Final version – January 19, 2015

R&D Evaluation Methodology and Funding Principles

First Interim Report: the R&D Evaluation Methodology

evropský sociální fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

R&D Evaluation Methodology and Funding Principles

First Interim Report – Draft version

January 19, 2015

Bea Mahieu, Erik Arnold – Technopolis Group

Table of Contents

Table of Exhibits

List of Abbreviations

ARC	Australian Research Council
BOF	Bijzonder Onderzoeks Fonds (Special Fund for Research, Belgium)
CNRS	Centre national de la recherche scientifique (National Centre for Scientific Research, France)
CR	Czech Republic
DG	Directorate-General
EARTO	European Association of Research and Technology Organisations
EC	European Commission
EM	Evaluation Methodology
ERA	European Research Area
ERC	European Research Council
EvU	Evaluated Unit
FTE	Full-time equivalent
GBER	General Block Exemption Regulation
IPR	Intellectual Property Right
HEI(s)	Higher Education Institute(s)
NESTI	OECD Working Party of National Experts on Science and Technology Indicators
NIS	National Innovation System
NZ	New Zealand
OECD	Organisation for Economic Cooperation and Development
PRFS	Performance-based Research Funding System
PRI	Public research institutes
R&D	Research and Development
RAE	Research Assessment Exercise (UK)
RD&I	Research, Development and Innovation
RD&I IS	RD&I Information System
REF	Research Excellence Framework (UK)
RI	Research Infrastructure
RO(s)	Research Organisation(s)
RQF	Research Quality Framework (Australia)

RTO(s)	Research and Technology Organisation(s)
RU	Research Unit
SSH	Social Sciences & Humanities
UK	United Kingdom
VQR	Valutazione della Qualità della Ricerca (Evaluation of the Quality of Research, Italy)
WOS	Web of Sciences

Executive Summary

This report is the final First Interim Report of a study developing an evaluation methodology and institutional funding principles for the R&D system in the Czech Republic. It is focused on the design of the R&D evaluation methodology. The second interim report will describe the funding principles.

Evaluation constitutes a key component of the policy cycle. In order to be effective and reach the desired results, an evaluation system should be well understood and trusted – by all stakeholders, and the evaluation procedures need to be built on consensus. The design of the evaluation methodology in this study is therefore to be conceived as a first step in a longer-term process, based on consultation of all actors involved and final policy decision-making. The Evaluation Methodology that we propose sets the basis for a new research evaluation and funding system and provides a concrete model for its implementation. It is based on a thorough analysis of international practice while taking into proper consideration the applicability of methods and tools in the context of the Czech R&D system.

The objectives for this study are to design an R&D evaluation methodology with a pronounced formative function, providing strategic information to actors at all levels in the R&D system, as well as informing the performance-based funding system (PRFS). It was expected to include all research organisations (ROs) while taking into account the differences among types of research organisations and disciplinary cultures.

The research community in the Czech Republic is composed of researchers employed in universities, public research institutes (including those 'grouped' in the Academy of Sciences and 'sectoral' public research institutes), private research institutions, and industry. Key actors in public research are the public universities and the research institutes of the Academy of Sciences. Only those institutions that are recognised as research organisations (ROs) are entitled to public institutional support for research.

Concepts and approaches to evaluation and its design in the international practice that are of particular interest for this study are:

- Evaluation is an integral part of the policy cycle. The national R&D policy is therefore a major factor influencing the design of a national evaluation methodology.
- Evaluations at the national level cover (also) the performance of research institutions and their management. However, the depth of assessment is more limited than in evaluations conducted at the institutional level, and the focus of the assessment is determined by the needs for strategic information at the national level as well as the purpose of the evaluation, e.g. to inform a PRFS
- The evaluation should take account of the differences among research organisations in terms of their "mission in society'. For this purpose, a categorisation of the research organisations in relation to their function in the National Innovation System is needed
- Bibliometrics and statistical data analyses require a minimum number of data to ensure robustness. Seeing the fragmentation in the Czech system, this implies the need for a minimum threshold for participation to the evaluation (but not for the institutional funding system)
- Informed peer review, i.e. the combined use of expert panels and bibliometrics, is common best practice

- Also peer reviews have their shortcomings and risks, but these can be overcome through the use of bibliometrics, combined with specific rulings and structural measures
- Submission of research outputs for review by research groups or departments rather than individual researchers avoids the creation of considerable negative effects on career prospects, 'employment markets' and the R&D system as a whole
- There are considerable differences among scientific disciplines and even sub-disciplines. They are expressed in terms of output types, main publication patterns, channels and timelines, citation behaviours, language of publication, collaboration behaviours and needs, intensity of the use and need of (human and financial) resources and research infrastructure, and last but not least, their societal mission and the way they create and transfer knowledge
- Discipline-based panels have the highest level of capacity to take field-specifics in concern during the assessment. There must, however, be consistency in assessment among disciplines and only a limited number of field adjustments should be allowed
- Inter-disciplinary research is a challenge for any assessment method, but less so for peer reviews
- Indicators and assessment criteria for research performance assessments are closely related to the theory of knowledge. Knowledge (both codified and tacit) is the major outcome of science and research; it also constitutes its major value – for research, industry and society alike.
- An evaluation system that intends to understand research performance in its broader sense, i.e. not limited to the size and scientific quality of research outputs, focuses on assessing the intensity of the knowledge transfer mechanisms. Knowledge transfer mechanisms are understood as *pathways to impact*, i.e. those aspects that are critical for the *creation* of impacts
- A basic concept of evaluation is that indicators should cover the various sequential components of a policy intervention, i.e. the inputs (financial and human resources) for the implementation of activities that are expected to lead to outputs, outcomes and impacts
- There is a clear trend in international practice, including indicator-based PRFS, to extend the focus of investigation from outputs (only) to outcomes, and in some cases even impacts
- The most appropriate and effective way to avoid unintended effects that some indicators may cause, especially in PRFS (gaming), is to use a **mix** of quantitative and qualitative indicators to inform an assessment criterion, judged by a peer
- National evaluations are a costly endeavour and the higher the level of sophistication, the costlier an exercise it becomes, and the lower the balance cost/benefit

The Evaluation Methodology that we propose has the following key principles and characteristics:

The Evaluation Methodology reflects the strategic **policy objectives** for the Czech R&D system.

The key function of the evaluation system is to support public R&D governance in the attainment of its strategic objectives, i.e. to strengthen R&D capacity, to foster excellence in research as well as its alignment with the societal needs, and to support the growth and competitiveness of the Czech Republic.

The role of evaluation as defined in the policy documents is to assess past research performance as well as to support future performance improvement. For this purpose, the Evaluation Methodology is designed to provide strategic information for policymaking and research management as well as incentives for positive change in the R&D system.

Its primary function is to act as source for **strategic information**, at all levels in the RD&I system

The outcomes of the evaluation will constitute quality information for R&D policy making at the national and funding bodies' level as well as for R&D management in the single research organisations, institutes and university faculties.

It is the **detailed and comprehensive** approach that gives quality. The evaluation covers all dimensions of the research activities, i.e. the research quality, research strategy, the research environment (i.e. the institutional conditions), the research outputs, outcomes and impacts. The evaluation results will give a view on the specific strengths and weaknesses of the different actors against all of these dimensions. It will allow for the identification of the factors upon which action is needed in order to improve research performance, at the national as well as institutional level.

The evaluation will assess the evaluated actors' role, positioning, and competitive value in the **national** R&D and innovation system as well as in the **international** R&D landscape. As research is becoming more and more international, and the competition as well as collaboration in research is at a global level, understanding the position of the research actors in the Czech republic relative to the international level is a critical factor in the assessment. For this purpose, the evaluation adopts as its primary unit of analysis the elements that worldwide constitute the fundamental structure of research, i.e. **scientific fields.**

The evaluation will directly inform public **institutional funding for research**.

The intent of a performance-based research system is to act upon previously identified failures in the R&D system and to steer research behaviour in order to overcome these failures by providing incentives. The indicators used in the evaluation and in particular the structuring into a set of **assessment criteria,** is therefore guided by the policy objectives. The intent is to achieve a balance between assessing and ultimately rewarding research quality, capacity development, and excellence, while taking into account the value of the activities - for research and for society.

The **Research Unit** is the primary unit of assessment.

A *Research Unit* (RU) includes all individual researchers in an EvU (across the organisation structure) that conduct research in a single scientific field. Researchers

need to be assigned to research units in their major field of research; each researcher can be assigned only to one research unit in an evaluated unit.

An *Evaluated Unit* (EvU) is a research organisation, except for the public HEIs where the Evaluated Unit is a Faculty or Institute or any other organisational unit at that level such as Centres

In other words, a Research Unit is a sub-set of an EvU does not necessarily represent a coordinated or collaborating research group. All researchers in an EvU have to be allocated to an RU.

The evaluation system allows for the registration of Interdisciplinary Research Units, under specific conditions.

The evaluation covers **all research organisations of a minimum size**

The evaluation is designed to assess the research performance of any organisation that is recognised as a research organisation in the Czech R&D system.

However, a minimum volume of research outputs is required in order to ensure the robustness of the assessment. The threshold for a research organisation's participation in the evaluation – and for the registration of a Research Unit by the EvU - is 50 research outputs within 1 field of research over the evaluated period (i.e. 5 to 6 years).

The Research Organisations in the Czech Republic have different missions in society, independently of their legal status. The Evaluation Methodology distinguishes between Scientific Research Institutions, Research and Technology Organisations (RTOs), Public Service Research Organisations, and National resource/Infrastructure Research Organisations.

The EM uses assessment criteria that are relevant for **all** types of RO, no matter the type of research they perform. The assessment criteria cover the conditions that enable for quality research to occur in any type of research organisation (the research environment, including the research strategy and management, and the membership of the global and national research community), the key criteria one quality of the research performance (scientific research excellence and overall research performance, including research output and competitiveness in research), and the activities that constitute pathways to impact – on research and the society at large, the latter illustrating the societal relevance of the research .

The evaluation is a process of **informed peer review**.

The structure of science and its most consolidated assessment practices, i.e. peer review and panels, constitute the basis of the evaluation system.

Expert panels at the level of scientific fields are at the core of the Evaluation Methodology. Partly working remote, they will draw on a **mix** of appropriate quantitative and qualitative data to support their professional judgement. This information will be based on international bibliometric data, data included in the national RD&I system, and quantitative and qualitative data provided by the evaluated Research Units, including a self-assessment. A key principle in this evaluation is that metrics inform, but do not substitute for judgment. Expert review is therefore paramount in all phases of the evaluation process.

The Evaluation Methodology defined specific rules governing conflicts of interest, against nepotism and 'clientelism' as well as auditing mechanisms and rules for punishing cases of fraud. Such punishment will be both public and severe.

In order to maintain the quality and depth of the strategic information provided to the Evaluated Units and Research Organisations, the panels' **evaluation results** will consist in an overview of the assessments of each Research Unit in the Evaluated Unit

against each criterion, accompanied by explanatory texts and conclusions and recommendations.

The panel evaluations will result in a panel report per RU, an overview panel report for EvUs with more than 1 RU; an analytical report per field and per disciplinary area.

These reports will allow the institutional management and national funding bodies and policy makers to reach an improved view of strengths and weaknesses and the competitive positioning in the national and international context of the research organisations and research as a whole in the Czech Republic - from a research, development and innovation perspective.

The evaluation will be conducted in full transparency. Information on the evaluation criteria, the names of panel chairs and members, and the evaluation results will be made public.

The evaluation is a **fair and egalitarian** system.

The Evaluation Methodology has defined a single framework for assessment across all disciplines and research organisation typologies.

This approach combined with the structure of the evaluation system as such, i.e. based on the field-specific expertise of the expert panels, enables **full comparability** of the evaluation results, independently of the scientific fields in which the research is conducted. Expert panels will apply standards of assessment consistently, working under guidance and supervision of the main panels.

The expert panels will assess the research units' performance against the different assessment criteria on an equal footing - wherever that research is conducted. A detailed Evaluation Protocol will provide the standard definitions and set the common procedures. The Evaluation Methodology allows for a reasonable level of field- and RO typology-specific variations to the common generic indicators and assessment criteria.

The **cost and burden** of the evaluation will be the minimum possible to deliver a robust and defensible process.

The total costs (direct and indirect) will not exceed 1% of the public institutional funding for R&D over a five-year period.

It should be noted that in this study we use the term 'research' in the broad sense, i.e. encompassing research, development and innovation. We use the term 'scientific research' when referring to research in the narrow sense of the word.

1. Introduction

This document constitutes the final version of the First Interim Report for the study developing a new evaluation methodology and funding principles for the R&D system in the Czech Republic. The report describes the outcomes of the first phase of the study, i.e. the design of a new R&D evaluation methodology.

This report is the fruit of the joint efforts by a large study team. It builds on the work of:

- Jan Dvořák, Tomáš Chudlarský (Infoscience Praha)
- Gunnar Sivertsen, Liv Langfeldt, Kyrre Lekve (NIFU)
- Michal Pazour, Zdeněk Kučera, Tomáš Vondrák, Jiří Vaněček, Ondřej Pecha, Ondřej Pokorný, Vladislav Cadil, Tomáš Ratinger (Technology Centre ASCR)
- Erik Arnold, Tomas Åström, Oliver Cassagneau-Francis, Kristine Farla, Barbara Good, Tobias Fridholm, Elina Griniece Zsuzsa Javorka, Malin Jondell Assbring, Peter Kolarz, Bea Mahieu, Göran Melin, Anke Nooijen, Fritz Ohler, Martijn Poel, Xavier Potau, Caspar Roelofs, Tammy-Ann Sharp, Brigitte Tiefenthaler, Geert van der Veen, Frank Zuijdam (Technopolis Group)

The study team was supported by field experts (Christofer Edling, Milena Horvat, Ron Perrott, Roland Pochet, Naomi Segal, and Ken Thomson) and an Advisory Panel composed of Diana Hicks, Paul Hubbard, Keith Jeffery, and Gunnar Sivertsen.

1.1.1 The Evaluation Methodology: Work in Progress

Reflecting the Terms of Reference (ToR), the study is structured in three phases: first we focus on drafting the Evaluation Methodology to then focus our attention on the funding principles. The last phase of the study is dedicated to the finalisation of both the Evaluation Methodology and the funding principles.

This report describes the outcomes of the first phase of the study, i.e. the design of an evaluation methodology. It constitutes a first step in a **process** towards the definition of a final evaluation methodology, which will continue beyond the time frame of this study.

In the context of this study, this process is made explicit in the efforts dedicated to the consultation with the IPN project team responsible for this study, the implementation of a pilot evaluation exercise, the consultation of all actors in the RD&I system, and last but not least, the very structure of the study itself in specific phases. The combination of these components of the study imply that a final proposal of both the Evaluation Methodology (EM) and the funding principles will be delivered only at the end of the study, i.e. at the end of the month of May 2015.

The Small Pilot Evaluation, conducted in the months of November/December 2014, constituted an important tool for the testing of the processes that were defined for the implementation of the research assessment, as well as the adequacy and feasibility of the defined assessment criteria and indicators. The results of the pilot evaluation and the reflections on the 'lessons learned', based on the feedback from the evaluation panels and their secretariats as well as from the participating research organisations, will be reported in the Third Interim Report of this study. That report will be delivered in February 2015.

Actors in the Czech RD&I system were given the opportunity to provide their feedback to the draft version of the current report – both in written form and during the conference on January 7. We take this opportunity to express our appreciation and

gratitude for the strong participation of the RD&I community in both of these occasions and the many valuable comments we received.

Several of these comments have been taken into account in this final 1st Interim Report, focusing in particular on the Assessment Criteria so that the 2nd Pilot Exercise can take the revised criteria into account. More will be considered for the future developments of the Evaluation Methodology and we will cover this topic more in detail in a separate communication to the actors in the RD&I system by the end of January.

In relation to the structure of this study, inevitably the division of the work in the first two phases is somewhat artificial: the direct link between the Evaluation Methodology and the funding principles implies a direct reciprocal influence of the one on the other. The EM was designed taking into account the (potential) features of the funding principles; the final proposal of the funding principles, following the consultations, will influence the final characteristics of the EM. An important factor for a final decision-making on the EM, its frequency and characteristics, is also the estimated cost of the assessment exercise, including the 'indirect' costs for the evaluated research organisations. This topic will be covered in the study Final Report.

1.1.2 The Evaluation Methodology proposed in this report

The Evaluation Methodology (EM) that we propose in this report defines the key principles for the future evaluation methodology and sets its basic components. In accordance to the process described in the previous section this means that the work presented here is still 'work in progress' and the elements of the design proposed are liable to be changed during the remainder of the project.

A fundamental principle for any evaluation methodology is that it should reflect the specific policy objectives and needs. These policy objectives define the purpose and function of the evaluation, which constitutes a key factor in the choice for its core elements, i.e. the scope of the assessment, the assessment criteria and indicators, the methods used, and the depth and breadth of the evaluation exercise.

