead' contributions to the universities' fixed costs but only paid the marginal costs of research. Non-

[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.

academic funders such as industry were expected to pay the marginal costs plus an 'overhead' contribution.

From 2005, a principle of 'full economic costing' of research was introduced across UK higher education, and this principle is also spreading internationally. It 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 not have enough money to pay their infrastructural costs of research. Increases in Swedish institutional funding in 2008 and 2012 were in part politically justified as a way to rebalance internal and external funding sources after a period of rapid growth in external funding.

A systemic issue is therefore 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 a particularly acute problem in systems that have a high ratio of external to institutional research funding. In the UK, there is a strong correlation between performance in the PRFS (the Research Excellence Framework, 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 assessment 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 tends to be willing to pay the full costs of research. The amount of this funding is driven by the amount of external funding the individual university obtains from charities and industry.

Internationally, 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 greater pressure on universities' institutional research funding, which needs to support a growing volume of externally-funded research.

Government university funding for research at the national level today tends to comprise the blocks shown in Figure 3. University autonomy means that the state has to provide incentives to encourage particular types of performance by the universities rather than being able to tell the

universities what to do in detail. Thus, in the state system an education ministry typically provides both institutional funding and external money for research, the latter through a research council. The institutional funding is infrastructural in nature: it provides a basis for strategy and planning and for maintaining capacity to do research in the universities. The research council money is intended to encourage the performance of high-quality research.

Innovation agencies are normally funded by an industry ministry and fund research of relevance to business. They are generally complemented by the research funding activities of ministries with other mission responsibilities, such as transport, environment, health and defence which fund research needed to meet the state's need for research in these specific sectors or the production of public goods in the form of applied knowledge. 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. (Often, they also use captive government labs or contracts with industry to do at least some of the research needed.)

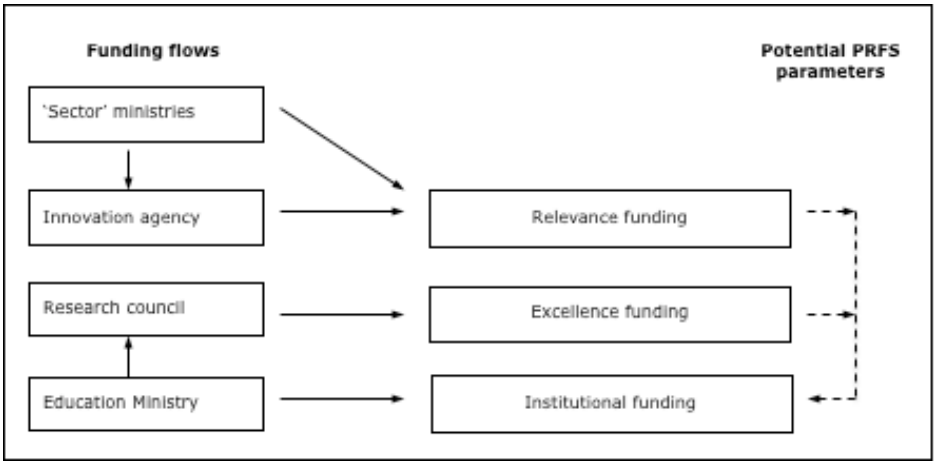
Universities also have other sources of research funding, typically from charities and foundations, industry and from abroad (for instance via the Framework Programme), providing a further funding stream for university research, sometimes connected to specific social purposes but sometimes simply providing more 'excellence' funding. The external funders thus provide incentives for particular kinds of research. The balance among all the flows would be expected to influence the overall shape of the university research system.

Many PRFS now use the amount of research funding from particular 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 of internationalisation via participation in the EU Framework Programme of research and technological development. That creates the feed-back loops shown at the right-hand side of Figure 3.

Since the money flows illustrated in Figure 3 are governed by several different actors, the balance of the research incentives provided simply reflects the budgets the various ministries make available. There is little evidence that countries explicitly develop policies about the desirable balance among these incentives. That would require a coordination mechanism, which is usually absent. Hence, the actual balance of incentives emerges bottom-up. Arguably that is a good mechanism, but it does make it difficult to prioritise among alternative incentives or deliberately to reallocate money from one category of incentive to another.

There is, however, some evidence that coordination can be attempted. The Finnish Research and Innovation Council and its predecessors played a strong role in coordination and prioritisation before the 2008 financial crisis, and may do so again. Norway and Sweden have a tradition of producing research and innovation bills for consideration by parliament every four years. These trigger extensive consultation and coordination across ministries and allow explicit discussion of issues such as the balance between institutional and state external funding for research. Following the 2016 Research Bill in Sweden, a committee was set up to consider the balance among incentives and especially whether there were sufficient incentives for the third mission.

Figure 3 Stylised state university research funding system



Source: Technopolis Group

Just as the balance of state external funding for research should reflect national needs, the internal incentives provided by performance-based funding also need to be in balance so that universities prioritise each of their three missions to an appropriate degree. The UK system offers an example of what happens when these incentives get out of balance. The fact that in the past a performance incentive was offered only for research quality meant that teaching became a second-class activity. This became especially problematic when the government allowed the universities to increase fees to £9,000 per year and students increasingly complained about the amount and quality of teaching they obtain. The Higher Education Funding Council for England (HEFCE) reacted by implementing a Teaching Excellence Framework – an exercise parallel to the REF. In 2017, the government announced that it would launch a Knowledge Exchange Framework addressing the third mission, in order to bring the

research, teaching and third mission incentives into a more appropriate balance.

While PRFS often contain other components, it is generally the part of the PRFS focusing on research quality that gets most attention, and especially the part that involves judgements about the quality of individual outputs, from policymakers, universities and the academics themselves. This is probably because it bears directly on academic prestige and influences the way universities recruit and manage academic staff.

Table 1 shows how those states participating in this MLE that were able to supply the relevant data allocate national funding to the university system. (As earlier indicated, private, charity and foreign income – including from the EU Framework Programme – are not included here, though EU Structural Funds are treated as quasi-national and included in the national figures.) We have additionally included data for the UK[3] and Finland, to emphasise the diversity in the ways this may be done. The distinction between block and performance-based teaching funding is in some cases hard to make. Most universities receive teaching funds against an expectation that they will deliver a certain amount of education, increasingly expressed in terms of outputs.

Austria, Italy and Norway do not differentiate between teaching and research in their block grant arrangements. In Italy, from 2014 part of the institutional funding is allocated on the basis of teaching performance. Italy and (to a lesser degree) Austria also stand out for their heavy reliance on institutional funding and limited use of state external funding for research.

The Estonian PRFS is special in that until recently there was no institutional funding for research in the universities. The Estonian PRFS has two stages: a first stage that determines whether a research group is good enough to qualify for institutional funding at all and a second stage, which annually allocates that institutional funding based on past performance. There is an ambition further to increase the proportion of institutional funding in universities' income from its present low level. The first stage existed before the funding stage was introduced. We treat the two together in this report, since some other PRFS include an eligibility check and because – logically – in Estonia one cannot benefit from the funds-allocation process without also having passed the eligibility check.

[3] The data relate to the Higher Education Funding Council for England, rather than the whole of the UK, since Scotland, Wales and Northern Ireland have their own funding councils, each using a different PRFS. The External research funding number is estimated, based on England's share of Mainstream QR funding from HEFCE, which is known to correlate strongly with research council income

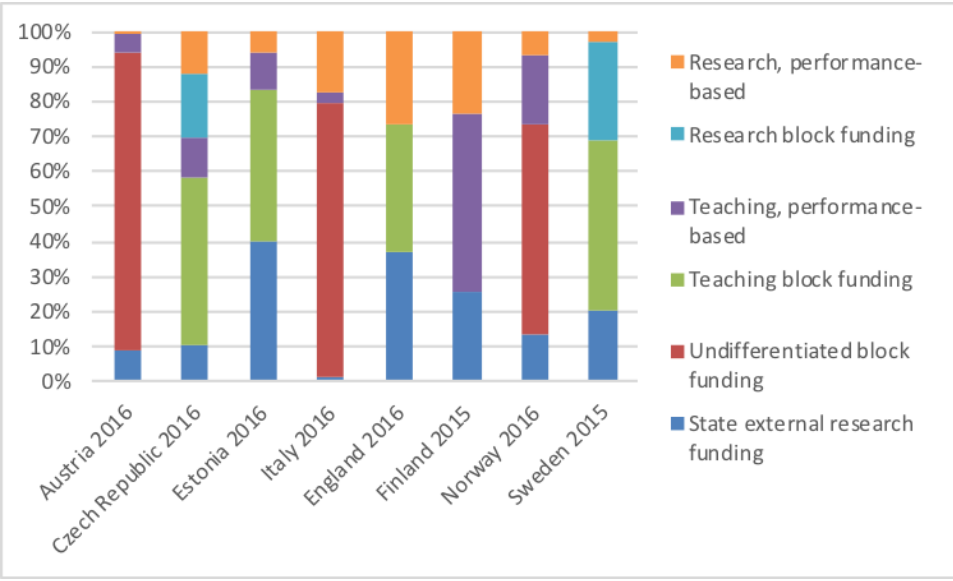
The bottom line of Table 1 indicates the degree to which university research funding is 'contested': either through external competition or via a PRFS. The Czech Republic, the UK and Finland all have highly contested systems that apply high levels of competitive pressure to the universities.

Table 1 Distribution of universities' total income from the national state

	Austria 2016	CzechIia 2016	Estonia 2016	Italy 2016	UK 2016	Finland 2015	Norway 2016	Sweden 2015
State external research funding	9%	11%	40%	1%	37%	26%	13%	20%
Undifferentiated block funding	85%	0%	0%	78%	0%	0%	60%	0%
Teaching block funding	0%	47%	44%	0%	36%	0%	0%	49%
Teaching, performance-based	5%	12%	10%	3%	0%	51%	20%	0%
Research block funding	0%	18%	0%	0%	0%	0%	0%	28%
Research, performance-based	1%	12%	6%	17%	27%	23%	7%	3%
Total funding	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Of which, total institutional funding	91%	89%	60%	99%	63%	74%	87%	80%
Sum of state external research funding plus PRFS as a % of total research funding	10%	22%	46%	19%	64%	49%	20%	23%

Source: Analysis of data supplied by participating states plus estimates by the experts. * Estimate of research council funding in England. Since there are also other state external research income sources, this is likely to be an underestimate of total external state funding.

Figure 4 Distribution of universities' institutional income from the national state



Source: Table 1

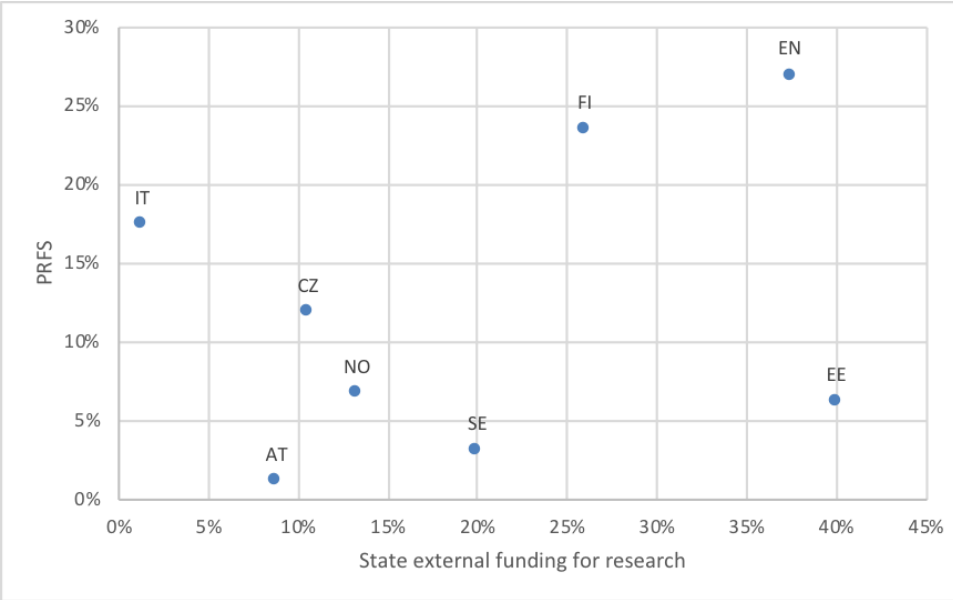
Figure 5, below, shows the proportions of universities' state external research funding and PRFS-based income in the countries that provided data. The UK, the Czech Republic and Finland make intensive use of both approaches so that all research funding is in some way 'contested'. A wholly metrics-based Czech PRFS was introduced in 2009 as an attempt to improve the evaluation culture by utilising a well-functioning national research information system. (The system was later abandoned because it generated funding instability and arbitrary allocations.) Italy, Austria, Sweden and Norway have strong block funding traditions. Italy uses almost no state external research funding but quite a strong PRFS, which may have compensated to some degree for the lack of external competition but which would appear more likely to preserve the status quo within the universities. In Austria both state external research funding and PRFS funding are low. While a lot of use is made of state external research funding in Sweden and Norway, the proportion of PRFS funding has been kept low. It is a little higher in Estonia, where the PRFS governs all institutional funding for research, in the context of a policy ambition to raise further the amount of institutional funding the universities receive.

Estonia, Finland and the UK do not provide block funding for research at all. Estonia does not do so because it is transitioning its university system from Soviet-style teaching universities to research universities by introducing institutional funding and is making institutional research funding conditional on performance. The UK and more recently Finland have abandoned their earlier tradition of block funding in order to improve performance.

Figure 5 plots the share of universities' income from state external research funding in their total income against the share of PRFS funding. In effect, it shows the balance different countries have chosen between quality-assured state external research funding and institutional research funding, as well as the extent to which research funding is contested.

Teaching income is not shown, hence the percentages are all well below 50%. However, there is clear polarisation between countries where research income is highly contested and those where it is not. Estonia is an outlier. Its heavy reliance on state external research funding reflects the availability of European Structural Funds and the fact that it is building up the share of institutional funding. Since all institutional funding is governed by the PRFS, Estonia is likely to move towards the middle of the diagram over time. Italy, too, lies far from the diagonal, with state external research funding playing almost no role – an unusual position compared with other countries.

Figure 5 External funding and PRFS in university income



2.1.4 PRFS serve policy objectives

PRFS are adopted in order to pursue policy objectives in the national context. Clearly, the mechanisms of the PRFS used need to be consistent with the wider set of incentives and to avoid – either on their 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 have introduced a 'Teaching Excellence Framework' to counterbalance the negative effects of the REF and its predecessors on the status and focus on teaching in the universities.

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

1. Enhance the quality of research and the country's research competitiveness
2. Steer behaviour in order to tackle specific failures in the research system
3. Strengthen accountability
4. Provide strategic information for research strategy at institutional and/or national level

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 – which no longer uses a PRFS) 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. In The Netherlands, the universities have evolved the so-called Standard Evaluation Protocol (SEP) based on an informal agreement between the government and the universities that a PRFS would not be introduced, provided that the universities themselves organised robust assessments.

Adopting the SEP in response to the threat that the government would impose a PRFS has to be understood against the importance of the so-called 'polder model' in Dutch culture. This is the idea of solidarity and the sharing of pain as well as gains. The universities devised the SEP as an alternative to a PRFS and persuaded the government to let them demonstrate that they could improve collective quality in this way. That demonstration has been successful and kept the idea of a PRFS at bay. All the Dutch universities are now high in the international rankings and

regard themselves as working together to support high quality rather than being in competition, as universities inevitably are when a PRFS is used to redistribute money from higher- to lower-performing universities.

Most 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). However, there are also examples of countries using PRFS to promote systemic changes. The commonest is probably the inclusion of indicators of external funding income. For example, Croatia, Estonia, the UK, Italy, Finland and Norway all reward external research income from industry and other societal stakeholders, with the aim of improving academic industry research links.

While rewarding 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. However, in 2016, the Swedish government decided that it would not approve the revised system. In countries such as Italy, where university policy is set at the level of the central state rather than the regions and concentration is generally felt to be undesirable, PRFS are used 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.

PRFS do not yet distinguish between 'quality' in the sense of the overall or average level of research quality and 'excellence' in the sense of having researchers who are among the very best in the world in their field. In the world of bibliometrics, this distinction is reflected in a move from typifying quality in terms of field-normalised citation performance compared to benchmarks such as the world average towards interest in those researchers whose publications are among the top-10% or top-1% most highly cited articles in their field worldwide. The distinction is important, for example, in Sweden, where there is a concern that while the average quality of research is good, there are very few Swedish researchers among the global leaders (Benner & Öquist, 2012). A revised PRFS design has been proposed, which would have tackled this and other concerns, but has been rejected by government.

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 Standard Evaluation Protocol process appears to exert a strong social influence over performance. The visibility and transparency of research output via the Croatian national CRIS 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 research-based industry. Table 2 shows more detail about the PRFS policy objectives and aims of states participating in this MLE. It confirms the importance of quality as the main policy purpose. This may also be wrapped up in the idea of competitiveness – the categories of policy purpose and aims used in descriptions of PRFS are not orthogonal. But the other purposes shown are also important. Transparency and accountability are also important in a number of places, presumably with the intention of demonstrating to decision-makers and taxpayers alike that the national investment in research is worthwhile.

The more detailed aims shown are largely consistent with the policy purposes: a focus on improved research performance and quality; a well-functioning system for allocating funding; and in quite a number of cases an aim to trigger reforms in the research institutions or the research system as a whole.

Table 2 Policy purposes and aims of participating states' PRFS

	AM	AT	HR	CZ	EE	IT	MD	NO	PT	SI	SP	SE	TR
Problems and policy purposes													
Quality	X			X		X	X	X	X		X	X	
Productivity								X					X
Overall competitiveness	X	X					X						X
Efficiency	X			X					X				
Transparency & Accountability		X	X	X		X							
Systemic changes	X				X					X			
Aims													
Improved research performance	X	X	X	X		X	X	X	X		X	X	
Behavioural change in research (commercialisation, internationalisation)	X					X			X				
Reform within the institutions					X	X		X	X				
Reform of the research system	X			X		X			X	X			X
Improved institutional funding allocation		X	X	X		X		X		X			X

2.1.5 Assessment criteria

Table 3 shows what criteria are considered in the PRFS systems of the countries considered. There is broad compatibility – but not a one-to-one match – between the policy purposes and aims on the one hand and the assessment criteria on the other. All but Austria consider the quantity and quality of scholarly outputs. Many consider external, competitive research funding but these vary in the degree to which indicators focus on specific types. Three countries use this indicator specifically to encourage internationalisation, consistent with their policy priorities (and Finland further emphasises it using an indicator for the proportion of foreign staff). A further three include international income within a catch-all external income category. Three specifically consider innovation outputs or societal impact, again reflecting policy priorities. Only one (Croatia) specifically mentions popularisation of science, although this is a potential element in the UK 'impact statements'. Most PRFS do not use specific

thematic focus as a criterion, but Estonia specifically encourages research in a small number of topics of national importance (mostly related to Estonian identity and culture).

The UK REF considers the 'environment' – that is the adequacy of resources and systems in the part of the university where the research is being produced. This kind of criterion might more naturally appear in a research assessment with formative aims (such as recent exercises in Latvia and Lithuania), rather than in the UK's summative process, which does not aim to give advice to those assessed. None of the participating states uses such a criterion.

Table 3 Criteria used in PRFS

	AT	HV	CZ	EE	IT	UK	FI	NO	SE
Academic output & quality		X	X	X	X	X	X	X	X
Any external funding		X		X	X		X	X	
National grant funding	X							X	X
International funding	X						X	X	
Business/charity funding		X							
Innovation outputs/impact			X*	X		X			
PhDs	X			X	X				
Personnel					X				
Popularisation of science		X							
Topics of national importance				X					
Environment						X			

*Innovation outputs were dropped from the Czech PRFS from 2017

2.2 Design options for PRFS

PRFS design is important not only in itself but also because there is a tendency for the design to be institutionalised into the way universities are funded and manage themselves, creating lock-in. Hence, there is a premium on getting the overall design right first time as once it is implemented it may be politically difficult to make more than incremental changes to it.

2.2.1 The structure of PRFS

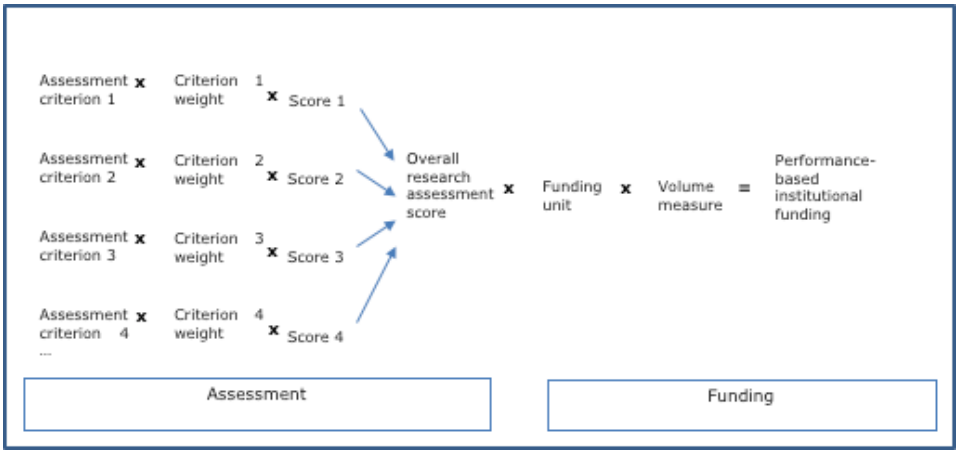
A PRFS has two components. First, it has an assessment process, which judges research outputs based on their scientific quality and increasingly also other criteria. The results of the assessment feed into the second component, a funding formula. This is an algorithm for allocating institutional funding for research among the universities, based on

performance. Common to these formulae is that they move money away from the universities that have obtained low ratings in the assessment and towards those that have been rated more highly.

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.

PRFS are implemented in many different ways and can be complex. However, they share an underlying structure illustrated in Figure 2, below, which shows how a PRFS allocates institutional funding to an individual university. If it is to be transparent, accepted by the academic community and useful for strategic decision-making, it is important that the PRFS strictly separates assessment from funding. For example, if there is a need to provide higher institutional funding per researcher to high-cost subjects such as the hard sciences, this adjustment should be done explicitly in the funding formula and not achieved by adjusting the criterion weights in favour of certain subjects.

Figure 6 How a PRFS allocates funding to a university



The assessment component uses one or more criteria to judge the 'goodness' of university research. Early PRFS focused exclusively on goodness in terms of scientific quality; more recent ones include more dimensions. Logically, the assessment criteria should reflect the policy purposes of the PRFS, which may change over time. Assessment systems have to decide how good the work is against each assessment criterion

and to express this in the form of a score. The importance of the individual criteria may not be the same. If there is more than one assessment criterion then the relative importance of the criteria needs to be made explicit by assigning a weight to each criterion. Multiplying the scores by the weights and aggregating the results can produce an overall assessment score, allowing the universities to be ranked and providing an arithmetical way to drive the funding component of the PRFS.

The overall assessment score indicates how 'good' an individual university is according to the assessment criteria. Where the assessment process takes individual universities or departments as its unit of analysis, it has to take account of the fact that different entities are of different sizes. (It would make no sense to give more money to an excellent theological college with 50 members of faculty than to a good, large university with 5000 faculty members.) The funding formula therefore needs to have some volume measure that takes account of the sizes of the entities being assessed. This can be tricky to define. Some systems try to count the number of researchers involved, though this runs into problems in defining the boundary between researchers and teachers (since many but not all academics do both). The UK system counts the number of authors for whom the universities submit outputs to be assessed in the REF.

A PRFS normally distributes a fixed sum of money or budget, so that amount has to be split into individual funding units by dividing it by the sum of the products of each university's overall assessment score and its volume measure. That allows the amount to be allocated to each individual university to be calculated by multiplying together the overall assessment score, funding unit and volume measure in each case.

Other systems (for example, the Norwegian one) consider individual research outputs aggregated to the institutional level. Each is allocated a score ("publication points" in the Norwegian terminology, see Appendix C section 8.3.3). The funding unit is determined by dividing the available budget by the sum of the scores (or publication points). The volume measure is effectively set to "1" because the university is rewarded for the total number and quality of its outputs, irrespective of the number of people producing them.

Table 4 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 (ie peer review, bibliometrics or a combination of both), affects the granularity and the periodicity of the PRFS.

Table 4 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	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

(Arnold, et al., forthcoming 2018)

2.2.2 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

Peer review versus bibliometrics (and beyond)

The choice between metrics and peer review is contentious. 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.

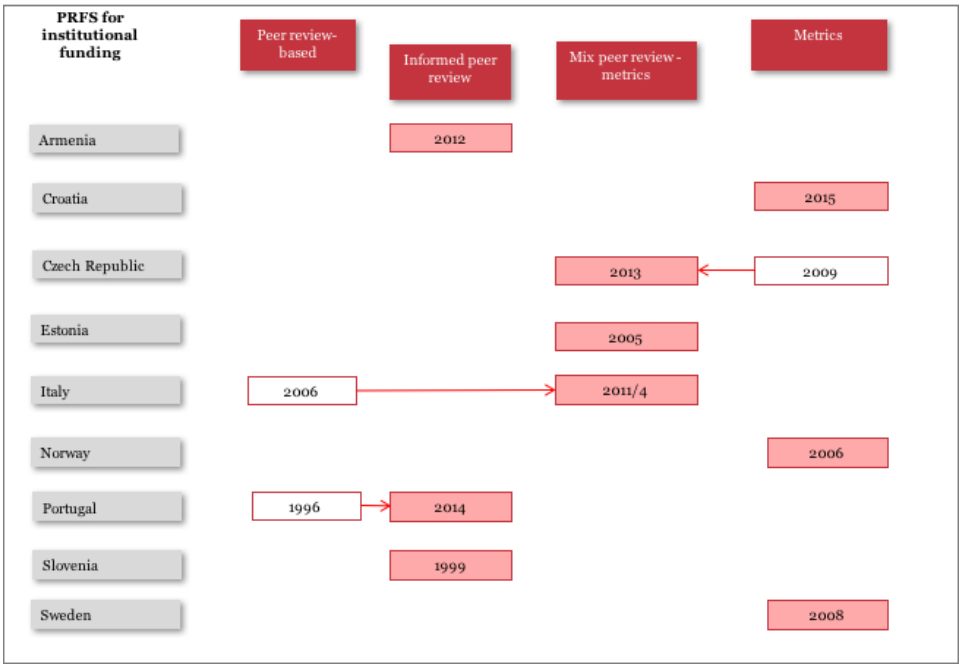
Cost, however, is a major factor. PRFS that rely entirely on metrics are generally considered to be less expensive to administer and less compliance-heavy than systems that use peer reviews, which are seen as cost-intensive and time consuming. Both peer review and metrics suffer well-known weaknesses, which are addressed in Chapters 3 and 4.

International practice changing over time

Many of the PRFS among the participating states in this MLE are in their first generation. Here, we look more widely to see how some of the countries with longer experience (including some in this MLE) changed aspects of their PRFS designs over time.

A historical analysis of PRFS systems indicates a continuing search for improvement (Figure 7). 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 (where the government finally decided **not** to add an informed peer review component to the PRFS in 2017), 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. The Czech Republic moved from a wholly metric system to one that involves a measure of peer review as well as metrics in response to a perception that the metrics system was causing instability in funding – especially because it was eventually intended to drive all of institutional funding. Portugal added bibliometric information into its peer review approach in 2014, to improve the basis for peer judgement.

Figure 7 Trends in the models for research performance assessment



Notes: In the Czech Republic, Norway and Sweden the PRFS system is currently being reviewed

Belgium (Flanders), 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 institutional or discipline-based evaluations using informed peer review that have a more formative character.

In Belgium/Flanders, the BOF bibliometric indicators started in 2003. The indicators used were gradually refined, complementing the use of the Thomson-Reuters Web of Science (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 licences.

An increasing number of systems combine the two approaches. The 'mixed peer review-bibliometrics' model uses both bibliometrics and peer review. In Australia the ERA 2015 used 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. In Italy for example, the last two PRF exercises, VQR 2013, which covered the period 2004-2010, and VQR 2017, covering 2011-2104, use an informed peer review process the social sciences and humanities, based on outputs submitted by the research institutions, and bibliometrics in the hard sciences. The results of the these assessments are used to allocate units of assessment into broad quality bands, and the product of the average score obtained and a volume measure drives the allocation of the funds.

2.2.3 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 information 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, ie 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 field evaluations of the social sciences and humanities at the national level in Norway.

2.2.4 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 (see Table 5, which considers the same broader set of countries as Figure 7). 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.

Table 5 Indicators used in PRFS in selected countries

	Austria (2016)	Croatia (2016)	Czech Republic (2016)	Estonia (2016)	Italy (VQR, 2014)	Norway (2017)	Sweden (2008)	Belgium /FL (2009)	Denmark (2009)	Finland (2015)	New Zealand (2003)	UK (REF 2014)
Output indicators												
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,												√

	Austria (2016)	Croatia (2016)	Czech Republic (2016)	Estonia (2016)	Italy (VQR, 2014)	Norway (2017)	Sweden (2008)	Belgium /FL (2009)	Denmark (2009)	Finland (2015)	New Zealand (2003)	UK (REF 2014)
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 (eg spin-offs)			√	√	√*			√				√
Popularisation of science		√										

Sources: Participating states, MLE secretariat

*This indicator is measured but does not influence funding

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 assess the productivity, quality and value of non-academic outputs and innovation-related ones such as patents directly
- 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 assessing the universities' broader research capacities and activities

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 5, 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 (eg 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.

Participants in this MLE argued that the use of excessive numbers of indicators is problematic in PRFS design, not only because of reporting difficulties but also because the people being assessed struggle to conform to all of the desiderata implied by the indicators. The large number of indicators means that it may no longer be clear to them how they are expected to behave.

2.2.5 Assessment criteria

Peer review involves qualitative judgement to make an explicit transition from expert observation to numerical grading or ranking. This involves using quite complex information to generate a small set of scores (and in some systems also feedback to those assessed). Table 6 shows four examples. The scores used are not only few in number but composite in nature.

In contrast, metrics-based systems can produce larger numbers of scores and therefore handle more separate criteria. However, since the transition from indicators to assessment results is essentially arithmetic, these assessments are less nuanced. 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. There is normally no scope for qualitative judgement. (An interesting exception

is the use of expert panels in the Italian VQR to quality-assure and potentially adjust arithmetically-derived scores.)

Table 6 Criteria used in peer review based assessment frameworks

Australia – ERA 2015	Netherlands – SEP 2015	Italy – VQR 2013 and 2017	UK REF 2014
OUTPUTS			
Volume and activity; publishing profile; peer review; citations; research income	Research quality	Originality, scientific rigour and scientific impact	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	

2.2.6 Granularity

Unit of analysis

The unit of analysis in a PRFS assessment process can be the individual researcher, a research group (defined by the field in which it works), 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 assessment 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 (which may be the same as the unit of assessment or the sum of several unites of assessment).

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).

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. Often, they reflect to some degree the PRFS calculations. 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[4]. 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

[4] 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)

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 in both approaches. Comprehensive inclusion of staff may for instance obscure the presence of a 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 by others who have not been included in the exercise. Normalising publication output by staff poses various measurement challenges (Glänzel, Thijs, & Debackere, 2016).

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 its own bibliometric analyses or consulting 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 differences among scientific disciplines

Differences among fields present a major challenge for research assessment. Two models emerge in international practice. (in the informed peer review system in Italy they are both used.)

- Some bibliometric-based systems try to overcome field differences in publication patterns by introducing a system of weights, field normalisations or field-independent indicators that balance the differences in publication behaviour and subsequent patterns and 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 solve the problem by using discipline- or field-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. 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 suitably to 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.