In its current version, the EM and its principles for implementation reflect the policy objectives and needs as defined in recent R&D policy documents in the Czech Republic. It also responds to the requests expressed in the ToR for this study.

- Reflecting the policy objectives, the current EM is centred on the assessment of the institutional conditions enabling the conduct of quality research (now and in the future), scientific research excellence, overall research performance, and the activities that constitute pathways to impact – on research and the society at large
- Reflecting the ToR for this study, the EM has a pronounced formative function, providing strategic information to actors at all levels in the R&D system, as well as informing the performance-based research funding system (PRFS). It is a national evaluation methodology, allowing for performance assessment of all types of research organisations (ROs) and disciplinary cultures, while taking into account their differences. It is set up so that total costs do not exceed 1% of public institutional support for R&D in the evaluated period

During the design process our ambition was to maintain an appropriate balance between the many objectives and dimensions that the evaluation methodology needed to cover, within the defined cost limits. Inevitably, this was a process of finding compromises.

Future revisions of the EM depend on the outcomes of discussions among all actors in the RD&I system, and ultimately a policy decision, whether the EM should more explicitly take into account the differences in the types of research organisations, assess more in depth quality of research rather than focusing on scientific research excellence, focus on evaluating impact rather than assessing the conditions that allow for impacts to occur (i.e., the 'pathways'), etc.

1.1.3 Structure of the report

This report is structured as follows:

First, we set the context for the Evaluation Methodology. The Evaluation Methodology takes into account the background and context of the R&D system in the Czech Republic, covered in Chapter 2, and builds on a thorough analysis of the concepts for evaluation in international practice, set out in Chapter 3.

In Chapter 4 we describe the Evaluation Methodology, its key principles and core elements, while we present the processes for its implementation in Chapter 5.

Two alone-standing background reports are provided as appendixes to this report:

- The Guidelines for the Evaluated Research Organisations, i.e. the guidelines for the registration and submission of the information for evaluation. These guidelines will give the reader a view on the type of information requested from the evaluated RU and its description. This is of particular interest in the context of the changes made to the EM in this final version of the 1st Interim Report
- Evaluation systems in international practice (country analyses)

2. Background to the evaluation methodology

In this Chapter we describe the developments in the Evaluation Policy in the Czech Republic (Chapter 2.1), give a brief overview of the Czech R&D system (Chapter 2.2), and set the evaluation methodology within the R&D policy context (Chapter 2.3).

2.1 The Evaluation Policy in the Czech Republic

Currently, in the Czech Republic, there is a single Evaluation Framework for the evaluation of performance in the research organisations as well as efficiency in the design and implementation of competitive research funding programmes. The Evaluation Methodology focuses exclusively on research outputs and combines two functions: it is both a mechanism for evaluating research and for allocating institutional funding for R&D, with a direct, automatic link between the two. The evaluation results therefore directly drive the Performance-based Research Funding System (PRFS).[1]

2.1.1 Historical context

The current evaluation framework in the Czech Republic has its roots in the National Policy on Research & Development (R&D) for the years 2004 – 2008, which included an attempt to improve the quality of the evaluation system. It said that a stronger 'evaluation culture' was needed and stressed the importance of evaluations as inputs to policy development and decision-making. Some of the conclusions were built into Government Resolution No. 644 on the evaluation of R&D and its results (June 2004). The objective of the Resolution was to tackle the perceived failure in evaluation quality in the R&D system. It set the basis for evaluating R&D institutions, programmes and final evaluations of projects, including the 'research intentions' through which institutional funding was at that time awarded.

The 2004 Evaluation Methodology (further: Metodika) introduced to the Czech Republic the concept of a metrics-based quantitative results evaluation, seen as a tool – and only one of the main criteria – to prove the quality of research performance. It also stressed the importance of respecting the differences between disciplines when evaluating research results.

A major shift occurred with the 2008 Reform of the RD&I System. The Metodika 2009, which implemented the principles of the 2008 Reform, marks the launch of an evaluation system that was profoundly different from the 2004 evaluation methodology. Fundamental changes, in the EM 2009 and its subsequent versions, were:

- A **narrowing of the function** of evaluation, abandoning the previous attempts to instil 'evaluation culture' and embed learning in the system and replacing them

[1] This section builds on three reports that were published as a result of the International Audit of the R&D&I System in the Czech Republic, i.e. Arnold, E. (2011), International Audit of the R&D&I System in the Czech Republic, Synthesis report, Technopolis Group; Arnold, E. et al. (2011), The Quality of Research, Institutional Funding & Research Evaluation in the Czech Republic and abroad, International Audit of R&D&I in the Czech Republic - Final Report - 3, Technopolis Group; Arnold, E., Mahieu, B., Horvath, A., (2011) R&D Governance in the Czech Republic, International Audit of R&D&I in the Czech Republic - Final Report - 2, Technopolis Group

with the idea of evaluation as a component in a performance-based resource allocation system

- A progressive **restriction of the scope** of the evaluation guidelines: while the 2004 evaluation methodology covered all the different layers of the research system, the Metodika 2009 focused almost solely on the quantification of research outputs for the evaluation of research organisations and research programmes
- An increasing **breadth of coverage** of the Metodika: the Metodika 2009 established the use of the metrics-based evaluation of R&D results for institutional funding at the level of funding bodies. The Metodika 2010 enforced and expanded the use down to the level of research institutions

The 2011 International Audit of the RD&I System in the Czech Republic (further: Audit) strongly criticised the Metodika. The study recognised that the evaluation methodology was driven by 'good ambitions' in addressing the issue of low productivity in some research organisations and intended to improve the quality of research outputs.

However, it identified important flaws in the evaluation system from the perspective of the quality of the evaluations and the role of evaluation in the policy cycle. Weaknesses that were identified included:

- An exclusive focus on the immediate outputs of the research system rather than assessing (also) whether it produces the intended societal effects
- A reduction of the complexity of performance to an overly simple category of outputs
- A lack in consideration for disciplinary differences (both in types of outputs among disciplines and in the costs of producing them)
- All institutions are treated in the same way, regardless of their missions, by using output indicators that are in practice arbitrary
- The concept of the evaluation system as being intrinsically part of a policy cycle is not perceived. Policy requirements such as the national thematic priorities are not taken into account and there is little to no effort for an effective measurement of the extent at which policy interventions achieved the expected effects on S&T fields

Broadly, the Audit considered the need for the Czech RD&I system to build up an evaluation culture and for policy-makers and research organisations to come to see evaluation as a tool for learning and improving research.

The International Audit concluded[2]:

> *"The Evaluation Methodology is not fit for purpose. It introduces structural and behavioural distortions and impedes many aspects of the NRIS' development. The Evaluation Methodology should be replaced by a system of performance contracts that have both prospective and retrospective components, supported by a combination of objective indicators and international peer judgement.*
>
> *The system of R&D evaluation in the Czech Republic more broadly focuses on counting outputs at the expense of understanding policy interventions and their impacts. It therefore provides information that is at best of limited relevance. Evaluation practice should be the subject of root and branch*

[2] Arnold, E. (2011), International Audit of the R&D&I System in the Czech Republic, Synthesis report, Technopolis Group

reform, refocusing on outcomes and impacts in addition to outputs and contributing to policy and programme development and planning."

2.1.2 Recent developments

Following the International Audit, changes were made in the methodological approach to evaluation, leading to the Metodika 2013-2015. The most important changes consisted in the improvement of the research assessment method through the introduction of a panel review component, more precise definitions for the research outputs and restrictions in the eligible typologies to contrast gaming, and efforts to increase the trust of the research community in the fairness of the system.

While this was a positive development, it corrected only partially the methodological flaws as an evaluation system that the Audit identified: Metodika remained close to exclusively focused on research outputs.

Five years after the introduction of the Metodika 2009, the negative effects of the evaluation and funding system on the evaluation culture in the RD&I system, leading to a misconception of the role of evaluation, are increasingly apparent. These include:

- The number of "RIV points" attained, i.e. the points attributed to the different research outputs in the 'evaluation', are considered as an indication of research quality and a tool for rewarding/punishment throughout the entire R&D system, down to the level of the individual researcher
- The direct link between evaluation and funding heavily dominates the evaluation culture and has affected profoundly the Czech R&D evaluation and policymaking system
- The discourse on evaluation is detached from any discourse on policy and strategy related to the national R&D system

2.2 The R&D system

2.2.1 R&D Governance

In recent years, the Czech Republic (CR) set the fundaments for a radical change in its R&D&I governance structure. These were based on the 2008 Reform and the subsequent National Research, Development and Innovation (further R&D&I) Policy document (2009) for the years 2009 – 2015 and other necessary legislative interventions. The current RD&I governance structure in the CR can be depicted as in Exhibit 1, below.

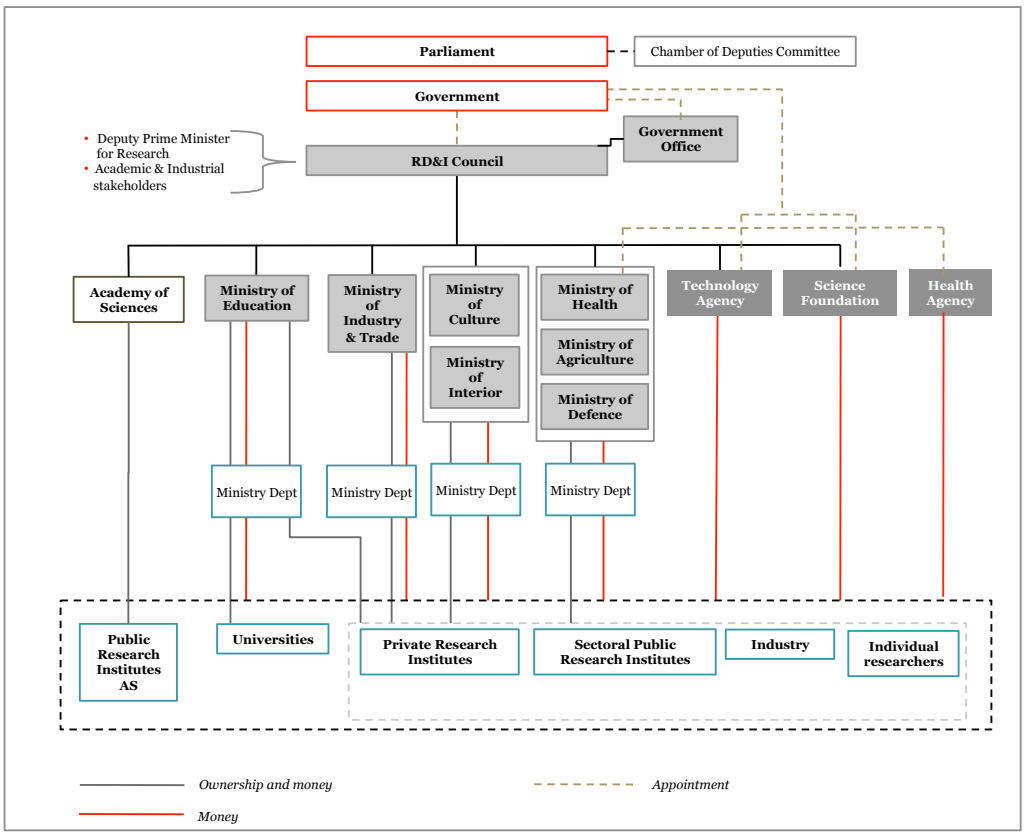
At the first level in the RD&I Governance system, the **RD&I Council** (further: Council) acts as an advisory body to the Government of the Czech Republic. It has 16 members (not including the Chairman) and is governed by a Board. The Deputy Prime Minister for Research acts as the Chairman of the Council, thus enforcing its legitimacy. Members of the Council are members of the different RD&I communities and are nominated by the Government on proposal of the Chairman, with a mandate of 4 years (once renewable).

The Council covers a broad range of tasks in the national governance of the RD&I system, including the definition of overall directions and priorities across the National Research and Innovation System, long-term strategy development, the preparation of a proposal of a very detailed budget for research and development, monitoring and evaluation. It is supported by 3 disciplinary advisory Expert Committees and 2 Advisory Commissions, i.e. the Commission on Bioethics and the Commission for Evaluation.

A set of ministries, the Academy of Sciences, and 3 agencies, responsible for the implementation of the RD&I policy, constitute the second 'intermediary' level.

The involvement of the Government and Parliament in the current RD&I System is considerable. Two of the agencies have a unique status; their governing bodies are nominated by the Government – upon proposal by the Council. The Government also nominates – or removes - the members of the Council and the Secretariat of the Council is part of the Office of the Government.

Exhibit 1 The RD&I governance system in the CR



Source: Technopolis, 2014

The National RD&I Policy 2009-2015, which implemented the 2008 Reform of the RD&I system, restructured the RD&I governance system. The number of Ministries and other public administration bodies with competences for R&D&I funding was reduced, limiting the national R&D budget chapters from 22 to 11. In total, 7 Ministries hold management responsibilities for national public R&D&I support: the Ministry of Defence, the Ministry of Health, the Ministry of Agriculture, the Ministry for Education, the Ministry for Industry and Trade, the Ministry of Culture, and the Ministry of Interior.

All of these Ministries manage the national institutional funding for the research organisations – public or private non-profit – in their area of competence; most of them also develop and manage competitive R&D programmes. Exception is the Ministry of Industry that officially does not have the responsibility for 'targeted' funding programmes, even though it currently runs such a programme that will last until 2017. None of these Ministries conducts evaluations for the assessment of its research organisations.

2.2.2 The R&D base

As shown in Exhibit 1, above, the research community in the Czech Republic is composed of researchers employed in universities, public research institutes (including those 'grouped' in the Academy of Sciences and 'sectoral' public research institutes), private research institutions, and industry.

Key actors in public research are the public universities and the research institutes of the Academy of Sciences.

- In the Czech Republic there are 26 public **universities**, 2 state universities (the Policy Academy and the University of Defence), and 45 private higher education institutions
- The **Academy of Sciences** has historically a special position in the Czech R&D system. It is an "organisational body of the Czech Republic" and its activities are financed directly by the state budget (it has its own budget chapter). It holds responsibility for the allocation of institutional funding to its 54 institutes. In this context, it organises internal institutional evaluations
- **Sectoral public research institutes** are public research institutes that were previously governed by specific Ministries (such as the Ministry of Agriculture or Transport) and gained the status of public research institutes in 2007. In several cases, these institutes have public administration as their target users of their products/services.
- **Private research institutes** include a broad range of private enterprises offering R&D services. These include industry-oriented research institutions that took up the role of public RTOs under communist times and survived the privatisation wave in the beginning of the 1990s

Only those institutions that are recognised as **research organisations (ROs)** are entitled to public institutional support.

Research organisations are defined in the Act No 211/2009 Coll. (a complete amendment of the Act No. 130/2002 Coll.) on the support of research, experimental development and innovation. The CR adopted the definition provided by the 2006 "Community framework for state aid for research, development and innovation", which states:

> *"Research organisation' means an entity, such as university or research institute, irrespective of its legal status (organised under public or private law) or way of financing, whose primary goal is to conduct fundamental research, industrial research or experimental development and to disseminate their results by way of teaching, publication or technology transfer; all profits are reinvested in these activities, the dissemination of their results or teaching; undertakings that can exert influence upon such an entity, in the quality of, for example, shareholders or members, shall enjoy no preferential access to the research capacities of such an entity or to the research results generated by it."*

Since July 2014, the new EU Commission Regulation (GBER) is in force, which uses a slightly different definition of RO, namely:

> *"Research and knowledge-dissemination organisation' means an entity (such as universities or research institutes, technology transfer agencies, innovation intermediaries, research-oriented physical or virtual collaborative entities), irrespective of its legal status (organised under public or private law) or way of financing, whose primary goal is to independently conduct fundamental research, industrial research or experimental*

development or to widely disseminate the results of such activities by way of teaching, publication or knowledge transfer."

A technical amendment to the Czech Act is in the process of approval, which will transpose all changes in the Community Framework into the Czech legislation (including the definition of ROs).

According to the R&D Information System (IS), there were **219 registered ROs** in 2014. Public research organisations accounted for the majority of the registered research organisations in 2014 (163 on a total of 219). The 56 non-public research organisations have different legal status, including

- 'Beneficial organisations' (various private agencies, think tanks, etc.) (12)
- Registered legal bodies (38)
- Other legal bodies (1), and
- Professional association/non-profit organisations (5).

However, only 164 of these research organisations actually received institutional funding in 2014. Close to 20% of the 219 registered ROs (43) gained their status of research organisation after 2011. The new research organisations are mostly state agencies (museums, hospitals, etc.), registered legal bodies (private universities and private companies) and beneficial organisations (various private agencies, think tanks, etc.).

Exhibit 2, below, shows the ministries responsible for the allocation of institutional funding (i.e. the 'funding bodies') and the number of ROs in their sphere of competence for the funding in 2014.