2.2.7 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.)

2.3 Other design issues

This section addresses three other design-related issues: cost, complexity, and PRFS in small countries.

2.3.1 Cost

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, which covered six years, 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.

Table 7**,** below, shows the roughly calculated annual direct costs for the research assessments in Italy, UK, Slovenia Portugal and Estonia. In all cases only direct costs are reported; costs carried by the research community for the submission process are not included. Since some systems operate only at intervals of several years, the total costs per exercise range from as little as €7,000 up to €10m because of the considerable variations in assessment and reporting levels (single University versus whole country).

Because of a 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 annual 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 (Farla & Simmonds, 2015) is added.

Table 7 Cost per researcher of PRFS

	Yearly direct costs in Mio. €	Researchers FTE in HEI, 2015	€ per researcher
Italy	2,0	52,677	38
UK	2.86	188.434	15
Slovenia	0.17	2.555	67
Portugal	0.30	27.766	11
Estonia*	0.0073	3000	7.1

*The Estonian system covers research institutes as well as universities and this is reflected in these figures. The cost covers both the 7-yearly accreditation and the annual allocation of funding addressed by the PRFS

In designing a PRFS, it is well to pay attention to the things that drive cost. These include

- The 'model' used. Peer review involves a lot of personal judgement and therefore time. It involves work to identify, recruit, manage and reward the peers. Usually, the monetary incentives given to the peers are less than their cost to their institutions, so there is an additional hidden cost associated with their time. Peers become especially expensive if they are international, although the use of foreign peers is absolutely necessary in small systems and advisable for at least some of the peers in large systems
- Formative elements also add cost, especially if they require site visits

In contrast, systems that extract and manipulate data from existing, quality-assured databases are inherently low cost

- Use of qualitative data in the assessment requires the use of human expertise to make judgements, creating a need for peers or other experts, together with the associated costs
- Scope can affect costs. The more people and units of assessment assessed, the greater the cost. The absolute difference is small in metrics-based systems but is large in peer review based systems
- Using a national CRIS involves high set-up costs, but once this is done it provides a useful – and sometimes indispensable – source of information, especially in smaller countries that use 'small' languages
- The proportion of universities' research income that is affected by a PRFS varies great deal among countries. Understandably, in those countries where the PRFS steers a large proportion of the money the academic community pays closer attention to the quality and precision of the assessment methods used. An extreme example is the UK REF, which requires repeated and extensive consultations with the community

2.3.2 Complexity

As this report demonstrates, there is a great deal of variation in the complexity of PRFS systems used.

- Peer reviewers will not tackle many assessment criteria at a time, so peer review systems can be simpler than some based on metrics
- Complex metrics-based systems involve a greater amount of desk analysis time than simple ones, but the cost increase associated with this kind of complexity is low provided the data required are accessible
- Using large numbers of indicators means that each indicator contributes little to the whole. As weights become smaller, so their values also become increasingly arbitrary because it is difficult to track through from the weights to the behaviour induced. The experience for the researcher is of being told to do a large number of things at once
- The objective of most PRFS is to increase quality and relevance by embedding the qualities needed to achieve those into university culture. This argues for a small number of relatively simple indicators and a cognitively comprehensible weighting scheme

2.3.3 PRFS in small countries

A number of issues with PRFS are particular to small countries, as many of those participating in this MLE are.

- As the discussion in section 2.3.1 indicates, the scale of a country is an important factor in driving cost. While metrics-based systems can tackle many categories (eg fields) at small scale, scale limits the number of peer panels that can economically be assembled
- Using peer review may nonetheless be especially important for small countries that need to upgrade or develop their research systems as these tend to need a formative element in the PRFS
- As with any other kind of peer review, small countries have a particular problem in that 'everyone knows everyone else' – and often has social obligations to them. This means it is better wherever possible to use foreign peers
- However, many small countries speak 'small' languages. While the 'hard sciences' to a considerable extent now operate in English globally, obtaining foreign peers with the necessary linguistic, cultural and historical knowledge to work in the humanities and parts of the social sciences can be hard
- The 'small language' issue also presses small countries to maintain national CRIS that are in some way quality-assured, since a proportion of the research that should be evaluated will be written in the national language
- Small countries can also have a problem in assembling the capacity needed to run a national research assessment. It is unusual to contract this out, but it has been done (for example in Latvia and partly in Lithuania)

2.4 Lessons learnt

Key messages from this Chapter are:

What PRFS are

- In general, universities obtain income partly through 'institutional funding' provided by a ministry of education or a similar government Department and 'external funding' for individual projects or programmes, won in competition with others from specific sponsors. Institutional funding may be allocated separately for teaching and research or for both together

- Institutional funding may be provided as a 'block grant', which is essentially unconditional, or as 'performance-based funding', which as the name suggests is provided on the basis of past performance
- Performance-based research funding systems (PRFS) are two-part systems that assess the 'goodness' of university research on various dimensions and use the results in a funding formula to allocate part or all of the institutional funding for research provided to universities. In addition to finding ways to measure various kinds of quality, they must also relate the quality measurement to the different sizes of institution assessed through some kind of volume measure
- Most PRFS focus on the quality of research and its scientific impact. In more recent systems, a growing number of other criteria are being used, especially in relation to societal impact
- PRFS are normally used to allocate a fixed pot of money or budget among institutions. They rarely trigger increases in total funding, though sometimes they allocate increases in funding
- In practice, few PRFS contain prospective elements
- The fundamental units of analysis used in PRFS are individuals or research groups but whichever approach is taken, they do not report at the levels of individuals. Analysis of both individuals and groups can be aggregated to the institutional level via a funding formula. Innovation is also possible in the scope of PRFS, for example by stretching across institutional boundaries

Systemic role of PRFS

- A PRFS is one of several mechanisms available to policymakers to encourage the improvement of quality, relevance, impact and other desiderata in university research. The other mechanisms include: providing or changing the amount of state external research funding provided; altering the ratio between institutional funding and external funding; increasing international collaboration in research; reforming and modernising university governance; increasing overall investment in higher education research
- There is wide divergence among countries in their balance of funding mechanisms and research funding. There is no clear theory available about how the individual components, their balance or changes in these relates to overall performance and – given the differences among national contexts – it is therefore unlikely that there is some 'optimum' mix of incentives that will work in all countries and at all times

- As with some of the other policy options, the introduction of a PRFS can administer a shock to the research system, changing culture and behaviour
- In introducing a PRFS, it is important to consider its systemic implications in the sense of its influence on the overall pattern of funding and incentives and whether the incentive system is then balanced so that universities are encouraged to fulfil all their three missions well and not improve one at the cost of the others
- PRFS are one of several mechanisms available to policymakers to encourage the improvement of quality, relevance, impact and other desiderata in national university systems
- They can play a part within the narrower domain of research funding policy, where they not only can play a role in pursuing the same desired policy objectives but also help steer the balance of the universities' research effort among fields, between 'basic' and applied research and various aspects of relevance
- PRFS are normally used to allocate a fixed pot of money or budget among institutions. They rarely trigger increases in total funding, though sometimes they allocate increases in funding
- In addition to finding ways to measure various kinds of quality, they must also relate the quality measurement to the different sizes of institution assessed through some kind of volume measure
- Crucially, PRFS exist to help reach policy goals. Despite widespread fascination in the academic community with their details, they are rarely 'just' mechanisms for distributing money
- Just as in any other kind of funding, it is therefore crucial that the assessment criteria used (and the corresponding indicators) reflect the policy objectives

PRFS design

- Key parameters for PRFS designers include: the assessment model; the scope of research included; types of indictors used; assessment criteria; the granularity of the analysis; and the frequency (periodicity) with which the PRFS can be run
- Early PRFS were wholly focused on scholarly quality. Over time, non-scholarly criteria have become more important
- PRFS design is important not only in itself but also because there is a tendency for the design to be institutionalised into the way universities are funded and manage themselves, creating lock-in. In practice some countries have become more or less culturally wedded to a particular broad design. Hence, there is a premium on getting the

overall design right first time as once it is implemented it may be politically difficult to make more than incremental changes to it

- Some more experienced countries learn over time, changing their models, but the changes tend to be incremental
- A PRFS has two components: an assessment process; and a funding formula. It is important to keep these separate if the PRFS is to be transparent and therefore to avoid putting elements of the funding calculation into the assessment process or vice versa
- There is continued disagreement about whether peer review or metrics approaches are 'best'. Key issues in the choice between these models are: cost; whether a formative element is needed; periodicity, with peer review tending to be done infrequently while metrics approaches can be taken annually
- Inter-field differences ranging from the methodological and epistemological norms of the epistemic communities involved through to publication tradition make it hard to compare fields directly. Peer review systems can do this through the use of common quality scales interpreted by practitioners from the respective fields. (Biblio)metrics systems must do this through field normalisation calculations
- The fundamental units of analysis used in PRFS are individuals or research groups but whichever approach is taken, they should not report at the level of individuals. Analysis of both individuals and groups can be aggregated to the institutional level via a funding formula
- In practice, few PRFS contain prospective elements
- The costs of running a PRFS appears to be driven by: scale, the model used, whether there is a formative element; use of qualitative vs quantitative data; the number of people in scope to the assessment; whether there is a quality-controlled national CRIS; the degree of opposition form the academic community and the consequent need to justify and consult about methodology (particularly where the PRFS governs a large proportion of research funding)
- Metrics approaches can be much more complex than peer review systems tend to be, though it is not necessary that metrics systems should be complex. There is probably a trade-off between complexity and the PRFS' ability to influence behaviour, suggesting that a good design principle (here, as in science) is 'as simple as possible but no simpler' and hence limiting the number of indicators used
- Small countries face particular design issues: the costs associated with small scale; the limited number of fields that can be addressed in peer review systems, as a result of which the few fields defined have to be broader than in big systems; the need to use foreign peers; the

constraints of 'small' languages on peer recruitment and the corresponding need for a quality-assured national CRIS; national capacity to run a research assessment exercise

- While the direct costs of those PRFS for which we could obtain data are rather modest, there was considerable concern among the participating countries about the importance of constraining that cost. Going beyond these countries, the cost of the process appears particularly to be an issue in the UK system, which not only involves large numbers of peer reviewers but also triggers a large effort by the universities to present their work in the best light

3 INFORMATION USED IN A PRFS

PRFS process large quantities of data in order to make assessment and award funding. The quality and reliability of these data is therefore key to the way that they themselves perform. This Chapter discusses the choice and sourcing of the information used in PRFS. We start by discussing the role of current research information systems (CRIS) then move on to sourcing bibliometric indicators, materials for peer review and information about societal impact. Finally, we address the role of self-assessment and self-reporting.

3.1 The role of a CRIS and other information systems

Current Research Information Systems (CRIS), so named by their co-ordinating organisation EuroCRIS but sometimes known as RIS, are increasingly being used to standardise and ease documentation, communication and administration of research. In the most advanced versions, CRIS help produce integrated data from what used to be documents for separate purposes: individual applications for funding, institutional annual reports, project reports, CV's, publications lists, profiles of research groups, project reports, information for media and the general public, etc. Searchable bibliographic references may lead on to full texts in local repositories.

If CRIS data are structured and quality-assured for statistical purposes, research performing and funding organisations may also use them for monitoring and evaluating research activities and outputs, allocating funding, supporting decision making on their policies and strategies, tracking researchers' careers, and describing their systemic role to policy-makers, stakeholders and the public. With broad coverage and sufficient completeness, data quality and standardisation, CRIS systems can also provide data for studies of research.

A European COST action, ENRESSH, in collaboration with the ECOOM group at University of Antwerp recently conducted a survey across Europe and found that 23 out of 41 countries have a national CRIS or a bibliographic database or repository with a similar function (Sile et al, 2017). Among the 23 countries are the following participating states in this MLE: Croatia, Czech Republic, Estonia, Italy, Norway, Portugal, Slovenia, and Sweden. Among these eight countries, the Czech Republic and Norway use the CRIS for bibliometric indicators that are directly applied in the PRFS. The example of Norway is given in this report's Appendix C. Croatia, Estonia and Sweden have for different reasons decided not to use their CRIS to produce indicators for their PRFS. They rely instead on data from the Web of Science or Scopus. These examples

are also covered in Appendix C. Italy, similarly included in Appendix C, uses a nationwide database of publications indirectly in the PRFS. In Portugal and Slovenia, the national database of publications has instead been used to inform panels in evaluations. However, from 2017, Portugal will no longer use CRIS information in that way.

An advanced CRIS is not the same as a repository of publications or a bibliographic database. A CRIS unites all information sources that are relevant for the administration of research activities in one dynamic interrelated system. It is by constantly interrelating bibliographic information and other types of information representing the factors influencing scientific production that a CRIS can break new paths in studies and evaluations of research activities. A CRIS has information about identifiable persons (not only authors) and institutional affiliations, titles and organisational roles (not only published addresses) as well as more complete economic information than is available in funding acknowledgements in publications.

Another strength of CRIS from the same perspective, is the possibility of completeness in the coverage of the published literature. Coverage can go beyond the existing bibliographic data sources such as the Web of Science and Scopus (Sivertsen & Larsen, 2012). This may be especially important in research areas such as engineering science, social sciences and the humanities, which tend to make less use of the kind of scientific journals indexed by the commercial databases and which are therefore insufficiently represented in them. While many types of bibliometric analysis, such as studies of collaboration and output profiles, can be performed by using CRIS data alone, there is both a limitation and a new option for citation analysis. The limitation is that CRIS do not index citation themselves. The data will need to be matched to another data source. The new option, however, is that a CRIS makes it easier to attribute non-indexed publications and citations provided by Google Scholar or Microsoft Academic to persons and affiliations.

The use of CRIS for local institutional purposes has become widespread during the last decade and is now served by a commercial market of several professional providers of CRIS solutions. The development of CRIS for studies and evaluations of research, however, is still in an early phase. Lack of data sharing options and comparability is still a major limitation. The data need to be available and comparable across local CRIS. This is so far only the case in a few countries that have managed to establish a national, non-commercial CRIS or a system for aggregating data from local CRIS. This has sometimes been supported by giving the CRIS a role in the national PRFS.

It is on the European agenda to establish internationally integrated CRIS with comparable data. The feasibility of establishing a European Integrated Research Information Infrastructure, mainly for administrative purposes, has been demonstrated in a report to the European Parliament (Mahieu, Arnold, & Kolarz, 2014). In 2016, Science Europe published a Position Statement on Research Information Systems inviting all research organisations to develop resilient information systems by adopting certain core principles and technical recommendations (Science Europe, 2016). Quality control of the data contained in these CRIS would need to be addressed in order to make this useful for PRFS.

Some bibliometricians argue that creating a bibliometric database as a public good would make it possible not only to obtain a uniformly high level of quality assurance of the data input but also make that process and the various calculation processes transparent as well as enabling wider access to bibliometric data than is possible when the databases are private property.

3.2 Bibliometric data sources and indicators

3.2.1 Introduction

This section focuses on bibliometric data sources and indicators that are available for use in PRFS. Possible uses of bibliometrics for PRFS are covered in chapter 4.1. Here, we explain what bibliometrics is in general and its main areas of use. It is important to involve bibliometric expertise if bibliometrics will be included in the development of a PRFS design.

The Frascati Manual (OECD, 2002) defines bibliometrics as a statistical analysis of books, articles, or other publications. Bibliometrics is particularly relevant in the research domain because the publication and communication of new results in peer-reviewed scientific and scholarly publication channels is often seen as an integral and necessary part of the research process itself. Bibliometrics therefore covers the research activity in general by tracing publications and bibliographic records that it leaves behind.

Originally, bibliometrics was limited to collecting data on numbers of scientific articles and publications, classified per authors and/or institutions, fields of science, country, etc., in order to construct simple 'productivity' indicators for academic research. Subsequently, more sophisticated and multidimensional techniques were developed, based on citations in articles (and more recently also in patents). The resulting citation indexes and co-citation analyses are used both to obtain more sensitive measures of research quality and to trace the development of fields of science and of networks.

Bibliometric analyses use data on numbers and authors of scientific publications and on articles and the citations therein (as well as in patents) to measure the 'output' and 'scientific impact' of individuals/research teams, institutions, and countries. The purpose is to identify national and international networks, and to map the development of new (multi-disciplinary) fields of science and technology.

Bibliometrics is an international field of research with its own conferences and journals. The most important conferences are the annual STI ENID conferences and the biannual conferences of the International Society for Scientometrics and Informetrics (ISSI). Examples of main journals are Journal of the American Society for Information Science and Technology, Journal of Informetrics, Research Evaluation, Research Policy and Scientometrics. Introductory courses to bibliometrics are held by CWTS at Leiden University and by the European Summer School for Scientometrics.

In recent years, bibliometric methods have been extended beyond scholarly publishing to include activity and impact in social media, particularly those media that researchers most often use. This evolving branch of bibliometrics is often called **altmetrics** or **webometrics**. As with traditional bibliometrics, the main sources of data and indicators are produced by commercial suppliers. Altmetrics is further discussed in Sections 3.4 and 4.3.

Bibliometrics is mostly used for statistical and monitoring purposes with macro indicators at the level of countries and institutions. Bibliometrics is also often used to inform research evaluation or funding decisions, sometimes directly in funding formulae in PRFS. Bibliometrics also has a long tradition of being used professionally by libraries to monitor their information resources, and by scientists to study the dynamics of science itself. We find the most widespread uses of bibliometrics in

- Statistical reports and strategic documents on R&D at the national or international level
- Information provided to research assessments at the national or institutional level
- Information used by local research management at research institutions
- The assessment components of PRFS, where indicators may be used directly in assessment or to inform peer review assessment
- Informal daily practice among researchers. This has become the most widespread use of bibliometrics in recent years, particularly after the

introduction of specialised social media where researchers upload and share their publications

Most researchers will know and use bibliometric data sources such as the Web of Science, Scopus and Google Scholar as resources for literature search in their daily activities. Many researchers may also know how to use them as the basis for bibliometric indicators of research performance. However, as we have seen, bibliometrics is also a well-defined area of expertise and a distinct field of research. The design of a PRFS will have economic consequences throughout the whole academic community. Our advice is therefore to involve bibliometric expertise (independently of commercial suppliers of data) in the design of a PRFS if bibliometrics will be used for this purpose. It will make the technical design more robust and less susceptible to criticism. The use of expertise should not exclude the involvement of scholars who are knowledgeable non-experts, but represent the best standards in different areas of research. They will be ready to judge the design of bibliometric indicators from the researcher's point of view.

Bibliometric data can be extremely useful but suffer from a number of well-known limitations, which is one of the reasons why it is important to involve experts when using them for a PRFS. There are techniques that can mitigate some of the known limitations. These limitations include

- A known bias in the commercial bibliometric databases towards journals published in English and based in developed countries
- Differences in publication and citation practices among fields and disciplines, leading to a need to field-normalise data in some kinds of comparative analysis
- These differences also mean that the extent to which the databases cover all the outputs varies dramatically among fields. Broadly, coverage is high in the natural sciences, lower in the applied sciences, lower still in social sciences and lowest in the humanities
- The databases only index citations within and between journals included in the database
- Inclusion criteria for journals are not always transparent. To the extent that they rely on numbers of citations, they can be gamed

3.2.2 Bibliometric data sources

The most used bibliometric data sources are the *citation databases*. These are bibliographic databases where publications are linked whenever they refer to each other in the reference list. This method demands a recording of the full reference list of each publication. Citation indicators are only possible if this method (citation indexing) is used.

There are two large commercial citation databases: *Web of Science* (WoS), which is now provided by Clarivate Analytics, and *Scopus*, which is provided by Elsevier. They are built on the same principles, i.e. a selection of the most important scientific journals for indexing, and recording a full range of bibliographic data (e.g. including all published author affiliations). They can be seen as competing products. The two citation databases are now available as toolboxes for local research management, i.e. *SciVal* from Elsevier, and *InCites* from Clarivate Analytics.

Google Scholar (GS) and Microsoft Academic (MA) are also based on citation indexing, but here the method involves algorithms that automatically identify documents on the web as 'scholarly' and discover possible links in the references. There is no restriction to certain journals, the coverage of the published literature is much wider, and GS and MA are so far free for users. Hence, there is a widespread use of them among researchers. However, the algorithms and the databases themselves are not available. Search results are unstable (in part because they analyse what is currently on the Web rather than the totality of what has been published). GS and MA have therefore not yet become standard data sources for professional bibliometrics, but they are being researched and tested within the field.

Several bibliometric indicators, mainly those representing productivity, research profiles (relative composition of disciplinary fields) and collaboration in publications, do not depend on citation indexing. In these instances, other international databases such as PubMed or Latindex may be used. They can be used instead of or as a supplement to citation databases, which are, of course, publication databases as well, but less complete in their representation of the literature.

Of particular interest for bibliometrics in relation to PRFS are local or national bibliometric databases or current Research Information Systems (CRIS, see chapter 3.1 above) that aim to give a more complete representation of the research output. These databases are often produced by or in collaboration with the research institutions themselves. Examples in Croatia are the Croatian Scientific Bibliography (CROSBI) and Hrčak, the central portal of Croatian scientific journals. The R&D Information System in the Czech Republic is an example of a current research information system (CRIS) serving the purpose of a PRFS. The two most used CRIS systems so far that serve the PRFS purpose and have been used in bibliometric studies as well, are the VABB-SHW database in Flanders, Belgium, and the Current Research Information System in Norway (CRIStin). The latter is further explained in Appendix C. CRIS databases may include peer-reviewed scholarly publications that are not covered by the commercial databases. Table 1 below compares

the coverage of scholarly publications from Norway in Scopus and Web of Science to their coverage in CRIStin (100 per cent).

Table 8 Proportion of CRIStin entries also listed in Web of Science and Scopus

	Scopus	Web of Science
Health sciences	79 %	75 %
Natural sciences	84 %	78 %
Engineering	67 %	58 %
Social sciences	38 %	22 %
Humanities	30 %	11 %

The main differences between the CRIS system and the two main commercial databases (Scopus and Web of Science) are the limited coverage in the latter of other publications than journal articles, of scholarly publications in national languages, and of international journals in the social sciences and humanities.

The price of access to the commercial databases has come down substantially in recent years and their coverage has improved. Universities in richer countries often buy access to these databases, though this tends to be expensive. Some lower-income countries have addressed the high cost of the databases by negotiating access for the whole university or research sector – a practice that could be of benefit in better-off countries too. This has been done in Moldavia, Estonia, Romania and the Czech Republic. The Estonian PRFS uses bibliometric data from Clarivate but requires the individual universities to validate their data.

Data from the commercial databases can be provided for a national PRFS in several ways. Most appropriate for this specific purpose, but also most costly, is to set up a replication database in agreement with the provider. This alternative allows for full control of the data and independent research-based design of the indicators. Sweden is an example of this alternative. Less costly is to agree with the provider to provide raw data for only the articles with author addresses in a given country. Norway is an example of this alternative by importing and validating the commercial

data in the national current research information system (CRIS). The least costly alternative is to ask each institution to provide (necessarily a limited set of) data from their ordinary searchable online version of the database. Croatia is an example of this alternative. An intermediate solution is to purchase a bibliometric toolbox (InCites from Clarivate Analytics or SciVal from Elsevier) with ready-made indicators and online only access to data. Neither of the two latter alternatives will provide the flexibility in indicator design and the control and transparency of the data that might be needed for the purpose of bibliometrics for a PRFS.

3.2.3 Most used citation indicators

The most commonly used citation indicators in relation to PRFS are

1. Total number of citations
2. Total number of citations compared to the average in the field
3. Proportion of publications among the most cited in the world in the same field

The first of these indicators is *size-dependent*, the two other indicators are not. The third of the indicators has had increased use in recent years, perhaps because it is understood as representing 'excellence'. But, just as with 'quality', the concept is not easy to measure in one dimension.

An interactive source of bibliometric information in which one can study the typical well-established field-normalised citation indicators by selecting particular countries and universities is the Leiden Ranking: http://www.leidenranking.com/. Technical explanations for the indicators are also given there.

Citation indicators are not very easy to construct. Citation analysis is an expert domain. Caution has to be exercised because the number of citations will depend on several factors such as

- The year of publication (time to get cited)
- The type of publication (e.g. review articles are more cited than original articles)
- The field of research – see next section

3.2.4 Field normalisation of citation indicators

Citation frequencies and productivity vary much across fields, partly because publication and referencing practices are different across fields, and partly because the fields may be covered to different degrees in the database used for the measurement. Indicator 1, above, is not field

normalised and therefore cannot be used to compare publications from different fields, which is often an implicit condition of a PRFS. Indicators 2 and 3, above, are examples of field-normalised indicators. With caution, these can be compared across fields, but a very detailed and sometimes problematic classification is often needed, e.g. distinguishing between neurosciences and clinical neurology.

3.2.5 To be handled with care: the JIF and the h-index

Two easily available indicators are widely used in research assessment, at least informally, but are **not** recommended to be used in *research evaluation* where the purpose is to assess the quality of the researchers' contribution. Even if the purpose is only to *allocate funding* at macro level in a PRFS, one should be careful with the application because there is a risk with these indicators of allocating funding to research that is below the standards that the PRFS aims to improve.

The Journal Impact Factor (JIF) measures the average number of citations to articles in a certain journal within a certain time period. The JIF was developed by the founder of Web of Science, Eugene Garfield, for the purpose of monitoring journal selections in libraries. While the JIF, and the alternative, SNIP (based on Scopus), may be **valid indicators of journal impact**, they **cannot be used to assess the citation performance of individual publications or authors**. The reason is that average citation impact of a journal is only a weak predictor of the citation impact of individual publications in the journal because article citedness tends to be highly skewed among the publications. While a few publications receive many citations, most publications are seldom or never cited. This has been demonstrated in several bibliometric studies since the 1990's.

Nevertheless, the Journal Impact Factor is widely used for the evaluation of individual researchers and their articles. This practice has recently resulted in a series of well-organised reactions from the scientific communities. First came the *San Francisco Declaration on Research Assessment* (ascb.org/dora), which was initiated by the American Society for Cell Biology and now has more than 13,000 signatories across the world. Then, published in Nature in April 2015 by experts in bibliometrics and research evaluation, came the **Leiden Manifesto for research metrics**, an annotated list of ten principles to guide research evaluation (Hicks, Wouters, Waltman, de Rijke, & Rafols, 2015). A few months later appeared **The Metric Tide** report (Wilsdon et al, 2015), which provided the Higher Education Funding Council for England with an independent review on the role of metrics in research assessment and management. All of these documents agree with Eugene Garfield, the inventor of the JIF (Garfield, 2001).

It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the journal impact factor as a surrogate. The latter practice is fraught with difficulties, as Seglen and others have pointed out.

Some countries, e.g. Flanders (Belgium), Italy and Spain, use the JIF explicitly as an indicator of **journal performance**, not of individual performance, in their funding systems. The indicator is thereby not replacing indicators of performance at the article level, but supplementing them with extended information. The reason for using the JIF in such contexts may be to incentivise publishing in world-leading journals.

Another indicator often used informally is **the h-index (or Hirsch index),** which is defined as follows: A research unit has index h if h of its publications each have at least h citations and the other publications each have no more than h citations. The h-index is 3 if at least three publications have at least three citations. Perhaps because of its simplicity, the h-index has recently become a popular bibliometric indicator among amateurs.

In practice, the h-index is almost never field-normalised and therefore contains the problems discussed above under field-normalisation. Most often, it is used with whole counts on individual author level, disregarding the contribution of other authors. The dynamics of the h-index imply that you get the highest score at the end of your career, and that the indicator is likely to increase even after you cease to do research. The above-mentioned Leiden Manifesto warns against using the indicator in research evaluation:

The older you are, the higher your h-index, even in the absence of new papers. The h-index varies by field: life scientists top out at 200; physicists at 100 and social scientists at 20–30. It is database dependent: there are researchers in computer science who have an h-index of around 10 in the Web of Science but of 20–30 in Google Scholar. Reading and judging a researcher's work is much more appropriate than relying on one number. Even when comparing large numbers of researchers, an approach that considers more information about an individual's expertise, experience, activities and influence is best.

Again, unfortunately, there is widespread unofficial use of the h-index in research evaluation. There are, however, many reasons for not installing the h-index in a PRFS for institutional funding. One of them is that it is difficult to aggregate the indicator from the individual level. Another is the risk of harming young careers in the research system. Advanced citation analysis based on actual citation counts per article is preferable.

Some of the most-frequently used indicators used in bibliometrics are based on publication data.

1. Total number of publications
2. Total number of publications compared to input variables (resources for research)
3. Distribution of publications among fields (research profile) compared to the general distribution in the database or in other units of assessment
4. Share of publications with co-authors in certain collaborations, e.g. with other countries or with industry
5. Frequencies of co-authored publication in certain relations

Of these, the first two are most often used in relation to PRFS. The first is size-dependent, the second is not. The three other indicators reflect properties of publication data (field of research, co-authorship) that need to be taken into account whenever productivity indicators are constructed for the purpose of PRFS.

Productivity indicators reflect *research activity*. Sometimes, the aim of a PRFS using productivity indicators is just to allocate funding according to research activity. Productivity indicators in a PRFS may aim at and have the effect to *increase productivity*. They can also increase the focus on stimulating research at the level of *research management*. Using them can demand a data source with such correctness and completeness that the data can support a *better overview of and insight into* ongoing research. Such overview can *create comparisons of research profiles* if the data are properly field-classified. Depending on how the productivity indicators are constructed, they can also stimulate or discourage *research collaboration*.

Just as with citation indicators, the indicators listed above are not easy to design while still producing the balanced representation of fields and incentives to the research and publishing practices that a PRFS aims to stimulate. One consideration is that co-authorship practices differ widely among fields. The average number of authors may differ, and also the norms with regard to the sequence of authors and the importance of some positions, e.g. the first or corresponding author. As a consequence of these differences, productivity at the level of individual researchers or at department level cannot be compared directly. This is often neglected, even if field-normalised citation indicators are applied.

In professional bibliometric studies on the macro level, e.g. when comparing countries or institutions, both publications and citations are distributed among contributing authors and affiliations by using *fractional*

counts. The alternative, however, *whole counts* of publications to each contributor, may be relevant for other purposes, such as assessing individual researchers by their CV. Intermediate solutions are also possible (see the example of Norway in Appendix C Section 8.3.3).

Particularly in any *direct* use of bibliometrics in PRFS, all aspects of field-normalisation and its possible consequences need to be simulated, discussed and resolved.

3.3 Papers for peer review

The amount and sort of papers and data needed as an input for peer review-based PRFS depend on the goals and criteria of the assessment. In general, the scope of assessment in peer review contexts is broader than in bibliometric approaches, which means that input data are more and more heterogeneous. **Scholarly publications** (journal articles, books, conference proceedings) are at the very core of information used in a peer review based PRFS. In most contexts other research outputs as videos, exhibitions or objects are allowed to be submitted too. Furthermore, material like lists of publications, statistics, indicators, impact statements as well as self-assessments may constitute important input data into a peer review based PRFS.

The following questions are crucial for the decision about necessary input data.

1. **Selection of Research outputs**: Who decides what work is to be considered and how do they do it?
2. **Assessment criteria:** Do the assessment criteria concentrate on research quality only or are other criteria such as impact, strategy, management etc. included?
3. **Informed peer review**: To what extent is peers' judgements 'informed' by indicators, metrics and bibliometrics?
4. **Self-assessment:** Is the assessment carried out in two stages with a self-assessment followed by an external assessment (see Figure 9)?
5. **Prospective elements**: Does the assessment focus on past performance exclusively or are there any prospective elements involved?

3.3.1 Selection of research outputs

In order to keep the assessment manageable, the number of research outputs to be submitted is limited and the outputs therefore have to be selected by the unit of assessment. This may cause **biases** if risk-averse university managers are less likely to submit interdisciplinary or

heterodox work than disciplinary and high-status journal articles (Elsevier, 2015).

In UK REF 2014 the number of outputs, hence individuals included, scaled with the number of impact case studies submitted. Due to some evidence on long-term consequences to individuals who are not returned in the REF, Lord Stern's review of the REF recommended that "All research active staff should be returned in the REF" (Stern, 2016). Nevertheless, comprehensive inclusion of staff may not only increase the costs of the assessment but also obscure the presence of some outstanding researchers in a unit whose level is otherwise average.

In the context of the MLE, all the participating countries limit the number of papers and other inputs into the peer assessments whereas bibliometrics covers all research (at least in Italy[5], Norway, Spain, Slovenia and in the Czech Republic). Furthermore, unlike in UK practice, there are **no restrictions concerning the researchers themselves**.

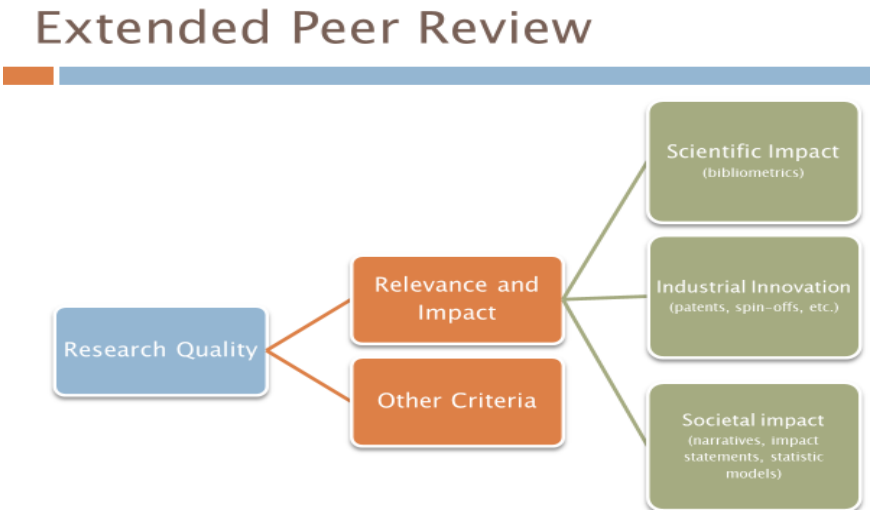
3.3.2 Assessment criteria in peer review

Most peer review based assessments include criteria apart from research quality, mostly impact criteria. However, the introduction of non-scholarly criteria such as impact makes new and often unfamiliar demands on peers. These affect the type of people who should make judgements, typically extending from scientific peers to include practitioners, or using additional information such as statements from researchers about the perceived societal impact of their work.

In a recent study, Derrick et al. investigated REF evaluators´ divergent opinions prior to the assessment. They showed that there was a wide range of evaluator views about impact, and suggested that these views could be depicted via a range of different positions along a conceptual evaluation scale (Derrick & Samuel, 2016). The following picture 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.

[5] Note that in Italy no bibliometrics is used in the humanities or in the social sciences apart from the economics and statistics fields.

Figure 8 Extended peer review



In the context of the MLE, nearly all participating countries focus on both **research quality and impact** as the most important **criteria for assessment**. In the case of Croatia, impact is ranked even more highly than research quality. Italy and Spain are the two countries that concentrate on research quality only.

Patterns in the use of other assessment criteria are less clear. Some countries include the promotion of young researchers, internationalisation and collaboration under various nomenclatures; 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.

Table 9 Assessment Criteria

Criteria	HR	EE	IT	MD	NO	PT	SI	ES	TR
Quality/Excellence/ Scientific impact	x	x	x	x	x	x	x	x	x
Sustainability/Human Resources		x		x	x	x			x
Impact economy/society	x	x		x	x	x	x		x

Criteria	HR	EE	IT	MD	NO	PT	SI	ES	TR
Management/Efficiency/	x				x	x	x		x
Interplay of research and education					x				

3.3.3 Informed peer review

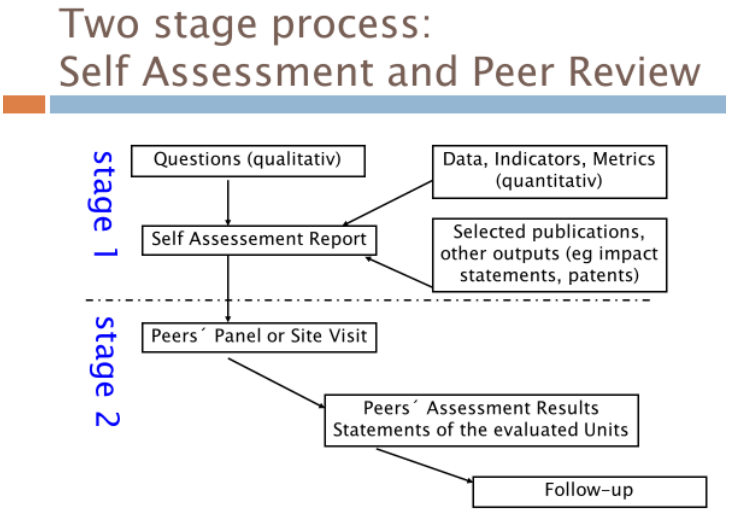
The idea that the **combination of peer review with metrics**, ie quantitative indicators and especially bibliometric indicators, may enhance the evaluation process lies behind the introduction of 'informed peer review', where peers get informed by metrics which allows comparisons and triangulation across methods. Some authors argue that bibliometric indicators can make peer review **more transparent and offer additional insight** in cases of diverging panel members´ opinions (Mahieu & Arnold, 2015). 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 (Fröhlich, 2016).

- Of the countries participating in the MLE (Figure 7), Slovenia, Italy, Turkey, Portugal and Spain use PRFS based on informed peer review. Others use combinations of peer review and bibliometrics in different forms
- **All countries use some indicators and metrics** to inform the peers, in most cases bibliometrics. In this context, one should note two specific cases
  - Portugal´s PRFS formerly had many characteristics of a 'pure peer review model' but in the most recent round peer review is informed by bliometrics
  - 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.

3.3.4 Self-assessment

In many cases, the assessment is carried out in two stages with a self-assessment followed by an external assessment. The self-assessment includes both qualitative and quantitative information and combines data with normative and formative elements.

Figure 9 Self-assessment and peer review as a two-stage process



In the context of the MLE, several of the participating countries (Croatia, Estonia, Moldova and Portugal) work with self-evaluation reports, often including qualitative evaluation. In the Estonian and Moldovan cases, these are associated with the period review of institutions' eligibility for institutional funding, while funding is reviewed annually based on metrics. 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 and reflect on meaning and relevance.

3.3.5 Prospective elements

Few PRFS contains prospective elements. In the UK REF, some opportunity for the presentation of future plans is given in the context of the impact template where submissions can describe "their approach to impact during the assessment period as well as their development of a strategy and plans for the future" (Submission Guidance 2012).

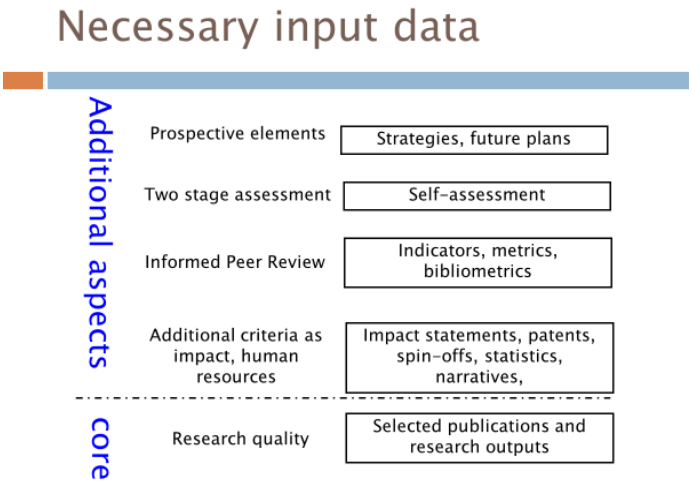
Prospective elements are inherent in the performance contracts used as steering instruments in, for example, Finland, the Netherlands, Austria and some German Länder. Even if direct links between PRFS and performance contracts are rare, peer review based research assessments often create the basis for the formulation of the agreements (eg in Austria or the Netherlands).

In the context of the MLE, nearly all participating countries include prospective elements in their wider peer review based research

evaluations (ie outside the PRFS). Prospective elements are especially important in contexts where self-evaluations are part of the process and where formative evaluation elements are included.

Figure 10 shows how different criteria and assessment aspects affect the amount and type of papers and other research outputs as well as data needed as inputs for peer review based assessments.

Figure 10 Input data for peer review- based PRFS



3.4 Data and indicators about impact

3.4.1 What is impact?

The word impact, as far (academic) research is considered, has been defined in many different ways, but basically refers to the effects of scientific work in the world outside the lab or office. This outside world can be divided in (i) other researchers and (ii) other people interested in (a particular sort of) research. Other researchers may refer to people working in the same field, but also in neighbouring fields, or even in quite different fields. The latter is often the case when multi- and interdisciplinary work is performed. This kind of impact, referred to as scientific or academic impact, can be measured through a large variety of indicators tied to the scientific literature and connected databases like Web of Science (Bollen, van de Sompel, Hagberg, & Chute, 2009). Well known examples are citation or other network analysis and the h-index.

The second category of impact related to other people interested in research may refer to a great variety of entities in society: individuals, small businesses, industry, governmental organisations, consumer groups, NGOs, and so on. Many of these entities may have researchers of even research departments, both in the private and the public sector. To measure this kind of impact, referred to as societal impact or broader impact, is much more complicated because of the many different contexts in which impact takes place and existing limitations on building robust databases (Bornmann, What is societal impact of research and how can it be assessed? A literature survey, 2013).

It is not always possible to make a sharp distinction between these two kinds of impact, first, because many research outcomes are relevant both to the academic community and society at large. Addressing most major problems in society these days requires input from many different academic disciplines as well as other expertise, witness for example the Grand Societal Challenges of the Horizon 2020 framework programme. Second, because the internet and its assorted search engines and databases enable ever wider audiences to have access to the scientific literature and thus be impacted by scientific work.

Having said this, most current evaluation practice is dominated by strong and accepted measurements for scientific impact but struggle to find ways to assess societal impact.

3.4.2 Assessing societal impact of research

In this section, we focus on the development of societal impact measurement by exploring three questions.

- How can we measure societal impact?
- Is it possible to collect reliable data about impact?
- And, if so, can we develop robust indicators?

3.4.2.1 How can societal impact be measured

In the light of the above, societal impact of research refers to the ways that work contributes to demands and questions coming from various groups in society, usually referred to as stakeholders. Evidence of the contribution to societal challenges can be delivered in various ways, some quantitative, other qualitative. Taken together we refer to the methods used in this area of research evaluation as 'third stream metrics' – but note that metrics may refer to both quantitative and qualitative instruments.

The development of robust third-stream metrics is not well advanced in most countries, yet it is "difficult but necessary" as the editor of the British Medical Journal, Richard Smith, wrote back in 2001. And currently, it is even more necessary, given the fact that research funding bodies at both national and at European levels are under pressure to provide more and more information about the societal impact or relevance of research.

While it is difficult to find robust metrics for societal impact measurement, some real progress has been made in this century. If we look at the literature on impact assessment, and at different evaluation research projects, and at national or institutional systems, we can distinguish at least three strands of impact measurement that show promising developments.

- Emulating what is done when scientific impact is measured, mostly using quantitative indicators. Initially, this strand focused on economic indicators such as cost-benefit analysis, but now we see a broadening of the spectrum to non-economic indicators that provide information about social, cultural, environmental, health, and other effects (Joly & Matt, 2017)
- Developing measurements that better fit the societal context. These are often more qualitative, such as case studies or impact pathways. In sectors like health research, agriculture and development studies, there is considerable experience with the impact pathway approach (Webb, 2013). Another part of this strand contains variations of network analysis, like the SIAMPI approach of productive interactions (Spaapen & van Drooge, 2011)
- Altmetrics, a relatively recent development using web-based metrics to assess the impact of scientific work in society, and sometimes also in scientific contexts (Bornmann, 2014)

We elaborate these three strands in terms of options for assessment in section 4.3.

3.4.2.2 Is it possible to collect reliable data?

The collection of reliable data for societal impact will always be more difficult than is the case when assessing scientific quality or impact. There is not one comprehensive and reliable data source (like the Web of Science or Scopus) for societal impact that has a reasonable coverage of all or most fields or disciplines. Rather, there are many different databases and other data sources which can be used for this purpose. Some of these are directed towards a certain sector of society (for example *lexis nexis,* which is oriented towards law and government policy), some are institutional where universities or institutions create websites with information about society oriented projects and some are

national (for example the impact case study database of the REF UK). The reliability and robustness of these database obviously varies, and making them more reliable and systematic is difficult because as a rule they aim to serve different audiences with different expectations.

Nevertheless, attempts are being made to build databases that can support the evaluation of societal impact, both at the institutional and the national level. A recent report on the impact of the Canadian Federation for the Social Sciences and Humanities (2017) gives an overview of the different kinds of impact scientific work can have on five different sectors in society, both in terms of scientific and societal merit. The authors divide scientific impact (referred to as 'academic' impact) into scholarship measured through bibliometrics and other indicators, and capacity measured through surveys and various qualitative measurements.

Societal impact is divided into three broad sectors, practice and policy, society and culture and economy. For each of these sectors, a number of indicators (or indicator categories) is offered, to be filled with quantitative or qualitative measuring instruments. The Canadian Federation encourages institutions to build supporting databases that can deliver information that can help the evaluation of societal impact.

In the Thematic Paper on third stream metrics developed in the context of this MLE, other examples are given of possible indicators for societal impact, for instance from the SIAMPI project on productive interactions. What becomes clear from looking at these indicators, or better indicator categories is that in some cases it makes sense to collect quantitative data while in other cases qualitative data are preferable. But in all cases, the development of reliable and sensible indicators is still at an early stage.

Table 10 Approaches to assessing impacts in the humanities and social sciences

Scholarship	Capacity	Economy	Practice and policy	Society and culture
Bibliometric indicators Downloads from Open Access repositories Citations in grant applications Acknowledgements Prizes and awards Reputation as	Number and quality of experiential learning/ research opportunities for students Surveys of students and	Advisory roles and board memberships Revenue opportunities and cost savings in the public, private and	Invitations to participate as an expert witness, an advisor, on an expert panel or committee Citations in government documents	Number and quality of partnerships between researchers and community groups Requests for consultancy/ advice from community groups Media coverage of

Scholarship	Capacity	Economy	Practice and policy	Society and culture
measured by survey Post-publication peer review (book reviews, dedicated symposia) Juried exhibitions and performances	alumni Employer surveys Integration of research as a learning outcome in courses	not-for-profit sectors resulting from research applied in practice Income derived from patents, patent licensing, copyright and trademarks Consulting contracts	Consulting for governments or think tanks Commissioned reports	research (newspapers, TV, online) Requests for media appearances Engagement of the public at events Research-related social media Public use of research-based web resources on social and cultural issues

Source: Canadian Federation for the social sciences and humanities 2017

3.4.2.3 How can we develop robust indicators?

Despite the difficulties, many attempts are being made to make progress in this area. Examples are wide ranging, also in geographical terms. The Canadian initiative mentioned above is one example, other examples are: the NSF in America, working on criteria and indicators for what they refer to as broader impact (Watts, George, & Levey, 2015); the REF UK database presenting systematic information on impact case studies for all different fields and disciplines; in the Netherlands a special database that includes indicators for societal impact has been developed for research in the humanities (www.qrih.nl); in sub-Saharan Africa, 22 of the region's universities started a process of institutionalising research uptake management (RUM), under the umbrella of the Development Research Uptake program in Sub-Saharan Africa (DRUSSA, 2017).

The common denominator in all these initiatives arguably is that there is no one-size-fits-all approach, and therefore a combination of methods and quantitative and qualitative indicators is the obvious outcome. For example, information regarding publications addressed to wider audiences coming from databases is combined with qualitative information based on interviews, and with policy documents or institutional plans.

Each of these data sources will have to be as robust as possible in order to play a meaningful role in the evaluation process, and specialists will have to work on making these data sources ever more systematic.

3.4.3 Experience of countries participating in the MLE – building-blocks for assessing impact in PRFS

Given the above, it is no surprise that most countries participating in the MLE when asked what kind of methodological developments can be distinguished with respect to third-stream metrics, do not mention concrete instruments. Many countries however expressed the need for such instruments. Some countries presented provisional attempts to develop a more systematic approach to evaluating the contributions of science to society. In many countries data collections exist at the national level, containing information relevant to answering questions regarding the societal impact of research. So while few concrete methodological developments have been made, there are many examples of potential building blocks collecting useful data and instruments which could be used in impact evaluation.

Evaluation criteria. Many countries have introduced criteria for societal impact, either in their national evaluation system or in targeted programmes, like for example the Estonian "Support for applied research in smart specialisation growth areas", which includes criteria considering project contribution to the cooperation between public R&D institutions and other organisations, direct contribution to the development of smart specialisation growth areas and economic impact and efficiency of the project. In Portugal, three broad sets of criteria are used.

- Quality, merit, relevance and level of internationalisation of the R&D performed during the period under evaluation, including publications, patents and technology transfer, but also young researcher training, promotion of scientific and technological culture and other activities – scientific, technological, cultural, artistic or economic, of special relevance to society
- Scientific merit of the team of researchers, also taking into account the researchers' level of national and international recognition
- Adequacy of the objectives, strategy, activity plan and organisation for the next 5 years, including plans to hire new researchers

In Croatia, criteria for societal impact also are part of an overall funding scheme which is divided into the following categories: scientific productivity – 60%; national and international competitive research projects and research mobility – 25%; collaboration between research and business sector, as well as collaboration with local and regional government and the non-governmental organisation (NGO) sector – 10%;

popularisation of science – 5%. Analysis is based on the individual performance of each researcher.

Countries not having a PRFS may also use metrics for the evaluation of quality or impact. Moldova, for example, focuses on teaching activities, using indicators about the teaching process, teaching staff, research, and so on. This additionally serves as a reminder that educating young people is an important impact on society. Moreover, Moldova has an ambition to put more focus on other quality and impact aspects like gender equality, young researchers and how to get a more even distribution of research among institutions.

Experimental metrics. Countries may experiment with metrics for third stream activities. For example, in Italy ANVUR is developing, on an experimental basis, a new system of metrics designed to support evaluation of third-stream activities and the impact of academic research. Third-stream activities have been divided in two main areas, respectively involving the economic valorisation of research and the production of public and social goods. Research valorisation indicators are related to intellectual property management (patents and plant varieties), academic entrepreneurship (spin-offs), third party activities and intermediation activities. Indicators for the production of public and social goods include indicators addressing the management of cultural activities and the cultural heritage (museums, archaeological excavations and cultural heritage), clinical trials, continuous education and public engagement. Evaluation is based on peer review, informed by these indicators.

Specialist research centres. Some countries have set up research centres, like the Osiris centre at the University of Oslo, to improve understanding of the ways in which research impact is generated. The centre applies both qualitative and quantitative approaches to their research questions. (http://www.sv.uio.no/tik/english/research/projects/osiris/About/research/)

Infrastructures. The pan-European distributed research infrastructure SHARE (*Survey of Health, Ageing and Retirement in Europe*) is a multidisciplinary and cross-national longitudinal database of micro-data on people aged 50+ and their partners regarding health, socio-economic status, social and family networks and other issues collected from more than 85,000 individuals from 20 European countries. SHARE allows researchers and state administrations to understand the consequences of demographic changes and formulate public finance, labour market, health care and pensions policies.

Countries also build national databases for research and higher education, like the Norwegian CRIStin CRIS. This includes a wide range of indicators, for instance business disclosures, patent applications, licence agreements and start-ups, income from business, NGOs and the public sector as well as indicators for the popularisation of research.

Also at the national level, countries use statistics that cover different aspects of impact. For example, in Sweden the innovation agency Vinnova has developed and piloted a non-metrics model in close collaboration with universities and others (Wise, et al., 2016). It was intended that this model would be used for distributing 1/3 of the performance based part of universities' institutional funding, together with bibliometric indicators (1/3) and External funding (1/3) but that intention was subsequently shelved.

Government policies to stimulate research with societal impact. In many European countries we find policies that focus on co-operation with partners in society, most often industry, such as the top sector policy in the Netherlands. Among the participating countries, Austria created incentives for third-party funding in the area of EU programmes: increased Framework Programme funding; high co-operation with industry (due to their size, many EU projects can only be achieved meaningfully in cooperation with companies/industry); international competition; global networking; efficient knowledge production through labour and cost sharing; and focus on societal challenges.

In Armenia there are three main funding instruments, (1) basic funding (with criteria including novelty of the research, importance, feasibility, human and technical resources, expected results, (2) Thematic funding for projects (with similar criteria), and (3) targeted programmes (novelty etc, but also criteria related to state policy, expected results and impact).

In Turkey the so-called Entrepreneurial and Innovative University Index has been developed. The index introduces a dedicated performance evaluation and monitoring system for tracing the activities of universities that benefit the private sector. The theoretical framework, indicators, data availability and accessibility, calculation and normalisation method, and weighting schemes are defined in consultation with relevant government organisations and universities. This is intended to ensure that the final index scores provide a meaningful distillation of the available information.

We can see a need in most countries for funding instruments that aim to reduce the gap between academic research and wider society. In most cases, the focus is either on business or on state policy. There are no cases where comprehensive third-stream metrics are integrated in PRFS or institutional evaluation methods. But we can see a number of

approaches to evaluate specific elements of societal impact. They range from sometimes very complex and demanding systems (Turkey, Italy in development) to fairly simple indicators (Croatia, Estonia) and everything in between. We also see experimental methods, the most elaborate in Italy. Sometimes a case study approach appears to be very informative. Some countries express a clear need for better instruments to measure societal impact.

In summary, countries do not have third stream metrics fully integrated in PRFS, but the overview of building blocks above shows that there are building blocks at all levels of the higher education and research systems that can be used to further develop societal impact measurement instruments.

3.4.4 Lessons

For third stream metrics, there are still many uncertainties regarding robust indicators. Currently, the best option is to allow for experimental development, and to improve the communication among the various attempts being made in different European countries. Practically speaking, it is best to rely on a combination of qualitative and quantitative methods, and use this information always in a broader review process where peers and other experts are involved. Both types of methods need further sophistication and the same goes for the supporting data infrastructures.

There are developments in third-stream metrics in many places but few research centres devoted to the subject. Similarly, many institutes conduct research on bibliometrics but we barely have similar institutes with regard to societal impact measurements. The Norwegian Osiris project is a new initiative here. CWTS in Leiden and the Rathenau institute in The Hague, have part of their agenda devoted to societal impact assessment.

3.5 The role of self-assessment

While self-assessment is in widespread use both among R&D evaluators and the broader evaluation community, rather little use is made of it in PRFS. Self-evaluation is more generally used in formative than summative contexts, so we would expect to see it mainly in PRFS that have a formative dimension. This also means that there is likely to be an element of peer review in the PRFS, since qualitative information from self-assessment is hard to integrate into a metrics-only framework. Self-assessment is a component of the Dutch Standard Evaluation Protocol (SEP), which is a formative research assessment that is not strongly connected to funding.

Among the states participating in this MLE, Croatia, Estonia, Moldova and Portugal use self-assessment within the PRFS, requiring research groups to report on their strategy as well as to provide other factual information. The quality of the strategy is explicitly considered in these peer review based PRFS.

The UK REF is another example of a peer-review based PRFS that includes qualitative self-evaluation data (in the 'environment' submission). It contains the following elements, which are typical of those found in self-evaluation requests elsewhere. (For example, a very similar set of questions was asked in both the Latvian and Lithuanian research assessment exercises, which were formative as well as summative, providing specific feedback about the self-evaluation statements as well as judgements about research outputs.)

- Overview, explaining what the research group does and pointing out key achievements
- Research strategy
- People: staffing strategy, staff development, various aspects of PhD training, gender equality and diversity, international recruitment
- Income, infrastructure and facilities

In some situations, it is necessary to ask in self-evaluations for 'hard' data such as numbers of people, lists of publications, numbers of PhD students etc. But wherever there are reliable national information systems (including where relevant a national CRIS) these should be used for data collection as self-evaluation reports are burdensome to produce and data are often unreliable.

Any time a self-assessment questionnaire is used, it needs to be accompanied by an explanatory completion guide containing definitions, in order to minimise differences of interpretation about the facts to be collected. Experience from the Latvian and Lithuanian research assessment exercises as well as from other situations (for example recent evaluations of the Humanities and the Social Science Research Institutes in Norway) makes it clear that in some situations there is a need to alert people to how to deal with providing such information in the context of evaluation. The extent of 'evaluation culture' among different groups of researchers differs across both countries and subjects. It is easy to disadvantage groups with little evaluation experience.

3.6 Lessons learnt

Key messages from this Chapter are as follows.

Bibliometrics

- A quality-assured CRIS is a very useful tool to support PRFS, especially in countries using 'small' languages, which tend to be under-represented in the commercial bibliometric databases and where a recognised national authority can categorise outputs in terms of quality
- Care must be taken in populating a CRIS to ensure that data are quality assured, complete and extend to those categories of non-scholarly data needed for research assessment
- Bibliometric tools are now widely available. However, generating and using robust bibliometric indicators still requires skill and considerable understanding of the limitations of the relevant databases and the characteristics of different fields of research. Hence, professional bibliometricians (independent of the commercial database vendors) should be involved in the design and implementation of any PRFS that uses bibliometric information
- Journal impact factors (JIF) and h-indices are easily accessible indicators but should be interpreted carefully
- Details matter in designing bibliometrics for a PRFS. For example, the way multiple authors and their institutions are credited affects incentives for collaboration
- Limitations of bibliometric data need to be understood. These include bias towards English-language journals, differences in the amount and pattern of publication among disciplines, the extent to which research outputs figure in the available databases, the self-referential nature of those databases and the lack of transparency and opportunities to game the criteria governing which journals are indexed
- Where bibliometric indicators are used in PRFS, they should be field-normalised, taking account of differences in productivity, use of publication channels, citation practice, authorship and co-publication conventions

Peer review

- The choice of information provided to peer review panels depends upon the purposes of the PRFS and the criteria used

- The choice of peers and panel members influences the outcome of assessment and needs to be considered in order to obtain a fair assessment across disciplines and schools of thought
- Most PRFS focus on the quality of research and its scientific impact. However, a growing number of other criteria are being used, especially in relation to societal impact. However, peer reviewers are not necessarily good judges of the impact dimension
- Peer review in PRFS is increasingly 'informed' by bibliometrics. This has the benefit that panel members can work with relatively robust indicators, rather than themselves looking up crude measures such as JIFs or h-indices
- In the absence of reliable impact indicators, peers have proven themselves able to assess narrative impact statements though it is not yet known whether these judgements are reliable or reproducible
- Peer-based PRFS often make use of self-assessment. These make it possible to introduce a formative element into the PRFS

Societal Impact

- Societal impact comprises many different things and can occur through a wide range of mechanisms. A useful categorisation is to distinguish among policy and practice, society and culture, and economic impact
- There is no uniform or comprehensive set of impact indicators available, despite the considerable research effort that continues to be devoted to the task. PRFS therefore have to use more or less ad hoc measures, preferably combining quantitative and qualitative methods
- Nor is there any case of systematic integration of third stream indicators into a PRFS in the countries studied
- Impact assessment methods are being developed but have to be customised to the individual context and tend to be labour-intensive. It is not evident that these are yet ripe to be used systematically across fields in the way necessary in PRFS
- There are nonetheless promising developments in societal impact assessment, both in terms of qualitative assessments (for example impact pathways) and quantitative assessments (for example productive interactions). Mutual learning is strongly advised.

Self-assessment

- Self-assessments are necessary for collecting qualitative information, for example on strategy, human resource management and research infrastructure. Where other facts – such as lists of publications or

numbers of researchers – are needed it is preferable to obtain them from quality-assured sources. When self-assessment is used, the PRFS manager must ensure that the information needs are well explained and understandable not only by people used to being evaluated but also by those with little or no experience

- Qualitative aspects of self-assessment can only be analysed by peer or expert panels. They have no place in metrics-only systems

4 THE PROCESS OF ASSESSMENT

This Chapter considers how assessment is done respectively in bibliometrics- and peer review-based models and considers the state of the art in societal impact assessment in relation to PRFS.

4.1 How is assessment done using bibliometric methods?

4.1.1 The purposes of the PRFS and the use of bibliometrics

This report distinguishes between four major design alternatives

- Peer-review, in which all judgements are made by peers
- Informed peer review, in which bibliometric and other indicators are taken into consideration by the peers, but the peers themselves make all the assessments
- Mixed peer review and metrics, where peer judgements drive parts of the performance-based funding while metric indicators drive other parts
- Metrics-based systems, where metrics alone drive funding allocations and there is no role for peer review

Among the countries participating in the MLE, Moldova does not have a PRFS while Austria uses performance contracts with a small performance-based incentive based on income from external funding from the research council (FWF) rather than a more specific research assessment. Among the eleven remaining countries, Italy, Portugal and Slovenia use informed peer review (alternative 2) while Armenia, Spain[6] and Turkey use a mix of peer review and bibliometrics (alternative 3). In all these countries, Web of Science and/or Scopus are the main bibliometric information sources, both for citation-based and publication-based indicators. The five remaining countries, Croatia, Czech Republic, Estonia, Norway and Sweden, have PRFS with direct use of metrics (alternative 4). The Czech Republic and Norway build the bibliometric indicators on data from their national research information systems while the other three countries make use of data on publications and citations in the Web of Science or Scopus. Italy (alternative 2, informed peer review) informs certain panels using an advanced bibliometric model that may be of interest also in countries with a more direct use of bibliometrics. In order to cover the variation in the use of bibliometrics for PRFS in a representative way, we

[6] The Spanish Sexenio System is regarded as a PRFS in this context.

have chosen to present five of the participating countries (Croatia, Estonia, Italy, Norway, Sweden) with more details about the motivation, design and effects of their use of bibliometrics in Appendix C.

In a wider international perspective, New Zealand and the United Kingdom are the only countries using alternative 1[7]. Several other countries can be placed in each of the three other categories. Many European countries have introduced performance based research funding systems (PRFS) for institutional funding. An increasing trend is evident when comparing three overviews of the situation at different times (Geuna & Martin, 2003) (Hicks, 2012) (Jonkers & Zacharewicz, 2016). There are more countries using alternatives 3 or 4 than 1 or 2. To understand the widespread and different use of bibliometrics for PRFS, there is need to distinguish between the two main purposes of a PRFS and, related to this distinction, the two main forms of use of bibliometrics in PRFS: namely, informing panel evaluation with bibliometrics; and the direct use of bibliometric indicators in the funding formula.

The two main purposes of PRFS are *research evaluation* and *funding allocation*. In principle, the two main purposes can be difficult to distinguish. Hicks (2012) defines PRFS as related to both purposes; they are "national systems of research output evaluation used to distribute research funding to universities". Nevertheless, in national contexts, one of the two purposes can be more relevant than the other for understanding the design of the PRFS. We will use the examples of four countries to illustrate this.

United Kingdom was the first country to introduce a PRFS in 1986. Peer review of individual research outputs was the chosen *method* for funding allocation (Geuna & Martin, 2003). Gradually, the method has become the more important *purpose*. The national research assessment exercise is now inextricably bound up with UK research culture and policy. This is the specific context in which the Metric Tide report (Wilsdon et al, 2015) concludes that peer review is needed: "Metrics should support, not supplant, expert judgement." This recommendation, however, is not followed by the majority of other European countries in their design of PRFS.