Exhibit 2 Funding bodies allocating institutional funding in 2014

Funding body	**Number of RO**
Academy of Sciences CR	54
MEYS (Ministry of Education)	41
MoA (Ministry of Agriculture)	20
MoH (Ministry of Health)	15
MoC (Ministry of Culture)	13
MIT (Ministry of Industry & Trade)	10
MoI (Ministry of Interior)	8
MoD (Ministry of Defence)	3

The RD&I IS also indicated that out of these 219 research organisations, **200 ROs** registered at least one scholarly research output[3] in the period 2008-2012 (Exhibit 3, below).

Only few of the private universities published scholarly outputs in that time period, (university) hospitals were particularly 'active', and close to 30 public research

[3] The four types of publications: J - Journal Article, B - Book (Monograph), C - Book Chapter, D - Conference Paper

organisations - beyond the Academy and the public universities – published scholarly outputs. Also, the 200 'active' ROs accounted for 384 institutes or faculties.[4]

Exhibit 3 Research institutes and public HEI faculties registering scholarly outputs in 2008-2012

	Nr of Research Organisations
ASCR	53
HEI – public	26
HEI – private	7
(University) Hospital	13
Ministry Interior/Defence (HEI & institutes)	10
Other institutes/Centre	62
Government Agency/Museum/Library	29
Total	200

Source: Public data of the Czech RD&I Information System (www.isvav.cz), Technopolis analysis

In terms of **size** of the ROs, the latest available data from the Czech Statistical Office illustrate the growing importance of the Higher Education sector in research. In 2011, this sector accounted for approximately 30% of the total FTE researchers in the country (Exhibit 4).

Exhibit 4 Number of researchers in the research performing sectors - 2011

Sector of performance	Total	Men		Women	
		Number	% from total	Number	% from total
Government sector					
Registered number of employees at 31 December (HC)	8,220	5,088	61.9%	3,132	38.1%
Full-time equivalent (FTE)	6,235	3,964	63.6%	2,272	36.4%
Higher Education sector					
Registered number of employees at 31 December (HC)	20,732	13,548	65.3%	7,184	34.7%
Full-time equivalent (FTE)	10,289	6,986	67.9%	3,303	32.1%
Business enterprise sector					
Registered number of employees at 31 December (HC)	16,698	14,157	84.8%	2,541	15.2%
Full-time equivalent (FTE)	13,958	11,913	85.3%	2,045	14.7%
Private Non-Profit sector					
Registered number of employees at 31 December (HC)	251	172	68.5%	79	31.5%
Full-time equivalent (FTE)	199	123	61.6%	77	38.4%
CZ Total					
Registered number of employees at 31 December (HC)	45,902	32,966	71.8%	12,936	28.2%

[4] The analysis was taken down to the level of faculty as this was the envisaged level of scope for the evaluation. Faculties are considered only for the public HEIs . In total we counted 210 faculties

Sector of performance	Total	Men		Women	
		Number	% from total	Number	% from total
Full-time equivalent (FTE)	30,682	22,985	74.9%	7,696	25.1%

Source: Czech Statistical Office, 2014

Comparable data on FTE researchers at the level of HEI faculty are not available.

Using the authors of the scholarly outputs as a proxy for the number of researchers (HC) in the research institutes and public HEI faculties, we noted that there is a high diversity among the research institutes/faculties in terms of size: in more than 20% of the institutes/faculties, less than 20 researchers registered scholarly outputs in the period 2008-2012 (Exhibit 5).

Exhibit 5 Size of the institutes/faculties based on the number of scholarly output authors (2008-2012)

	Nr of faculties/institutes	% of total
Less than 20 authors	76	20%
20 – 50 authors	67	17%
50 – 100 authors	79	21%
100 – 200 authors	59	15%
200 – 300 authors	37	10%
300 plus authors	66	17%
Grand Total	384	100%

Source: Public data of the Czech RD&I Information System (www.isvav.cz), Technopolis analysis

2.2.3 Scientific focus of the research

We analysed the publication profile of the Czech Republic using as framework the field classification defined by the OECD. The use of this classification system allows for the international comparability of data on research activities in specific fields and their outcomes and will therefore increase the capability for benchmarking and use of international datasets.

The data in the Czech R&D IS on scholarly publications[5] in the time period 2008-2012 (Exhibit 6) show the following publication profile of Czech research

- The Czech Republic is particularly active in the Engineering & technology and the Physical sciences (Mathematics, physical sciences, chemical sciences and earth & environmental sciences)
- A second group of disciplines in terms of size in activity is Medical and health sciences and Social sciences
- Disciplines of medium activity are Humanities and Biological and Agricultural sciences

[5] The four types of publications: J - Journal Article, B - Book (Monograph), C - Book Chapter, D - Conference Paper

The R&D IS data also show that we can estimate the annual total output of original peer-reviewed scholarly and scientific publications to be 22,000.

Exhibit 6 Publications in Disciplinary Areas and Fields in the Czech Republic, 2008-2012

	No publications from 2008-2012
1. Physical sciences	**51,501**
Mathematics	8,326
Physical sciences	18,736
Chemical sciences	12,885
Earth and related environmental sciences	11,554
2. Engineering & technology	**59,247**
Civil engineering	10,569
Electrical engineering, electronic engineering, information engineering	14,952
Computer and information sciences	8,593
Mechanical engineering	3,801
Chemical engineering	1,796
Materials engineering	10,203
Medical engineering	241
Environmental engineering	2,939
Industrial Biotechnology	826
Other engineering and technologies	5,327
3. Medical and Health sciences	**43,513**
Basic medicine	3,942
Clinical medicine	31,662
Health sciences	3,765
Other medical sciences	4,144
4. Biological and Agricultural Sciences	**29,421**
Biological sciences	19,160
Agriculture, forestry, and fisheries	6,523
Animal and dairy science	1,755
Veterinary science	1,521
Other agricultural sciences	462
5. Social sciences	**43,164**
Psychology	2,282
Economics and business	17,886
Educational sciences	9,581
Sociology	2,462
Law	6,594
Political Science	2,539
Social and economic geography	446
Media and communications	265
Other social sciences	1,109
6. Humanities	**33,846**
History and archaeology	11,833

	No publications from 2008-2012
Languages and literature	11,222
Philosophy, ethics and religion	5,122
Art (arts, history of arts, performing arts, music)	5,669

Source: RD&I IS, 2014

2.3 The policy objectives of this study

This study is part of the IPN project 'Effective system for the evaluation and funding of RD&I', which has as objective

> "To propose a performance evaluation system as a whole and funding mechanisms for the entire system so that public support would contribute to an increase in the excellence of Czech RD&I, act as an incentive for all players, and support the growth of competitiveness of the Czech Republic."

In relation to the funding principles, in the online description of the project the IPN team indicates the strategic objective "to suggest a motivating funding system for RD&I with emphasis on excellence and medium-term planning."

The project also set the policy context for its activities: in the project progress report of August 2013 it stated,

> "The IPN project is based on the objectives and measures of the strategic documents of the Government (the Reform of RD&I, National RD&I policy for the years 2009 - 2015, International Competitiveness Strategy for the period 2012-2020, National Innovation Strategy), the recommendations of the IPN International audit, the Long-term principles of evaluation and funding approved by the RD&I Council, and good international practice.

An analysis of the currently valid strategic policy documents in the Czech Republic[6] allowed us to identify the key policy objectives of the evaluation methodology and funding principles, summarised in Exhibit 7, below.

Exhibit 7 Policy objectives of the evaluation methodology and funding principles

Objective category	Objectives
R&D capacity	To improve research and development management, at all levels
	To improve human resource development, reflecting the needs of the knowledge economy of the CR
	To strengthen cooperation between the RD&I actors at the national level, i.e. academic research, universities, applied research and the application sphere
	To strengthen international cooperation
Excellence in R&D	To motivate research organisations (ROs) to excellence
Societal relevance	To motivate ROs for collaboration with industry
	To motivate ROs for the transfer of knowledge to practice

[6] National RDI policy for 2010 – 2015 with an outlook to 2020; Operational programme Research, Development and Education 2014 – 2020; Operational programme Enterprise and Innovation for Competitiveness 2014 -2020 (MIT); National Reform Programme 2014 (Office of the Government); International Competitiveness Strategy 2012 – 2020 (MIT); National Priorities of Oriented Research, Experimental Development and Innovations (Office of the Government); Policy statement of the government of the Czech Republic 2014

Objective category	Objectives
	To stimulate ROs to research corresponding to the needs of society and the business sector

The National RD&I policy 2010-2015 also specified: "The new EM will take into account differences among different types of ROs and research fields/groups of research fields".

Based on the available documents and discussions, our reflections and conclusions on the policy objectives for our study, and in particular for the design of the Evaluation Methodology (EM) are:

- A first requirement is that the assessment informing the PRFS is trusted by the RD&I community. History suggests that this may be hard in the Czech Republic; the key elements needed are transparency and the use of disinterested peer reviewers
- The current system through which ROs are officially recognised as research performing and therefore entitled to some level of institutional funding is seen as undemanding and leads to the registration of quite a large number of small (arguably, under-critical) research entities. It is not the task of an evaluation methodology to take (implicit) policy decisions from this perspective
- A requirement is that the system should provide institutional research funding for all parts of the RD&I system that do research and are officially recognised as research organisations. Without this, the overhead and infrastructural needs of researchers will not be met and it will in practice be hard to sustain research. In the former British terminology, this funding was described as enabling ROs to provide a 'well-found laboratory', in which research can be undertaken using external or 'targeted' funding. Clearly, there has to be a mechanism that allocates such funding at any point in the system where it is agreed that research may be done. Inherently, some of this funding has to be provided ahead of research being done, so to build capacity it has to be allocated prospectively. This can be done through performance contracts. Once research capacity is in place, of course, its effectiveness can be tested using output and performance measures.
- At the same time as building and sustaining research capacity right across the RD&I system, there is a desire to identify areas of research excellence and to concentrate resources on these, building up capacity in areas of particular research strength. This implies an additional excellence incentive that redistributes some institutional funding, perhaps in the non-linear style of the UK RAE/REF
- Research is inherently a rather long-term activity, so there is a particular need for the institutional funding system to provide a degree of stability and freedom to make strategic investments (a) through the mechanism of performance contracts and (b) by damping excessive movements in institutional funding between funding periods. This latter element has to be built into the structure of the funding model: what proportion of institutional funding should be contestable in any one funding round?
- Increasing the quality, relevance, productivity and internationalisation of Czech RDI are goals across the whole system. In effect, they are extensions of the general capacity building goal, so they can be encouraged through the use of competitive incentives (in practice, using the metrics module of the assessment to trigger the peers)
- The desire to introduce a formative, advice-giving element into the assessment methodology by looking at the health of a selected number of research organisations or fields. The intention is to confine this component to a small number of leading organisations. In the absence of other criteria, cost may need to

be a factor in deciding how and how many ROs and fields to involve in this exercise

- The final goal of the EM is to provide information for policymakers, which can be achieved through additional analysis of the information the evaluation collects

Our conclusions specifically in relation to the role and function of the EM are

- The EM is expected to assess performance but is also conceived as a tool to steer the behaviour of the research actors and inform institutional funding
- The EM has a pronounced formative function. It is expected to act as a source of information for strategic management of RD&I - at the level of government, support providers, programmes, and research organisations
- The EM will include all research organisations in the Czech Republic of a minimal size that are officially recognised as such
- The EM will take into account the differences among types of ROs and disciplinary cultures

3. Concepts and core elements of evaluation in international practice

We reflect on the role of evaluation (Chapter 3.2), and then describe international practice in relation to the core elements that influence the design of any evaluation methodology, i.e. the *scope* of evaluation (Chapter 3.3), the evaluation *method* (Chapter 3.4), the evaluation *structure and level of analysis* (Chapter 3.5), the *focus* of evaluation (Chapter 3.6), indicators and assessment *criteria* (Chapter 3.8), and the risks and *risk management* (Chapter 3.8.6).

3.1 Introduction

In the chapters below, we set the context of the Evaluation Methodology and describe the basic concepts and international practice related to its core elements, i.e. the scope of evaluation, the evaluation method, the evaluation structure and level of analysis, indicators and assessment criteria. Finally, we cover the risks that are associated in particular to the use of the evaluation results, and describe the measures that are typically taken to mitigate them.

3.2 The role of evaluation

The European Commission defines evaluation as "a judgment of interventions according to their results, impacts and needs they aim to satisfy."[7] It highlights that evaluation is "a process that culminates in a judgment (or assessment) of an intervention" and indicates as main purposes

- To contribute to the design of interventions, including providing input for setting political priorities
- To assist in an efficient allocation of resources
- To improve the quality of the intervention
- To report on the achievements of the intervention (i.e. accountability)

The function and purpose of the evaluation is a key element that influences the choice for its core elements, i.e. the scope of the assessment (institutions, individual researchers etc), the assessment criteria and indicators, the methods used (peer reviews, metrics, etc), and the depth and breadth of the evaluation exercise.

In international practice, evaluation has increasingly become an integral part of the **policy cycle**. Depending on the policy needs, evaluation is expected to foster learning and improvement as well as to ensure accountability. It is expected to provide information to help design better policies and/or to assess performance in order to legitimise past initiatives. In most cases, evaluations therefore have both a *summative* and a *formative function*: they assess past performance (the summative function), analyse the factors that facilitated or hindered the achievement of the policy intervention objectives, and recommend changes to the intervention or new interventions (the formative function).

The term "policy cycle" does not refer only to national R&D policies. In fact, the increased institutional autonomy of research organisations in the new public management model, distributing the responsibility for research performance over the

[7] http://ec.europa.eu/dgs/internal_market/evaluation/evaluation/index_en.htm

various levels of the research system, led to an expansion of the *scope and use* of evaluation. Actors involved in the development of information for evaluation and interested in the use of its outcomes now span over **all levels** of the research system, from policy makers to the researchers themselves. This trend has led to the creation of a system of distributed intelligence; it also implied a diversification of the expectations and needs for evaluation.[8]

In relation to purpose, a major distinction needs to be made between evaluations that have as *only* purpose to assess research performance, and evaluation systems that are intended (also) to inform a performance-based research funding system (PRFS).

- Assessment systems, whose results do not directly feed into the allocation of institutional funding, have a pronounced 'formative' dimension in the sense that they provide research-performing organisations with feedback that they can use to improve performance while also providing research policymakers with strategic intelligence about the national research-performing system.
- Evaluations informing PRFS have a pronounced summative function. Steering research behaviour is a key function of a PRFS and in practice PRFS have different objectives, depending on the national context and policy strategies

The different purposes of these evaluation exercises and their potential effects define methodological choices in particular in relation to the indicators to be used, and the processes for the design of the methodology as such. We summarise these differences in Exhibit 8.

Exhibit 8 The difference in purpose and effects of 'general' evaluations and evaluations in PRFS

'General' evaluation	**Evaluation in a PRFS**
Informs policy making on the failures in the system and **recommends** possible policy interventions	Is part of a **policy intervention**: it acts upon previously identified failures and steers research behaviour to tackle these by providing incentives
Has **no** effects directly linked to the evaluation	Is **intended** to create effects
Has **no** consequences of gaming or unintended effects	Has **inevitable** consequences of gaming and may lead to unintended effects
Gives **information** on the positioning of the evaluated objects in the national/international context	Sets the evaluated objects in **competition** to each other

In the course of this study we analysed the national evaluation system in the 5 'comparator' countries that were selected for this study (Austria, the Netherlands, Norway, Sweden, UK). We also looked into the specific aspects of the practice in 5 more countries: Australia, Belgium/the Flanders, Finland, Italy and New Zealand (see the Country Analyses background report).

Exhibit 9, below, categorises these national evaluation systems in terms of their **purpose** (to assess research performance or to guide institutional funding or both), and their **function** (formative or summative or both). It highlights the strong variety

[8] Mahieu, B., Arnold, E., Kolarz, P., *Measuring scientific performance for improved policy making*, European Parliament, STOA, 2014

in the evaluation systems, determined by the specific policy needs in the national R&D context as well as the background of the R&D governance system they are part of.