Sweden recently decided between a metrics-based and a peer review-based alternative. A UK-inspired model for resource allocation based on

[7] The Stern Report (Stern, 2016) suggests however that a number of panels have used bibliometrics to inform their assessment in the 2014 exercise and that the 2018 exercise may be characterized by a more systematic and transparent use of bibliometrics

expert panels, FOKUS (Swedish Research Council, 2015), was designed but the government decided not to implement it, mostly for reasons of cost, but also because the universities are concerned about their institutional autonomy and want to organise research evaluations themselves (Swedish Government, 2016). Some also objected that peer review was redundant because the publications considered in the existing PRFS indicator were already peer reviewed. This objection rather misses the point that peer review in journals involves a judgement about whether articles exceed the threshold of quality that makes them publishable while peer review in a PRFS distinguishes among high-order levels of quality. Sweden consequently continues with the solution it has used since 2009, where a small part of the resource allocation for research is based on indicators of external funding and of productivity and citation impact within Web of Science. The understanding in Sweden is now that the purpose of *research evaluation* must be achieved by other means than the PRFS, while the funding model continues to be mainly about funding allocation. The emerging alternative for the purpose of evaluation is that each university runs a research assessment exercise by itself and with the help of international panels of experts. As an example, Uppsala University is currently running a research evaluation named "Quality and Renewal" where the overall purpose is to "analyse preconditions and processes for high-quality research and its strategic renewal".

Norway not only has a metrics-based PRFS but also runs periodic field evaluations, which are not used for funding allocation. The purpose of having subject-specific evaluations at long intervals is to provide critical reviews of parts of the Norwegian research system in an international perspective, and to provide recommendations about measures to improve the quality and efficiency of research. Norway's PRFS, on the other hand, is designed for other purposes that also typically may motivate these systems (Jonkers and Zacharewicz, 2016): increased transparency of the criteria for funding, enhancing the element of competition in the public funding system, and the need for accountability coupled to increased institutional autonomy. Norway's PRFS includes several performance indicators representing both research and educational activities. Hence, the idea of replacing the indicators with panels performing research evaluation has never been discussed. Research evaluation takes place via field evaluations and resource allocation is the single purpose of the PRFS.

The Netherlands is similar to Norway in the sense that there is a research assessment exercise at certain intervals which does not influence resource allocation. The Dutch exercise is self-organised by each of the universities and coordinated on the national level by a Standard Evaluation Protocol. With this autonomous self-evaluation system in place, there is an agreement with the government that performance

indicators representing research should *not* be part of the performance-based funding system, which covers other activities.

As these four examples show, PRFS need to be examined in their national contexts to understand their motivations and design. In the following, based on experience and discussions in the MLE on PRFS, we give specific advice for each of the two main forms of use of bibliometrics in PRFS: namely, informing panel evaluation using bibliometrics; and direct use of bibliometric indicators in the funding formula.

We distinguish between the two sets of recommendations for the two main forms because the typical problems and considerations that arise in the design of direct use of bibliometrics for PRFS are to a great extent different from those connected with bibliometrics for informing panels in research assessment. For example, the problem of evaluating interdisciplinary research is inherent in assessments by disciplinary panels but is not as urgent in bibliometrics, where there are methods to circumvent it. On the other hand, disciplinary panels have less need to consider field differences in publication and citation practices, while bibliometric indicators certainly need to do so.

4.1.2 The use of bibliometrics for informing panel evaluation within a PRFS

The Metric Tide report (Wilsdon et al, 2015) on the use of metrics in research evaluation is a good basic source of discussions and guidelines whenever the decision is taken to base the PRFS on panel evaluation that is partly informed by bibliometrics. The report is the result of thorough discussions of the experiences with a PRFS based on research evaluation in the UK in combination with an international outlook and the involvement of independent experts. The report is independent of the Research Excellence Framework (REF) but gives advice to the government for its further development.

Italy and Portugal are two other countries that have approaches to PRFS similar to the UK. We recommend learning directly from such countries if a 'UK-style model' is chosen. As examples, Italy has managed to replace some of the labour-intensive work of the panels with an advanced bibliometric assessment model, while Portugal allows the bottom-up formation of units of assessment across institutions (a practice possible also in the Dutch SEP where the exercise only indirectly affects institutional funding).

It is generally acknowledged among bibliometric experts that no one indicator can alone cover all aspects of research. The h-index is an example of a simple indicator that removes, rather than condenses, the

information that is already there about the productivity of the unit of assessment and the citation impact of each of the publications. Expert panels in research evaluation can be informed by several sources of information. They can be informed via a larger set of bibliometric indicators covering different aspects of performance, and not just one indicator. Examples of sets of bibliometric indicators can be found in reports from professionally conducted institutional or national research assessments.

Apart from these general recommendations, our advice is to follow the **Leiden Manifesto for research metrics** (Hicks, Wouters, Waltman, de Rijke, & Rafols, 2015) and use its annotated list of ten guiding principles as a check list in the design process of a PRFS with bibliometrics informing panel evaluation. The ten principles are

1. Quantitative evaluation should support qualitative, expert assessment
2. Measure performance against the research missions of the institution, group or researcher
3. Protect excellence in locally relevant research
4. Keep data collection and analytical processes open, transparent and simple
5. Allow those evaluated to verify data and analysis
6. Account for variation by field in publication and citation practices
7. Base assessment of individual researchers on a qualitative judgement of their portfolio
8. Avoid misplaced concreteness and false precision
9. Recognise the systemic effects of assessment and indicators
10. Scrutinise indicators regularly and update them

In addition to the publication of the Leiden Manifesto in *Nature*, there is a video freely available with explanations of each of the ten principles: https://vimeo.com/133683418

4.1.3 The direct use of bibliometrics in the funding formula for a PRFS

A direct use of bibliometrics in the funding formula may be seen as in conflict with the first principle of the Leiden Manifesto (quantitative evaluation should not replace qualitative, expert assessment) if the emphasis of the PRFS is on research evaluation. Only if the purpose of the PRFS is mainly funding allocation, and the purpose of research evaluation is taken care of by other means, can direct use of bibliometric indicators

be defended. In addition, the ninth principle cannot be disregarded: Recognise the systemic effects of assessment and indicators. One of the main challenges with designing bibliometrics for direct use in PRFS is to avoid negative systemic effects. Simulation, testing, and evaluation after implementation, will be needed. Problems with acceptance in the academic community can arise and need to be addressed.

Based on observations from the material submitted for the MLE, and on the plenary and group discussions of the MLE, below is a list of ten typical problems and considerations in the design of bibliometric indicators for direct use in a PRFS.

- Bibliometrics is not 'objective'. The idea that one indicator based on a particular data source might solve the problem or promote the best research should be regarded as subjective as long as it is not tested and discussed. As shown in chapter 3, the available data sources and indicators for bibliometrics may be problematic already at the outset if the aim is to give a balanced representation of performance in all areas of research and all types of research organisations. Main consideration: avoid power games in the design process
- More than one expert is needed to do a good design. Our advice is to design the bibliometrics for PRFS in dialogue between the funder and the organisations funded and to represent all areas of research in the process. Advice from bibliometric experts, independently of the commercial providers of bibliometric data, is also necessary. Main consideration: participation and transparency
- The strength of the incentives connected to the indicators will partly depend on the relative economic influence of the bibliometric indicators within the total PRFS, and the relative influence of the PRFS on total institutional revenues. The bibliometric indicators are most often part of a set of performance indicators in the PRFS, and the PRFS will not be the only source of funding or revenues. Main considerations: the perceived importance of the bibliometric indicators and their effects as incentives will partly depend on their economic importance, partly on other incentives in the research system, by which they can be strengthened
- Data sources. Some countries use only WoS or Scopus; others add other sources; and yet other countries construct national databases to cover the research output from the institutions more comprehensively. Main considerations behind the solutions are: data quality; disinterested data production; incentives for internationalisation; costs; comprehensiveness; balanced representation of all fields; the representation of national language publishing

- Definitions and delimitations. The types of publication included in the indicators must be defined, and the set of publication channels included must be delimited. A chosen data source, e.g. WoS, represents a definition and delimitation. The same is true for any combination of external data sources. National databases created for the purpose need an explicit definition, a set of reporting instructions and some monitoring of the reporting practices. The main considerations behind the solutions will be much the same as in point 2 above
- Indicators. Publication, collaboration and citation indicators are all used in PRFS designs. The main considerations are connected to the dimensions of performances that the different indicators represent, and whether they are available and valid across all fields
- Field normalisation. The well-established field normalisation methods for citation indicators need to be supplemented with a balanced representation of productivity across fields. This may depend on how the indicators are designed – see the remaining points below. It can also be solved with balances in the funding formula itself or by separating field-specific streams of funding. The main considerations are that institutions with different profiles of specialisation, e.g. a technical university versus a general university, need to be treated equally, and that the funding mechanism should be acceptable across fields
- Counting methods. How to count publications with multiple authors and affiliations is an often-overlooked problem in the design of bibliometric indicators for PRFS. The main considerations will be: to balance the indicators across subfields with different co-authorship practices, and to incentivise collaboration without stimulating the inclusion of authors with minimal contributions
- Weighting of publication types. If more than one type of publication (e.g. journal articles) is included in the indicators, such as peer reviewed conference papers, book chapters and monographs, these must be weighted against each other. The main considerations will be: to balance the indicators across subfields with different publication practices, and to incentivise a favourable development of those practices
- Ranking of publication channels. Publications of the same type may be given different weights depending on where they are published. This can be done by using journal impact factors, journal 'quartiles", the delimitation of certain respected data sources, or panel evaluation of publication channels. The main considerations will be: to incentivise internationalisation or publishing in certain important publication channels; to balance between research quality and research productivity; to provide legitimate incentives that do not discriminate

against national language publishing in the social sciences and humanities; to respect the DORA declaration[8] on the use of bibliometrics in evaluation

Even if one succeeds in taking all of these considerations into account, one should be aware that bibliometric indicators designed for the macro level (institutions) are often not adequate at the level of research groups or individuals. The application of bibliometrics at the micro level should instead follow the ten principles of the Leiden Manifesto given above.

4.2 How is assessment done using peer review and informed peer review?

4.2.1 Introduction

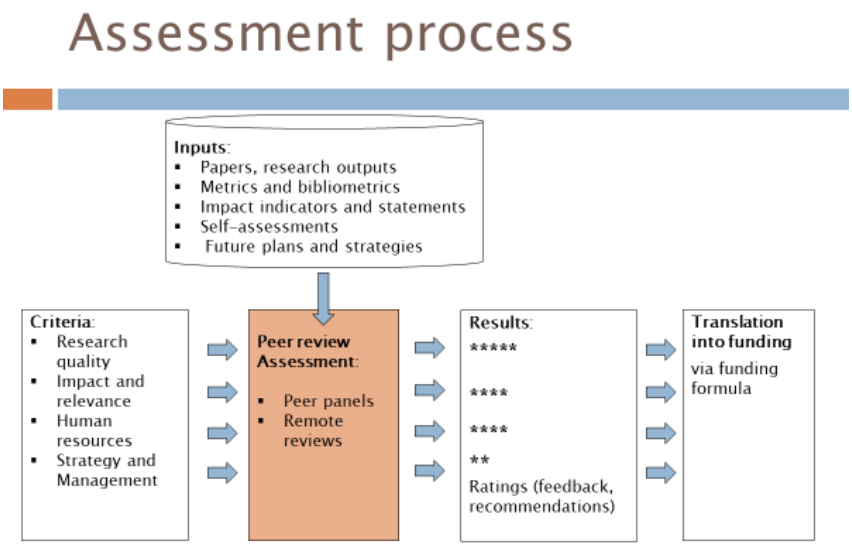
In the context of peer review, the process of assessment works as follows. Specific criteria and demands are used to turn the judgement of peers (and experts) into quantitative rankings and ratings. These may subsequently be used to drive funding via a specific funding formula. This transformation is done with the help of various inputs, mainly papers and scholarly research outputs. Inputs may be extended by

- Metrics and bibliometrics ('informed peer review')
- Indicators and statements on relevance and socio-economic impact ('extended peer review')
- Self-assessments ('two stage assessment')
- Strategies and future plans ('prospective and formative elements')

Figure 11 illustrates the assessment process and how it feeds into a funding formula.

[8] San Francisco Declaration on Research Assessment (ascb.org/dora)

Figure 11 Assessment process



4.2.2 Necessary elements and specific challenges

4.2.2.1 Design Options: Assessment model, scope and criteria

The various forms of peer review-based assessment reflect policy objectives and priorities. Nevertheless, the degree of complexity depends on the availability of appropriate data and resources, especially in the context of small countries. Table 11 lists the various design options; these are varyingly interlinked (e.g. if the institution's mission is one of the assessment criteria the assessment must include some formative elements).

Table 11 Design Options

Dimension	Options
Peer review model	'pure peer review', 'informed peer review', mixed model
Function	Summative only, in addition: formative and/or prospective elements
Assessment criteria	Research quality only, in addition: relevance/impact, sustainability/human resources, strategy/management, others
Granularity	Total HEI, institutional sub-unit, research groups, individual researcher

Peer review model

The UK, Portugal and New Zealand are the only countries relying nearly exclusively on peer review ("pure peer review"). Most countries combine peer review with metrics in an **"informed peer review"** model or in similar mixes. Some authors argue that bibliometric indicators can make peer review more transparent and offer additional insight in cases of where panel members´ opinions diverge (Mahieu & Arnold, 2015). 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 haphazardly googled h-index. Critical voices claim that erroneous and distorted peer reviews cannot be counterbalanced by adding bibliometrics because both are prone to include the same distortions (mainly conservatism). Furthermore, the '**Matthew Effect'** will be strengthened when quantitative and qualitative methods are combined (Fröhlich, 2016).

Function

A PRFS is by definition mainly summative and backwards-looking. The UK REF provides some space for the presentation of future plans is the assessment of the research 'environment' at the university and in the impact template where submissions can describe "their approach to impact during the assessment period as well as their development of a strategy and plans for the future" (Submission Guidance 2012). In the Scandinavian countries, the PRFS is complemented by national evaluations using informed peer review with a more formative character and including prospective aspects.

In systems relying on performance contracts as steering instruments, prospective elements are inherent. This is the case for example for Finland, The Netherlands, Austria, and some German Länder. Even if direct links between the PRFS and the performance contracts are rare, peer review-based research assessments often provide a basis for the formulation of the agreements (e.g. in Austria or in the Netherlands).

Assessment criteria

While early PRFS focused mainly on research quality in terms of publications and scholarly outputs, the current trend is to encompass also aspects of **relevance and impact beyond academia**. The UK REF 2014 introduced the criterion of 'impact' across all disciplines – mainly as a justification against competing demands for resources in times of austerity cuts. 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/)

Peer review is likely to ignore wider social and economic effects owing to its strong scientific orientation. Therefore, in its pure forms, peer review tends to be inappropriate for impact evaluation (Ruegg & Jordan, 2007). To solve this problem, peers can be informed by analysis of socioeconomic needs and priorities, which is mostly done by providing 'impact statements' as narratives (compare 3.4).

Further criteria often reflect different political priorities and change over time. Sustainability (in combination with the research environment or in combination with human resources development) is a common argument. Some countries define quite long lists of many different criteria, others concentrate on few (Jonkers & Zacharewicz, 2016).

Example: Peer review and metrics in UK´s REF

The REF was initially proposed as a metrics-based evaluation to replace the Research Assessment Exercise (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. (2015) 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 did not 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.* (Stern, 2016)

4.2.2.2 Selection of Peers and Panel architecture

After the decisions concerning scope, design and criteria are made, the **composition of various panels** as well as their **appropriate staffing** is the next major challenge when designing a peer review-based research assessment approach. There are a lot of trade-offs at stake, as shown in Table 12, below.

Table 12 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

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 are predominantly from outside the academic sector and represent a variety of private, public or charitable sectors that either use university-generated research, or commission or collaborate with university-based researchers (Derrick & Samuel, 2016).

Selection of peers and panellists should be done in the knowledge that the act of selection can determine outcomes, for example by in- or excluding particular schools of thought or research traditions (Langfeldt, 2004) (Lee & Harley, 1998). ANVUR uses a 'snowball' method to select peers, appointing panel chairs and then letting them do the recruitment. This has the advantage of separating the choice of peers from administration and funding but it does also expose the assessment to the risk of excluding relevant perspectives. Procedures similar to those normally used by research councils in appraising proposals need to be in place to manage conflicts of interest – essentially to make sure that peers do not have to assess their friends, relatives, colleagues or collaborators.

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.

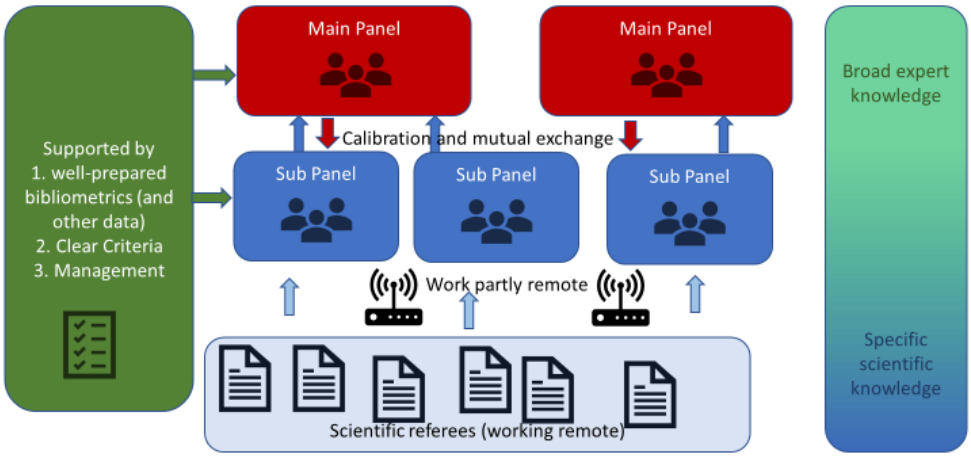
In contrast to this broad range of requirements, the need to control **costs and coordination tasks** implies that the number of persons, interest-groups, aspects and criteria involved should be limited.

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. Given a well-structured division of labour between the different panel stages, even very subject-specific aspects can be considered without losing track of the main assessment issues. Not only the UK, but many other countries such as Norway and the Czech Republic use sub-panels in research assessment and evaluation.

Another partial solution is to introduce **'remote'** panel evaluation and to avoid on-site visits (Mahieu & Arnold, 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 held. Such an approach has been adopted, for example, in the UK RAE/REF exercises and in disciplinary evaluations in Norway.

Figure 12 below shows a possible architecture for the panels.

Figure 12 Panel architecture



4.2.2.3 Social dynamics and management

Panel meetings 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 & Tindale, 2004) (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. **Tacit negotiations and compromises** can affect the decision and a few panel members can dominate the discussion (Langfeldt, 2004).

In a critical study of the evaluation of research grants application, the Swedish research Council argues that panel evaluation is not **gender neutral** and proposes some measures to address this (Ahlqvist, Andersson, & Söderqvist, 2014). Further problems occur if panel members are poorly prepared, are unsure about their role or are simply shy.

4.2.3 Peer review in the MLE participating countries

Few countries – both in the context of this MLE and in general – use peer review as an evaluation tool in their PRFS. The scope of this analysis 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 (e.g. the Standard Evaluation Protocol in the Netherlands). 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 section.

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 also have other important characteristics that are common to comprehensive research assessments.

- Research is assessed at **different levels** (mostly panels and remote reviews). This means that there has to be 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 where 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

4.2.3.1 Design Options: Assessment model, scope and criteria

Peer review model

The following peer review models are applied by the countries participating in the MLE (Table 13):

- Of the countries participating in the MLE: Slovenia, Italy and Turkey apply PRFS based on informed peer review. Portugal relies close-to-uniquely on peer review.
- 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 as funding processes, and 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 to a PRFS, and in fact, some authors eg 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 collective assessment

Table 13 Peer review in the MLE participating countries

Peer Review in PRFS as well as in systematic and comprehensive Research Assessment			
Integrated in the PRFS	Outside PRFS		
Pure peer review, informed peer review or mixed methods including peer review • Czech Republic • Italy • Portugal • Slovenia • Turkey	Subject-specific evaluations and other assessments: • Norway • Spain	Accreditation (eligible for public funding): • Croatia • Estonia • Moldova	At local (university´s) level only: • Austria • Sweden

Function

Formative and prospective elements are included in the peer review models of countries with research assessments additional to a PRFS (such as Croatia, Estonia, Moldova, Sweden and Norway). Especially in contexts where self-evaluations are part of the process and where formative evaluation elements are included, prospective elements become significant. Peer review based PRFS in the countries considered here sometimes consider prospective aspects (eg in the UK) but do not offer formative feedback. However, there are examples elsewhere (for example in Latvia and Lithuania) of peer review-based PRFS that do offer such feedback, which may be especially useful in countries with less developed research systems.

4.2.3.2 Selection of Peers and Panels' architecture

In line with the broad range of criteria in use, the required qualification profile for reviewers encompasses not only scientific and 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.

The number and architecture of the panels vary widely and range from one main panel in the case of Croatia, Moldova and Slovenia up to six in the Czech Republic. The UK REF is more complex than the systems of states participating in thins MLE. It contains four main panels and 36 sub-panels.

These important differences are partly explained by both Croatia's[9] 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. Portugal, Croatia and Estonia include on-site visits, most others do not.

Norway applies a two-step approach with sub-panels where necessary in its field evaluations, while the other countries generally follow a one-step approach, partly assisted by remote evaluations. Interestingly enough, the panel architecture of RCN evaluations is in the process of being reorganised into a multi-step approach. Physical meetings are supplemented by remote interventions and supported by SharePoint as a platform.

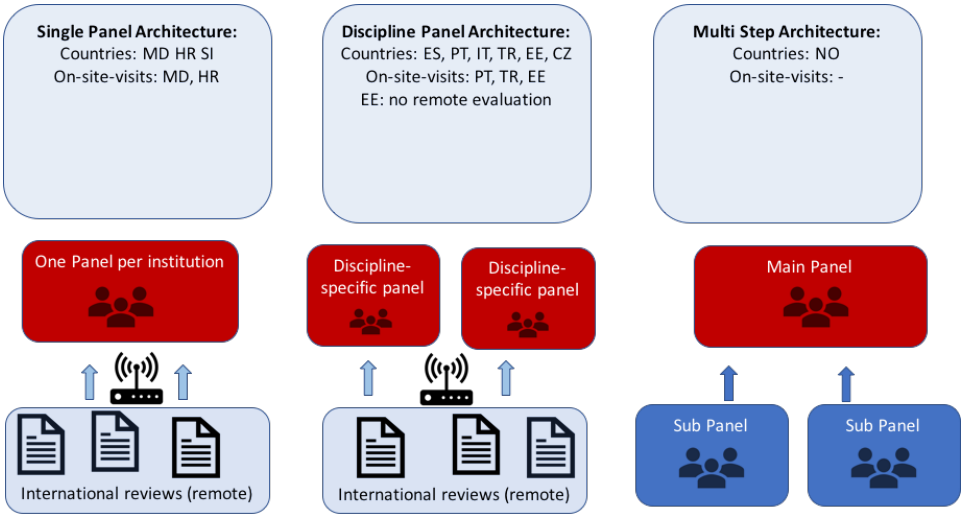
Abstracting from some country-specific particularities, three different panel models can be distinguished.

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

[9] If universities and research organisations work on similar research topics they are often evaluated together.

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

Figure 13 Different Panel solutions



4.2.4 Lessons and Conclusions

The experiences of the MLE participants differ widely due to differences in size, research orientation and national evaluation culture. 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 the research in question
- 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 non-scientific peers have been 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 used
- A well-balanced **mix of different measures** (e.g. 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 their 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 should not be ignored

Some specific lessons can be drawn from the participating states' experience.

- Countries with informed peer review-based PRFS (Italy, Slovenia and Turkey) struggle with the data base for the bibliometrics involved. They should think about how they can improve and simplify their models by implementing a national CRIS, as for example Norway has done. **Portugal** relies on some numeric input into the peer review process but is not including bibliometrics. Based on the experience of the other countries in the MLE, this could be a future option.
- **Austria and Sweden** decided not to implement peer review-based PRFS for similar reasons: the universities worried about their institutional autonomy and wanted to organise research evaluations themselves. As a result, most universities run internal assessments but comparisons between various research groups in the country are not made. One task of the central government in such a system could be the organisation of mutual learning among the universities as well as the formulation of certain minimum standards. Further, Austria could take Sweden as an example where nationally-created and maintained data and bibliometrics can be used by decentralised evaluation exercises
- All the small countries with some peer review models in place (**Croatia, Estonia, Moldova, and Slovenia**) struggle with the implementation of a proper peer review-based PRFS at reasonable costs[10]. The significant amount of fixed costs may discourage the use of a peer review-based assessment system. High costs of peer review can be mitigated by introducing a proper panel system with elements of remote review as well as by international cooperation both in the design and in the application phase

[10] The current Czech system involves peer review of about 10% of scholastic output, so cost is not a significant problem

4.3 How is societal impact assessed?

Assessing the societal impact of a state intervention such as funding research is beset with difficulties. Terminologies vary among countries and evaluators. However, many evaluation approaches rely on the idea of mapping out a chain of causation leading from the policy intervention to impact in society.

- A policy intervention, such as research funding provides **inputs**, often in the form of money. These enable people and organisations to undertake **activities** such as research
- **Outputs** are produced as direct results of the activities. These might for example be new knowledge, new methods, scientific publications, patents, PhD graduations. They are normally quite easy to count and to attribute to the intervention. Most PRFS focus on outputs
- The outputs may trigger **outcomes**, such as improved career prospects for people involved in the research or the use of the new knowledge in developing products or policies. The further along the causal chain one looks, the more difficult these become to track and the harder it is to attribute them to the original research. Often that research is necessary but not sufficient for the outcomes to arise – other factors (even other research) comes into play and there is no logically coherent way to attribute the credit for the outcomes among the several factors that made them possible
- Outcomes in turn may trigger **impacts** at the societal level: things like increased employment, sustainability, health, growth or other social desiderata. Few evaluators make a strict distinction between outcomes and impacts. However, since policy interventions are normally intended to benefit society rather than just their direct beneficiaries, it can be convenient to think of this distinction as marking the point at which the private benefits of the intervention trigger societal benefits. Attributing impacts to an individual intervention such as a research project can be extremely challenging

The attribution problem is further complicated by the fact that at all stages beyond the production of outputs, chance becomes an important factor. If PRFS are to address outcomes and impacts, this raises the question of whether to try to assess the outcomes and impacts themselves or to give the researchers credit for doing activities such as knowledge exchange and dissemination or engaging in 'productive interactions' with potential users of the knowledge irrespective of whether they are lucky enough to be successful. A further quandary involves defining which research 'counts' in generating impact. The UK REF, for example, only considers impacts of research that has previously been given a high rating in the UK assessment exercise. Poorly-rated research does not count, even if its societal impacts turn out to be much bigger.

Some PRFS have used very simple output metrics to indicate the potential to create outcomes and impacts. However, these are easily gamed (Good, Vermeulen, Tiefenthaler, & Arnold, 2015) and few now use them. Sophisticated combinations of indicators and qualitative accounts appear promising for analysing societal impacts of research more widely but are massively labour-intensive compared with what is possible in a PRFS. The UK and (in field evaluations) Norway have adopted a practice of peer reviewing qualitative impact narratives, which is said to work reasonably well but is by no means transparent.

4.3.1 The difficulties of assessing societal impact

How societal impact is assessed depends on many factors inside and outside academia. Regarding internal factors, disciplinary and institutional differences need to be considered since some fields or institutes are more oriented towards society than others. While most universities now routinely include societal impact in their mission statements ((van den Akker & Spaapen, 2017), it does not mean that all these institutions also routinely assess the societal impact of their research. The main reason for this is that it is much more complex to develop reliable data systems and robust indicators for societal impact than is the case when scientific quality or impact has to be assessed. And, a main reason for that is that the societal impact of research can take place in so many different societal contexts, and therewith is part of a multifaceted research and innovation process in which various non-academic partners also contribute to the goals of the project at hand.

Gibbons, (1994) and Nowotny and colleagues (2001) who made a distinction between traditional academic research – mode 1 - and research in the context of application - mode 2 -, noticed this and referred to the first kind of knowledge production as discipline-oriented and set in a homogenous context, while the second kind of knowledge production takes place in a heterogeneous and socially distributed context, for the better part outside academia. They see it as inevitable that novel quality control mechanisms need to be developed in collaboration with the latter largely external context. They refer to the kind of knowledge developed in and with the wider context as socially robust knowledge: knowledge that links up with diverse practices, institutions and actors, and it addresses audiences that are never solely composed of fellow experts, whose expectations and modes of understanding reflect the heterogeneous experience of mixed audiences (Nowotny, 2003) (Kuhlmann & Rip, 2014) go even further when discussing the EU Grand Societal Challenges. Unlike society-orientated programmes in the past that targeted a particular field or goal, the large society-oriented programmes like H2020 are about systems transformations, such as more sustainable ways to produce goods and services. Such programmes are much more open ended and goals and

targets may evolve over time. Ultimately, they involve changes to social, economic, and technical systems.

4.3.2 How to assess societal impact in such a complex context

What does this mean for the assessment of societal impact? Given the context dependence of societal impact measurement, and the fact that the societal networks in which researchers operate are often very dynamic, societal impact measurement will be difficult at the individual level. If we look at the perspective of an individual researcher who is working in a large collaborative project with lots of other stakeholders, the focus of an evaluation is likely to be on the collective goals of the project and on the ways these goals are to be reached. For the individual researcher, it is possible to look at his or her contributions to the larger project, which might be in writing or forms of collaboration, or in knowledge diffusion or engagement of the public. Clearly, these are varied activities that lend themselves more to qualitative (case studies, impact pathways) than quantitative approaches. Think of organising meetings with stakeholders, conducting training workshops, providing expert advice to policy or other stakeholders, establishing a start-up, outreach activities etc. Especially when looking at the research project level, and the ways in which these activities fit into the larger endeavour, societal impact evaluation is best be directed towards short- or medium-term goals ('intermediate goals' or 'outcomes'), in particular since goals in the larger network may change over time. In the Thematic Paper on third-stream metrics produced in the context of this study, several examples are given of of how this can be done, for instance the SIAMPI model of three types of productive interactions.