- Austria and the Netherlands are examples of countries where research assessment is performed with no explicit link to institutional funding. In both countries, the key focus of the evaluations is on informing institutional R&D management
- Norway and Finland are examples of purely indicator-based PRFS (not only based on bibliometrics). However, they are complemented by a broad range of other national or institutional evaluations, providing the 'formative' information needed. To be noted is also that the PRFS drives only a small part of the institutional funding
- Belgium/the Flanders distributes the BOF fund for bottom-up basic research by means of a formula predominantly based on bibliometrics (the BOF key). This fund is additional to institutional funding and linked to a light-touch performance agreement. Since 2008 the BOF key is used also for the formula component of the universities' institutional funding. In 2011, it drove about 15% of the universities' institutional funding
- In 2003 and 2011, Italy has run an evaluation exercise similar to the UK in terms of size and depth. Both Australia and Italy make a selective use of bibliometrics instead of the panel system, i.e. bibliometrics for the hard sciences and peer review for the others
- Sweden is designing a new PRFS system, with the ambition to use bibliometrics only to inform the panels
- The UK and New Zealand are similar in that they both use firmly peer review based evaluation methodologies. A distinction is that the UK REF focuses on research excellence, the New Zealand RAE on research quality

Exhibit 9 Main characteristics of national evaluation systems

	Australia (2014)	Austria (2014)	Belgium (FL) – BOF (2014)	Finland (2014)	Italy (2014)	Netherlands (2014)	New Zealand (2014)	Norway / evaluations (2014)	Norway / PRFS (2014)	Sweden (2014)	UK – (2014)
Purpose											
Performance assessment	X	X	X	X	X	X	X	X	X	X	X
Inform funding			X	X	X		X		X	X	X
***Main* function**											
Formative	X	X		X		X	X	X		X	
Summative			X		X				X		X
Formative function											
National R&D governance	X		X		X			X	X		X
Institutional R&D management	X	X		X		X	X	X		X	
Summative function											
R&D quality	X	X	X	X	X	X	X	X	X	X	X
R&D capacity building	X		X	X	X		X	X	X		X
Research excellence	X										X
Societal relevance	X				X	X					X
Notes											

3.3 The scope of evaluation

One of the first steps in any evaluation methodology design is to identify and focus the scope of the evaluation, i.e. the object of evaluation and the actors it involves.

In the Chapters below, we first consider the distinction between evaluations at national and institutional level, their use of methods and focus of the evaluation results that can be expected. We then describe the efforts made in international practice for the definition of research organisation typologies and finally cover a minor but nevertheless important element for evaluation design, i.e. the use of thresholds for participation in evaluations.

3.3.1 Evaluations at national versus institutional level

A primary element for decision-making in evaluation design that is of particular relevance for this study is whether the evaluation should assess performance at the institutional or national level. This guides the choice of methods and indicators - depending on goals, objectives and therefore needs, as well as the focus and depth of the analysis on specific topics.

In an increasing number of countries, evaluations at the national level include also information on the performance of research institutions and their management. However, the depth of the analysis at institutional level is more limited than in evaluations conducted at the institutional level, and the focus of the assessment is determined by the needs for strategic information at the national level as well as the purpose of the evaluation, e.g. to inform a PRFS.

- **Evaluations at the institutional level** typically have the goal to help the universities' and research institutions' management understand their strengths and weaknesses and measure their research competitiveness, at an aggregate level and the level of their departments and research groups. The *focus* can be on internal quality assessment, evaluation of HR management processes, evaluation of internal research projects, or the identification of excellent research groups, depending on the needs. *Results* of these analyses feed into institutional strategy making (and eventually internal fund allocations) and help the institutions in their monitoring and reporting on outputs and impacts. Institutions also use this information for publicity purposes, i.e. to help student and academic recruitment, aid research partnerships (with private enterprises or other research institutions), and initiate or sustain investments. From a *methodological* perspective, these evaluations typically involve (informed) peer reviews with on-site visits
- **Evaluations at the national level** can have two functions or a combination of the two: a) the collection of information for the comparison and benchmarking of the national performance at an international level, and b) the collection of strategic information for the definition and monitoring of research policy strategies and interventions. Evaluations focusing on the latter can be at institutional, thematic, or disciplinary level, depending on the objective. The *evaluations* conducted in Norway are an example of the latter (see the Country Analyses Background Report). Methods used are metrics and bibliometrics, and peer review or a combination of the two. Seeing the scale of the exercise, peer review rarely includes on-site visits (see also Chapter 3.4.1, below)

3.3.2 Typologies of research organisations

Research organisations are an obvious scope for research performance assessments.

In the context of this study, the identification of different typologies of research organisations is of particular relevance. This is linked to both the purpose of the evaluation and its objectives, one of which is to take account of the differences among research organisations in terms of their "mission in society' (see Chapter 2.3, above)

From a methodological perspective, handling the difference in missions of the evaluated objects for the assessment of performance is a common feature of the evaluation practice. It is a common request in evaluations of policy initiatives to take account of the roles and missions of the different actors involved. It is common practice in the context of programme evaluations; also in disciplinary evaluations such as the ones implemented in Norway, different actors in the system are covered with due consideration for their positioning in the National Innovation System (NIS). In all of these cases, the organisations are assessed in relation to their **function** in the R&D system and/or the NIS, and their performance is measured against their related objectives and expected impacts – be they explicitly stated or not.

The issue becomes more radical for the PRFS, i.e. the institutional funding allocation. To the best of our knowledge, there is no PRFS that defines performance criteria for a broad range of actors. The typical approach is that separate budgets are established for the different categories of organisations and different criteria are defined for the formula, depending on their missions. An example is again Norway, where the institutional funding of three different types of organizations, higher education institutions, research institutes (mainly RTOs and public service labs), and research hospitals, is driven through three different PRFS with different funding principles (see the Country Analyses background report).

The rationale lies in the very nature of PRFS, i.e. its function of supporting the particular aims of the research organisations in question. Different actions may be required for different actors. Even if the same assessment criteria can be applied, the weights of the different criteria should differ, reflecting the functions in the R&D system of the different actors. We will cover this topic further in the Second Interim Report, focused on the Funding Principles.

In the official statistics, research-performing organisations are grouped into three main categories: Higher Education Institutes (HEIs), Public Research Institutes (constituting the 'government' sector), and private research organisations. This categorisation is based on the organisations' **legal status**.

Such categorisation is of little use when considering the societal mission of research of the non-university research organisations. A recent OECD study[9] highlighted the broad variety in activities and missions of these research organisations, ranging from institutes performing 'blue sky' science to institutes with a more short-term market-oriented research and those providing access to knowledge to industry, government, and other societal actors. The 2011 OECD report also shows that the distinction public/private research institutes, based on the legal status, is more often due to historical developments in the specific countries rather than indicating proper business models. Non-university research organisations may have different legal forms at different times and places. Some are foundations; others are limited liability companies that do not aim to distribute profits to shareholders. Others are associations or even state agencies. In some systems the institutes have changed their legal form without changing their social and economic function.

The picture that emerges from the 2011 OECD report very much reflects the one that can be seen also in relation to the Czech R&D base, as described in Chapter 2.2.2, above.

Another distinction among research organisations that is typically made is based on the **type of research** that the research organisations conduct, i.e. basic research,

[9] OECD (2011), Public Research Institutions – Mapping Sector Trends

applied research or development. In this context, the EC Expert Group[10] warns against simplification of a more complex discourse:

> *"*The growing complexity of knowledge and society has corresponded with blurring boundaries between vocational and classical higher education institutions, and between research and development. Simplistic distinctions between basic and applied research have been replaced by greater emphasis on strategic, regional and/or field specialisation. This diversity of research mission is reflected in the wide range of research outputs and outlets mapped across the full spectrum from discovery to knowledge transfer to innovation.*"*

In other words, categorisations of research organisations along the spectrum of basic research to experimental development cannot properly reflect a situation where many research organisations conduct a mix of research typologies. In this context, Czech readers will remember the change that has taken place in their research system over the last decades, with an increasing role of universities in the conduct of (also) basic research.[11]

A third criterion for categorisation is the **function** of the research organisation in the National Innovation System (NIS). Several studies[12] have taken this approach. Also the OECD and its Working Party of National Experts on Science and Technology Indicators (NESTI) is looking into the feasibility of using this approach for future statistical purposes, in order better to capture the contributions made by non-university research organisations to developments in research and innovation. The 2011 OECD report on Public Research Institutes is to be placed in the context of this work.

In summary, non-university research organisations typically are categorised into three main typologies: Scientific Research Institutes, Research and Technology Organisations (RTOs), and Government laboratories.

Scientific research institutes are research organisations such as the Max Planck institutes in Germany, CNRS in France or the institutes of the national academies of science in various of the new member states Historically, some of them have their origins in Research Councils or Academies of Science, which were simultaneously research-funding and research-performing organisations. They largely do the same kind of research as universities and correspondingly get a high proportion of their income in the form of institutional funding. In many parts of Western Europe, the funding and performing functions of Research Councils have been separated some decades ago.

Research and Technology Organisations (RTOs) tackle the needs of industry for knowledge and a range of knowledge-related services. Large-scale examples include VTT Finland, the Fraunhofer Society in Germany or TNO Netherlands, but there are also smaller and more specialised institutes. Their origins are often as testing

---

[10] Expert Group on Assessment of University-Based Research (2010), *Assessing Europe's University-Based Research*, European Commission, DG Research

[11] Arnold, E. (2011), International Audit of the R&D&I System in the Czech Republic, Synthesis report, Technopolis Group

[12] Examples are the studies that constituted the main reference for this section, i.e.: Tomas Åström et al, International Comparison of Five Institute Systems Faugert & Co Technopolis Ltd. 23 December 2008; Arnold, E., Barker, K., Slipersaeter, S., Research Institutes in the ERA, Technopolis Group/MIoIR/NIFU, 2010; Arnold, E., Clark, J., Jávorka, Z., (2010) Impacts of European RTOs - A Study of Social and Economic Impacts of Research and Technology Organisations, A Report to EARTO

laboratories, product and process developers for industry or branch-based research associations. From an innovation-systems perspective, RTOs place themselves in-between the university sector and industry and they tend to market themselves as intermediaries, interpreters or "bridge builders" between the two "sides".

The function of RTOs is rooted in the economics of research and the idea that 'market failure' makes it difficult for companies to invest in general forms of knowledge. Typically, the role of the RTOs is to assume some of the risks of industrial innovation, helping companies to go beyond what they would be able to do, based on their technological capabilities. Since the overriding purpose of RTOs is to promote industrial competitiveness by technological means, they can only do their job if they in fact are technologically capable and can offer firms inputs that are in advance of or otherwise superior to those available on accessible commercial knowledge markets. What in practice sets an RTO apart from a regular consultancy is the constant need for renewal of competence and capabilities as well as society's expectation that it is to work with unprofitable customers.

The 2011 OECD report as well as all other studies emphasise the RTOs' critical role for the countries' innovation and economic performance through their activities in creating, discovering, using and diffusing knowledge.

Government laboratories focus on producing public goods to meet knowledge needs of the state or wider society. Sometimes referred to as 'sector' institutes, they are generally owned by the state and their main function is normally to deliver services and policy-relevant information to government. Examples include nuclear research, marine institutes (which mix counting fish stocks with more fundamental work in marine biology) and metrology. Generally, the bulk of their income comes from the ministry whose policy mission they support[13]

Their typical role is in providing fundamental research in strategically important areas (*e.g.* nuclear research or public health), supporting public policy through precautionary research (*e.g.* into sustainable development or food safety), policy design and monitoring, supporting the building of technical norms or standards, and constructing, maintaining and operating key facilities.

Some government laboratories strongly support innovation, through certification, testing, monitoring and measurement, finding new uses of existing knowledge, creating links between scientific fields and establishing multidisciplinary knowledge bases (such as gene banks and quality-assured scientific collections).

The 2011 OECD report identified another category of non-university research organisations, i.e. ROs that have research only as a secondary function. These include entities with strong public-service goals (e.g. hospitals) or a strong cultural focus (e.g. museums and libraries). In some countries, these institutes are considered as integral part of the research system or research infrastructure (e.g. the research-oriented hospitals in Italy and the museums and libraries in Denmark).

Whereas the entities with public-service goals can be considered as a sub-category of the Government laboratories, those with a cultural focus are typically considered as part of the R&D system in their function of providers of infrastructure for research. They therefore constitute a fourth RO category.

The categorisation above is not perfect: especially larger "national" RTOs, which play an important "infrastructural" role in their respective country have several distinct missions. These RTOs typically combine, for example, advice to government, public laboratory services (e.g. assaying, norms and standards), condition surveillance (e.g.

[13] Simmonds, P. (2008) *Activities of the EU Member States with Respect to the Reform of the Public Research Base*, Report of the ERAWTCH ASBL, Brussels: European Commission, ERAWATCH service

environmental monitoring), facilities hosting as well as strategic research and contract R&D for enterprises.[14]

In this context, **Higher Education Institutes** (HEIs) are considered integral part of the scientific research component in the R&D system, including the university hospitals. They perform a wide range of roles, responsibilities and activities, and cut across different economic, political and social networks.[15]

Their primary mission, however, is education and in this context they have an additional role to play in the NIS, i.e. the education of the future researchers - to the benefit of the research and the industry sector alike. Many governments and universities therefore strongly support the interconnection between teaching and research as one of the core principles of higher education, and encourage stronger partnerships and knowledge exchange between business and academia. The concept of the Entrepreneurial University is an illustration of this approach.

3.3.3 The use of thresholds

When considering the scope of research assessment systems, an aspect that should be considered is the size of a unit of evaluation, especially as measured by research outputs. There are various factors that need to be taken into account:

Firstly, a proper evaluation exercise is a particularly **time and resource-consuming** endeavour. This is true for the body carrying out the assessment, as it means that a considerable larger number of units of evaluation would be assessed if only those of a certain minimum size were considered. However, this resource factor is even more significant for the units of evaluation themselves, especially relatively small units. For small-scale research groups or organisations with minor research components in a certain field, the resources involved in drawing together a submission for a research assessment can provide a potentially insurmountable burden, or at the very least call into question whether participation in the assessment is worthwhile at all.

Especially in the context of sophisticated assessment systems, the presence of small units can present a problem also to the assessors themselves: for certain types of assessment data, units with a small number of overall outputs can decisively hamper the **robustness** of results. This is especially true in the case of metrics and bibliometrics. Regardless of whether the institution, the department or the field is the unit of evaluation, there is a certain minimum for robustness and validity. Below such a minimum, it becomes harder to identify statistical outliers and a single output can decisively skew the overall result.

In the case of bibliometrics, 50 research outputs generally presents a suitable minimum threshold for meaningful analysis to take place. This threshold is, for instance, always used by CWTS, Leiden University, in their bibliometrics to support institutional evaluations across Europe.

International practice highlights some approaches to minimum thresholds as eligibility criteria for participation in evaluation, in the overall majority of cases expressed in terms of research outputs. Australia takes a possibly interesting

---

[14] Research and Technology Organisations in the evolving European Research Area – a status report with policy recommendations, European Association of Research and Technology Organisations - EARTO, 2013

[15] Jordi Molas-Gallart et al, (2002) Measuring Third Stream Activities, Final Report to the Russell Group of Universities, SPRU, University of Sussex

approach: the ERA2015 defines Low Volume Thresholds (50 indexed apportioned journal articles or apportioned weighted outputs) at the level of field **or** sub-field. It hereby accommodates two needs: smaller units may reach the threshold at the field but not sub-field level, while large departments may want to be broken up for assessment purposes at the sub-field level (see the Country Analyses background report).

However, it should be noted that many countries do not apply thresholds for participation in evaluation. The reasons are simple:

- The R&D system is sufficiently concentrated and there is no significant presence of small units (meaning that there is no need for thresholds for robustness purposes), or most often,
- The national evaluation covers only the major actors in the R&D system, typically the universities. Other types of research-performing organisations are assessed separately (and differently).

3.4 The evaluation method

In this chapter we cover the three main methods for evaluation of research performance: peer review, bibliometrics, and informed peer review.

We first reflect on the use of expert panels and the strengths and weaknesses of this method. We then reason on bibliometrics, its use in international practice and the different approaches and issues arising. Finally, we cover the third method, which combines the two methods, i.e. informed peer review.

3.4.1 Peer review

There are good reasons why expert panel evaluation is still the preferred practice in the UK RAE/REF and other research assessments. Notions such as quality and impact are best **judged** by experts rather than metrics. This relates in particular to the capacity of expert panels adequately to assess the performance of actors in the different fields (and sub-fields) of science (see Chapter 3.6, below). Peers possess the needed specific knowledge and understanding to take into account the specifics of the disciplinary cultures, ranging from different publication profiles to the needs for research infrastructure, as well as the roles of the different actors in the R&D system for the field, and can position it in an (international) quality framework.

The EC Expert Group considered:

> "Assessing research quality requires a detailed understanding of the knowledge cluster, in order to evaluate the methodological soundness of the research and the (potential) significance of its contribution to knowledge. Only peers tend to have such an understanding, and this is why peer review has always been an important quality control instrument."[16]

However, with the expansion of the research system and the assessment procedures, peer review is under considerable strain. A practical problem widely discussed is that the explosion in peer review effort required by research funders, research evaluators and the overall growth in the numbers of scientific journals and publications is placing high demands upon the capacity of the community to undertake peer review.

A partial solution to this issue is the **'remote'** panel evaluation. As mentioned in Chapter 3.3.1, above, panel evaluations are common practice in institutional

[16] Expert Group on Assessment of University-Based Research (2010), *Assessing Europe's University-Based Research*, European Commission, DG Research

evaluations. Typically, they include on-site visits as an opportunity for interaction with both the organisation's management and its academics, which allows the expert panels to deepen their understanding of the institutional environment.

This use of on-site visits is seldom the case in national evaluations covering all themes and fields, due to the scale of the exercise. Also in the case of other evaluations, including institutional ones, an increasing number of countries started adopting the *'remote'* panel evaluation model. This reduces the costs of the evaluation as will as the burden on the reviewers in terms of time investment. In most cases, the interaction among the experts themselves is kept because of its importance in the decision-making process, so a (minimal) number of physical panel meetings are foreseen. Such an approach has been adopted, for example, in the UK RAE/REF exercises, disciplinary evaluations in Norway, and the newly developed Swedish FOKUS.

Peer review also has its inherent problems: it is costly and time consuming, prone to bias, leaves no audit trail and can even be open to abuse. A study of review panels in operation[17] highlights other important weaknesses:

- Selection of panel in/excludes 'schools' of thought
- Time limits set by the organisers affect outcomes
- Tacit negotiations and compromises affect decisions - disagreements among peers get swept under the carpet
- Those who feel they have less knowledge rate more positively
- Division of labour within panels means some judgements are made by individuals, not the full panel

These weaknesses can be neutralised by means of procedural guarantees. We cover international practice from this perspective in Chapter 3.9, below.

3.4.2 The use of bibliometrics

The weaknesses of the panel method are exactly the kinds of problems that bibliometrics seek to address.[18] However, there may be problems with the bibliometric data coverage in certain fields of research, and there is no general agreement that indicators based on citation counts can capture what is understood as research quality.

In this context one should note that in the countries using metrics-based PRFS (the Flanders, Finland and Norway – see the Country Analyses background report), these systems are **not** regarded as research evaluation and they are instead complemented by other evaluation systems of a more formative character that are based on informed peer review.

Our sample of international assessments informing PRFS and their use of bibliometrics also suggests that policymakers generally fail to adopt many of the more sophisticated indicators the bibliometricians can provide. We summarise our findings below.

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

---

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

[18] For a thorough professional discussion of the possibilities and limitations in the use of bibliometrics for research assessment, see: Moed, Glänzel, and Schmoch: Handbook of Quantitative Science and Technology Research (Kluwer 2004) and Moed: Citation Analysis in Research Evaluation (Springer 2005).

this context we refer to the San Francisco Declaration on Research Assessment (2013), published by the international bibliometrics research community, which warns against the inappropriate use of Journal Impact factors, at all levels of the R&D system. A theme that runs through these recommendations is the need to assess research on its own merits rather than on the basis of the journal in which the research is published.

The use (and abuse) of bibliometrics for research performance assessments spurred to bibliometric community also to the development of the Leiden Manifesto, which states in its draft version of September 2014):

1. Metrics should be properly used to support assessments; **they do not substitute for judgment.** Everyone retains responsibility for their assessments.
2. It is easy to underestimate the difficulty of constructing accurate data. Spend the **time** and money required to produce data of high quality. Those mandating use of metrics should be able to provide assurance that the data is accurate.
3. Metrics should be **transparent**, the construction of the data should follow a clearly stated set of rules. Everyone should have access to the data.
4. Data should be **verified by those evaluated**, who should be offered the opportunity to contribute explanatory notes if they wish.
5. **Sensitivity to field differences** is important. Metrics will differ by field. Humanists will not be able to use citation counts; computer scientists will need to ensure conference papers are included; and chemists will look the best in raw metrics constructed from Web of Science data. The state-of-the-art is to select a suite of possible indicators and allow fields to choose among them.
6. **Normalize** data to account for variation in citation and publication rates by field and over time.
7. Metrics should align with **strategic goals**

International practice in this context is as follows:

- The Belgian (Flemish) BOF counts numbers of publications and citations in the Web of Science (WoS). It weights publications using JIFs as part of the funding formula. A Flanders-specific database of scholarly articles in journals, articles in books, and books, in the social sciences and humanities is used to extent the list of 'approved' journals beyond those internationally indexed
- Denmark, Finland and Norway, all of them using the "Norwegian model", also extend the data coverage to all scholarly peer-reviewed publications in journals, series, book chapters, and books, but makes no separate calculation for WoS publication (as in Flanders). The publications are weighted into "publication points", taking into account both publication type, number of authors and the level of the publication channel: Level I (the "normal" level, representing 80 per cent of the publications and level II (only for publications in the most prestigious international channels in each field). The weighting system is supposed to balance between different publication practices, thereby reducing the need for field normalisation.
- Italy (VQR) uses an informed peer review process for the greater part of the funding, based on outputs submitted by individual research organisations to the evaluation process, but complements this with indicators to allocate the balance. The indicators used are ScImago journal rankings and Journal Impact Factors and number of citations per article relative to the average in its field. The results of the peer review and bibliometric exercises are used separately to allocate units of assessment into broad quality bands, and the combination of these bands with the volume of output then drives the funding provided

- Sweden's system is being revised as we write, but the most recently applied system drives institutional research funding 50% based on external research income and 50% based on output indicators. The latter is based on WoS-data only and comprise: number of publications; the percentage of publications that are not cited (or cited only by the authors themselves); the proportion of self-citations; field-normalised citation rates; the percentage of publications in the 90th, 95th and 99th percentiles for citations in their fields; journal-normalised citation rate; journal to field normalised citation rate
- UK – REF uses peer review, informed in the case of those panels that desire it by citation counts and contextual analysis to help clarify citation behaviour and patterns in the relevant field. Use of journal impact factors and other bibliometric indicators not supplied through the REF administration is forbidden. One panel (Computer Science and Informatics) had planned to use Google Scholar data as a way to capture more of the conference activity that is central to the way that field communicates, but was defeated by inability to collect the needed data

3.4.3 Informed peer review

It is reasonable to say that both peer review and bibliometrics (as about any other evaluative technique used in relation to R&D) have their weaknesses and strengths. As the Dutch Committee on Quality Indicators in the Humanities stated,

> 'The most serious objections to peer review can be neutralised by means of procedural guarantees and by utilising external indicators [bibliometrics] that give an inter-subjective basis to the judgment of peers.'[19]

Since at least three decades, the international practice has been to **combine** the two approaches, with bibliometrics informing the peer review. The evaluation methodology hereby exploits the ability of indicators to represent large sets of data in a simplified overview 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.

This 'triangulation' approach is standard in modern R&D evaluation practice. It is common practice for the professional evaluator to use multiple evaluation techniques and methods in parallel and compare their results before reaching a final judgment.

3.5 The structure and level of analysis

In this chapter we first describe factors influencing the definition of the unit of evaluation and subsequently, the approach in international practice to the inclusion in the analysis of the individual researcher.

3.5.1 The unit of evaluation

The unit of evaluation can be the individual researcher, a research group (field defined), the faculty, the department, or the institution.

The selection of the most appropriate unit of evaluation is to a large extent dependent on the purpose of the research assessment. In 'general' evaluations, the deciding factor is the most suitable level for the collection of the information that is required for policy

[19] Committee on Quality Indicators in the Humanities (2011), *Quality indicators for research in the humanities*, Interim report, KNAW - Royal Netherlands Academy of Arts and Sciences, NL

making or governance; in PRFS it is driven by the level at which the funding is allocated.

In research performance assessments, there are various sets of considerations that leave most systems with the task of balancing emphasis on institutions on one hand and scientific field or research group on the other.

A major factor for decision-making in this regard is the evaluation model. In fact, bibliometrics and panel evaluations are variably flexible from this perspective.

Metrics-only systems can collect data about outputs at the level of the individual researcher, as is most obviously done in the countries that operate a national research information system for this purpose, such as Norway and the Czech Republic.

To our knowledge, nobody allocates institutional funding at this individual level – that would be unreliable and erratic. 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). However, allocating institutional funding to intra-organisational entities would challenge the principle of the autonomy of universities and other research organisations, so this tends not to be done.

Peer review systems could have this same flexibility only if they had the capacity to assess all the output produced by each individual researcher – which is impossibly resource intensive. To our knowledge, the only country adopting such system is New Zealand (see the Country Analyses background report).

Research organisations therefore **select** what outputs they submit for peer review, so this becomes an act of research management rather than one of individual performance. Where assessment systems ask for contextual information (such as the appropriateness of the available research equipment, group research income and so forth) the unit of analysis also has to be a collective rather than an individual. Since peer review assessment works using discipline or 'field' panels, it cannot relate directly to the overall performance of a research organisation –though the funding outcome certainly is organisationally connected.

3.5.2 The inclusion of individual staff

Different evaluation systems take different approaches to the question whether and how all of the individual researchers should submit a selection of their research outputs for review.

Broadly, there are two different approaches:

- The evaluation is comprehensive, i.e. all researchers at an institution must submit a selection of their work. Systems that take this approach typically specify clear inclusion criteria, including most often a minimum level of professional attachment to the institution (eg at least 0.2 full-time equivalent contract in New Zealand), as well as other criteria detailed below where applicable. This approach allows for a relatively representative overview of the outputs, quality and/or impact of research within the unit of evaluation. Exhibit 10, below, shows that this approach is taken in Australia, Finland, Italy and New Zealand
- In the UK RAE, the units of evaluation are expected to identify a smaller selection of its researchers who will then submit their work. This normally means - implicitly or explicitly - that the 'best' researchers' work will be put forward. This reduces the burden on the evaluators, as the overall amount of submitted work is smaller. Conceptually, this approach does not give a representative overview of all research activity that has occurred in an evaluation unit, but instead indicates the maximum standard that the unit is capable of

There are weaknesses to both approaches: comprehensive inclusion of staff may for instance obscure the presence of a select few outstanding researchers in an overall average unit, whilst selection of the best examples may obscure a bulk of relatively poor quality research also happening in the unit.

Importantly, it should be noted that other evaluation systems avoid setting the level of assessment down to the individual researcher level. Research outputs are submitted for peer review based upon a selection of the work at the level of research group or department, not at the level of the individual researcher.

A common concern is expressed in all countries where individual researchers and the results of their research activities constitute the level of analysis. This is especially the case in evaluations that inform a PRFS, such as in the UK RAE, Australia and Italy.

Even though the assessment of individuals is not an objective in any of these evaluations, the inclusion of results at the individual level has created unintended and negative effects on career prospects and the R&D system as such.

In consideration of these negative and unintended effects, the UK REF changed its approach and now considers the work of research groups rather than individual researchers.

Exhibit 10 Individual researchers included in evaluations - Australia, UK, Finland, New Zealand and Italy

		Australia	**UK**	**Finland**	**Italy**	**New Zealand**
Academics included in evaluation	All academics*	Yes		Yes (indirectly)	Yes	Yes (for submission, not evaluation)
	Selected sub-group of academics		Yes			
	Criteria for selection	There are detailed eligibility although the objective is comprehensive submissions	Units select the academics to be included. Beyond that: 'Category A' specifications must apply, special circumstances for early career researchers; minimum 0.2 FTE		Academic staff (gradations apply depending on rank/ seniority	All eligible academics are included in the assessment of the TEO. The EO only excludes the academics (with low score) that do not add value to the overalls score of the university department

3.6

3.7 The focus of evaluation: scientific disciplines

This chapter focuses on the differences among the scientific disciplines and their implications for research performance assessment.

Differences among disciplinary cultures derive from the history of the disciplines or research fields and are influenced by their size and the way in which research is conducted. Schmoch (2010) worded it: "The scientific production process has a complex structure which is shaped by *technical* and *social* influences."[20] They are expressed in terms of output types, main publication patterns, channels and timelines, citation behaviours, language of publication, collaboration behaviours and needs, intensity of the use and need of (human and financial) resources and research infrastructure, etc.

In the sections below we cover three major topics that are of particular relevance to this study: the differences in collaboration patterns, publication and citation practices, and societal missions. In the final sections we briefly describe how these differences among the scientific disciplines are handled in evaluation practice, in general and more specifically in relation to interdisciplinary research.

3.7.1 Collaboration patterns

Building upon data collected in the first national research assessment in Italy, Franceschet and Costantini (2010)[21] studied how scholar collaboration varies across disciplines in science, social science, arts and humanities and the effects of author collaboration on impact and quality of co-authored papers. Their analysis showed that collaboration intensity neatly varies across disciplines:

- The intensity of research collaboration is negligible in arts and humanities: the set of paper co-authors is frequently a singleton
- Social scientists often work in team, sharing competencies and other resources, but collaborations are smaller in scale and formality compared to science disciplines
- By contrast, collaborative work is heavily exploited in science, in particular in physics and medicine
- Collaboration is, however, moderate in mathematics, computer science, and engineering

There are differences also *within* the broad disciplinary areas: in the social sciences, collaboration in research has become the norm in psychology while philosophers are more inclined to work alone; a 'collectivistic' practice is noted in some subsectors of physics (e.g., high energy physics), where the scale and the complexity of research projects imply a professional organisation in large teams.

These differences in collaboration behaviour have important implications for the 'natural' organisational unit of research for assessment. The EC Expert Group on Assessment of University-based Research (further: EC Expert Group) considered:

> "In the life or physical sciences, the basic unit tends to be the research group. In contrast, in many parts of the humanities, research tends to be conducted on an individual basis, and the individual constitutes the natural unit of

[20] Schmoch, U. et al, (2010) How to use indicators to measure scientific performance: a balanced approach, *Research Evaluation*, 19(1), March 2010, pages 2–18

[21] Franceschet, M., Costantini, A. (2010), The effect of scholar collaboration on impact and quality of academic papers, *Journal of Informetrics* 4 (2010) 540–553

research. In clinical medicine, the unit tends to be a multi-disciplinary project group, and one individual can participate in several groups."[22]

3.7.2 Publication and citation practices

Research fields show a high level of heterogeneity also in their publication practices and profiles Some fields (especially in the humanities) publish in monographs or books; others (notably the basic sciences) in journals. Applied scientists and engineers often communicate more via conference proceedings than through learned journals. Mathematicians write few but extensive articles; chemists produce many, short articles. While in the biomedical sciences hardly any researcher publishes a book, historians publish about 60% of their research in books rather than journals. Also, a good biomedical researcher will be able to publish around five articles from a given research project, whilst for instance in engineering this ratio is significantly lower. Some fields have a large canonical literature that needs to be cited, others – often the newly emerging ones – do not have this. Additionally, when new sub-fields or interdisciplinary areas of interest arise, they rarely fit into the established publication channels and high impact factor journals of their overarching discipline, whilst their own niche-journals can take time to become established.

The differences in communication practices among the scientific disciplines regard a broad set of aspects, ranging from the preferred form, outlet and publication channels to publication propensity and citation practices.

The EC Expert Group mapped the *primary* forms of communication in the major discipline groups as shown in Exhibit 11, below. In this context it highlighted that an evaluation focus on journal articles only cannot do justice to the contribution of all scientific disciplines.

Exhibit 11 Primary Form of Written Communications by Discipline Group

	Natural sciences	Life sciences	Engineering sciences	Social sciences and humanities	Arts
Journal Article	X	X	X	X	X
Conference Proceedings			X		
Book chapters				X	
Monographs/Books				X	
Artefacts					X
Prototypes			X		

Source: Expert Group on Assessment of University-Based Research (2010)

The most obvious and most frequently cited examples in the literature in terms of **publication practices** are the differences that exist between the natural and physical sciences on the one hand, and the humanities, arts and social sciences on the other. However, the picture is more complex and differences in publication behaviour have been identified also at the discipline and sub-discipline levels. Mutz et al (2013) saw great differences in particular *within* the natural sciences and humanities and concluded: "There are not only differences between scientific disciplines in the research output profiles; there is also great heterogeneity of research output profiles within disciplines and segments of disciplines, respectively."[23] Butler (2007)[24] came to

[22] Expert Group on Assessment of University-Based Research (2010), *Assessing Europe's University-Based Research*, European Commission, DG Research

[23] Mutz, R., Bornmann, L., Daniel, H-D., (2013), Types of research output profiles: A multilevel latent class analysis of the Austrian Science Fund's final project report data, *Research Evaluation* 22 (2013) pp. 118–133

a similar conclusion and noted that, for example, in high-energy physics, theorists tend to publish more frequently than experimentalists.

Considerable differences exist also in the **citation practices**. Sandström and Sandström (2009)[25] state, "It is well known that medical researchers tend to produce more, often shorter papers where methodology and prior knowledge is codified in citations; and engineering scientists produce less frequently and have fewer cross-references." Referring to the journal impact factors (published, for example, by Thomson Reuters in its Journal Citation Reports), the EC Expert Group mentions, "In mathematics, a journal impact factor of 1.0 is high whereas in biochemistry journals with an impact factor of 1.0 is in the lower range. In the social sciences and humanities, journals tend to have impact factors below 1.0."