Table 14 Examples of indicators for societal impact in three categories

Direct interactions (people)	Indirect interactions (media)	Material interactions
• face-to-face meetings • double functions, other mobility arrangements • phone/skype conferences • email • social media • videoconferencing	• academic journals • professional journals • non-academic journals • popular media • exhibitions • artefacts, models • films • master theses,	• research contracts • facility, instruments sharing • start ups • contribution "in kind" (people) • IPR arrangements, patents, licenses • Professional training • Other stakeholder

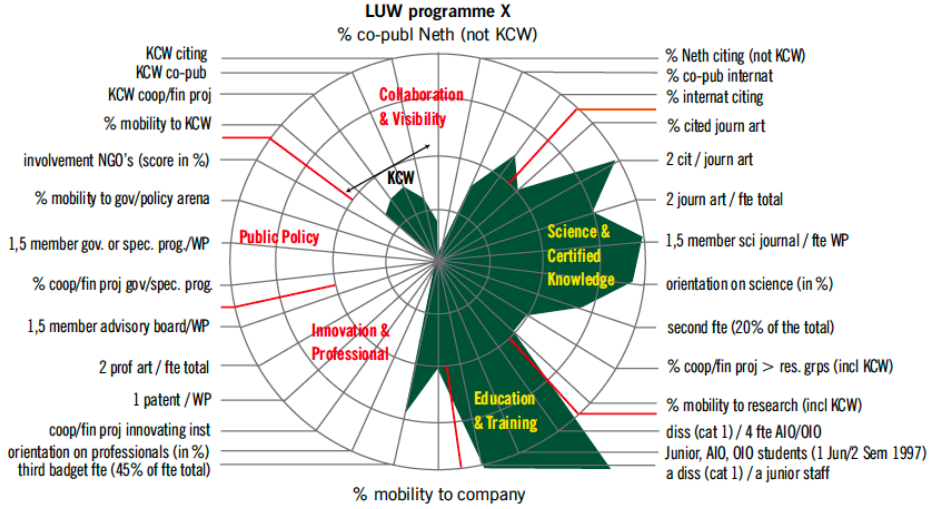
Direct interactions (people)	Indirect interactions (media)	Material interactions
• public debate • radio, tv, internet	graduate projects • standards, protocols • social media	interest

Researchers aiming at societal impact for their work will from time to time engage in several of the activities listed above, sometimes writing an article for a scientific journal, at other times collaborating with other individuals inside or outside academia. As in the Canadian example given above, it is possible to develop metrics for each of these indicators, some quantitative, others qualitative, but underpinning them with robust data remains difficult.

Still, while the development of robust third-stream metrics for societal impact measurement is still in its infancy, some real progress has been made in this century – see for example (Bornmann, 2013). As mentioned above, we can distinguish at least three strands of impact measurement that show promising developments:

1. Emulating what is done when scientific impact is measured, mostly using quantitative indicators. Initially, this strand focused on economic indicators such as cost-benefit analysis, but now we see a broadening of the spectrum to non-economic indicators that provide information about social, cultural, environmental, health, etc. effects (Joly & Matt, 2017). There is a wide variation of efforts that range from single metrics to comprehensive research project assessments. An example of the latter can be found in Spaapen et al (2007) who developed for the Agricultural university in Wageningen a so-called *Research Embedment and Performance Profile* (REPP), in which a variety of indicators are depicted in a single graphic representation along five dimensions: (a) science and certified knowledge, (b) education and training, (c) innovation and professionals, (d) public policy and societal issues, and (e) collaboration and visibility. For each of these dimensions five quantitative measurements were developed that together gave a complete picture of the work done by the research groups. Figure 1 shows an example of a REPP.

Figure 14 Research embedment and performance profile



Source: Spaapen et al (2007)

This kind of profiles were used by the university to present their research groups to international committees that evaluated their work using the SEP protocol that values scientific and societal aspects of research in a balanced way. While it was well received by both the university and the panel members, the REPP was not used further because there were problems with the quality of the underlying data and the calibration of the different indicators within a single model.

There are many examples of single indicator metrics – economic ones like return-on-investment, bibliometric ones like popular publications and specific ones, depending on the sector. There is a clear risk in using a single indicator as third-stream metric. For example, to measure the effects of research on the quality of life, a large variety of indicators is available (via Eurostat) related to aspects like health, education, living conditions, natural environment, etc. It only makes sense to combine different indicators in an assessment because of the potential trade-offs among the different aspects.

2. Developing measurements that better fit the societal context, often more qualitative, like case studies or impact pathways. In sectors like health research, agriculture, development studies, there is quite some experience with the pathway approach (Webb, 2013). Another part of this strand are variations of network analysis, like the SIAMPI approach of productive interactions (Spaapen & van Drooge, 2011). One of the most developed examples of the impact pathways

approach is arguably the REF UK. The Research Excellence Framework introduced impact case studies in which research units could demonstrate their impact on the economy and society at large. The studies are structured in five chapters.

- Summary of research
- Underpinning of research
- References to research
- Details of impact
- Sources to corroborate impact

These case studies counted for 20% in the assessment (as opposed to 65% for scientific research output and 15% for the research environment). The impact case studies are short (max. 4 pages). They are assessed by two criteria: *Reach* – "the spread or breadth of influence or effect on the relevant constituencies"; and *Significance* – "the intensity or the influence or effect". There is a complete database on the REF website where all these case studies can be seen. A review study showed that much of the research was multi- and interdisciplinary, that the societal effects were diverse and wide-ranging, with over 60 unique 'impact topics' identified, and more than 3,700 unique pathways leading from research to impact (King's College and Digital Sciences, 2015).

3. Altmetrics, a relatively young development of web based metrics to assess the impact of scientific work in society, and sometimes also in scientific contexts (Bornmann, 2014). The gist of altmetrics is that there are all kind of traces in the digital world that may inform researchers about the uptake of their research. With altmetrics it is possible to track mainstream media, public policy documents, blogs, social media, online reference managers (Mendelay f.e.), citations, etc. Through the combination of these methods it is possible to get information on both the uptake in the scientific world and in society at large. (Holmberg, 2017); see also https://www.altmetric.com/).

   While still a young development, some believe that altmetrics will quickly become more important as a tool for research assessment, because more and more research is quoted in popular social media such as Twitter and Facebook. Also, major newspapers like the Guardian have become very active on Twitter, communicating about scientific research. More telling perhaps is the fact that that the big publisher Elsevier, which produces the SCOPUS bibliometric database, has established a partnership with the start-up company Altmetric, and professionalised the interface (https://www.altmetric.com/) . Bornmann points out that altmetrics fits an existing trend where platforms follow the output of research in the social web and develop

alternative metrics. He refers to Priem & Hemminger (2010) who mention bookmarking, reference managers, recommendation services, comments on articles, microblogging, Wikipedia, and blogging as platforms that provide insights into the research process.

While with the help of altmetrics potentially more aspects of scientific output can be measured (and more than just publications), there also drawbacks. Bornmann mentions the following: commercialisation, data quality (not always clear what is being measured), manipulation (easier than with traditional bibliometrics), missing evidence (not many reliable studies on altmetrics).

4.3.3 Experiences of countries with regard to analysis and rewarding impact

While most participating countries have set up policies or instruments to stimulate connections between science and industry or broader society to achieve goals deemed important for the national economy and policy, there is little experience with the analysis of the success of these measures. Some countries are setting up experiments but there are no cases where comprehensive third-stream metrics are integrated in PRFS, and neither are there institutional evaluation methods with this purpose.

What we can see is a number of experimental approaches to evaluating societal impact. Sometimes these experiments are wide ranging, sometimes they are very limited and specific. Turkey has developed the Entrepreneurial and Innovative University Index as a dedicated performance evaluation and monitoring system for tracing the activities of universities orientated towards the private sector. In Italy, ANVUR is developing a new system of metrics designed to support evaluation of third stream activities and impact of academic research, though this is not sophisticated enough to be integrated into the funding formula. Estonia is starting to use a baseline funding formula where, after publications and patents, contract research and work on topics of national importance is rewarded. Norway asked OSIRIS (Oslo institute for research on the impact of science) to study and develop indicators for societal impact. Impact outside academia is also included as a criterion in the field evaluations carried out by the Research Council Norway. Sweden has national statistics that cover different aspects of impact but these are not brought together to create an aggregate picture. A non-metrics model has been developed and piloted by the innovation agency Vinnova. They have the task to create a model for evaluation of third stream activities (Wise, et al., 2016).

In Croatia, the individual performance of each researcher is analysed using the following criteria: scientific productivity; national and

international competitive research projects and research mobility; collaboration between research and business sector as well as collaboration with the units of local and regional governance and non-governmental sector; activities of the popularisation of science.

In summary, countries do not have third stream metrics fully integrated in PRFS, but some have partial quantitative methods, and others are developing and experimenting.

Table 15 Third stream metrics in use or emerging

Third stream metrics in countries participating				
Part of PRFS, institutional or national level		Emerging demand outside PRFS		
Quantitative measurements:	Qualitative methods	National demand:	Institutional level:	Disciplinary level:
Croatia, Estonia, Turkey Italy (experimental)		Sweden Norway/Osiris Czech Republic	Austria/universities	Norway (social sciences)

4.3.4 Conclusions

The answer to the question how societal impact is assessed is the same as the answer to the question how to handle a porcupine: with the utmost care. Neither the methodological development nor the underlying data is yet very sophisticated. This certainly goes for the quantitative approaches. Regarding qualitative methods, these seem to be a safer bet. Case studies, impact pathways, narratives, as long as they are developed in a systematic way, often present insightful knowledge about what the impact of a research project is. If quantitative methods are used, they should only be used in a broader context of evaluation where peers and other experts play the leading role. The assessment of societal impact is best done above the individual level. The contributions of individual researchers can be charted, but at this moment in time, they should only be used to understand the interactions between researchers and context.

What would help the assessment of societal impact is the development of reliable databases that include both quantitative and qualitative information (like the REF impact case studies). That would not only be helpful for evaluation purposes, it would also be an excellent way to inform society.

4.4 Lessons learnt

- While early PRFS tended to be wholly based either on peer review or metrics, a growing number rely on a combination of these approaches to produce higher-quality judgements at lower cost
- Innovation is also possible in the scope of PRFS, stretching across institutional boundaries

Bibliometrics

- The Leiden Manifesto provides a good set of principles and should be consulted before doing PRFS design, especially when metrics are used
- Important considerations about bibliometrics
  - Bibliometrics is not 'objective' but contains many in-built assumptions and limitations
  - Stakeholder consultation and advice from disinterested bibliometricians are preconditions for using bibliometrics in PRFS
  - The strength of the incentives provided via bibliometrics depends on their economic importance as well as interplay with other incentives
  - Whatever bibliometrics data sources are used, the main considerations are data quality, disinterested data production, incentives for internationalisation, costs, comprehensiveness, balanced representation of all fields and the representation of national language publishing
  - The principles used for including publications in databases must be understood and considered in the light of the PRFS' objectives
  - Publication, collaboration and citation are typically analysed using bibliometric indicators. Care must be taken to ensure consistency and comparability across fields
  - Traditional field normalisation must be balanced by a proper representation of productivity across fields
  - Counting methods for multi-author and multi-institution papers affect collaboration behaviour. This must be taken into account in PRFS design
  - Different types of publications (eg books versus journal articles) need to be differently weighted, in order to take account of differences in publication behaviour among epistemic communities
  - Rating publication channels as opposed to individual research outputs can be misleading not only because a JIF does not

predict the quality of an individual article but also because of differences in practice among fields. Respect the DORA declaration

Peer review

- Peer review is indispensable when the assessment of quality requires a detailed understanding of research in question
- New forms of peer review considering multiple criteria and relying at least partly on non-scientific peers have been innovated to help extend peer review beyond scientific excellence, to address wider socio-economic criteria like relevance, impact, and new forms of interaction with society
- 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 used
- A well-balanced mix of different measures (e.g. 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 their use by peers or panels are important in this context
- Using a cost-effective mix of panel and remote reviews demands some aggregation mechanism. In order to ensure consistency some kind of calibration between different disciplines, interest groups and different panels is needed, and social dynamics should not be ignored
- 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. When selecting peer panels, it must be understood that the choice of people affects the outcomes of the exercise
- Panels are social entities, in which individual behaviour (shirking, bullying, leadership) affects outcomes
- It is impractical to run a PRFS with a single panel. Sub-panels are needed
- Peer review is inherently labour-intensive and therefore expensive. Separating assessment from panels work and using remote assessment can reduce cost
- International peers should always be preferred, especially in small countries
- Extending peer review beyond scientific quality calls for the involvement of non-scientists. Scientific peers are not necessarily good at judging societal impact, for example

- A CRIS can be a useful support for generating data to inform peer review. The Norwegian CRIStin offers a useful model
- Mutual learning among universities can be considered as an alternative to peer review as a source of formative assessment – though it is naturally less binding
- Potentially high costs of peer review can be mitigated by introducing a proper panel system with elements of remote review as well as by international cooperation both in the design and in the application phase

Societal impact

- As regards societal impact, the diversity of impact pathways and mechanisms combined with a lack of a standard way to describe or measure impact makes implementation of impact assessment in PRFS difficult
- Appropriate metrics for societal impact can be developed but are context-specific, making comparisons hard. Techniques include extending cost/benefit analysis to take in additional parameters and developing context-specific indicators
- Societal impact assessment in PRFS requires even greater caution than quality assessment and should not be taken lightly by PEFS designers

5 USING THE RESULTS OF THE ASSESSMENT

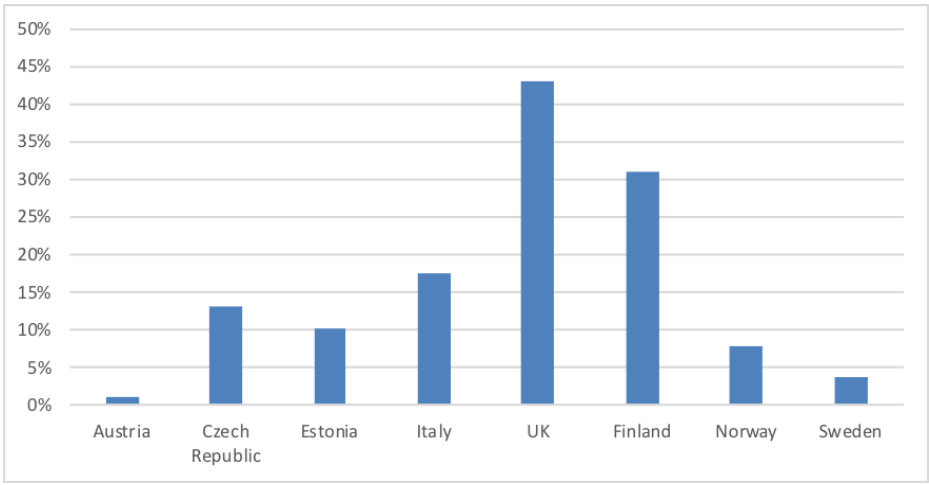
The principal use of PRFS assessments is to determine all or part of the institutional funding for research to be awarded to individual universities. However, they can also produce strategic intelligence of use to policymakers and the universities themselves.

5.1 The funding formula

A funding formula connects the results of research assessment to the amount of institutional funding to be awarded. But such formulae operate within the broader policy mix. The proportion of research income provided through the formula has to be balanced against the incentives and disincentives created by the formula itself.

There are marked differences in the proportion of different countries' university systems' institutional income[11] that is governed by PRFS systems, as Figure 15 shows, and therefore in the degree to which assessment is a dominant tool in managing university performance.

Figure 15 PRFS as a percentage of institutional income, 2016 (or 2015 where not available)



Source: Data reported by participating states. UK and Finland from official statistics

[11] It would be more interesting to be able to do this analysis based on the proportion of PRFS in universities institutional funding for research. However, some of the countries covered here do not distinguish between the teaching and research elements of their block grant, so we are obliged to consider total institutional funding instead.

The share of PRFS in institutional funding is a fairly marginal factor in Austria, Norway and Sweden. In the UK and Finland the PRFS governs all institutional income for research. This is also true in Estonia, which has only recently introduced institutional funding for research and where the PRFS is therefore very important despite the fairly low value shown in the Figure. Italy occupies an intermediate position. While few PRFS have been evaluated, there is evidence from Norway that they can be effective in changing behaviour even when they govern only a small proportion of research funding (Bloch & Schneider, 2016).

Table 16 shows how the PRFS funding relates to individual assessment criteria. As we will illustrate later in this report, PRFS affect – and are intended to affect – universities' and researchers' behaviour. Funding formulae tend to have multiple components, so they simultaneously encourage different types of behaviour, as should be evident for example from Table 16. Appendix B provides a more detailed account of the way individual funding formulae operate.

Table 16 Importance and structure of PRFS

Austria	
PRFS as % of institutional funding	*1.2%*
• Grants from Austrian Science Fund (FWF) and EU	• 78.9%
• Number of PhD students	• 21.1%
Croatia	
PRFS as % of institutional funding	• NA
• Scientific productivity	• 60.0%
• External research funding obtained	• 25.0%
• Research-industry collaboration	• 10.0%
• Popularisation of science	• 5.0%
Czech Republic	
PRFS as % of institutional funding	*26.2%*
• Academic outputs	• 85.0%
• Non-academic outputs	• 15.0%
Sweden	
PRFS as % of institutional funding	*3.9%*
• Bibliometrics-driven funding	• 50.0%
• Funding based on external competitive grant income	• 50.0%

Estonia	
PRFS as % of institutional funding	*10.2%*
• Publications and patents	• 38.0%
• Contractual and foreign income	• 47.5%
• No of PhDs graduated	• 9.5%
Support to topics of national importance	(22)
Italy	
PRFS as % of institutional funding	*17.0%*
VQR	*76.5%*
• Output quality (all products)	• 65.0%
• Share of external research funding obtained	• 5.7%
• Share of PhDs, med. students, research fellows post-docs	• 5.7%
• Output quality recent hires	• 23.5%
The UK	
PRFS as % of institutional funding	• 43.0%
• Mainstream QR	• 67.8%
• Research Degree Programme Charity Support Fund	• 15.2%
• Business Support Fund	• 12.5%
• National Research Libraries	• 4.1%
	• 0.4%
Norway	
PRFS as % of institutional funding	*8%*
• Number of PhD candidates	• 22%
• Publication points	• 26%
• EU income	• 24%
• RCN and Regional Research Fund income	• 14%
• Contract research income	• 14%

Table 17 focuses on the share of PRFS in institutional funding and then the percentage of the PRFS resources devoted respectively to publications and external funding. It is striking that the systems that govern only small parts of institutional funding (Austria, Norway, Sweden) place heavy reliance on external funding indicators. These systems are simple, in keeping with the fact that they govern only small amounts of money. Most likely, these comparatively unrefined systems would not be tolerated by the academic community if they were to be expanded to cover a much bigger fraction of institutional funding.

In contrast, systems that govern large parts of institutional funding tend to be more elaborate. Croatia, the Czech Republic, Italy and the UK all have complex and sophisticated ways to assess quality (whether using metrics, peer review or both). Perhaps for this reason their designers feel they have less reason to lean on the proxy of external funding. Estonia and Finland occupy a different position, having made the choice to use a comparatively simple system for assessing publications while at the same time placing significant funding emphasis on publications as well as on external income.

Table 17 Importance of PRFS, focus on publications versus external funding and cost drivers in funding formulae

	PRFS as % of institutional funding	% for publications	% for external funding*	Use field normalised costs	Volume of publications drives funding
AT	1%		79%		X
HV	NA	60%	35%	X	X
CZ	13%	85%			X
EE	10%	38%	48%		X
IT	18%	89%	6%		X
UK	43%	83%	17%	X	
FI	31%	42%	29%		X
NO	8%	25%	75%		X
SE	4%	50%	50%	X	X

*Includes both state and other external research funding

Most of the funding formulae encourage an increased volume of production. Only the UK focuses on quality rather than trying to increase the number of publications produced.

The PRFS funding formulae have in common that their task is periodically to distribute a fixed budget among institutions. Mostly they do this in a more or less linear way, so that money may be distributed pro rata publication points, external income etc. The REF component of the UK system and the Finnish publications indicator work in a non-linear way, focusing resources on those who get high scores. This reflects a design choice in parts of these systems to focus resources on excellence (and potentially on excellent institutions) rather than to lift the performance of the university system as a whole. In the case of Italy, where concentration is felt to be undesirable, the design choice tends to produce limited differences: in 2017 only 11% of the overall PRFS funding was allocated in a way that deviates from a distribution based on the sizes of the universities.

Another notable difference among formulae is that some (Croatia, the UK, Sweden) explicitly take account of the different cost levels associated with different disciplines (Table 17), in an attempt to compensate for the fact that some fields (typically the 'hard' sciences) are inherently more expensive than others (such as social sciences and the humanities).

Where the PRFS provides only a small proportion of institutional income there is likely to be more organisational slack available so that the universities can handle field differences in costs by themselves. the UK and Sweden use field-normalised assessment methods that focus on quality and not productivity, so it can be argued that taking explicit account of costs is a sensible approach. The other countries use formulae which are partly driven by the volume of scientific production, tending to favour disciplines with a high propensity to publish over others. To some extent, the expensive disciplines are those that publish most and this will produce some implicit compensation for their higher cost levels. The 'publication points' tradition of Norway, Sweden and Finland awards high marks to books and monographs, in part to compensate for their lack of visibility in scientific journals. The approach is somewhat rough and ready, however. Table 18 aims to capture the spirit of the different formulae. Most aim to increase both the volume and the quality of scientific output. Croatia encourages institutions to improve their performance on both dimensions relative to the Croatian average – which may in practice have the same effect since it should raise the average level. the UK and Sweden use formulae for funding output that do not have components that encourage volume, though Sweden backs this up with an external funding indicator that is strongly volume driven. The UK formula encourages universities to hire more excellent researchers, rather than encouraging existing researchers to produce more. Finland and Italy encourage both volume and quality, but Finland uses a non-linear scale in relation to quality that skews funding towards excellence while the way the Italian system sets the boundaries that use the position of articles in international ranking to generate different numbers of publication points tends to focus the highest rewards on excellence in a non-linear way.

Table 18 Main thrust of funding formulae

Ctry	Main thrust of PRFS funding formula
AT	Encourages PhD training and acquisition of external funds, especially for basic research
HV	Rewards performance relative to average performance in Croatia
CZ	Encourages volume and quality of output
UK	Focuses funds on excellence,. Increased volume is not an evident objective
EE	Encourages volume and quality of output
FI	Encourages volume and quality of output, with a focus on excellence
IT	Encourages volume and quality of output
NO	Encourages volume and quality of output
SE	Focuses on quality, but also with a volume incentive via external funding

5.2 Other uses of the assessment

Depending upon the details of their design, there are opportunities for the kind of research assessment associated with PRFS to provide significant amounts of strategic intelligence, both to government and to universities. Unfortunately, there is no systematic analysis of this aspect available in the literature. Nonetheless, it is possible to make some observations, based on experience.

Purely bibliometric assessments probably provide the least added value compared with the information already available to university managers and policymakers, most of whom have access to the commercial bibliometric databases and buy the tools or capacity to benchmark themselves in various fields. Other, free tools are available, such as the Leiden ranking, which allows the user to rank some 900 universities on a range of criteria, and there is a plentiful supply of world university rankings (such as the Shanghai ranking and that produced by the THE (Times Higher Education) magazine. Universities use information from various rankings in their marketing and may also publicise good scores in national research assessments. There is little evidence, however, that PRFS-derived scores play much of a role in students' choice of institution.

These tools can be less relevant in smaller systems, especially in countries which do not speak a 'world language' and where a significant part of the national research output is hard for bibliometric methods to address. Countries such as Norway, Finland and the Czech Republic that maintain a quality-assured national CRIS are in a better position here.

It is not clear that summative[12] peer-review systems have much value either in providing forms of strategic intelligence. Their results correlate closely enough with those of bibliometric tools that they bring little new information and they are inherently not reproducible, so little can be done with the information they provide. They also run infrequently. As a number of UK university managers commented to Technopolis during a review of the REF (Arnold, et al., forthcoming 2017), it would a pretty incompetent university management that was willing to wait five or six years for a source of strategic intelligence. This is precisely why universities invest in bibliometrics facilities. There is interest in both the UK and Italian universities in understanding how institutions' work is rated but this interest is driven by a desire to understand how to maximise the returns from the PRFS rather than to understand quality in

[12] In evaluation, a distinction is made between summative and formative exercises. A summative exercise simply judges. A formative one is designed to generate insight into why those who are evaluated perform well or badly and to provide feedback to them about how to improve

a more traditional sense. The universities may also be interested in feedback about performance at the level of the individual researcher. However, except in the very special case of the Spanish Sexenio, such information is not normally provided.

More formative peer review exercises have greater value as strategic intelligence. The key question is the extent to which the exercise inquires into the circumstances in which research is done, such as the facilities, strategy, human resource polies and so on of the institutions and groups assessed. The REF required universities to provide a standard set of information about the research environment. However, each submission is simply awarded a score: there is no feedback. The environment statements provide a dataset that can be used for analysis – for example, of equality and diversity (Mellors-Bourne, Metcalfe, & Gill, 2017) – but it appears to be little used for policy development. In contrast, the recent research assessment exercise in Latvia (Arnold, et al., 2014) – which explicitly considered the production process as well as the outputs of research – has enabled the government to start a radical programme of mergers and closures aimed at eliminating weakly performing groupings and concentrating the national research effort into fewer, better run and more effective ones while providing explicit advice to the groups and institutions concerned. The next such exercise seems likely to be connected directly to a funding formula. As the research system develops, the formative element is likely to become less necessary.

An experience from the REF and more recently from an evaluation of Norwegian social science research institutes is that the institutions re-use their impact statements in marketing. HEFCE has made the REF impact statements available on the Web and triggered a small industry in academia and the consulting world to do analysis of them.

The states participating in this MLE make little use of PRFS information beyond the allocation of funding. In Austria, Croatia and Norway the relevant ministry uses it as strategic intelligence for wider policymaking and funding. In Italy, the ANVUR agency uses the information in the process of accrediting PhD courses. In Norway, it is published as part of an annual overview of the state of national research intended in part to provide 'soft steering' of the universities by publicising their performance levels.

5.3 Lessons learnt

- The proportion of universities' research income driven by PRFS varies considerably among countries. In combination, the PRFS and the other parts of the policy mix need to address policy goals in a complementary manner

- The funding formula is the mechanism that translates research assessment results into money. It too must be consistent with the policy objectives
- Different formulae provide incentives for different kinds of behaviour. The use of skewed formulae – where the best performers are rewarded disproportionately – is a way to concentrate resources on 'excellence'
- The greater the proportion of universities' research income governed by a PRFS, the more robust its methods need to be in order to withstand scrutiny by the beneficiaries
- Overhead and infrastructure costs vary among fields. Consideration should be given to weighting the formula in order to take this into account. This appears to be especially important where the PRFS governs a large proportion of institutional income, leaving universities little 'slack' in using the block grand to tackle variations in costs among disciplines
- The opportunities to use PRFS results as a source of strategic intelligence for policy are often under-exploited and are not necessarily considered in PRFS design. Explicit consideration should be given to whether and how to exploit information generated by the PRFS in research policy and management more widely
- While PRFS results are widely claimed to provide the universities with strategic intelligence, they are at best an incomplete source of it. Universities should be monitoring their own quality on frequent basis in any case
- Peer review based PRFS can be designed so as to provide formative information to universities, but can be done only at long intervals. This is especially useful in emerging or developing research systems but becomes less so as systems mature

6 WHAT ARE THE EFFECTS OF PRFS?

There is little evaluation literature relating to PRFS. What there is suggests positive effects on the amount and quality of research output. There is not yet evidence about the dimensions of societal relevance and impact but it appears reasonable to expect these to appear and such criteria become increasingly embedded in PRFS. However, these benefits may not be costless. PRFS design therefore needs explicitly to consider potential effects of PRFS on the overall system of funding and governance, on system performance, the nature of the research done and the implications of the PRFS for research career.

6.1 … on system performance

Internationally, studies tend to associate PRFS with increased productivity and quality (measured as citations) (Moed, 2005) (Butler, 2003) (Jiminez-Contreras, Anegon, & Lopez-Cozar, 2003) (Sivertsen, 2016) (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) (Schneider, Aagaard, & Bloch, 2016) (Jiminez-Contreras, Anegon, & Lopez-Cozar, 2003) (Sivertsen, A performance indicator based on complete data for the scientific publication output at research institutions, 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) (Schneider, Aagaard, & Bloch, 2016).

The former Australian PRFS incentivised increased production of research outputs without at the same time offering an incentive to maintain or increase quality. The result was an increased volume of publications but a reduction in average citation impact (Butler, 2003). The current Norwegian system also encourages increased publication, but it is based on entries in CRIStin, the national CRIS, which classifies journals into two quality 'tiers', with publications in the upper tier triggering more income from the PRFS than those in the lower tier. It appears that this combination of a volume and a quality incentive has been effective, in that the volume of outputs has increased while the overall citation impact remains about the same (Schneider, Aagaard, & Bloch, 2016).

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 Research Selectivity Exercise – the early

form of the REF – (consistent with Martin and Whitley's (2010) assertion that the universities did not take the first exercise in 1986 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 (not a PRFS) 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. It may also be the case that policy changes – including but not only the introduction of PRFS – may induce a kind of Hawthorne Effect, where behaviour and performance change because these things receive greater attention.

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 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 probable therefore that PRFS contribute to improved performance 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 and had few negative effects (PREST, 2000). It was not able to measure effects in quantitative terms. The citation impact of UK research has increased continually since the end of the 1980s, after the second research selectivity exercise (Adams & Gurney, 2014). It is therefore highly likely the effect of the UK PRFS in the 1980s was to put research quality onto an upward trajectory. It should however be noted that while UK institutions report positive effects of the PRFS, a considerable body of scientific literature as well as surveys of UK academics point to negative effects of the PRFS on the character of research and on research careers (Arnold, et al., forthcoming 2017).

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

- In Italy, the adoption of PRFS has stimulated the university system to adopt performance-oriented policies: there is evidence that the recruiting mechanism, local allocation of funding at the Department level, organisation of graduate and PhD courses have been heavily and positively influenced by the new system. On top of that, new policy instruments aimed at funding "excellence" both at the Department and individual level have been gradually introduced, thanks to the availability of PRFS results; finally, there are some signs of increased convergence in the country in terms of reducing the North/South divide (Checci, Mazzotta, & Momigliano, 2017)
- In Norway, where effects on the volume of publications were visible especially for the applied research universities
- In Portugal, where the evaluation system (ie the units for evaluation) has created a dynamic in how the higher education system organises itself
- In Estonia, where PRFS not only distributes institutional funding but also tests whether research organisations should be eligible for it, irrespective of the legal form of the organisation

Other effects mentioned were:

- In Portugal, an effect on education was created thanks to the introduction of PhD programmes as one of the PRFS indicators
- In Norway, pushing colleges to publish encouraged young researchers outside the established mainstream universities
- In Italy, part of the PRFS is specifically orientated towards young researchers, making the system more performance-oriented

6.2 … on the character of research

While university managers tend to focus on the positive, performance-enhancing effects of PRFS, scientific literature and surveys of academics tend to focus on negative effects. This may simply reflect the positive attitude of managers who feel that they have been given more powerful management tools and the negative attitude of those who feel they are being managed. This section explores what the literature has to say about effects on the character of research while the next one looks at research careers, again using the literature.

The 2010 OECD report on performance based research funding provided a list of dimensions of research on which there was concern that PRFS had a problematic 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)

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 have 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 'anti-disciplinary 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). These effects result from the behaviour of UK university research managers concerned with maximising economic returns from the PRFS. There is no evidence that the REF assessment process itself discriminates against interdisciplinarity. Lee and Harley (1998) show that the RAE has marginalised heterodox forms of economics.

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 potential 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. The thrust of the literature is that researchers feel driven by the PRFS requirement for a continuous flow of high-status publications and therefore become risk-averse, since a failed transformative project will generate a hole in their publication output. Opinion surveys support this idea, but there appears to be no systematic evidence about behaviour either way.

A related assertion is that a PRFS discourages the construction of instrumentalities such as infrastructure, databased and methods, because these generate little credit in publication channels. Again, there appears to be no systematic behavioural evidence that can be used to test this claim.

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, 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).

There is no clear evidence about whether the potential for PRFS to discourage applied research (OECD, 2010) is actually realised. To the extent that such research 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. Studies in New Zealand (Edgar & Geare, 2010) and the UK (Henkel, 1999) point to the tension between PRFS' tendency 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.

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)

6.3 … 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 (Hartley, 2002) (Bence & Oppenheim, 2005). Moreover, as the outputs need be of a certain perceived 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), where that is not explicitly rewarded. Many governments provide incentives and programmes for popularisation of science outside the PRFS.

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 in 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 influencing 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).