Other topics covered by the EC Expert Group in this context include the **scope of research**, which inevitably influences the incentive to publish internationally ('national' disciplines, e.g. studies on history, literature, language, law, versus 'global disciplines'), the **language** of publication (English for the natural, life and technical sciences; the national language for certain parts of social sciences and humanities), and the **time span** of relevant research, i.e. the time span over which the research is relevant and cited by other researchers (in the natural and life sciences, normally 5-10 years; in the social sciences and humanities sometimes 10 years is considered too short).

The data below give a more detailed view on the differences across research fields in terms of expected types of output and forms of publications, including their coverage in the international bibliometric databases. The data refer to scholarly publications in Norway, defined as "Scholarly publications, usually peer-reviewed before publishing, where the author(s) present their own new and original research findings in a format that allows for a critical assessment and use of the findings by other researchers in further research."

Exhibit 12, below, shows the large variations in publication patterns across research fields and major areas with regard to use of foreign language, publication type, and Web of Science coverage.

Exhibit 12 Scholarly publication patterns

Major area	Field	WoS coverage of all publications	WoS coverage of journal articles	Foreign language	Articles (ISSN)	Articles in books (only ISBN)	Books
Engineering	Engineering	63 %	86 %	97 %	74 %	26 %	0 %
Health Sciences	Biomedicine	97 %	98 %	100 %	98 %	2 %	0 %
	Clinical Medicine	94 %	95 %	83 %	99 %	1 %	0 %
	Dentistry	57 %	57 %	64 %	99 %	1 %	0 %
	Neurology	95 %	99 %	99 %	97 %	3 %	0 %
	Nursing Sciences	40 %	47 %	54 %	86 %	14 %	1 %
	Pharmacology and Toxicology	88 %	91 %	93 %	98 %	2 %	0 %
	Psychiatry	79 %	84 %	92 %	94 %	5 %	1 %
	Psychology	49 %	65 %	68 %	76 %	22 %	2 %
	Social Medicine	63 %	72 %	79 %	87 %	12 %	1 %

[24] Butler, L., (2007), Assessing university research: a plea for a balanced approach, *Science and Public Policy*, 34(8), October 2007, pages 565–574

[25] Sandström,U., Sandström, E., (2009) The field factor: towards a metric for academic institutions, *Research Evaluation*, 18(3), September 2009, pages 243–250

	Social Work and Health Care	9 %	20 %	33 %	43 %	51 %	6 %
	Sports Sciences	62 %	79 %	91 %	79 %	20 %	0 %
	Surgery	93 %	96 %	100 %	97 %	3 %	0 %
	Veterinary Sciences	87 %	88 %	89 %	98 %	2 %	0 %
Health Sciences	All subfields	75 %	84 %	81 %	90 %	9 %	1 %
Humanities	Archaeology	11 %	22 %	50 %	49 %	47 %	4 %
	Architecture and Design	5 %	8 %	44 %	59 %	36 %	6 %
	Art History	9 %	18 %	44 %	51 %	39 %	10 %
	Asian and African Studies	9 %	21 %	89 %	45 %	48 %	7 %
	Classical Studies	7 %	11 %	50 %	65 %	31 %	4 %
	English Studies	18 %	51 %	86 %	35 %	59 %	6 %
	Ethnology	4 %	9 %	34 %	47 %	46 %	7 %
	Gender Studies	6 %	14 %	31 %	43 %	56 %	1 %
	Germanic Studies	10 %	27 %	96 %	38 %	54 %	8 %
	History	16 %	33 %	36 %	48 %	46 %	7 %
	Linguistics	21 %	36 %	75 %	59 %	38 %	3 %
	Literature	10 %	17 %	28 %	58 %	39 %	3 %
	Media and Communication	3 %	8 %	52 %	38 %	55 %	8 %
	Music	8 %	16 %	34 %	51 %	45 %	5 %
	Philosophy	7 %	12 %	38 %	58 %	34 %	9 %
	Religion and Theology	7 %	14 %	39 %	48 %	45 %	7 %
	Romance Studies	18 %	45 %	82 %	40 %	51 %	9 %
	Scandinavian Studies	0 %	1 %	12 %	30 %	64 %	6 %
	Slavic Studies	6 %	12 %	86 %	50 %	44 %	7 %
	Theatre Studies	9 %	14 %	50 %	60 %	39 %	2 %
Humanities	All subfields	9 %	18 %	44 %	47 %	47 %	6 %
Natural	Biology	85 %	89 %	97 %	96 %	4 %	0 %
	Chemistry	96 %	99 %	100 %	97 %	3 %	0 %
	Geosciences	92 %	96 %	99 %	95 %	4 %	0 %
	Informatics	22 %	55 %	93 %	40 %	59 %	1 %
	Mathematics	75 %	85 %	96 %	88 %	11 %	1 %
	Physics	94 %	96 %	99 %	97 %	3 %	0 %
Natural	All subfields	81 %	90 %	97 %	90 %	10 %	0 %
Social Sciences	Anthropology	12 %	22 %	65 %	56 %	37 %	7 %
	Business and Administration	18 %	32 %	61 %	58 %	38 %	4 %
	Economics	55 %	69 %	78 %	80 %	18 %	1 %
	Educational Research	7 %	14 %	33 %	49 %	45 %	5 %
	Geography	35 %	44 %	76 %	78 %	19 %	2 %
	Law	2 %	3 %	27 %	64 %	29 %	8 %
	Library and Information Science	33 %	39 %	93 %	85 %	14 %	1 %
	Political Science	27 %	60 %	64 %	45 %	51 %	4 %
	Sociology	12 %	26 %	39 %	45 %	50 %	6 %
Social Sciences	All subfields	18 %	30 %	49 %	60 %	36 %	5 %
All fields	All subfields	48 %	67 %	71 %	72 %	25 %	3 %

Notes: Based on the registration of 70,500 scholarly *publications in Norway's Cristin-system since 2005*

Source: Sivertsen, G. (2014)

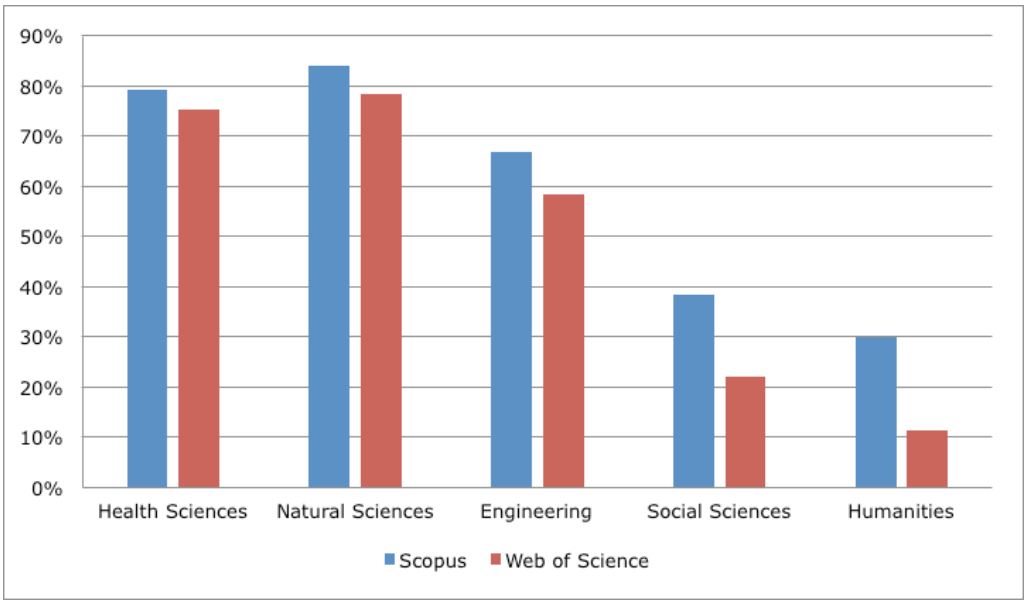
Publications in books are more frequent in the humanities and social sciences, but there are also large field variations within these major areas. In engineering, articles in books (ISBN) represent the strong tradition of publishing in peer-reviewed conference proceedings. In the other major areas, journal articles are the most frequent publication type.

The use of foreign language will depend on the international relevance versus the local societal relevance of the research being performed. The use of the national language is more frequent in the humanities and social sciences (again with large variations

among fields) and in the health sciences (with the presence of scientific articles also in national professionally-oriented journals).

With regard to Web of Science (of Thomson Reuters) coverage, there are large differences among areas and fields that mainly follow the same patterns as with publication types and use of foreign language. This is because the core idea of Web of Science is to index the scholarly literature that is covered in a core of international journals with articles that often are cited in the same literature. Scopus (of Elsevier) is known to have a wider coverage of journals; however, as shown in Exhibit 13 (based on the same data), the patterns of deficiencies are the same.

Exhibit 13 Coverage in Scopus and Web of Science



Notes: Based on the registration of 70,500 scholarly *publications in Norway's Cristin-system since 2005*

Source: Sivertsen, G. (2014)

3.7.3 The societal missions

A primary element that influences the *scope* for the societal relevance of research is the focus of the investigation. Different fields of science have different contributions to make for the solution of societal challenges that national systems, Europe or even the world are facing. These range from threats to the European social welfare model as costs rise because of ageing and the ever-increasing demands on health and care services, to longer-term global challenges which will continue to increase in significance but which also require immediate attention. First amongst these are global issues like climate change, sustainability and the efficient management of limited resources.

Whereas all fields of research (and types of research) have a societal mission to fulfil, the major differences among the disciplines are in the **directness** of the effects on society and the **time span** needed for these effects to occur.

A key function of science and research is to create and transfer **knowledge**. In a report for the Russell Group, Molas-Gallart, J. et al. (2002)[26] highlighted the differences among the disciplines related to the ways in which knowledge can be applied and used outside academia. They mentioned:

- Differences between applied disciplines (like mechanical engineering, business administration or medicine) and fundamental theoretical disciplines (like theoretical physics or philosophy) are that in applied sciences direct channels of application may exist; in the theoretical disciplines, impact on the economy and social welfare is likely to be more long-term and indirect
- Emerging areas of economic activity are characterised by low market entry barriers and a direct connection can be made between scientific activity and the exploitation of scientific discoveries (e.g. through industrial start-ups and university spin-offs). This is common in science-based emerging sectors such as biotechnology and information technologies
- In more mature sectors, where barriers to entry are very high, academic discoveries of direct relevance to industry are likely to be protected through patents, and then commercialised. Studies show that IP exploitation is of greater importance for a few science-based sectors, such as pharmaceuticals, chemicals and parts of electronics, i.e. the sectors that are highly dependent on scientific advances and rely on patents as a source of competitive advantage. Patenting and patent commercialisation activities can be expected to be higher in some fields (e.g. pharmacy) than in others (e.g. industrial engineering)
- Channels of diffusion and application of knowledge that are common in, but not unique to, the social sciences are the use and exploitation of the capabilities through the application of skills and tools to specific societal problems
- Many of the activities of Medical Schools can be classified as 'societal-oriented', in particular, the linkages with university hospitals and the role that these hospitals play in the training of future doctors and in the running of research projects

The patterns for publication of the non-scholarly outputs[27], i.e. *outputs that provide for societal or commercial use of research*, are a useful indicator to illustrate field differences in research missions, independently of the type of research. In research that is funded and expected directly to meet societal needs, non-scholarly outputs, e.g. commissioned reports or publications for wider audiences, can have high relative importance and be even more frequent than scholarly publishing.

- In general, scholars in the social sciences and humanities more often publish directly (as authors of their own publications) for a wider audience in the societies and cultures that they relate to in their research. These publications are both popularisations of research in general (most frequent in the humanities) and professional communication of specific results, e.g. in commissioned reports (most frequent in the social sciences). In addition, nationally adapted textbooks for students are often preferred over international standard editions in the social sciences and humanities. Consequently, their scholars more often appear as

---

[26] Molas-Gallart, J. et al. (2002), Measuring Third Stream Activities - Final Report to the Russell Group of Universities, SPRU, University of Sussex

[27] There is no comprehensive quality-assured data set allowing for a quantitative comparison of non-scholarly outputs across fields. The knowledge of these patterns comes from various surveys, evaluation processes, or overviews of the contents of national information systems such as the R&D IS.

authors of textbooks and other educational material. Note that there are also field differences: History usually has a wider national readership than classical studies or general linguistics, and experts in economics or law are more often asked to produce commissioned reports than experts in anthropology or media studies

- The societal impact and relevance of the natural sciences, the health sciences and engineering may be high indeed, but not so easy to trace back in non-scholarly publications authored by the researchers themselves. In addition to what may be counted apart from publications, a report on extra-mural activities and collaboration in the self-evaluation may be just as important information for an evaluation panel
- One should also expect field differences in non-scholarly outputs because of the different roles they take in society. Some fields have a more widespread societal use in education (mathematics or Czech reading and writing versus geology or pharmacology), in professional practice (dentistry and law versus astrophysics and limnology), or in societal and commercial use (informatics or oncology versus palaeontology or philosophy)

3.7.4 Implications for evaluation

Handling the complexity of the differences among the different scientific fields is a challenge for any national evaluation exercise and for obvious reasons, the topic is sensitive especially in the context of evaluations that drive PRFS.

In the international practice (see the Country Analyses background report), two models emerge:

- Systems like the UK Research Assessment Exercise (RAE) and a number of others, instead, solve the issue by make explicit allowance for field-specific characteristics through the establishment of discipline-based panels
- Many bibliometric-based systems try to overcome field-differences in publication patterns by introducing a system of weights (publication points) that balance the differences in publication patterns and most important, by presenting scholarly publications with complete data from research information systems (not only relying on commercial data sources). The latter is in order to compensate the differences in the coverage of scientific fields in the commercial data sources, i.e. WoS and Scopus (see also Exhibit 13, above)

The approach taken in the UK is worth considering in this context. Throughout its history, the UK RAE has gone a long way in allowing the disciplinary panels to establish variations in assessment methodologies within an overall 'egalitarian' framework.

- It allowed the subject panels to develop field variations of the generic evaluation indicators defined. Possible variations regarded, amongst others, the types of outputs, the specific information and evidence required (e.g. on HR management and equipment), indications for the topics of the impact narratives, and (in the REF) the use of bibliometrics as support for the assessment of the research outputs
- Panels were expected to define their 'field-specific' interpretation and understanding of the generic instructions for the attribution of different scores against the assessment criteria.
- The field-specific variations included the definition of weights for the different main indicators that were jointly to define a score against an assessment criterion. An example in the REF is the assessment criterion Environment for which the panels were to cover 4 main indicators: the research strategy; staffing strategy and staff development (including PhD students); income, infrastructures and facilities; collaboration and contribution to the discipline or research base.

In the long run, however, this drive for perfection has resulted in an increased complexity and most important, a high financial burden – both on the RAE/REF administration and the submitting universities. Butler (2007) worded a common reflection in the evaluation research community on this topic as follows:

> "While disciplines have their distinctive characteristics, it is not practical or desirable to develop a discrete set of quantitative measures for every distinct discipline or group of similar disciplines. There must be consistency in assessment among disciplines, while allowing for *sensible* adjustments to generic indicators."[28]

3.7.5 Handling inter-disciplinary research

A topic closely linked to the choice of method is the ability suitably to assess and reward interdisciplinary research.

The benefits of interdisciplinary research are universally acknowledged. Importantly, different areas of interdisciplinary research emerge over time, and eventually can become established disciplines of their own. Interdisciplinary research can produce innovative new perspectives and lead to entirely new disciplines emerging.

It is essential that interdisciplinary research is not systematically dis-incentivised through the workings of a research assessment system. This becomes especially problematic in assessments that require any work to be submitted within the context of disciplines. Emerging interdisciplinary fields may then have to compete with the established field, within which they are categorised. Recognising these issues is crucial for any research assessment system, and all the more so in metrics-led assessments.

Both broad approaches to assessing research outputs and quality – bibliometrics and peer review – respectively contain dangers in this respect.

Overall, interdisciplinary researchers often publish in a range of journals and their outputs are spread over several different fields, thus weakening a university's claim to have concentrations of excellent research. The result is that more narrowly specialised research is encouraged at the expense of cross-disciplinary innovation.

In peer-review based systems, interdisciplinary work likewise poses a challenge, especially so, as peer review necessarily requires grouping of panellists into areas of expertise. Hence the structuring function of disciplines required for peer review clashes directly with work that seeks to transcend these structures. Langfeldt (2006)[29] notes:

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

In most of the discipline-based evaluations (in the UK, Australia and NZ), inter-disciplinarity is accounted for by including a broad range of experts in panels and if

---

[28] Butler, L., (2007), Assessing university research: a plea for a balanced approach, *Science and Public Policy*, 34(8), October 2007, pages 565–574

[29] Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation*, *15*(1), 31–41. doi:10.3152/147154406781776039

necessary cross-referring to additional panels. In Australia, submissions at area rather than field level are allowed, thus making research from a mix of reasonably closely related fields un-problematic. Panel members may be assigned between the various panels to bring appropriate expertise to bear on the evaluation.

This approach goes some way to tackle the problem of assessing interdisciplinary research. However, simply conducting reviews according to a 'mix' of the contributing fields does not recognise that such research may not just combine but transcend established norms in any of those fields. Moreover, it does not solve the issues around metrics use for interdisciplinary work discussed above, at least where metrics are designated as an assessment tool in contributory fields.

This problem has been tackled in the case of Italy's VQR: here, every submitted research output needs to be accompanied by several pieces of information, including the specification that the product is an outcome of research in emerging areas or in areas of high specialisation or inter-disciplinary character. In addition to the 14 disciplinary sectors through which research is then assessed, there are 6 additional inter-disciplinary sectors work flagged in this way. Though overall the VQR has a broad range of assessment tools, in these sectors peer review is preferred to metrics due to the limited availability of bibliometric indicators in these often small and emerging fields, as well as in response to the problems outlined above.

Whilst many systems therefore give the evaluated institutions the opportunity to flag interdisciplinary research in some way, so as to trigger various special processes to assess it, the challenge around inclusion of interdisciplinary research in research assessment systems is yet to be met with a solution satisfactory to all. A few general points are worth highlighting at the outset:

- Interdisciplinary fields that have established themselves to a reasonable extent, or fields that combine relatively closely related disciplines are easier to assess than embryonic fields or fields comprising normally distant disciplines
- Metrics are generally not advisable for interdisciplinary research, or should at least be used with extreme caution
- Interdisciplinary work can be a challenge for output assessment, but it can simultaneously be a systemic indicator
- Some degree of qualitative deliberation, either by the assessors or provided by the evaluated institutions usually needs to accompany interdisciplinary outputs in order to then decide on a suitable assessment procedure

3.8 Indicators and assessment criteria

In this Chapter we first introduce some key concepts of indicators and their use in research performance assessment, and then present the different indicator categories, providing a short description and their use in international practice. We also discuss briefly the different approaches in the use of assessment criteria. In the final section we cover the risks related to the use of certain indicators – in particular in PRFS, and how these can be mitigated.

3.8.1 Introduction

Indicators and assessment criteria for research performance assessments are closely related to the theory of knowledge. Knowledge is the major outcome of science and research; it also constitutes its major value – for research, industry and society alike.

An evaluation system that intends to understand research performance in its broader sense, i.e. not limited to the size and scientific quality of research outputs, focuses on assessing the intensity of the **knowledge transfer mechanisms**. Different knowledge transfer mechanisms transfer different types of knowledge:

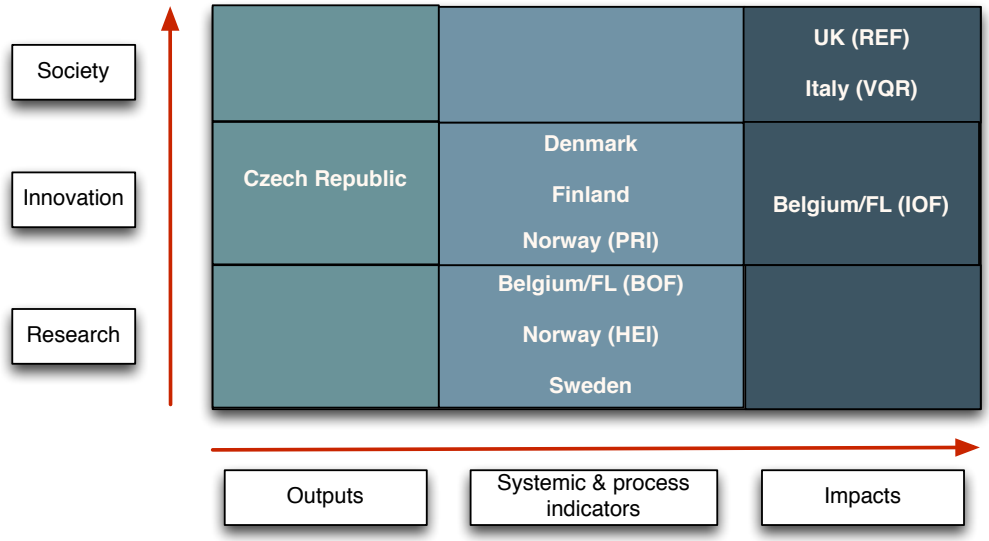
- Publications and patents transfer codified (written) knowledge

- More interactive mechanisms, such as contract and collaborative R&D, transfer both codified and tacit knowledge (know-how, skills). [30]

Knowledge transfer mechanisms are understood as *pathways to impact*, i.e. those aspects that are critical for the *creation* of impacts – in the form of increased knowledge and potential use of the research outputs for advancements in research or innovation. They typically constitute the 'outcomes' of practice of research activities and are assessed through the use of two categories of indicators: process indicators and systemic indicators (see further below).

A basic concept of evaluation is that indicators should cover the various sequential components of a policy intervention, i.e. the inputs (financial and human resources) for the implementation of activities that are expected to lead to outputs, outcomes and impacts. [31] We note a clear trend in international practice, including indicator-based PRFS, to extend the focus of investigation from outputs (only) to outcomes, and in some cases such as the REF, even impacts (Exhibit 14).

Exhibit 14 Indicators used in European PRFS (2013)



An ever-returning theme is the importance of the **strategic policy objectives** for the design of the evaluation methodology. The function of the evaluation and its policy objectives constitute the criteria for the selection of the indicators.

Different types of indicators reflect different possible sets of aims behind research assessment systems and the wider policy rationales and strategic priorities underpinning them.

- Indicators of research productivity and research quality or excellence are particularly stressed in those R&D systems where the lack of productivity is cause for concern (eg Italy, the CR) or where the need is felt for concentration of the research funding (the UK)

---

[30] Arnold, E. et al. (2012), Knowledge Transfer From Public Research Organisations

[31] Arnold, E., Mahieu, B., Horvath, A. (2011), R&D Governance in the Czech Republic, International Audit of the RD&I System in the Czech Republic, Technopolis Group

- Pathways to impact (research esteem, collaborations, etc) are focus points for those systems where a major intent is to steer research behaviour in order to overcome specific systemic failures (eg in Norway an enhanced institute-HEI collaboration in the PRFS for the research institutes) or foster the societal relevance of research
- Input from external (competitive) funding sources are selected as indications of quality of research (competitive strength) and the value of the research activities for research and/or society (i.e. responding to the needs as expressed in public competitive funding programmes, contract research etc)

Whilst some countries have separate evaluation systems for different goals (eg Norway's field evaluations, institutional evaluation and PRFS), which then consequently draw on different indicators to suitably address them, other countries (e.g. UK and Italy) have single, complex assessment systems, which typically draw on a relatively broad range of these different indicator types discussed here, in order to satisfy a wide range of policy needs and strategic priorities and ambitions.

Exhibit 15, below, illustrates the indicator types discussed here and the aspects and features of a national research system they most directly relate to:

Exhibit 15 Overview of indicator types and use

	Research productivity	Research quality	Relevance of research	Efficiency/ value for money	Quality/ sustainability of national research systems
Input Criteria		X	X	X	
Systemic indicators			X		X
Process indicators	X		X		X
Research outputs	X	X	X		
Impact indicators			X	X	

There are two categories of indicators: quantitative indicators and qualitative indicators, the latter are collected in narratives.

3.8.2 Input and output indicators

Input Criteria help to clarify and assess the context in which research happens. Firstly this involves basic features such as number of staff in an EvU and relatedly, expenditures on staff and activities. On one hand, indicators of this type can highlight value for money, especially when input indicators are directly contrasted with outputs. Additionally, these indicators can also act as an audit of financial efficiency.

A further form of input indicators concerns the amount of funding received. This can simply be levels of institutional funding, the significance of which depends strongly on the rules attached to institutional funding in a particular country. But more importantly, external funding in the form of national or international competitive funding, corporate funding, contract research, etc are important inputs enabling the corresponding outputs, whilst also demonstrating that the activities or strategy of a unit of evaluation is aligned with wider national or international scientific, social or economic concerns and priorities. Whilst to some degree successful competition for these types of external funding can additionally highlight research quality as well as presence of additional resources (inputs), these measures above all highlight **relevance** on a unit of evaluation to wider contexts.

	NL*	NO**	SE	UK	IT***	FI
Input criteria						
Institutional funding	B					
Third-party funding						
• **National competitive funding**	B	All	X	X	X	X
• **International competitive funding**	B	All	X	X	X	X
• **Contract research**	B/S	INS, ISH	X	X	X	
• **Non competitive funding**	B	All		X		
Research staff (FTE)	B				X	X
Total staff incl. supporting (FTE	B				X	X

*For Netherlands: The SEP provides a non-exhaustive list of suggested indicators, some commonly provided additional indicators (B), and the suggested indicators (S). The indicators are not mandatory. The idea is that each research unit chooses indicators that fits the discipline and corresponds to the mission and strategy.

**For Norway: We mark the cells with INS for institute evaluation for natural sciences (incl. medicine and technology), ISH for institute evaluation for social sciences and humanities, SENS for subject specific evaluation for natural sciences (incl. medicine and technology) and SESH for subject specific evaluation for social sciences and humanities.

***For Italy: there are plans taking shape for additions of indicators to future VQR exercises. We indicate these here (FUT)

Research outputs are the most long-established forms of indicators in research assessment and are used to some extent at least by all countries considered here. Research outputs primarily fulfil the purpose of assessing the productivity of researchers. In the first instance research outputs can fall into various categories, including peer reviewed journal articles, other articles, books/ monographs, PhD theses and book chapters. Some countries count just some of these, other all of them, extending also to a range of further possible types of outputs. If the assessment system moves beyond 'counting' outputs, or makes further classifications, such as 'high impact factor' journal articles, the measurement of outputs then also goes beyond measuring productivity and also encompasses some degree of research quality assessment.

Whilst these types of indicators provide a certain check on productivity and quality, they do not contain consideration to either the resources going into the production of research outputs, not to their impact or the context in which they are produced. Furthermore, beyond the production of research as such, they say little about contribution to the integration and strategic development of a country's overall research and innovation landscape.

Exhibit 16 Indicators on research outputs

	Evaluations			PRFS		
	NL*	**NO****	**UK**	**SE**	**IT*****	**FI**
Refereed articles	B/S	All	X	X	X	X
Non-refereed yet important articles	B/S	All	X	X		
Books	B/S	All	X	X	X	X
Book chapters	B	All	X	X	X	
Conference papers	B	All	X	X	X	
PhD theses	B/S	All	X	X		
Professional publications	B/S	All	X	X	X	
Publications aimed at general public	B	All	X	X	X	
Policy reports	S					
Other research output <specify>	B/S	SENS, INS, ISH	X	X	X	
Research (cultural)				X		
IPR	S	INS	X		X	
Other innovation outputs		INS	X		FUT	
	NL*	**NO****	**UK**	**SE**	**IT*****	**FI**

Co-publications

The issue of how to treat co-publications in research assessment systems centres for the most part on the question of whether co-publications should be "deduplicated" (counted only once) or "fractionized" (counted as shares).

Fractionalising can take place at two levels: between units and between authors (depending on the level of contribution to the publication). Furthermore, there are additional and related questions:

- Fractionalising for co-publications within the same unit versus only between different units
- The level of sophistication to take levels of contribution into account for points allocation
- International co-publications

The decision-making on this topic is related to two issues: policy objectives in terms of fostering collaborations and the need to prevent gaming.

- In Sweden, only first authors and corresponding authors are included, and the publications are split if these researchers come from different institutions
- Norway pays attention to international co-publications, thus implicitly encouraging internationalisation, other publications are generally fractionalised, which effectively discourages, at the best does not encourage co-publication
- In Italy, the notion of not fractionalising in order to actively encourage inter-unit collaboration is explicit: in case of co-authorship across institutions, the research output can be submitted by each of the institutions. Research outputs with more than one author within a single institution, instead, can be submitted only once
- The UK has an overall similar approach. Joint submissions in a unit of evaluation by two or more HEIs are also possible in the instance that this is the most appropriate way of describing research they have developed or undertaken

collaboratively. A joint submission will be considered in the same way as a single submission and the outcome will be a single quality profile

- New Zealand deals with the issue of individual co-author contribution in qualitative detail: panels will assess joint research on a qualitative basis. To enable this, the staff member should include information on their contribution (relative to other co-authors or equivalent) . Panels are solely concerned with the quality of the output and the relative contribution of the researcher. Co-authors or co-producers do not need to be aware of one another's submissions of the same research output, but they are encouraged to confer about the details of their contributions, to ensure that there is no conflict in the information provided

There are several different key observations that transpire from this overview:

- Within institutions, co-publications tend only to count once
- All authors of a co-publication can have equal weighting, but more commonly, systems seek to represent larger and smaller contributions to an output.
- Between institutions, marks for co-publications can either be split, which results in a more accurate picture of the overall research landscape, or it can be double-counted. The latter is the more common approach in the countries we reviewed, and where this happens, it is sometimes done explicitly with the intent of encouraging cross-institutional collaboration
- International co-publication do not typically lose points to the institution external to the national evaluation system, especially due to internationalisation being desired and encouraged by many systems.

3.8.3 Systemic and process indicators

Systemic indicators are a response to the fact that research outputs are in and of themselves a poor measure of the overall health and quality of a research system, especially in terms of collaboration, mobility and the consequent sharing of knowledge and expertise. There is a distinction between indicators of national and international scope as well as for the different components in the NIS (industry, education etc). These indicators can indicate a certain level of esteem and quality. Especially in the case of international competitive funding, internationalisation also highlights a degree of relevance to internationally recognised concerns and priorities.

However, it is crucial to consider that indicators of internationalisation do not just reflect the extent and quality of outward projection of a nation's research system, but simultaneously indicate the extent of outside knowledge and expertise that is brought in. This is the fundamental reason why internationalisation is a *systemic* indicator. Alongside national mobility, training and collaboration, these **'outside influences'** are critical to ensuring the development and overall health of national research systems. This is especially the case in smaller countries, hence we see a strong emphasis on internationalisation for instance in many of the Nordic countries.

Exhibit 17 Systemic indicators

	NL*	NO**	SE	UK	IT***	FI
International cooperation						
• **In general**	S	All		X		
• **Within research community**	S	All		X	X	
• **International mobility**		All			X	X
National cooperation						
• **Within research community**		All		X	X	
• **Science-industry**		SENS, INS		X	FUT	
• **Research-education cooperation**		SENS, SESH		X		
• **National mobility**		SENS, SESH				

Process indicators are a further area of assessment worth considering: research outputs only reflect the 'end product' of the research process, but in themselves give little indication of the collaborative, dialogic or training activities that feed into research. These indicators help to ensure that even large numbers of high-quality outputs are not created in an insular fashion with little **knowledge transfer** benefits to the wider research or industry community.

More than any other set of indicators, these areas of interest contain implicitly the realisation that the process of research itself can, to varying extent, be shared with other researchers: the resulting dialogue and transparency may then allow for new perspectives to be assessed and developed, regardless of publication or prestige of output channel.

Exhibit 18 Process indicators

	NL*	NO**	SE	UK	IT***	FI
Knowledge transfer to the research system						
• **Editorship in journals**				X		
• **Conferences etc**	S			X		
• **Intra-research collaboration**	S			X		
Knowledge transfer to education						
• **PhDs enrolment/success rates**	B	All		X	FUT	X
• **Postdocs**				X		
• **Graduate teaching**		SENS, SESH		X	FUT	X
Knowledge transfer to enterprises & society						
• **Collaboration research-industry**		SENS, INS		X		

3.8.4 Impact indicators

Impacts are broadly the most recently adopted type of indicators. More directly than the aforementioned input indicators on external funding, they highlight the extent to which a unit of evaluation's activities are aligned with wider societal, economic or other **strategic needs**.

Wider societal, cultural or economic impacts of research present a significant challenge to research assessors, primarily because there often is a long time delay between the publication and an impact of research outputs. Moreover, impact of research can occur either directly or through several proxies. The UK's REF 2014 is the first major concerted attempt to demonstrate research impact in a systematic way across all disciplines (by means of narratives), and the forthcoming results will highlight the extent of feasibility of measuring this particularly challenging yet pertinent aspect of a country's research system.

Exhibit 19 Impact indicators

	NL*	NO**	SE	UK	IT***	FI
Innovation (spinoff, incubators)		INS		X	X	
Societal impacts	B	INS, ISH		X		

3.8.5 Assessment criteria

Whilst in metrics-led research assessment there are many issues surrounding the suitability, strengths and weaknesses of the various indicators, they ultimately constitute a form of assessment that is numerical and measurable from the point of observation onwards.