6.4 Lessons learnt

- PRFS favour the adoption of more performance-oriented policies, providing correct incentive mechanisms for the governance of the University system; possible use of PRFS include: favouring better recruiting mechanism at the university level; greater transparency in the local allocation of funding at the Department level; providing important information to support the organisation of graduate and PhD courses; allowing the introduction of new policy instruments aimed at funding "excellence" both at the Department and individual level
- There is evidence from studies and evaluations that PRFS can increase the volume and quality of research. Since the trend towards including incentives for non-scholarly performance in PRFS is relatively new, it is too early to expect to see much evidence that these are affected though there is anecdotal evidence that incentives such as those encouraging PhD education are effective.
- While performance generally rises after the introduction of PRFS, it is also rising in countries with no PRFS and changes in performance are sometimes triggered by other policy changes than the introduction of a PRFS. It seems that PRFS can contribute to structural changes as part of larger processes of output-orientation rather than always being the unique cause of such changes. At this stage in our knowledge, it is probably best to regard PRFS as supporting performance improvement in the context of a wider set of measures intended to do so. This is also consistent with the perception that PRFS have systemic effects and that their design and introduction must also be considered in a systemic context
- Literature about the effects of PRFS indicates that university managers often believe PRFS improve performance while other academics are inclined to point to actual and potential negative effects on the character of research and on research carers.
- Key risks identified in connection with PRFS include
  - Discouraging interdisciplinary research, which is normally regarded as one of the key sources of progress in science, by encouraging 'safe' disciplinary work that will allow the researcher to produce publications during the period of assessment. There is little evidence that this is the case – though there is clear evidence that the risk-averseness of university managers in deciding what research to submit to the UK PRFS means that

interdisciplinary research is less likely to be submitted (implying negative effects on interdisciplinary researchers' career prospects)

- Similarly, discouraging 'blue skies', 'high risk' and 'transformative' research because these are risky to the career of the researcher. This would similarly be a loss in terms of scientific progress. Equally, research on infrastructures and instrumentalities are said to be discouraged. However, while the literature and researchers speculate that this might be the case there is little systematic evidence that these risk changing researchers' behaviour.
- Discouraging research in non-mainstream topics or using heterodox theories and methods. These may receive low ratings from peer reviewers from mainstream research 'schools' and be hard to publish in the mainstream journals. Here there is supporting evidence, at least from UK economics
- While discouragement of applied research has been raised as a concern, there is little solid evidence that PRFS affect the attractiveness of applied research, provided the approach takes account of differences among fields and the rating system is not skewed in such a way as to judge basic research as being better than applied research more or less by definition. This implies that the risk may be more serious in bibliometrics- than peer-based PRFS
- Reducing researcher autonomy, owing to a need to conform with university management's efforts to encourage research that will generate income from the PRFS. There is literature and survey evidence that this is in fact the case (Arnold, et al., forthcoming 2017)
- Discouraging collaboration. Whether PRFS do this or not depends on the detail of how research publications and activities are 'credited' to different people
- Discouraging popularisation of science and – depending on the system – potentially also third-stream activities. Again, the concern is credible but there is little evidence. We also lack evidence about whether changes in PRFS to include criteria related to impact and societal engagement has an effect on behaviour

- Effects of PRFS on universities and research are mediated by universities' human resource and research management practices. They are likely over time to change the composition of the faculty, building in any positive or perverse effects of the PRFS to the staffing and culture of the university

- University research managers are not well able to predict the results of a PRFS, so they encourage low-risk behaviour and use misleading proxies (such as JIFs) for quality

7 LESSONS AND RECOMMENDATIONS

This chapter first collects together the lessons from the MLE in a single place. Thereafter, it presents some recommendations and 'rules of thumb' intended to use the lessons as a source of learning and advice for countries implementing or considering the implementation of a PRFS.

7.1 Lessons from the MLE

What PRFS are

- In general, universities obtain income partly through 'institutional funding' provided by a ministry of education or a similar government Department and 'external funding' for individual projects or programmes, won in competition with others from specific sponsors. Institutional funding may be allocated separately for teaching and research or for both together
- Institutional funding may be provided as a 'block grant', which is essentially unconditional, or as 'performance-based funding', which as the name suggests is provided on the basis of past performance
- Performance-based research funding systems (PRFS) are two-part systems that assess the 'goodness' of university research on various dimensions and use the results in a funding formula to allocate part or all of the institutional funding for research provided to universities. In addition to finding ways to measure various kinds of quality, they must also relate the quality measurement to the different sizes of institution assessed through some kind of volume measure
- Most PRFS focus on the quality of research and its scientific impact. In more recent systems, a growing number of other criteria are being used, especially in relation to societal impact
- PRFS are normally used to allocate a fixed pot of money or budget among institutions. They rarely trigger increases in total funding, though sometimes they allocate increases in funding
- In practice, few PRFS contain prospective elements
- The fundamental units of analysis used in PRFS are individuals or research groups but whichever approach is taken, they do not report at the levels of individuals. Analysis of both individuals and groups can be aggregated to the institutional level via a funding formula. Innovation is also possible in the scope of PRFS, for example by stretching across institutional boundaries

PRFS in the policy toolkit

- A PRFS is one of several mechanisms available to policymakers to encourage the improvement of quality, relevance, impact and other desiderata in university research. The other mechanisms include: providing or changing the amount of state external research funding provided; altering the ratio between institutional funding and external funding; increasing international collaboration in research; reforming and modernising university governance; increasing overall investment in higher education research
- There is wide divergence among countries in their balance of funding mechanisms and research funding. There is no clear theory available about how the individual components, their balance or changes in these relates to overall performance and – given the differences among national contexts – it is therefore unlikely that there is some 'optimum' mix of incentives that will work in all countries and at all times
- The proportion of universities' research income driven by PRFS also varies considerably among countries. Effects on behaviour are visible at both low and high proportions of funding
- As with some of the other policy options, the introduction of a PRFS can administer a shock to the research system, changing culture and behaviour. Striking a balance between shock and continuity is another challenge for designers of PRFS
- The PRFS and the other parts of the policy mix need to address policy goals in a complementary manner
- PRFS also can play a part within the narrower domain of research funding policy, where – like other sources of research funding – they can help steer the balance of the universities' research effort among fields, between 'basic' and applied research and various aspects of relevance
- In introducing a PRFS, it is important to consider its systemic implications in the sense of its influence on the overall pattern of funding and incentives and whether the incentive system is then balanced so that universities are encouraged to fulfil all their three missions well and not improve one at the cost of the others

Linking the PRFS to policy design

- Crucially, PRFS exist to help reach policy goals. Despite widespread fascination in the academic community with their details, they are rarely 'just' mechanisms for distributing money
- Just as in any other kind of funding, it is crucial that the assessment criteria used (and the corresponding indicators) reflect policy objectives. The funding formula must similarly be consistent with the policy objectives

Issues in PRFS design

- PRFS design is important not only in itself but also because there is a tendency for the design to be institutionalised into the way universities are funded and manage themselves, creating lock-in. In practice some countries have become more or less culturally wedded to a particular broad design. Hence, there is a premium on getting the overall design right first time as once it is implemented it may be politically difficult to make more than incremental changes to it
- While early PRFS tended to be wholly based either on peer review or metrics, a growing number rely on a combination of these approaches to produce higher-quality judgements at lower cost. Some more experienced countries learn over time, changing their models, but the changes tend to be incremental
- It is hard for the university system to react appropriately to a PRFS unless it is transparent and the incentives provided by the system can easily be understood
- A PRFS has two components: an assessment process; and a funding formula. It is important to keep these separate if the PRFS is to be transparent and therefore to avoid putting elements of the funding calculation into the assessment process or vice versa
- Key parameters for PRFS designers include: the assessment model; the scope of research included; types of indictors used; Assessment criteria; the granularity of the analysis; and the frequency (periodicity) with which the PRFS can be run
- There is continued disagreement about whether peer review or metrics approaches are 'best'. In practice some countries have become more or less culturally wedded to one or the other. Key issues in the choice between these models are: cost; whether a formative element is needed; periodicity, with peer review tending to be done infrequently while metrics approaches can be taken annually
- Inter-field differences ranging from the methodological and epistemological norms of the epistemic communities involved,

different propensities to publish and publication traditions make it hard to compare fields directly. Peer review systems can do this through the use of common quality scales interpreted by practitioners from the respective fields. (Biblio)metric systems must do this through field normalisation

- Metrics approaches can be much more complex than peer review systems tend to be, though it is not necessary that metrics systems should be complex. There is probably a trade-off between complexity and the PRFS' ability to influence behaviour, suggesting that a good design principle (here, as in science) is 'as simple as possible but no simpler' and hence limiting the number of indicators used
- The greater the proportion of universities' research income governed by a PRFS, the more robust its methods need to be in order to withstand scrutiny by the beneficiaries
- While much of the discussion about PRFS focuses on the assessment processes, different funding formulae provide incentives for different kinds of behaviour. Different funding formulae provide incentives for different kinds of behaviour. The use of skewed formulae – where the best performers are rewarded disproportionately – is a way to concentrate resources on 'excellence'
- Overhead and infrastructure costs vary among fields. Consideration should be given to weighting the formula in order to take this into account
- Small countries face particular design issues: the costs associated with small scale; the limited number of fields that can be addressed in peer review systems, as a result of which the few fields defined have to be broader than in big systems; the need to use foreign peers; the constraints of 'small' languages on peer recruitment and the corresponding need for a quality-assured national CRIS; national capacity to run a research assessment exercise
- The costs of running a PRFS appears to be driven by: scale, the model used, whether there is a formative element; use of qualitative vs quantitative data; the number of people in scope to the assessment; whether there is a quality-controlled national CRIS; the degree of opposition from the academic community and the consequent need to justify and consult about methodology (particularly where the PRFS governs a large proportion of research funding)
- While the direct costs of those PRFS for which we could obtain data are rather modest, there was considerable concern among the participating countries about the importance of constraining that cost. Going beyond these countries, the cost of the process appears particularly to be an issue in the UK system, which not only involves

large numbers of peer reviewers but also triggers a large effort by the universities to present their work in the best light

Information sources – their uses and influences

- PRFS outcomes are influenced by the characteristics and quality of the information upon which the research assessment is based. Key issues include
  - Biases introduced when universities or other actors select a sub-set of research results for submission to the PRFS
  - Biases introduced by the way and extent to which sources (such as CRIS or commercial bibliometric databases) decide what should be included in, or excluded from, the database
  - The definition of who counts or does not count as a researcher, for the purpose of the PRFS
- Self-assessments are necessary for collecting qualitative information, for example on strategy, human resource management and research infrastructure. Where other facts – such as lists of publications or numbers of researchers – are needed it is preferable to obtain them from quality-assured sources. When self-assessment is used, the PRFS manager must ensure that the information needs are well explained and understandable not only by people used to being evaluated but also by those with little or no experience
- A quality-assured CRIS is a very useful tool to support PRFS, especially in countries using 'small' languages, which tend to be under-represented in the commercial bibliometric databases and where a recognised national authority can categorise outputs in terms of quality
- Care must be taken in populating a CRIS to ensure that data are quality assured, complete and extend to those categories of non-scholarly data needed for research assessment

Bibliometrics

- Bibliometric tools are now widely available. However, generating and using robust bibliometric indicators still requires skill and considerable understanding of the limitations of the relevant databases and the characteristics of different fields of research. Hence, professional bibliometricians (independent of the commercial database vendors) should be involved in the design and implementation of any PRFS that uses bibliometric information
- Limitations of bibliometric data need to be understood. These include bias towards English-language journals, differences in the amount and pattern of publication among disciplines, the extent to which research

outputs figure in the available databases, the self-referential nature of those databases and the lack of transparency and opportunities to game the criteria governing which journals are indexed

- Important considerations about bibliometrics
  - Bibliometrics is not 'objective' but contains many in-built assumptions and limitations
  - Stakeholder consultation and advice from disinterested bibliometricians are preconditions for using bibliometrics in PRFS
  - The strength of the incentives provided via bibliometrics depends on their economic importance as well as interplay with other incentives
  - Whatever bibliometrics data sources are used, the main considerations are data quality, disinterested data production, incentives for internationalisation, costs, comprehensiveness, balanced representation of all fields and the representation of national language publishing
  - The principles used for including publications in databases must be understood and considered in the light of the PRFS' objectives
  - Publication, collaboration and citation are typically analysed using bibliometric indicators. Care must be taken to ensure consistency and comparability across fields
  - Where bibliometric indicators are used in PRFS, they should be field normalised, taking account of differences in productivity, use of publication channels, citation practice, authorship and co-publication conventions
  - Counting methods for multi-author and multi-institution papers affect collaboration behaviour. This must be taken into account in PRFS design
  - Different types of publications (eg books versus journal articles) need to be differently weighted, in order to take account of differences in publication behaviour among epistemic communities
- Rating publication channels as opposed to individual research outputs can be misleading not only because a JIF does not predict the quality of an individual article but also because of differences in practice among fields. Journal impact factors (JIF) and h-indices are easily accessible indicators but should be interpreted carefully. Respect the DORA declaration
- H-indices are other easily accessible indicators but should also be avoided in PRFS because they generate misleading results

- The Leiden Manifesto provides a good set of principles and should be consulted before doing bibliometric PRFS design

Peer review

- Qualitative aspects of self-assessment can only be analysed by peer or expert panels. These have no place in metrics-only systems. Peer review is indispensable when the assessment of quality requires a detailed understanding of the research or field in question
- The right choice of information provided to peer review panels depends upon the purposes of the PRFS and the criteria used
- A well-balanced mix of different measures (e.g. 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 their use by peers or panels are important in this context
- Peer-based PRFS often make use of self-assessment. These make it possible to introduce a formative element into the PRFS
- 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. When selecting peer panels, it must be understood that the choice of people affects the outcomes of the exercise. Panels are social entities, in which individual behaviour (shirking, bullying, leadership) affects outcomes
- Peer review in PRFS is increasingly 'informed' by bibliometrics. This has the benefit that panel members can work with relatively robust indicators, rather than themselves looking up crude measures such as JIFs or h-indices
- Most PRFS focus on the quality of research and its scientific impact. However, a growing number of other criteria are being used, especially in relation to societal impact. However, peer reviewers are not necessarily good judges of the impact dimension. Extending peer review beyond scientific quality calls for the involvement of non-scientists
- In the absence of reliable impact indicators, extended peer panels have proved themselves able to assess narrative impact statements though it is not yet known whether these judgements are reliable or reproducible
- New forms of peer review considering multiple criteria and relying at least partly on non-scientific peers have been innovated to help extend peer review beyond scientific excellence, to address wider socio-economic criteria like relevance, impact, and new forms of interaction with society

- 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 used
- Using a cost-effective mix of panel and remote reviews demands some aggregation mechanism. In order to ensure consistency some kind of calibration between different disciplines, interest groups and different panels is needed
- Peer review is inherently labour-intensive and therefore expensive. Separating assessment from panels work and using remote assessment can reduce cost. Potentially high costs of peer review can be mitigated by introducing a proper panel system with elements of remote review as well as by international cooperation both in the design and in the application phase
- It is impractical to run a PRFS with a single panel. Sub-panels are needed
- Mutual learning among universities can be considered as an alternative to a peer review PRFS as a source of formative assessment – though it is naturally less binding

Societal impact

- Societal impact comprises many different things and can occur through a wide range of mechanisms. A useful categorisation is to distinguish among effects on policy and practice, society and culture, and economic impact
- The diversity of impact pathways and mechanisms combined with a lack of a standard way to describe or measure impact makes implementation of impact assessment in PRFS difficult. Societal impact assessment in PRFS requires even greater caution than quality assessment and should not be taken lightly by PRFS designers
- Many researchers are not used to analysing or reporting impact, so the quality of their self-reporting is likely to be variable
- The lack of a standard way to describe or measure impact makes implementation of impact assessment in PRFS difficult. Nor is there any case of systematic integration of third stream indicators into a PRFS in the countries studied. PRFS therefore have to use more or less ad hoc measures, preferably combining quantitative and qualitative methods
- Impact assessment methods are being developed but have to be customised to the individual context and tend to be labour-intensive. It is not evident that these are yet ripe to be used systematically across fields in the way necessary in PRFS

Using the results of assessment

- Overhead and infrastructure costs vary among fields. Consideration should be given to weighting the formula in order to take this into account. This appears to be especially important where the PRFS governs a large proportion of institutional income, leaving universities little 'slack' in using the block grand to tackle variations in costs among disciplines
- The opportunities to use PRFS results as a source of strategic intelligence for policy are often under-exploited and are not necessarily considered in PRFS design. Explicit consideration should be given to whether and how to exploit information generated by the PRFS in research policy and management more widely
- While PRFS results are widely claimed to provide the universities with strategic intelligence, they are at best an incomplete source of it and are often not designed to provide it. Universities should be monitoring their own quality on a frequent basis in any case
- Peer review based PRFS can be designed so as to provide formative information to universities, but can be done only at long intervals. This is especially useful in emerging or developing research systems but becomes less so as systems mature

What are the effects of PRFS?

- PRFS favour the adoption of more performance-oriented policies, providing incentive mechanisms for the governance of the University system. Possible uses of PRFS include: favouring better recruiting mechanism at the university level; greater transparency in the local allocation of funding at the Department level; providing important information to support the organisation of graduate and PhD courses; allowing the introduction of new policy instruments aimed at funding "excellence" both at the Department and individual level
- There is evidence from studies and evaluations that PRFS can increase the volume and quality of research. Since the trend towards including incentives for non-scholarly performance in PRFS is relatively new, it is too early to expect to see much evidence that these are affected though there is anecdotal evidence that incentives such as those encouraging PhD education are effective
- While performance generally rises after the introduction of PRFS, it is also rising in countries with no PRFS and changes in performance are sometimes triggered by other policy changes than the introduction of a PRFS. It seems that PRFS can contribute to structural changes as part of larger processes of output-orientation rather than always being the unique cause of such changes. At this stage in our knowledge, it is

probably best to regard PRFS as supporting performance improvement in the context of a wider set of measures intended to do so. This is also consistent with the perception that PRFS have systemic effects and that their design and introduction must also be considered in a systemic context

- Literature about the effects of PRFS indicates that university managers often believe PRFS improve performance while other academics are inclined to point to actual and potential negative effects on the character of research and on research carers.
- Key risks identified in the scientific literature in connection with PRFS include
  - Discouraging interdisciplinary research, which is normally regarded as one of the key sources of progress in science, by encouraging 'safe' disciplinary work that will allow the researcher to produce publications during the period of assessment. There is little evidence that this is the case – though there is clear evidence that the risk-averseness of university managers in deciding what research to submit to the UK PRFS means that interdisciplinary research is less likely to be submitted (implying negative effects on interdisciplinary researchers' career prospects)
  - Similarly, discouraging 'blue skies', 'high risk' and 'transformative' research because these are risky to the career of the researcher. This would similarly be a loss in terms of scientific progress. Equally, research on infrastructures and instrumentalities are said to be discouraged. However, while the literature and researchers speculate that this might be the case there is little systematic evidence that these risk changing researchers' behaviour.
  - Discouraging research in non-mainstream topics or using heterodox theories and methods. These may receive low ratings from peer reviewers from mainstream research 'schools' and be hard to publish in the mainstream journals. Here there is supporting evidence, at least from UK economics
  - While discouragement of applied research has been raised as a concern, there is little solid evidence that PRFS affect the attractiveness of applied research, provided the approach takes account of differences among fields and the rating system is not skewed in such a way as to judge basic research as being better than applied research more or less by definition. This implies that the risk may be more serious in bibliometrics- than peer-based PRFS

- Under-valuing research published in languages other than English. This concern reflects a clear bias in the way the scientific literature is indexed but has not been systematically tested in relation to PRFS. Some of these concerns, such as English-language bias and the maintenance of incentives for collaborative research have been addressed in individual PRFS designs
- Reducing researcher autonomy, owing to a need to conform with university management's efforts to encourage research that will generate income from the PRFS. There is literature and survey evidence that this is in fact the case (Arnold, et al., forthcoming 2017)
- Discouraging collaboration. Whether PRFS do this or not depends on the detail of how research publications and activities are 'credited' to different people
- Discouraging popularisation of science and – depending on the system – potentially also third stream activities. Again, the concern is credible but there is little evidence. We also lack evidence about whether changes in PRFS to include criteria related to impact and societal engagement have an effect on behaviour

- Effects of PRFS on universities and research are mediated by universities' human resource and research management practices. They are likely over time to change the composition of the faculty, building in any positive or perverse effects of the PRFS to the staffing and culture of the university
- University research managers are not well able to predict the results of a PRFS, so they encourage low-risk behaviour and use misleading proxies (such as JIFs) for quality

7.2 Recommendations

PRFS in the policy toolkit

PRFS are a useful addition to the set of instruments and policies governments can use to improve the quality and performance of university systems. The evidence is that they help improve system performance but that they do so in interaction with other policy instruments and aspects of the context. They are not alone sufficient for performance improvement and there are alternative ways to achieve the same end but can usefully be a component in a wider policy for such improvement. A PRFS that governs a high proportion of institutional funding risks making a dangerously high proportion of funding contestable. Many countries have obtained performance improvements using PRFS that govern only a small fraction of institutional funding

- **Depending upon the specific developmental needs of the university research system, states should therefore consider adopting a PRFS or an appropriate alternative if the national university system's research performance is in need of improvement**
- **The scope and design of the PRFS should be consistent with other policy measures operating and with the national context**
- **In particular, the proportion of institutional funding governed by the PRFS should be given careful consideration, based on national policy needs and the likely interplay between the PRFS and other policy instruments**
- **Where PRFS risks triggering abrupt changes in funding, use damping mechanisms that limit these changes to levels that are sustainable**

A PRFS is intended, in combination with other policy instruments, to help reach not only new but also often older policy goals that remain relevant.

- **Both the assessment criteria and the funding formula used in the PRFS should encourage behavioural and institutional changes that reflect the policy goals**
- **Care must be taken in introducing new goals that old ones are not impeded, for example by creating a focus on research performance that is not balanced by incentives for universities to pursue their two other missions of education and knowledge exchange with society**

PRFS design

Getting the design of a PRFS right is particularly important because international experience shows that such systems quickly become 'locked in' to the way a national research system and its institutions operate. PRFS design can be difficult, both because of the interactions between PRFS on the one hand and other policies and instruments on the other and because there are technical difficulties inherent in designing and implementing such systems without generating undesirable distortions in behaviour.

- **Designers should anticipate that it may be hard to make more than incremental adjustments to the design once the PRFS has been fully implemented**

- **Design requires expertise in the design and implementation of assessment and funding systems and in technical aspects of the assessment process, such as the use of bibliometrics, peer review and societal impact indicators**
- **The choice and weighting of criteria or indicators can affect researcher behaviour. The likely intended and unintended effects of the PRFS on behaviour should be anticipated and simulated as far as possible at the design stage**
- **While any PRFS design should use robust methods, the pressure from the research community for robustness and precision will be greater the greater its influence over universities' total institutional funding**

The evaluation community – and especially bibliometricians – has devoted considerable effort to clarifying principles for good research assessment and in identifying practices that lead to distortions.

- **PRFS designers should routinely consult both the DORA declaration and the Leiden Manifesto during the design process**

It is hard for the university system to react appropriately to a PRFS unless it is transparent and the incentives provided by the system can easily be understood.

- **PRFS should be simple to understand and easy to communicate to universities and researchers alike**
- **They should avoid using large numbers of criteria and indicators as these create conflicting incentives and researchers cannot satisfy them all**
- **The assessment process and the funding formula should be distinct and independent**

PRFS model

The choice between a peer-review based and a metrics-based approach (or some combination of the two) involves a number of important trade-offs. These include the cost of the system, the frequency with which it can operate, its likely credibility with the academic community and its ability to make qualitative judgements and provide formative information.

- **The choice of a peer review-based or a metrics-based system should be made based on the needs that the PRFS is intended to satisfy and the constraints under which it will operate**
- **A combination of the two approaches is often useful and should be considered**

The quality of a PRFS is strongly affected by the quality of the information that it processes. In order to obtain robust outcomes, the PRFS managers need to ensure that there are processes in place to quality-assure that information and that they understand the biases and omissions to which that information is prone.

- **Where a PRFS reviews a sub-set of research output, the biases created through selection must be understood**
- **The character and biases inherent in bibliographic and other databases as well as in other metrics should be understood and accounted for**
- **Self-assessment is a useful tool in research assessment, especially where a formative element is desired. Where self-assessment is used, it should involve transparent questions which have been clearly explained by the PRFS managers. Researchers not used to self-evaluation may need particular guidance. Self-assessment should not be used to collect data available from quality-assured sources elsewhere as this increases the burden on those completing self-assessments and introduces errors**
- **A national CRIS should be considered as a way to maintain a high quality of input data to the PRFS, especially in smaller countries that work in 'small' languages**

Generating and using robust bibliometric indicators still requires skill and considerable understanding of the limitations of the relevant databases and the characteristics of different fields of research.

- **Hence, professional bibliometricians (independent of the commercial database vendors) should be involved in the design and implementation of any PRFS that uses bibliometric information**
- **Building advanced bibliometric expertise requires a significant investment in a research centre in the field. There are a handful of such centres in Europe already. Especially small countries with limited resources should exploit existing**

expertise abroad rather than necessarily themselves making a heavy national investment in one

Constructing and managing robust peer review panels and assessment processes is a complex art. Several dimensions need to be considered.

- **Care must be taken in defining the membership of panels not to predetermine outcomes by excluding relevant disciplines or schools of thought and to ensure that members have a track record of being 'team players'**
- **Where criteria other than scientific quality are to be considered, panels should be 'extended' beyond academic peers to include relevant representatives of other parts of society**
- **Procedures should be in place to prevent conflicts of interest. In small countries, all peers should be international. A good proportion of international peers is also desirable in larger systems, to maintain contact with world science**
- **A PRFS cannot be run by a single panel. There have to be sub-panels and/or panels informed by remote peer assessment. Judgements have to be calibrated and made consistent across different panels and reviewers**

Societal impact

Assessing or trying to measure societal impacts of research is extremely difficult because of the importance of differences in context, differences in impact mechanisms (including the fact that some impacts involve things not happening), different conceptualisations of 'impact', the lack of a unitary 'measure' of impact that goes beyond economic to non-economic impacts and difficulties of attribution. While there is research in progress to improve approaches to impact assessment these tend to be context-dependent and labour-intensive. There are nonetheless promising developments in societal impact assessment, both in terms of qualitative assessments (for example impact pathways) and quantitative assessments (for example productive interactions). Mutual learning is strongly advised.

- **Societal impacts of research arise partly because of the value of new knowledge and partly through a range of other contextual factors, not the least of which may be luck. PRFS experimenting with assessing societal impacts should consider whether it is better to reward outputs and outcomes that can**

reasonably be expected to be steps on the way to impact or whether they want to reward impact itself

- **Impact metrics should not be used on their own in a PRFS. Their meaning is unclear and they are easily gamed**
- **At the current state of the art, human judgement is the only way to assess impact. This can be supported by metrics but these must be interpreted by people**
- **Extended peer review of impact narratives can be used, though defining how these are to be prepared and assessed in a consistent way is very labour-intensive. Judgements are subjective and it is not known whether they are reproducible**
- **Impact assessment should therefore be attempted with caution and can in certain cases be included in the assessment module of a PRFS while being excluded from the funding formula so that the incentive to generate impact is prestige rather than money**

Using the results of assessment

In a PRFS, the primary intended use of assessment is – by definition – to drive a funding formula. However, the assessment exercise also provides an opportunity to generate strategic intelligence for universities and policymakers.

- **Funding formulae should reflect the policy priorities of the PRFS. They can be skewed towards high-performing institutions if the desire is to concentrate funding in fewer places, they can drive quality or quantity or both, they can be used to adjust the relative incentives for researchers to perform along different dimensions (such as scientific quality versus societal impact), and so on**
- **In systems where the PRFS steers a high proportion of institutional funding, it is especially important that funding formulae take account of the need to align with influences on the institutional cost of doing research such as variations in cost among disciplines and the need to pay the overhead costs of doing research for external funders such as research councils, charities and sometimes industry that can or will not pay the full economic costs**
- **Assessment results can provide useful strategic intelligence for policymaking, such as offering a picture of national research performance in different fields. It is not necessarily**

difficult to combine and recombine assessment results at different levels to provide such strategic intelligence. PRFS systems should do so, where they are able to make a unique contribution to policy development

Intended and unintended effects of PRFS

There is evidence that PRFS in combination with other measures can improve university research performance on a range of dimensions, though the amount of evaluation that has been done in this area is modest. There is also a range of concerns about actual and potential negative consequences that appears in the scientific literature and in surveys of researchers, especially in the UK, but for which there is limited systematic evidence from changes in behaviour and performance, as opposed to opinions. Given the growing importance of PRFS in funding systems, this situation is clearly unsatisfactory.

- **Member States should consider evaluating their PRFS periodically, if possible aiming not only to describe the gross effect of the current range of policies on research performance but also the net effect of the PRFS itself and the extent to which it leads to behavioural and organisational changes that render the PRFS itself redundant**
- **They should also conduct specific studies in order to search for stronger evidence about potential negative consequences**
- **Member states should understand those potential risks and consider them when deciding on the mix of university funding instruments deployed and in the design and architecture of their PRFS.**
- **Internationally comparative studies and evaluations are also needed in order to start to disentangle the operation of PRFS from national contexts**

8 APPENDIX A: BIBLIOGRAPHY

(excludes references in Appendix)

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.

Ahlqvist, V., Andersson, J., & Söderqvist, L. (2014). *A gender neutral process?* Stockholm: Swedish Research Council.

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., Knee, P., Angelis, J., Giarracca, F., Grinice, E., & Jávorka, Z. (2014). *Latvia Research Assessment Exercise: Summary Report.* Riga: Ministry of Education and Science.

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.

Benner, M., & Öquist, G. (2012). *Fostering breakthrough research: A comparative study.* Stockholm: Kungliga Vetenskapsakademiet.

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).

Bollen, J., van de Sompel, H., Hagberg, A., & Chute, R. (2009). A Principal Component Analysis of 39 Scientific Impact Measures. *PLoS ONE, e6022*.

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the Americal Society of Information Science and Technology*, 217-233.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research?An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics, 8*, 895-903.

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.

Canadian Federation for the Humanities and the Social Sciences. (2017). *Approaches to Assessing Impacts in the Humanities and Social Sciences.* Ottawa: Canadian Federation for the Humanities and the Social Sciences.

Checci, D., Mazzotta, I., & Momigliano, S. (2017). Research assessment in Italy: have the results of universities converged? *Science, Technology and Innovation Indicators Conference.* Paris.

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.

Derrick, G., & Samuel, G. (2016). The Evaluation Scale: Exploring Decisions About Societal Impact in Peer Review Panels. *Minerva, 54*.

DRUSSA. (2017). *Development Research Uptake in Sub-Saharan Africa (DRUSSA) Programme Completion Report October 2012 – September 2016.* DRUSSA.

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.

Fröhlich, G. (2016). Ausgleich der Fehler und Verzerrungen? Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung? *Hochschulrektorenkonferenz*, pp. 193-204.

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.

Garfield, E. (2001). Impact factors, and why they won't go away. *Nature, 411*(6837), 522.

Geuna, A., & Martin, B. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva, 41*(4), 277-304.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge.* London: Sage.

Glänzel, W., Thijs, B., & Debackere, K. (2016). Productivity, performance, efficiency, impact—What do we measure anyway? Some comments on the paper "A farewell to the MNCS and like size-independent indicators" by Abramo and D'Angelo. *Journal of Informetrics, 10*, 658-660.

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.

Hartley, S. (2002). The Impact of Research Selectivity on Academic Work and Identity in UK Universities. *Studies in Higher Education, 27*(2), 185-205.

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

Hicks, D. (2012). Performance-based university research funding systems. (Elsevier, Ed.) *Research Policy, 41*, 251-261.

Hicks, D., Wouters, P., Waltman, L., de Rijke, S., & Rafols, I. (2015, April 23). The Leiden Manifesto for research metrics. *Nature, 520*, 429-431.

Holmberg, K. (2017). *[July – 2017] Mutual Learning Exercise: Open Science- Altmetrics and Rewards Different types of Altmetrics Thematic Report No 1.* Brussels: DG-RTI Policy Support Facility.

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.

Joly, P.-B., & Matt, M. (2017). Towards a new generation of research impact assessment approaches. *Journal of Technology Transfer*.

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.

Kerr, N. L., & Tindale, R. S. (2004). Group Performance and Decision Making. *Annual Review in Psychology - Annual Reviews, 55*, 623-655.

King's College and Digital Sciences. (2015). *he Nature, Sscale and Beneficiaries of Research Impact; an Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies.* London: King's College and Digital Sciences.

Kuhlmann, S., & Rip, A. (2014). *The challenge of addressing Grand Challenges.* STePS Group. Enschede: University of Twente.

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

Langfeldt, L. (2004). Expert panels evaluating research: decision-making and sources of bias. *Research Evaluation, 13*(1), 51-62.

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.

Mahieu, B., Arnold, E., & Kolarz, P. (2014). *Measuring scientific performance for improved policy making Study, IP/A/STOA/FWC/2008-096/Lot8/C1/SC13.* Brussels: STOA.

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.

Mellors-Bourne, R., Metcalfe, J., & Gill, A. (2017). *Exploring equality and diversity using REF2014 environment statements: Report to HEFCE by the Careers Research & Advisory Centre (CRAC).* Cambridge: CRAC.

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.* New York: Springer.

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

Nowotny, H. (2003). Democratising expertise and socially robust knowledge. *Science & Public Policy, 30*(3), 151-156.

Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-Thinking Science; Mode 2 in Societal Context.* Cambridge: Polity Press.

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

OECD. (2002). *Proposed Standard Practice for Surveys of Research and Development (Frascati Manual), 6th edition.* Paris: OECD.

OECD. (2010). *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings.* Paris: 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? *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. *16*, 79-91.

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.

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday, 15*(7).

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.

Ruegg, R., & Jordan, G. (2007). *Overview of Evaluation Methods for R&D Programs A Directory of Evaluation Methods Relevant to Technology Development Programs.* U.S. Department of Energy Office of Energy Efficiency and Renewable Energy.

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

Schneider, J. W., Aagaard, K., & Bloch, C. W. (2016). What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation, 25*(3), 244-256.

Science Europe. (2016). *Position Statement on Research Information Systems. Co-ordination by Science Europe Working Group on Research Policy and Programme Evaluation.* Brussels: Science Europe.

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

Sile et al, L. (2017). *European Databases and Repositories for Social Sciences and Humanities Research Output.* Antwerp: ECOOM & ENRESSH.

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.

Sivertsen, G. (2016). Publication-Based Funding: The Norwegian Model. In M. Ochsner, S. E. Hug, & H. D. Daniel, *Research Assessment in the Humanities: Towards Criteria and Procedures* (pp. 79-90). Zürich: Springer Open.

Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. *Scientometrics, 91*(2), 567-575.

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

Spaapen, J. B., Wamelink, F., & Dijstelbloem, H. (2007). *Evaluating research in context: A method for comprehensive assessment.* The Hague: Consultative Committee of Sector Councils for Research and Development.

Spaapen, J., & van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. *Research Evaluation, 20*(3), 211-218.

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

Stern, N. (2016). *Building on Success and Leaning from Experience: An Independent Review of the Research Excellence Framework.* London: Department for Business, Energy and Industrial Strategy.

Swedish Government. (2016). *Regeringens Proposition 2016/17:50: Kunskap i samverkan—för samhällets utmaningar och stärkt konkurrenskraft.* Stockholm: Swedish Government.

Swedish Research Council. (2015). *Research quality evaluation in Sweden – FOKUS Report of a government commission regarding a model for resource allocation to universities and university colleges involving peer review of the quality and relevance of research.* Stockholm: Swedish Research Council.

van den Akker, W., & Spaapen, J. (2017). *Productive Interactions: Societal impact of research in the knowledge society, position paper.* Brussels: LERU.

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.

Watts, S. M., George, M. D., & Levey, D. J. (2015). Achieving Broader Impacts in the National Science Foundation, Division of Environmental Biology. *BioScience, 65*(4), 397-407.

Webb, P. (2013). *Impact Pathways from Agricultural Research to Improved Nutrition and Health: Literature Analysis and Research Priorities.* FAO and WHO.

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

Wilsdon et al, J. (2015). *The Metric Tide Report of the Independent Review of the Role of Metrics in Research Assessment and Management.* Bristol: HEFCE.

Wise, E., Berg, M., Landgren, M., Schwaag-Serger, S., Benner, M., & Perez Vico, E. (2016). *09Evaluating the Role of HEIs´ Interaction with Surrounding Society Developmental Pilot in Sweden 2013-2016.* Stockholm: Vinnova.

9 APPENDIX B: FUNDING FORMULAE IN PRFS, VALIDATED AND APPROVED BY PARTICIPATING STATES

Austria. The PRFS is one component of a wider indicator-based system for allocating a small part of the universities' institutional funding. The other indicators relate to education: numbers of active students and graduates. The indicators are used to distribute a fixed budget. Some 4% of the budget rewards universities for the number of PhD students they have so a university gets

Budget * 4% * (University's No of PhD students / All PhD students in Austria)

The funding indicator covers income from the Austrian Science Fund (FWF) and from the EU and weights FWF funding by a factor of two, in order especially to encourage more fundamental research. It distributes 15% of the budget. Thus

Budget * 15% * (2 * FWF revenues + EU revenues)/(2 * all FWF revenues + all Austrian EU revenues)

Croatia. The Croatian funding formula distributes funding based on institutions' performance relative to the national average in a range of fields. It uses four groups of indicators

- Scientific productivity (60%)
- External research funding (25%)
- Research-industry collaboration (10%)
- Popularisation of science (5%)

The detail of the sub-indicators and their weights differs among each of the seven fields of research with which the PRFS operates.

- Scientific productivity includes
  - number of research papers published in journals indexed in WoS and SCOPUS databases
  - number of other papers relevant to career advancement in accordance with the national legislation
  - number of relevant books published
  - citation index in WoS and Scopus database

- National and international competitive research projects and research mobility includes
  - number of contracted national competitive projects
  - share of the budget of contracted national competitive projects in the total budget of the institution
  - number of contracted international competitive projects
  - share of the budget of contracted international competitive projects in the total budget of the institution
  - incoming mobility (minimum one month period)
  - outgoing mobility (minimum one month period)
- Collaboration between research and business sector as well as collaboration with the units of local and regional governance and non-governmental sector includes the number of contracted projects with business sector and/or units of local and regional governance and non-governmental sector
  - Popularisation of science includes international activities related to the popularisation of science
  - national activities related to the popularisation of science
  - local activities related to the popularisation of science

The volume unit is the number of FTE researchers. The funding unit is calculated by multiplying the number of FTE researchers per field by specific field weights. The overall budget is then divided by this number. Field weights are

1. Natural Sciences = 2.7
2. Technical Sciences = 2.5
3. Biomedicine = 2.7
4. Biotechnology = 2.6
5. Social Sciences = 1.2
6. Humanities and Arts = 1.2

The funding formula works by

1. Multiplying all indicators per field by the relevant field weight and dividing this by the number of PROs to establish a mean score per indicator

2. The scores each institution obtains on each indicator are then divided by their number of FTEs to obtain a score per FTE
3. For each indicator, the score per FTE (2) is divided by the mean score per indicator (1)
4. The results of (3) are then summed up for each institution to give an 'institution coefficient'
5. Finally, the budget is allocated as: volume unit * funding unit * institution coefficient

Czech Republic. The Czech PRFS allocates points based on individual outputs such as scientific papers and patents and provides a monetary reward per point to each institution. It is based on metrics based on data submitted by the institutions to the national Research Information System but also uses a limited amount of peer review. Each research organisation has to submit one output per CZK 10m of institutional support. These outputs are assessed by panels of national experts which select 20% of them as excellent.

The criteria are

- Academic outputs (85%)
  - Scientific publication in an impacted journal
  - Scientific publication in Scopus
  - Scientific publication in ERIH
  - Scientific publication in non-impacted journal (has to be indexed in an official list approved by the national Research and Innovation Council)
  - Book
  - publication in proceedings (has to be indexed in SCOPUS or WoS)
  - Outputs deemed excellent by panels of national experts
- Non-academic outputs (15%)
  - European, US and Japanese patents
  - Czech patents (a valid licence agreement must exist)
  - Other patents
  - Funds obtained for projects of applied research and from private sources

The PRFS works with 12 fields of research, each of which has its own method of constructing and weighting indicators in order to field-normalise them where relevant.

The funding formula works by allocating a number of points for each indicator. 85% of the available budget is distributed based on each institution's total number of Academic output points and 15% based on Non-academic outputs.

The UK. The UK PRFS has 5 subsystems, aiming to concentrate funding on excellent research.

- It distributes about two-thirds of institutional funding based on the Research Excellence Framework (REF), which uses peer panels' judgement of the quality of outputs submitted by the universities (65%), their statements about the impact of past research (20%) and information about the environment in which research is conducted (15%)
- The Research Degree Programme distributes a fixed budget for PhD education to selected university departments, based largely on REF scores
- The Charity Support Fund distributes a fixed budget to universities pro rata their shares of the UK universities' total income from research charities
- The Business Support Fund distributes a fixed budget to universities pro rata their shares of the UK universities' total income from research funded by business
- A small amount goes to five university libraries, which have been designated National Research Libraries because their collections have national and not only local importance

The REF operates every 5-6 years, completely reallocating 'Mainstream Quality Related' institutional funding each time. The support funds are allocated annually, based on four-year rolling averages.

The largest component of REF funding is 'Mainstream QR', which is 65% based on the quality of research outputs submitted by the universities to the REF, 20% on statements of the impact of past research and 15% for descriptions and data about the 'environment' in which research is conducted. All three dimensions are given 0-4 stars. In the funding calculation, the output of a 'unit of assessment' is split among these categories. The proportion of 4-star research is weighted by 4 in the

funding calculation and the proportion of 3-star work by 1. Lower grades are weighted 0, so count for nothing.

Further weights are calculated into the formula to address differences in costs across fields.

- A = 1.6, for the most expensive fields, which are typically laboratory-intensive
- B = 1.3 medium-cost ones
- C = 1.0 for low-cost fields such as the humanities

The resulting score is further weighted to reflect the higher cost of doing research in London. The scores of universities based in central London are multiplied by 1.12 and those in outer London by 1.08. The total of the scores allocated is used to divide the QR budget in volume units. The funding formula is

Mainstream QR = (% of 4-star research x 4 + % of 3-star research) x discipline cost weights x London weighting x volume unit

Estonia. The Estonian PRFS not only allocates institutional funding for research but also accredits institutions to entitle them to receive that funding as well as grants from the Estonian Science Council. It distributes money based on institutions' shares of publication points (38%), external research funding (47.5%), PhD graduations (9.5%) and participation in humanities research essential to sustain the Estonian language and culture (5%).

Publication points are allocated as follows. (There is no fractional counting.)

- An article which is included in the Web of Science, Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index and/or Scopus (except books) – one point
- Article with more than 100 authors, 0.5 points.
- Article with more than 1000 authors, 0.3 points
- Monograph published by the editorial office named in the list published in Estonian Research Information system and solely prepared by the author – 5 points.
- Monograph published by the editorial office named in the list published in Estonian Research Information system where one chapter is prepared by author – 1 point

- Monograph published by the editorial office named in the list published in Estonian Research Information system and 2 or more chapters but not the whole monograph are prepared by the author – 2 points
- Number of patents and patent applications: Registered patent application 2 points and patent 3 points

The funding formula for publications totals the points achieved per institution and allocates 38% of the funding budget pro rata. That for external contact research funding excludes regular state funding instruments but allocates 47.5% of the budget pro rata each institution's share of the external research funding received by all the institutions. That for PhD graduations allocates 9.5% of the budget pro rata the institution's share of total Estonian PhD graduations. The humanities funding is allocated pro rata universities' share of national humanities research income.

Finland. All institutional funding in Finland is allocated based on results, using 3-year averages. It effectively rewards PhD graduations, the achievement of publication points, the volume of external research funding obtained and the proportion of faculty from abroad.

Publication points are obtained from

- Refereed scientific publications
  - Articles, monographs, conference proceedings, calculated using the publication forum rating by the Federation of Finnish Learned Societies. The classification has three levels: 1 = basic; 2 = leading; 3 = top.
  - Other publication channels which have not received level 1 status are allocated to level 0
  - The translation of these levels into points is non-linear
    - Level 0 scores 0.1 points
    - Level 1 scores 1 point
    - Level 2 scores 3 points
    - Level 3 scores 4 points
  - Monographs score an additional 4 points
- Other publications: non-refereed articles, books, and conference proceedings, professional publications and publications intended for the general public score 0,1 points

The funding formulae for an institution are

- Share of national PhD graduations * 54.2% of the PRFS budget
- Share of national publication points * 37.5% of the PRFS budget
- Share of international external research income * 12.5% of the PRFS budget
- Share of national external research income * 25% of the PRFS budget
- Share of national total of foreign researchers * 8.3%% of the PRFS budget

Italy. The VQR PRFS considers the quality of institutions' published outputs, an assessment of the quality of research outputs by recently-hired personnel, institutions' shares of the national total of external research funding, their share of the national total of certain kinds of personnel. It is a component in a larger indicator system that also allocated performance-based funding based on education results.

The system uses 16 different scientific fields. Quality assessment is done using the criteria: originality; methodological rigour; and actual or potential impact on the relevant scientific community.

Evaluation is based on a system of "informed peer review". In STEM areas and, to some extent, in Economics and Statistics, peer evaluation is informed by bibliometric indicators concerning citations and journals' impact, extracted from the ISI/Web of Science and Scopus databases. In SSH (with the partial exception of economics and statistics), evaluation is based purely on peer review.

Where bibliometrics is used, for each field a journal impact factor and the number of citations are obtained for world publications and the articles ranked on both dimensions. In many panels, JIFs and citations are given equal importance. In some, however, one may be weighted more highly than the other in generating the ranking. Papers submitted to the VQF are then identified within the set of world publications and allocated a grade according to where in the world ranking they are positioned.

- Excellent (in the top 10%)
- Good (11%-30%)
- Fair (31%-50%)
- Acceptable (51%-80%)
- Limited (81%-100%)

Outliers (15.8% of cases; eg high JIF but few citations, or vice versa) are sent for peer review. Experts review also the bibliometrics scores and override them in 5.3% of cases. Where peer review is used, the panels allocate publications directly to the same 5 categories.

Publications in these categories obtain the following numbers of publication points.

- Excellent – 1.0
- Good – 0.7
- Fair – 0.4
- Acceptable – 0.1
- Limited – 0

The funding formulae are

- For the overall assessment of output quality, share of national publication points in each of the 16 fields * weight of each field * 64.0% of the PRFS budget
- For the assessment of outputs from recently hired staff, share of national publication points in each of the 16 fields * weight of each field * 25.0% of the PRFS budget
- For external funding, share of state external research funding income * 5.5% of the PRFS budget
- For the share of PhD students, medical students, research fellows and post-docs, share of the national total * 5.5% of the PRFS budget

Norway. The Norwegian PRFS rewards institutions for their shares of publication points, EU funding, national research grant funding and contract research income. Publication points are awarded as follows.

	Level 1 (normal)	Level 2 (20 percent)
Article in ISSN title	1	3
Article in ISBN title	0,7	1
Book (ISBN title)	5	8

All publications have to be registered in the national research information system CRIStin. National committees appointed by the association of universities decide which publications belong to the higher level (Level 2), which may make up no more than the 20% best publications in the respective field.

Publication points are multiplied by 1.3, where the work is an international co-publication. Points are fractionalised for multi-institution publications by multiplying by the square root of the institution's proportion of institutional participations. For example, if three institutions contribute to a publication, each receives the number of points allocated to the publication multiplied by the square root of one third. Calculations are done annually on the previous year's data.

The funding formulae are

- For the overall assessment of output quality, share of national publication points * 33% of the PRFS budget
- For EU external income, share of national EU research income * 30% of the PRFS budget
- For national income, share of Norwegian universities' total research income from RCN or the RFF * 30% of the PRFS budget
- For contract research income, share of Norwegian universities' total contract research income * 30% of the PRFS budget

Performance-based institutional research funding was worth the following to institutions in 2017.

- NOK 25,550 per publication point
- NOK 1.374 per krone of EU funding received
- NOK 0.12 per krone of RCN or RFF funding received
- NOK 0.102 per krone of contract research income

Sweden. The Swedish PRFS has two components: bibliometric and external-funding-based. The bibliometric component is based on a calculation for each university in each of 34 fields of field-normalised publication production multiplied by a field-normalised citation score. The sum of these calculations generates a score that determines the university's share of the PRFS funding pot.

Field-adjusted production is calculated by normalising the number of publications a university makes per field by a 'field factor' that reflects the propensity to publish in each field. If more than one institution is

involved, the number of publications is fractionalised so as to share the credit among the contributing organisations. Credit is only given for first authors and (where applicable) corresponding authors.

The field normalised citation score is the actual number of citations achieved by the university's publications in a four-year window divided by the world average number of citations in the respective field.

The funding-based component reflects the university's share of all Swedish universities' competitive grant research income.

An additional adjustment to both categories is made, to reflect different propensities to publish and to obtain external income among fields. For universities with large humanities and social sciences (SSH) departments, their bibliometric scores and grants are weighted by 2.0. Natural sciences are weighted 1.5 and medicine and engineering 1.0. Other fields are weighted by 1.1. (S Carlsson, Allocation of Research Funds Using Bibliometric Indicators – Asset and Challenge to Swedish Higher Education Sector, InfoTrend 64, 2009(4), 82-88)

10 APPENDIX C: EXAMPLES OF SOLUTIONS IN FIVE COUNTRIES

This Appendix covers the different motivations, designs, implementations and experiences with the use of bibliometrics for PRFS that we find at the national level in five of the countries participating in the MLE. They have been chosen as examples to illustrate the ten typical considerations presented in section 4.1.2. Four of the countries, Croatia, Estonia, Norway and Sweden, are examples of metrics-based PRFS with different designs. The fifth country, Italy, has a PRFS based on informed peer review within a national research assessment exercise.

10.1.1 Croatia

Background and motivation

In 2012, the Ministry of Science and Education of the Republic of Croatia started negotiations with Public Research Organisations (PROs) - which then included 7 public universities (now 8) and 25 public research institutes - about the development of a new model of multiannual funding based on performance indicators. The idea was to reduce state intervention in the decision-making processes of PROs and to shift the responsibilities for strategic planning and financing research activities to the PROs. The new model would also enable systematic monitoring of scientific activities and an allocation of financial resources which would be in accordance with the achieved results of the PROs. Performance indicators provided by a more transparent and merit-based system were agreed upon. The first PRFS was adopted on 6 June 2013 and used for funding in 2013-2015 (three year period).

The same model was used both in 2016 and 2017. Currently there is a discussion how to proceed with the model in the following years. The idea is to modify some indicators in order to address the impact of research on society and economic challenges as well as the research environment.

It is important to note that the Croatian PRFS reallocates only a small portion - EUR 6,6 million annually - of the resources made available for research. It is only meant to provide stable funding for regular research activities at the institutional level. Other direct state funding is provided for expenditures for employees' (salaries), the material costs of ongoing activities of a research organisation. In addition, competitive projects are funded by the Croatian Science Foundation.

The intention of the PRFS has been to focus on a small number of measurable performance indicators and use them to keep track of the development of certain research organisations and, based on the results achieved, continue to further finance research activities to a greater or lesser extent.

The PRFS model combines four sets of performance indicators and takes into account the specificity of each field of science and different types of PROs (with shares in the funding model):

- scientific activity - 60%
- national and international competitive research projects and research mobility - 25%
- collaboration between research and business sector, as well as collaboration with the units of local and regional governance and non-governmental sector - 10%
- popularisation of science - 5%

Taking into consideration varying research costs among fields of science, the following weights have been agreed upon on the national level: Natural Sciences = 2.7; Technical Sciences = 2.5; Biomedicine and Health Sciences = 2.7; Biotechnology Sciences = 2.6; Social Sciences =1.2; Humanities = 1.2; Arts = 1.2.

The funding formula starts by calculating the base amount per full time equivalent (FTE) of an employed researcher. The base amount per FTE is calculated on the basis of the total number of FTE researchers, weights for the fields of science and the total amount of funds foreseen in the State Budget for PRFS.

The amount for an institution is then calculated on the basis of the science performance indicators and the base amount per FTE. It takes into consideration the size of institution and its field of science.[13]

In this chapter on bibliometrics, we mainly consider the first set of indicators (scientific activity) mentioned above.

[13] For example, if the base amount is 100 HRK, a researcher in the biomedicine will receive 100 x 2.6= 260 HRK and a researcher in humanities will receive 100*1.2= 120 HRK. The funding formula takes into consideration science performance indicators, so depending on the performance, the institution receives more or less money per researcher (if the institution in humanities performs good they can receive much more than 120 HRK per researcher, or if it performs bad, less than that).

Data sources and indicators

The Croatian model uses a combination of bibliometric indicators that depends on the field of science. Overall weight of this indicator called Science activity is 60%.

Each institution gathers the data from all faculties or departments. The institution annually submits integrated lists for the indicators which are divided by the three main groups of field of science (I: natural sciences, technical sciences, biomedicine, biotechnology; II: Social sciences and humanities, and III: arts). For group I, only publications in Web of Science are included. Group II may include publications in Scopus plus selected publications according to according to the national Regulations on Selection Procedures for Academic Ranks. Regarding group III, see below.

Only original research publications should be selected from any of the approved data sources. Research organisations are given clear instructions on how to collect data on papers published in Web of Science and Scopus and how to count their citations. Citations in a year are counted regardless of the year when the cited paper was published.

Data is collected per research organisation and not per scientist, with the published address of affiliations as criterion. Research organisations are registered according to their field of science. All papers published at a research organisation are taken into account for the field of science the research organisation is registered in. Papers must be de-duplicated at the institutional level. Each list must be signed by the rector of the university or the director of the public research institute and made publicly available on the institution's website.

An agreement has been reached that Web of Science defines the eligible scientific output in the STEM fields. Two indicators are given the following weights, as shown in Table 19:

Table 19 Scientific activity indicators for STEM fields

Indicator	Data Source	Weight
Journal article	Web of Science	0.5
Citations	Web of Science	0.1

The indicators in the social sciences and humanities are based on a wider range of data sources and publication types (Table 20):

Table 20 Scientific activity indicators for SSH fields

Indicator	Data Source	Weight
Journal article	Web of Science	0.12
Journal article	Scopus	0.10
Research article	Regulations on Selection Procedures for Academic Ranks	0.08
International book[14]	Public Research Organisation (PRO)	0.08
Book	PRO	0.06
Edited book (with article)	PRO	0.03
Textbooks, etc.	PRO	0.03
Citations	Web of Science	0.06
Citations	Scopus	0.04

There is no list of international and other publishers that define the book indicators. However, it is clearly indicated in the instructions that the books should be published with leading international publishers and in one of the major world languages.

In addition to the indicators based on scientific and scholarly publications, the Croatian model has an indicator set for scientific activity in *the arts*. It includes indicators and weights for artistic production (works, performances, exhibitions), nominations and awards, artistic research

[14] In international languages and published by leading international publishers.

projects, and contributions to congresses, evaluation committees or festivals. This part of the Croatian model is unique in the context of European indicator-based PRFS. Some other countries (e.g. Flanders/Belgium and Norway) have tried, unsuccessfully so far, to create indicators for artistic research.

The set of indicators related to popularisation (weighted 5 per cent) is unique as well in metrics-based systems. Systems based on peer review, however, often allow outputs from artistic research or popularisation to be submitted.

Effects of the bibliometrics for the PRFS

The effects of the Croation solutions for PRFS has not been studied, but it is currently being evaluated for a possible redesign.

Documentation, references and links

Domagoj Karacic, D., Miskulin, I., Serdarusic, H. (2016.) State investment in science and scientific productivity of universities. UTMS Journal of Economics, 7(1): 37–48.

http://public.mzos.hr/Default.aspx?sec=3521

http://narodne-novine.nn.hr/clanci/sluzbeni/2013_06_69_1367.html

https://www.google.hr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0ahUKEwiD9f-KxtPTAhWFAxoKHThsC5oQFgg5MAM&url=https%3A%2F%2Frio.jrc.ec.europa.eu%2Fen%2Ffile%2F9514%2Fdownload%3Ftoken%3D-8JG6aKx&usg=AFQjCNH8uBpJ9zvMO9L5xPwR4K-S170ZnQ&sig2=SCFU_zyJeRCdgvoeiKeXJQ

http://public.mzos.hr/fgs.axd?id=24811

10.1.2 Estonia

Background and motivation

The funding of research in Estonia is mainly based on competitive grants from the Estonian Research Council and from similar sources abroad, and from contract research. However, in 2005, Estonia introduced a direct institutional funding for research which is called the baseline funding. It

represented 13 per cent of the total R&D funding in the beginning, but the share has been increasing up to 30 per cent lately.

All kinds of research organisations may apply for the baseline funding, but they have to pass an institutional evaluation every seven years to be eligible. The evaluation is organised at the national level and based on peer review by international experts.

The annual baseline funding itself is almost completely determined by performance indicators. The exception is 5 per cent of the budget, which is allocated to support research that is particularly relevant for the Estonian language, culture, history and society. The large part is allocated on the basis of three types of performance indicators:

1. 50 per cent (40 per cent until 2016) is allocated according to indicators of external revenues, in practice mainly from contract research, but also from competitive grants.
2. 40 per cent (50 per cent until 2016) is allocated according to bibliometrics and indicators of patenting.
3. 10 per cent is allocated according to the number of doctoral graduates.

With regard to the first type, the indicators of external revenues are carefully defined to represent only so-called R&D services for the private or public sector. They involve only research activities, not other services such as consulting, commissioned by the third party and for the benefit of the third party, not of the research organisation itself. Funding from EU framework programmes is also included as evidence of quality and cooperation with high level partners.

With regard to the second type, registered patent applications are weighted 2, and registered patents are weighted 3, compared to articles in international journals, which are normally weighted 1. In the following, we will concentrate on the bibliometric indicators.

Data sources and indicators

The data sources for the bibliometric indicators are Web of Science and (more recently) Scopus. In addition, data from scholarly book publishing is selected from the Estonian Research Information System (ETIS).

Only publications count; citation indicators are not used. Publications are only attributed to the institutions if they are mentioned as an affiliation in the publication. Publications that can be attributed to more than one

institution will count once for each of them. There is, however, an element of fractionalisation in that publications with more than 100 authors will be given only half of the weight, while publications with more than 1,000 authors will be given a third of the weight. Data from the preceding three years is used every year to calculate the indicators for next year's budget. The indicators are the same for all fields of research. They are shown in Table 21.

Table 21 Estonia: Bibliometric indicators and their weights

Indicator	Data Source	Weight
Journal article	Web of Science or Scopus	1
Monograph	ETIS	5
Article in book	ETIS	1
Two or more articles in book	ETIS	2

Books will be included only if they are published by an international scholarly publisher that is listed on a specific list in ETIS (e.g. Cambridge University Press or American Mathematical Society): https://www.etis.ee/Portal/Classifiers/PublishingHouseDetails?lang=ENG

ETIS is an information system that covers a wider range of research outputs than those included in the bibliometric indicators for the baseline funding. ETIS will for example record scholarly books or journal articles in the Estonian language. The reason for limiting the bibliometric indicators to internationally indexed journal articles or internationally published books is a concern about data quality and research quality. One should bear in mind, however, that research that is particularly relevant for the Estonian language, culture, history and society is funded in a separate part of the baseline funding.

Effects of the bibliometrics for the PRFS

There has been no direct study of the effects of the PRFS in Estonia, but the impression is that it has influenced an increased focus on performance at the universities. The allocation of funds has changed over time, also indicating that is has an effect. Several other incentives are connected to

publication performance – the effects of the PRFS are difficult to isolate. But there has been an increase in the interest of patenting at the universities.

Documentation, references and links

To some extent, The Research and Innovation Policy Monitoring Programme (TIPS Programme), commissioned by the Estonian Ministry of Education and Research, has relevant information in English:

http://www.tips.ut.ee/index.php?module=2&op=&dok_id=735
http://www.tips.ut.ee/public_funding

10.1.3 Norway

Background and motivation

When a PRFS for the Norwegian higher education institutions was first introduced in 2002, it covered both research and educational activities with a relatively light touch. At that time, 40 per cent of the total direct funding was reallocated according to performance indicators with 25 per cent representing education and 15 per cent representing research. Since all performance indicators were size-dependent, the distribution of funds across institutions did not change much after the implementation. Although the general aim of the PRFS was to increase the quality of all activities, the major immediate effect was to make the funding allocation more *explicit* and to provide both the funding and the funded organisations with a statistical overview of the activities.

The performance indicators representing research were indicators of external competitive funding, completed doctoral degrees and the number of FTE in academic positions. Since the latter indicator was partly dependent on educational activities and not a direct indicator of research activity, the universities and the government agreed that it needed to be replaced by research output indicators. The idea of using research evaluation with peer review for the same purpose was not discussed. For advice and formative purposes, Norway had already implemented a national research assessment system based on peer review by international panels, as it was already decided that this system should not be used for direct institutional funding.

Supporting the choice of direct use of bibliometrics for the PRFS was also the fact that the institutions for several years already had been reporting statistics from metadata representing scholarly publications in their local CRIS.. Since the institutions could see that international databases such

as Web of Science only partly covered their output with wide differences among their faculties, the ambition became to improve the data quality in the CRIS systems and develop a bibliometric indicator with a balanced representation of all fields. The Ministry of Education and Research asked the funded organisations, represented by the Norwegian Association of Higher Education Institutions, to realise this ambition. The result was the so-called 'Norwegian model', which was developed in 2003-2004 and implemented in 2005 in the budgets for 2006.

Data sources, indicators and overall design

After the implementation of the so-called 'Norwegian Model' (Ahlgren et al., 2012; Schneider, 2009; Sivertsen, 2016) in 2005, it has later on also been adopted at the national level by Denmark (2009) and Finland (2015), partly also by Flanders, Belgium (2009) and by several Swedish universities at the local level. It has three components:

1. A complete representation in a national database of structured, verifiable and validated bibliographical records of the peer-reviewed scholarly literature in all areas of research;
2. A publication indicator with a system of weights that makes field-specific publishing traditions comparable across fields in the measurement of 'publication points' at the level of institutions;
3. A performance-based funding model which reallocates a small proportion of the annual direct institutional funding according the institutions' shares in the total of publication points.

In principle, component C is not necessary to establish components A and B. The experience is, however, that the funding models in C support the need for completeness and validation of the bibliographic data in component A. We will return to the Norwegian variant of component C in the next section.

With regard to component A, the local CRIS systems were united to one integrated national system, CRIStin, in 2010. Around 160 institutions from the higher education sector, the independent institute sector and the hospital sector are participating. The publication indicator now serves PRFS for institutional funding of research in all three sectors. CRIStin covers all kinds of research activities and outputs and is running for several other purposes than the PRFS, such as CV's, applications to research councils, evaluations, annual reports, internal administration, bibliographies for Open Archives, links to full text, etc. The data for the publication indicator and the purpose of the PRFS is delimited by a definition of scholarly and scientific publishing. According to this definition, a scholarly publication must:

1. present new insight
2. in a scholarly format that allows the research findings to be verified and/or used in new research activity
3. in a language and with a distribution that makes the publication accessible for a relevant audience of researchers
4. in a publication channel (journal, series, book publisher) which represents authors from several institutions and organises independent peer review of manuscripts before publication.

While the first two requirements of the definition demand originality and scholarly format in the publication itself, the third and fourth requirement are supported by a dynamic register of approved scholarly publication channels. Suggestions for additions can be made online at any time through the same web page.

Component B is the publication indicator itself. In the measurement for the funding formula by the end of each year, are weighted as they are counted. The intention is to balance between field specific publishing patterns, thereby making the publication output comparable across research areas and institutions that may have different research profiles. In one dimension, three main publication types are given different weights: articles in journals and series (ISSN), articles in books (ISBN) and books (ISBN). In another dimension, publication channels are divided into two levels in order to stimulate publishing in the most prestigious and demanding publication channels within each field of research. The highest level is named 'Level 2'. It includes only the leading and most selective international journals, series and book publishers. There is also a quantitative restriction, since the publication channels selected for Level 2 can only in total represent up to 20% of the world's publications in each field. The weighting of publications by type and channel is shown in the table below.

	Level 1 (normal)	Level 2 (20 percent)
Article in ISSN title	1	3
Article in ISBN title	0,7	1
Book (ISBN title)	5	8

Publication points are measured at the level of institutions, not at the level of individual researchers. Publications with multiple authors representing several institutions are counted only once. Their points are divided between the institutions by multiplying the points with the square root of the institution's proportion of the authors. This is an intermediate solution between so-called fractional counts and whole counts that has been simulated to balance well between the types of productivity in different fields and at the same time incentivise collaboration. As an extra incentive to international collaboration, a factor of 1,3 is used in the multiplication of points for publications with co-authors in other countries.

The list of journals, series and book publishers on 'Level 2' is revised annually in collaboration with national councils in each discipline or field of research. These councils propose changes to an interdisciplinary National Publishing Board, which governs the process on behalf of all institutions and has the final decision. Bibliometric statistics (world production versus national production in channels on both levels, and citation statistics for publication channels) are used as an aid in this process, but not as criteria in themselves.

Bibliometrics as a component in the funding system

Norway is generally characterised by generous direct funding of higher education institutions as compared to external competitive funding from research councils and other sources. In addition, the PRFS only represents 8 per cent of the total of direct institutional funding, as seen in tables Table 17 and Table 18 in section 5.1. The publication indicator is one of four indicators in the PRFS. If measured as **a share of the total direct institutional funding** (excluding competitive external funding), the publication indicator only reallocates 1,6 per cent of the total budget.

One publication point now represents around EUR 2,500, a little more than half of what it represented in 2005. Since then, the publication activity has almost doubled while the economic basis for reallocation by the indicator has remained almost the same.

Effects of the bibliometrics for the PRFS

The Norwegian model has been extensively documented and evaluated, also in the international journal literature. Initiated in 2012 by the Norwegian Association of Higher Education Institutions (representing the funded organisations) in collaboration with the Ministry of Education and Research (the funding organisation), the bibliometrics for the funding model was evaluated extensively in Norway in 2013. An independent Danish team of researchers studied its design, organisation, effects, and

legitimacy (Aagaard et al. 2014). As well as advising improvement and further development, the exercise provided the basis for four in-depth studies of internationally relevant questions (Aagaard 2015; Aagaard et al 2015; Schneider et al. 2015; Bloch & Schneider 2016).

The overall results of the evaluation showed an increase in productivity beyond what could be expected from increases in resources, particularly in fields and at institutions that had been less active in research before. However, the evaluation pointed at three major areas for improvement. Since 2014, the funded and funding organisations have collaborated on following up the evaluation in order to improve the model and its practices in these three areas.

One of them was the design of the indicator itself. The former indicator had a possible disincentive to collaboration (Bloch & Schneider 2016) and an imbalance in the representation of research fields (Aagaard et al 2015) which have been solved by a redesign of the indicator (the new indicator is presented above).

In 2016, the Ministry of Education and Research asked for a further redesign of the publication indicator with which it can include a citation indicator. There has been a consultation on the proposal for a solution. The results seem to imply that the publication indicator will continue as it is, and that the added citation indicator will not be used in the PRFS, but for analytical purposes only.

A second area for improvement was the fact that the indicator is also used locally (i.e. within the institutions themselves), in some contexts for purposes where it is not appropriate and can do harm (Aagaard 2015). This has been followed up establishing inter-institutional learning arenas for proper managerial use of the indicator and by agreeing on national recommendations for good conduct on the local level.

The third problem was partial lack of transparency of the process for the definition of the high-level publication channels (Level 2). This has been remedied by introducing a new interactive national portal for the indicator where all members of the committees are listed and where they explain their decisions and respond to comments and suggestions. The portal has a bilingual interface with information also in English: https://npi.nsd.no/.

Documentation, references and links

The design of Norwegian model in 2004 and its evaluation in 2014 is documented in two publications from the Norwegian Association of Higher

Education Institutions that can be downloaded at the portal of the Norwegian Publication Indicator.

The design: *A Bibliometric Model for Performance-based Budgeting of Research Institutions* (2004, in English translation): https://npi.nsd.no/dok/Vekt_pa_forskning_2004_in_english.pdf

The evaluation: Evaluation of the Norwegian publication indicator (2014, summary in English of a report in Danish):

https://npi.nsd.no/dok/eval2014/Evaluation_of_the_Norwegian_Publication_Indicator_2014_English_Summary.pdf

Updated information in English is available at the same portal:

https://npi.nsd.no/informasjon#dokumenter

Other relevant publications available in English are:

Aagaard K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. Science and Public Policy 42(5): 725-737.

Aagaard, K, Bloch CW and Schneider JW (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. Research Evaluation 24(2): 106-117.

Ahlgren P, Colliander C and Persson O (2012). Field normalised citation rates, field normalised journal impact and Norwegian weights for allocation of university research funds. Scientometrics 92(3): 767-780.

Bloch C and Schneider JW (2016). Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication level at the individual level. Research Evaluation 25(4): 371-382.

Schneider JW (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. European Political Science 8(3): 364-378.

Schneider JW, Aagaard K and Bloch CW (2015). What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. Research Evaluation. DOI: 10.1093/reseval/rvv036

Sivertsen G (2016). Publication-Based Funding: The Norwegian Model. In: Ochsner M, Hug SE, Daniel HD (eds). Research Assessment in the Humanities: Towards Criteria and Procedures. Springer Open: Zürich, 79-90. https://link.springer.com/chapter/10.1007%2F978-3-319-29016-4_7

10.1.4 Sweden

Background and motivation

Sweden's PRFS is used for distribution of increased block grants between universities and university colleges. It has also been used for redistribution of between 10 and 20 per cent of the direct funding of all higher education institutions. It was implemented in 2009 on the basis of a commissioned report to the government in 2007 on institutional funding for research (SOU, 2007). The general aim of the new partly performance-based funding model was to enhance research quality while at the same time increasing the transparency of the funding allocation and the relative autonomy of the institutions in their own resource allocation. The report proposed several performance indicators among which two were implemented, one based on bibliometrics and another based on external funding. Only the biblometric indicator will be considered here.

Data sources and indicator design

Sweden's solution to bibliometrics for a PRFS is unique in Europe in the sense that the data for the indicator are not produced in collaboration with the institutions, but instead derived from (and defined by) only one specific commercial data source, Web of Science (WoS, from Clarivate Analytics). The main argument for using WoS was to give incentives to increase quality and internationalisation (Sandström & Sandström, 2008 and 2009). The bibliometric model is operated on behalf of the government by the Swedish Research Council, using an annually updated replication database of WoS. The Swedish Research Council refines the data (disambiguation of institutional addresses, field classification, citation links) for the purpose of the PRFS and their own analytical needs.

The design of the bibliometric model for the PRFS was first proposed in an appendix to the above mentioned report (SOU, 2007) and then further developed by the same authors (Sandström & Sandström, 2008 and 2009). A short documentation in Swedish of the model as it actually works is available at the Swedish Research Council (Vetenskapsrådet, 2014).

The bibliometric indicator for the PRFS combines productivity and citation impact by using the number of publications and their received citations. Both are field-normalised and fractionalised according to the institution's contribution of authors to the article. The field normalisation of citations follows an advanced bibliometric procedure where actual citations are compared to the world average within the same field, year and publication type in the database.

As for the field normalisation of productivity (number of articles), there is no available standard bibliometric procedure since data representing the avaiable resources for research will be needed in addition to the bibliometric data. The solution in Sweden is to calculate a field factor, using the so-called Waring method, that estimates the average production of a researcher in each of the 34 subject fields of the model (Sandström and Sandström, 2009). The field factor takes into consideration the average productivity of Nordic or Swedish authors within the field, based on publications in the WoS database, while the Wahring method is a mathematical solution in which the same data is used to estimate the number of authors in the fields with publications not covered by the database. The latter estimates are particularly needed in the social sciences and humanities, where the output is insufficiently covered by the WoS.

An alternative to the Waring method (with special treatment of the social sciences and the humanities) was initially proposed by the Swedish Research Council (Vetenskapsrådet, 2009), but has not been implemented, in spite of the Waring method being documented to suffer from flaws in the context of the Swedish PRFS (Vetenskapsrådet, 2010).

An initial simulation of the effects using WoS compared to the comprehensive coverage of scholarly publications by the Norwegian model demonstrated that the reallocation of funds would be much the same (Sandström & Sandström, 2009). A change of model would mainly affect universities with medical faculties since medicine is relatively well represented with publications and citations in WoS.

Effects of the bibliometrics for the PRFS

The effects on the research system of the Swedish bibliometric model for the PRFS have not been studied, and the model itself has not been evaluated so far. It is generally considered a 'black box' with limited financial consequences for the institutions. The present government has even reduced its redistribution effects. However, using bibliometrics in the national PRFS has increased the local use of bibliometric information at the institutions, and a bibliometric component has been introduced in

many places for internal redistribution of resources. In a recent study, the institutions were asked whether they use the Swedish or the Norwegian model for local purposes. Most of them seem to prefer the latter (Hammarfelt et al., 2016).

A commission appointed by the government will start working in May 2017 with one of its aims to design the next generation of PRFS in Sweden.

Dismissed proposals for alternatives

Alternatives to the present bibliometric model have been discussed at two occasions in Sweden. In a commissioned report to the government, Anders Flodström (2011) proposed that the bibliometric indicator should be based on complete data produced in collaboration with the institutions. This solution would have been similar to the one initially proposed by the Research Council (Vetenskapsrådet, 2009) and to the parallel bibliometric indicators for PRFS in the other Nordic countries. Instead of following the proposal, the government assigned the Research Council to design a PRFS model based on peer review, inspired by the Research Excellence Framework in the United Kingdom. However, after the assignment resulted in a solution, the so-called FOKUS model (Swedish Research Council, 2015), the government decided not to implement it.

In 2016 the Research Council was assigned to suggest changes to improve the model (Vetenskapsrådet 2016) but these changes have not been implemented.

Hence, the WoS-based bibliometric indicator is still in place. Presently, the government is considering to extend the two existing indicators in the PRFS (WoS bibliometrics, external funding) with a third indicator representing (evaluation of) third stream activities (VINNOVA, 2016). This indicator would constitute the realisation of ideas presented also at the earlier stages (SOU, 2007; Flodström, 2011; Swedish Research Council, 2015).

Documentation, references and links

The design of the Swedish bibliometric model is well documented in English in Sandström & Sandström (2009). A short documentation of how it actually runs is found in Vetenskapsrådet (2014). All references and links are given here:

Flodström, A. Prestationsbaserad resurstilldelning för universitet og högskolor. Rapport till Utbildningsdepartementet (U2011/7356/UH).

Available at: http://www.ksla.se/wp-content/uploads/2012/03/7356-Rapport-Prestationsbaserad-resurstilldelning-f%C3%B6r-universitet-och-h%C3%B6gskolor.pdf

Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. Research Evaluation 25(3): 292-305.

Sandström, U. & Sandsström, E. (2008). Resurser för citeringar. Rapport 2008:18 R, Högskoleverket.

Sandström, U. & Sandsström, E. (2009). The field factor: towards a metric for academic institutions. Research Evaluation, 18(3): 243-250.

SOU. (2007). Resurser för kvalitet (Resources for quality in research). SOU 2007:81. Available at: http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2007/11/sou-200781/

Swedish Research Council (2015). Research Quality Evaluation in Sweden - Fokus: Report of a Government Commission regarding a Model for Resource Allocation to Universities and University Colleges Involving Peer Review of the Quality and Relevance of Research. Available at: https://publikationer.vr.se/wp-content/uploads/2015/05/VR1545_FOKUS_ENG_WEBB.pdf

Vetenskapsrådet. (2009). Bibliometrisk indikator som underlag för medelsfördelning. Available at: http://www.vr.se/download/18.72e6b52e1211cd0bba8800010145/bibliometrisk_indikator.pdf

Vetenskapsrådet. (2010). Kan man använda Waringmetoden för att uppskatta antalet forskare? Available at: https://publikationer.vr.se/produkt/kan-man-anvanda-waringmetoden-for-att-uppskatta-antalet-forskare/

Vetenskapsrådet. (2014). Bibliometriskt underlag för medelsfördelning. Available at: https://www.vr.se/download/18.5a947f0d145b21c1709f9d/1399897207262/Bibliometriskt+underlag+f%C3%B6r+medelsf%C3%B6rdelning_140512.pdf

Vetenskapsrådet. (2016). Regeringsuppdrag att justera indikatorerna vetenskaplig produktion och citeringar i modellen för resursfördelning.

Available at: https://www.vr.se/download/18.73bacb0c15490b79b0ac5eaf/1462782033410/Svar+på+regeringsuppdrag+U2016-01350-F.pdf

Vinnova. 2016. Evaluating the role of HEI's interaction with surrounding society. VR 2016:09. Available at: http://vinnova.se/en/Publications-and-events/Publications/Products/Evaluating-the-Role-of-HEIs-Interaction-with-Surrounding-Society/

10.1.5 Italy

Background and motivation

Italy is characterised by a central government university funding system that relies mainly on a single grant, the Fondo di Finanziamento Ordinario (Ordinary Financing Fund, FFO), for teaching, research, and other infrastructural needs. Non-government sources such as student fees and contractual funding are becoming more important, but account for only around a quarter of the financial resources. The FFO is allocated according to a mixed model in which the largest part (56% in 2017) is based on historical data; this part it has been steadily decreasing in recent years and it is planned to reduce further in the future. A smaller part is based on a formula-based adjustment that takes into account performance indicators for teaching and research. Other components have specific objectives and formulae.

Most of the formula-based adjustment of the FFO for research is determined by a recurring research assessment exercise based on peer review of submitted publications. This part is inspired by the Research Assessment Exercise in the UK. Italy introduced it in 2006 as the Valutazione Triennale della Ricerca (VTR), which looked back at the period 2001-2003. The outcome was used by the Government since 2009 to allocate a small but growing share of public funding, starting from 2.2% of total funding in 2009 and reaching 16% in 2014. The second assessment exercise, (VQR, Evaluation of research quality), looked back at the years 2004-2010 and was published in July 2013 by a new independent agency (Agenzia Nazionale per la Valutazione dell'Università e della Ricerca; ANVUR). The third research assessment exercise (VQR 2011-14) referred to the period 2011-14. It was published in February 2017 by ANVUR. The two most recent exercises had a similar framework.

Submitted output for peer review

Italy has a central online system for submitting and archiving research output for the evaluation exercise. It is administered by ANVUR in collaboration with the universities and with technical support from Cineca.

To take the example of the third round of evaluation, the VQR evaluated the research outputs of all permanent scientific staff on government contracts in 96 universities and 39 research organisations. A total of 52,677 researchers submitted their best outputs (2 for each university researcher, and 3 for each scientist employed in a public research organisation). All in all, 118,036 outputs were submitted, out of which 70 per cent were journal articles and the remainder other types of research outputs such as books, book chapters, conference proceedings (with ISBN codes); critical editions, translations, scientific comments; patents; compositions, designs, performance, work of arts, and others.

The outputs are classified by their authors in 16 research areas. ANVUR appoints a panel of experts for each research area. In the humanities and social sciences, a pure peer review system is applied with the auxilium of external (national and international) reviewers (almost 13,000 reviewers were used in the third round).

In science, technology, engineering and medicine (STEM), the same procedure is used, but in addition, bibliometric indicators are also produced by ANVUR to inform the panels. These indicators are limited to the part of the outputs that are covered by Web of Science or Scopus.

In any case, according to the Ministerial decree, at least 50% of the publication submitted for evaluation should be peer-evaluated.

Consequently, the system for evaluating outputs for the Italian PRFS can be characterised as based on peer review, but the burden of large-scale reviewing is somewhat relieved and its quality improved by the support of a systematic, centralized and professional use of bibliometrics in the STEM fields

Use of bibliometrics in STEM fields

The researchers have to indicate if they prefer to be evaluated using either WoS or Scopus. Indicators considered are the number of citations; the 5-years Impact Factor and the Article Influence Score (WOS database); the Impact per publication and the Scimago Journal Rank (Scopus database). All of the indicators are normalised and adjusted

according to the state of the art in professional bibliometrics (for details, see Anfossi et al, 2016)

In principle, then, all journal articles covered by Scopus and Web of Sciences are initially assessed on the basis of variants of two indicators representing their citation frequency and the citation impact of the journal in which they are published. If the two indicators are broadly coherent, they are translated into the five classes of merit: Excellent; Good; Fair; Acceptable; Limited, taking into account the world distribution of the performance of articles in each specific sector. The result is validated by the experts in order to determine the final score attributed to the publication. If the two bibliometric indicators are clearly diverging, the article is evaluated with traditional peer review.

Metrics for third stream activities

ANVUR is developing, on an experimental basis, a new system of metrics designed to support evaluation of third stream activities and impact of academic research. Third stream activities have been divided in two main areas, respectively involving research economic valorisation and the production of public and social goods. As for research valorisation, indicators are produced concerning intellectual property management (patents and vegetal varieties), academic entrepreneurship (spin offs), third party activities, and intermediation activities. As for the production of public and social goods, ANVUR considers indicators concerning the management of cultural activities and the cultural heritage (museums, archeological excavations and cultural heritage), clinical trials, continuous education and public engagement. Evaluation is based on peer review, informed by the aforementioned information.

Effects of the bibliometrics for the PRFS

There has been no systematic evaluation of the Italian PRFS as such, but some studies have been published discussing possible effects or changes in the system. Two examples:

After the second round of evaluation (the VQR covering 2004-2010), Benedetto et al. (2016) found that the ratings crucially depend on language and typology of publication and on the methodology adopted for evaluation. Younger researchers and full professors tend to receive, ceteris paribus, a better evaluation. Outcomes submitted by researchers working in the Centre-North of the country usually obtain better evaluations than those in the South. Finally, public funding to university research is found to be correlated with positive ranking in the assessment exercise, especially in natural sciences and engineering, where there is

particular need of appropriate funding for conducting laboratories experiments and research.

In addition to these observations, Ancaiani et al. (2015) and Alfò et al (2017) found a significant degree of concordance among peer review and bibliometric evaluations, supporting the choice of using both techniques in order to assess the quality of Italian research institutions. They also found that, in general, the results of the VQR (second round) were already highly informative about the existing strength and weaknesses of the Italian university research system.

Finally, Checchi et al. (2017) find that the dispersion in research quality across universities has significantly fallen in the second exercise compared to the first. They also find that convergence is largely due to changes in the relative productivity of researchers who participated to both exercises and to the hiring decisions of universities. The speed of convergence falls instead when they include the changes due to researchers' retirement (an event which is almost entirely determined by age). These results suggest that convergence may reflect changes in the behaviour of individuals and institutions induced by the monetary and reputation incentives created by the PRFS.

Documentation, references and links

Alfò M., Benedetto S., Malgarini, M. Sarlo S., (2017), On the use of Bibliometric information for assessing articles quality: an analysis based on the third Italian evaluation exercise, paper presented at the STI 2017 Conference on "Open indicators, innovation, participation and actor based STI indicators", Paris, September 2017

ANVUR. (2013.) Rapporto sullo stato del sistema universitario e della ricerca. Available at: http://www.anvur.org/attachments/article/644/Rapporto%20ANVUR%202013_UNIVERSITA%20e%20RICERCA_integrale.pdf

ANVUR (2015.) Evaluation of Research Quality 2011-2014 (VQR 2011-2014). Call for participation. Available at: http://www.anvur.org/attachments/article/825/Call%20VQR_2011_2014_11nov_~.pdf

ANVUR (2016.) Rapporto sullo stato del sistema universitario e della ricerca. Available at: http://www.anvur.org/attachments/article/1045/ANVUR_Rapporto_INTEGRALE_~.pdf

Ancaiani, A., Anfossi, A.F., Barbara, A., Benedetto, S., Blasi, B., Carletti, V., Cicero, T., Ciolfi, A., Costa, F., Colizza, G., Costantini, M., di Cristina, F., Ferrara, A., Lacatena, R. M., Malgarini, M., Mazzotta, I., Nappi, C.A., Romagnosi, S., Sileoni, S. (2015.) Evaluating scientific research in Italy: The 2004-10 research evaluation exercise. Research Evaluation, 24(3): 242-255.

Anfossi, A., Ciolfi, A., Costa, F., Parisi, G., Benedetto, S. (2016.) Large-scale assessment of research outputs through a weighted combination of bibliometric indicators. Scientometrics, 107(2): 671–683.

Benedetto, S., Cicero, T., Malgarini, M. (2016). Determinants of research quality in Italian universities: Evidence from the 2004 to 2010 evaluation exercise. Research Evaluation, 25(3): 257–263.

Checchi, D., Mazzotta, I., Momigliano, S. (2017), Research assessment in Italy: have the results of universities converged?, paper presented at the STI 2017 Conference on "Open indicators, innovation, participation and actor based STI indicators", Paris, September 2017

Geuna, A., Piolatto, M. (2016). Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). Research Policy, 45(1): 260-271.

11 APPENDIX D: SUMMARY OF INFORMATION FROM PARTICIPATING COUNTRIES

Problems and policy purposes

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Increase scientific performance Improve quality and relevance Concentrate resources Increase competitive-ness 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 were granted increased autonomy	Need to modernise the research assessment and funding system Desire for 'objective' assessment process	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 the levels of fields and of programmes	Universities are free to perform institutional evaluation, using data from the PRFS aggregated to the 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	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

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
	organisations								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 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	Bibliometrics Counts of innovation outputs, eg prototypes (dropped from 2017) Peer review of a small number of submitted outputs	Peer review for eligibility: Scientific impact of research, sustainability and potential of research, and societal importance of research Quantitative formula for allocation includes data on Publications Patents External contractual 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 CRIS (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 competitive grant funding

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
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

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
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	Societal relevance: usefulness or need	Scientific impact of research, sustainability and potential of research, and societal importance of research	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 CRIS, 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	National CRIS, patents, self-evaluation report, site visits, data on R&D funding and revenue, infrastructure	WoS: 5-year impact factor and Article influence score Scopus: Impact per publication and Scimago journal ranking	National CRIS (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	Field weights	Eight fields with different weights Bibliometric	Bibliometric indicators are field-normalised Six field	Not considered	16 field panels	Field-normalised bibliometric indicators	Not handled before 2014. Now, publications in different fields attract	Field panels	Bibliometric indicator is field-normalised

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Humanities panel decides how to treat non-indexed materials		indicators are not field normalised	specific panels to peer-review selected outputs				varying numbers of points		

Experience with metrics

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Scholarly quality									
Indicators accepted		Bibliometrics Competitive grants	Bibliometric indicators used for scholarly outputs	National CRIS, patents, self-evaluation report, site visits, data on R&D funding and revenue, infrastructure	Citations and journal impact are used outside SSH to inform peer review		PRFS increased 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 numbers of non-scholarly outputs	National CRIS, patents, self-evaluation report, site visits, data on R&D funding and revenue,	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

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
				infrastructure					indicators will weigh 1/3 each
Prospective dimension?									
Included in assessment system	Performance contracts	No	No	National CRIS, patents, self-evaluation report, site visits, data on R&D funding and revenue, infrastructure		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% for calculation	Bibliometric indicators are field-normalised Non-scholarly outputs are peer-reviewed and scored Scoring of universities into four grades. Funding determined in	Publications and patents 38% External contractual research income 48% PhD graduations 9% Support to topics of national	The PRFS covers 17% of institutional funding for research That 17% is distributed: 64% on general evaluation; 25% on results from researchers		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	

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
			negotiation with provider	importance 5%	hired or promoted in the period; 11% on the basis of competitive funding and PhD students				
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 accreditation of PhD courses Universities ay use for internal allocation		Aggregate data are published annually and used as strategic intelligence by policymakers		
Periodicity									
Annual	3 years	Annual	Annual	Annual	5 years, starting in the period 2015-19		Annual	3-6 year intervals	Annual

Benefits and Costs

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Benefits									
Increased competition	Transparency		Initially performance	Results are difficult to	I Increased accountability,	Expect quality	Increased quantity of	Stimulated flexibility and re-	Increased quality focus

Armenia	Austria	Croatia	Czech Republic	Estonia	Italy	Moldova	Norway	Portugal	Sweden
Concentration of resources Increased efficiency in research production	Increased performance		improved	dispute	adoption of more performance-oriented policies, better mechanisms for the governance of universities; better recruiting; greater transparency in the local allocation of funding	improvement	research output with no reduction in quality	organisation Promoted internationalisation and improved quality	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		Cost was not an issue	Relative cost is lower than in other countries; see table 8 Some difficulties in data cleaning Sometimes hard to define the boundaries of contract research	A national CRIS is being developed for the next exercise		The main cost was developing the CRIStin CRIS Compliance costs for researchers Some money moved from established to newer institutions	Expensive but seen as fundament-al 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, eg PhD education

Table 22 Policy goals and assessment criteria in PRFS

	Policy goals	Assessment criteria
AT	Stability of institutional funding Increase performance	Research income from the Austrian Science Fund (FWF) and the EU
HR	Increase the overall quality of the research system and to create a more transparent and merit-based system. Enable the systematic monitoring of scientific activities and allocation of financial resources in accordance with the PROs achieved results through the implementation of agreed performance indicators	Scientific productivity National and international competitive research projects and research mobility Collaboration between research and business, government and NGOs Popularisation of science – 5%
CZ	Contribute to an increase in the excellence of Czech R&D system Design an R&D evaluation methodology with a pronounced formative function Provide strategic information to stakeholders in the R&D system Provide information to the performance-based funding system Take into account differences among different types of research organisations and research fields	Academic outputs Non-academic outputs 'Excellent' outputs (peer reviewed)
EN	Inform the selective allocation of research funding to HEIs Provide benchmarking information and establish reputational yardsticks Provide accountability for public investment in research and demonstrate its benefits	Quality (Originality, Significance, Rigour) Impact on society Research environment
EE	Accredit research-performing institutions to make them eligible for institutional research and Estonian Research Council funding Allocate institutional funding	Evaluation criteria • Scientific impact of research • Sustainability and potential of research • Societal importance of research Calculation of institutional funding Publications and patents

		Contractual and foreign income Number of PhDs graduated humanitarian research essential for sustainability of Estonian language, culture and sovereignty
FI	Improving the quality and internationalisation of Finnish higher education and university research	Publication channel of scientific outputs National and international research funding Proportion of research personnel from abroad
IT	Enhance research quality providing benchmarking information and incentives for Universities' strategic planning Informing the distribution of Italian institutional funding for research Informing the accreditation of graduate and Phd courses	Quality of submitted research outputs (originality, rigour, actual or potential impact) Quality of research outputs submitted by researchers recruited in 2011-14 (since the last VQR) Numbers of PhD students, medical students, research fellows and post-docs
NO	To stimulate increased research activity and to distribute resources to units achieving good research results	Publication points (product of a calculation based on type of publication, quality level and number of external authors) Income from Research Council Norway and Regional research funds (the latter of minor monetary importance) Income from the European Commission Income from other external funding sources both for research and education Doctoral degrees awarded
PT	In PT the PRFS is seen as a part of the National Science and Technology System (NSTS), and not as a way to tackle any specific problem. The evaluation is seen as an opportunity to increase the quality of the research system, and to increase the efficiency in the allocation of resources.	Quality, merit, relevance and internationalisation level of the R&D activity of the research unit in the last 5 years Scientific merit of the team of integrated researchers, evidences of national and international recognition and, when applicable, the technical, cultural or artistic merit in the team Objectives, strategy, organisation and

		activity plan for the next 5 years
SE	Enhance research quality while at the same time increasing the transparency of the funding allocation and the relative autonomy of the institutions in their own resource allocation	Publication quality (defined in bibliometric terms) Volume of competitive research income

Table 23 Croatia – indicators and formula used in PRFS

Analysis is based on the individual performance of each researcher based on the following criteria:

1. **scientific productivity** which includes the following sets of entry data:
   - number of research papers published in journals indexed in WoS and SCOPUS databases
   - number of other papers relevant to career advancement in accordance with the national legislation
   - number of relevant books published
   - citation index in WoS and Scopus database
2. **national and international competitive research projects and research mobility** includes the following sets of entry data
   - number of contracted national competitive projects
   - share of the budget of contracted national competitive projects in the total budget of the institution
   - number of contracted international competitive projects
   - share of the budget of contracted international competitive projects in the total budget of the institution
   - incoming mobility (minimum one month period)
   - outgoing mobility (minimum one month period)
3. **collaboration between research and business sector** as well as collaboration with the u**nits of local and regional governance and non-governmental sector** includes the following sets of entry data
   - number of contracted projects with business sector and/or units of local and regional governance and non-governmental sector
4. activities of the **popularisation of science** includes the following sets of entry data
   - international activities related to the popularisation of science
   - national activities related to the popularisation of science

- local activities related to the popularisation of science

The MoSE designed the funding formulae in cooperation with the Institute of Economics, Zagreb.

- In the first step all indicators per field are multiplied with the weight factor of the respective field and then divided by the number of PROs in the field. The result is an average score per each indicator.
- In the second step, the indicator score is divided by the number of FTEs for the respective institution resulting in the second score.
- In the third step, the obtained second score is divided by the average score per each indicator.
- The final score - institution coefficient - for each institution is a sum of the above-obtained score for each institution.
- In the end, the funds for each institution in particular field are calculated as a product of FTE x base x institution coefficient.

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.

KI-AX-18-003-EN-N

To support countries in reforming their research and innovation systems, the Directorate-General for Research & Innovation (DG RTD) of the European Commission set up a Policy Support Facility (PSF) under the European Framework Programme for Research & Innovation 'Horizon 2020'. It aims to support Member States and associated countries in improving their national science, technology and innovation systems.

This report documents a mutual learning exercise on performance-based research funding systems (PRFS) for universities. Armenia, Austria, Croatia, the Czech Republic, Estonia, Italy, Moldova, Norway, Portugal, Slovenia, Spain, Sweden and Turkey all participated in the exercise, which was supported by a group of five experts.

The report reviews wider international experience with PRFS as well as the specific experiences of the participating states. It constitutes a 'handbook' for policy makers and PRFS designers, providing analyses of experience, a discussion of the policy role of PRFS, key parameters of PRFS design, strengths and weaknesses of different 'models' for PRFS design including discussions of peer review, indicators, and how to consider the societal impact of research. It concludes by drawing lessons and making policy recommendations about how and when to design and use PRFS for policymakers and PRFS designers

Studies and reports

Publications Office

doi:10.2777/644014

ISBN: 978-92-79-80413-7