In the case of peer review, it is different: whilst ultimately, peer review still needs to quantify, or at least rank in a coherent manner, the various units of assessment, i.e. produce a verdict that is essentially numerical/ quantitative. As such, unlike metrics, peer review requires at some point a transition from qualitative and often holistic expert observation to delineated rankings and categorisation. To a certain extent, this transition places reliance and trust in the expertise and judgement of the reviewers. Typically, considerable guidelines and criteria are put in place to lessen the subjectivity and 'blind' trust in the peer reviewers and systematise their qualitative judgements in a more detailed and rigorous fashion.

We summarise some key cases here. Some detailed approaches are different, but broadly the processes are the same.

In **Australia**, an artificial separation is created between the peer reviewers who directly engage with a unit of evaluation's outputs and the application of ranking criteria as such, done by the panels. The Peer Review Report consists of a textual response on the quality of the sample of outputs that they have reviewed, against the following broad categories:

- Approach, i.e. methodology, appropriateness of outlet/venue, discipline specific publishing practices etc
- Contribution, i.e. timeliness, originality, significance of the research question, subsequent use by others, contribution nationally and/or internationally etc.

Peer Reviewers do not provide a rating or ranking of any of the work reviewed.

In the **UK**, reviewers are to consider three aspects in each UoA's submission, with specific weightings:

- Outputs are judged in terms of their 'originality, significance and rigour', with reference to international research quality standards (weighting: 65%)
- Impact is assessed in terms of 'reach and significance' of impacts on the economy, society and/or culture (Weighting: 20%)
- Environment, i.e. the institutional conditions, is assessed in terms of its 'vitality and sustainability' (Weighting: 15%)

Using these three components and weightings, UoAs then receive a composite categorisation on an overall 5-level quality scale.

The **Netherlands** has a system of peer review criteria that is effectively broadly similar to that of the UK, with a comparable overall ranking system and main criteria, though with no additionally stated specific criteria for individual disciplines. Guidelines are broad enough to ensure applicability to all disciplines. The review panel evaluates the research based on three main criteria: research quality; societal relevance; and viability. The review committee provides both a qualitative judgment and a quantitative judgment. The committee ensures that the quantitative and qualitative judgements are in agreement, and that the criteria and judgement are related to the unit's strategic targets. In addition to these main criteria, the review also provides a qualitative judgement on the unit's PhD programmes and the unit's policy on research integrity.

The four-level scale alongside the main categories of 'Quality', 'Relevance' and 'Viability' (more-or-less analogous to the UK's 'Outputs', 'Impacts' 'Environment' are summarised below:

Category	Meaning	Research quality	Relevance to society	Viability
1	World leading/ excellent	The research unit has been shown to be one of the few most influential research groups in the world in its particular field.	The research unit makes an outstanding contribution to society.	The research unit is excellently equipped for the future.
2	Very good	The research unit conducts very good, internationally recognised research.	The research unit makes a very good contribution to society.	The research unit is very well equipped for the future.
3	Good	The research unit conducts good research.	The research unit makes a good contribution to society.	The research unit makes responsible strategic decisions and is therefore well equipped for the future.
4	Unsatisfactory	The research unit does not achieve satisfactory results in its field.	The research unit does not make a satisfactory contribution to society.	The research unit is not adequately equipped for the future.

In the **Italian VQR** 2004-2010, the peer reviewers were expected to assess the research products along the criteria listed below:

Criteria	Description
Relevance	Relevance, as added value for the advancement of knowledge in the field of science in general, as well as the induced social benefits also in terms of consistency, effectiveness, promptness and duration of the fallouts
Originality & innovation	Originality/innovation, as contribution to the advancement of knowledge or to new discoveries in the field
Internationalisation	Internationalisation and/or international standing, as positioning in the international scenario, in terms of importance, competitiveness, editorial spreading and appreciation from the scientific community, including explicit cooperation with researchers and research groups form other countries
Evaluation of technology transfer	Concerning patents, the judgment must also include the evaluation of technology transfer and development, and socio-economic fallouts (even though only potential)

Based on these criteria, evaluators translated their descriptive judgments into synthetic judgements, and provide all products with a **level of merit** ranging from A (Excellent) to L(Limited). The outcome of the peer review and the bibliometric assessment consist of for each research product a score for the quality with range [1,-2].

3.8.6 Risks and risk management

Risks related to indicators and assessment criteria are related mainly to three topics:

- The quality of the data used
- The unintended effects that some indicators may cause, especially in PRFS
- The balance between costs and benefits

A challenge that most of these research performance assessment models are struggling with regards the **quality of the data**, in particular related to the research outputs.

It led to various measures implemented in the countries covered in order to ensure and enhance the quality of the data entered. Extended control checks have been put in place especially in the Czech Republic and the UK, where the PRFS attribute high importance to research outputs and more than 50% of the core funding is guided by the PRFS. (Exhibit 20)

The issue of data quality is directly linked to the process for the data entry into the system. In most countries, this is a manual process, with the researcher or his/her institution inserting data directly on the interface of the research information system collecting the information. The experience is that this system is prone to mistakes, omissions and duplications. It causes a burden on the researchers or institutions required to enter the data as well as on the processing public agency in charge of the cleaning and checking of the data.

This topic constituted a major driver for the development of national research information systems, foreseeing a *direct harvesting* of the data from the institutional information systems and/or Open Access Repositories and therefore creating a higher level of efficiency as well as ensuring data quality.

In Europe and at international levels, an increasing number of initiatives have been launched in recent years developing national research information systems (RIS). These information systems can take two forms:

- Research information systems set up by *funding agencies,* typically geared towards collecting information for accountability purposes to the benefit of a specific agency in the country. Examples are the VIPUNEN system created by the Finish Ministry of Education and the UK ROS system, capturing data for most of the UK Research Councils

- *Fully national research information systems*, capturing information on all research and its outputs/impacts in the country from a variety of sources. These RIS are often intended to act as platforms for the creation of visibility on outputs and outcomes in society. Examples are the FRIS system in the Flanders/Belgium and the CRIStin in Norway

Exhibit 20 Quality control measures in PRFS

Country	Quality control
Austria	Formal quality check and estimate of missing data
Belgium-Flanders	Close collaboration with the institutions (future harvesting)
Czech Republic	Control by the funding providers, IT-based control, peer review control
Denmark	Responsibility of the institutions (harvesting)
Finland	Publication Data Collection Manual
Italy / VQR	Guaranteed by submitter
Norway	Shared quality assurance system
UK - REF/RAE	Validation rules, REF audit team: Sample-based verification, Data comparisons, Panel instigated audit

Another topic that is more difficult to handle are the **unintended (negative) effects** of the PRFS. In this section we focus on effects that are directly linked to evaluation and the use of indicators.

'Gaming' is a common effect of the introduction of PRFS, and it is a well-known phenomenon in any PRFS informing a significant share of institutional funding,. In the Czech Republic, this phenomenon is additionally accentuated due to the pervasive use of the PRFS throughout the R&D system, down to the level of the individual researcher (see Section 2.1.2, above).

Salami-slicing, i.e. publishing several small and often overlapping papers on a particular research project or idea rather than a single comprehensive article, is a common effect in all countries where metrics used focus on raw publication counts. Researchers in various countries have found various ways to 'game' the journal indicators used, for example by publishing in fields 'adjacent' to their own, where it was easier to get into journals that gave more points in the PRFS, or where the 'competition' is less severe.

In the UK it was found that the on-going process of change in the RAE was an effective way of reducing the effectiveness of the gaming, as was the fact that until after the evaluation exercise, nobody knew how the evaluation results would guide the funding.

A golden rule in the design of an evaluation informing PRFS is that *any* indicator is vulnerable to gaming. There are however several proposed ways of **combining** different indicators in order to mitigate these effects as much as possible, in other words, the use of triangulation. This can be the combined use of quantitative and qualitative indicators to assess the same criterion, or the use of the same indicator against two criteria, with potential opposite effects for gaming strategies.

Two measures are the most effective against gaming practices:

- One indicator is only part of a **mix** of quantitative and qualitative indicators informing an assessment criterion
- The use of expert panels

Finally, there is the issue of the **balance cost/benefit**. National evaluations are a costly endeavour and the more a high level of sophistication is sought, leading to an increase in information needed, a high number of panels and sub-panels, and/or extensive peer review, the costlier an exercise it becomes. Examples os such more

'sophisticated' systems are the UK RAE/REF, the VQR in Italy and the RAE in New Zealand (Exhibit 21).

The high levels of sophistication led in particular to a high burden on the evaluated organisations (the indirect costs).

Exhibit 21 Total costs and 'efficiency' of three international PRFS

	UK RAE 2008	IT VQR	NZ QE 2006
nr ROs	159	133	46
nr FTE res	68,563	61,822	8,671
nr RU	2,363		
Nr products submitted	216,497	184,878	
Nr products peer reviewed	216,497	99,000	
Share of PRFS in institutional fund	23%	13%	10%
Years funding allocated	6	3	6
Funding allocated	**€ 11,970,000,000**	**€ 2,892,000,000**	**€ 1,165,600,000**
Total direct & indirect costs	**€ 89,460,000**	**€ 64,600,000**	**€ 46,960,000**
Efficiency	**0.75%**	**2.23%**	**4.03%**
Direct costs (staff & panels)	**€ 15,120,000**	**€ 10,570,214**	
Indirect costs	**€ 74,340,000**	**€ 54,029,786**	
indirect costs % total costs	83%	84%	

Restrictions to government budgets have led to a trend towards limiting the costs of national performance assessments in most countries, even though several studies highlighted that national assessments for funding allocations are considerably less costly than the processes for the allocation of competitive funding.

Efforts to reduce costs typically focus on the main **cost drivers** i.e.

- The high number of indicators and level of information requested from the evaluated research organisations
- The extended use of peer reviewers and the high number of panels and sub-panels

3.9 The implementation of evaluation through expert panels

Panel-based evaluation systems typically set up a hierarchical system of main panels, panels, and sub-panels. The number of panels and the existence of sub-panels depends on the size of the R&D system as well as the depth and complexity of the exercise, quite obviously influencing also its costs.

There are two main characteristics of 'quality' panel-based evaluations that all panel-based evaluations strive to comply to. This is typically done through the definition of specific rulings as well as more structural measures (e.g. the tasks of main panels, the panel secretariat, specialist advisors etc).

The **robustness** of the system is a first key topic of focus. It regards in particular two elements

- *Clear and universal guidelines for the evaluation process*: In a national peer review system many peer review panels are responsible for the primary assessment of the units of evaluation. It is crucial for the quality in assessment that all the panels have the same understanding of the criteria, scoring system, ways to assess the submitted output, etc.

  The panels must be guided by detailed instructions and clear and universal guidelines. The guidelines must minimise the differences in interpretation and shaping of the process between panels and ensure a thorough and consistent

review procedure. In most countries detailed guidelines for the peer review process are available and made public.

- *A professional management and support organisation*: A national peer review exercise is a complex and labour intensive exercise. This requires professional management and a professional support organisation. The management must prepare and coordinate the whole review and is also responsible for the communication to the research organisations. It is also important that the panels and the management are supported by staff, e.g. taking notes, prepare meetings, arrange facilities, etc. This also includes support tools like online systems and a database for statistics, publications and other outputs.

The **fairness** of the system is a second – very important – premise. There are a number of possible biases which can endanger the fairness of the system:

- *Clientelism & nepotism*: The peers should not have a conflict of interest. In most countries there are detailed procedures to avoid a conflict of interest, e.g. a declaration of unbiasedness or the obligation to record declarations of interest.
- *Scholarly bias:* This is an in important bias that is often discussed in literature on peer review processes. Scholarly bias results from the fact that judgement by panel members cannot be independent of their own 'disciplinary culture'. Within a disciplinary culture, certain values, interests, and expectations dominate, as well as research and publication practices, and perspectives on what constitutes high quality or highly relevant research.

  The most accepted practice to limit scholarly bias is the inclusion of a wide range of disciplines in review panels and preferably ensure overlap in competences to promote critical debate. Other practices are to frequently replace the reviewers, and to let those under evaluation have input on the reviewers (a direct dialogue between scientists being evaluated and experts evaluating them)
- *Appropriate representation***:** in order to ensure a thorough assessment there must be an appropriate representation in the peer review panels. This concerns for example gender balance and geographic distribution of the peers, but also the inclusion of industry, end-users and other stakeholders.

  Generally, great care is taken to ensure an appropriate representation in the panels. In Italy and Norway the guidelines contains detailed instruction for ensuring an appropriate representation. Appropriate expertise in the panel can also be ensured by appointing panel members in consultation with the evaluated unit.

  The inclusion of stakeholders and users of research is not very common practice.
- *Inter-disciplinary research:* The integration of diverging perspectives and criteria in inter-idisciplinary panels can be very difficult. This difficulty especially disfavours inter-disciplinary research (as well as multi- and trans-disciplinary research) as it is expected to meet diverging expectations and criteria of multiple disciplines.

  There are several ways to mitigate the risk of disfavouring inter-disciplinary research. In some countries, like Italy, designated inter-disciplinary panels are established. In many countries inter-disciplinarity is accounted for by including a broad range of experts in panels. Finally cross-referring to additional panels is used for a more appropriately assessment of inter-disciplinary research.
- *Consistency of the assessments*: In a common review framework it is important to ensure a common understanding of the assessment criteria, standards and the application of the quality scores. There could be divergence in perspectives between disciplines and even between peers, which poses an important challenge for a consistent assessment according to the generic guidelines. In order to ensure consistency several means could be implemented. Where sub-panels are used,

they are generally represented in the overarching main panels, which helps to ensure coherence between the sub-panels. In the UK the main panel undertake calibration exercises (attend some sub-panel meetings and discuss reports from the sub-panel chairs).

- *Transparency*: Finally transparency of the review process can contribute to the fairness and validity of the assessment. In the literature it is argued that transparency can help guard against scholarly bias and may also help to identify and address conflicts of interests. In most countries there is a large degree of transparency. Procedures are clearly laid out on websites and the outcomes of the reviews are made public. Also panel selection processes are transparent. One of the factors of success for the Italian Research Exercise was the high transparency of the panel selection process.

Finally, various suggestions are made in various sources to limit the **costs of the review** without sacrificing too much quality. In Norway, it is found that having an in-house secretariat to the panel, rather than an external secretariat, is also a way to mitigate costs to the evaluating bodies. In the ERA 2012 handbook[32], it is stressed that it is the responsibility of a panel member to ensure they adequately prepare for meetings to avoid unnecessary additional administrative costs and inconvenience to other committee members. In an OECD working paper on enhancing public research performance by means of evaluation[33], it is suggested that ways of minimizing evaluation costs must be sought in the process of application, selection of expert reviewers, and panel discussion of discipline committees. The following ways of controlling the cost of the review meeting are suggested: structuring the agenda so that panel members' time is used efficiently; and making maximum use of teleconferences, videoconferences, and other electronic media to prepare the review panel. In addition, it is suggested the part of the evaluation cost born by the subject of the evaluation should be minimized by simplifying administrative procedures and evaluation formats.

[32] "ERA 2012 Evaluation Handbook Excellence in Research for Australia," 2012.

[33] OECD, *Enhancing Public Research Performance through Evaluation, Impact Assessment and Priority Setting*, 2009.

4. The proposed Evaluation Methodology

The objectives of this study were to develop a national evaluation methodology that would provide strategic information for the actors at all levels in the R&D system (institutional and national) as well as inform the institutional funding for research organisations system

The expectations listed in the ToR for this study were that the evaluation methodology would

- Apply a peer-review evaluation process
- Fulfil formative and summative functions
- Actively involve assessed entities in the evaluation
- Cover outputs, impacts, and institutional projections of research development
- Take into consideration the different missions of research organisations within the research system
- Take into consideration field specifics
- Set up evaluation processes that will be resistant to clientelism and conflicts of interests
- Take into account 'gaming'
- Be set up so that total costs do not exceed 1% of public institutional support for R&D in a five-year time period

The Evaluation Methodology described in this chapter responds to these expectations.

The choices upon this Evaluation Methodology is based, were guided by a set of key principles, listed in Section 4.1, below.

In the remaining of this chapter, we set out the core elements of the Evaluation Methodology, i.e. the evaluation structure (Section 4.2), the scope of the evaluation (Section 4.3), the evaluation methods (Section 4.4), the assessment criteria (Section 4.5), how the Evaluation Methodology handles diversities among the research organisations and fields (Section 4.6), the evaluation results and their intended use (Section 4.7), and a reflection on data sources for the EM (Section 4.8)

4.1 Key principles

We defined the key principles of the proposed Evaluation Methodology as follows:

- The Evaluation Methodology reflects the strategic **policy objectives** for the Czech R&D system.
- Its primary function is to act as source for **strategic information**, at all levels in the RD&I system. It is therefore **comprehensive**, covering all the dimensions of the research activities and its outputs, outcomes and impact
- The evaluation results will also directly inform the **institutional funding for research organisations**
- The assessment of the research performance will take place at the level of field-defined **Research Unit** (RU) within an Evaluated Unit, i.e. a research organisation or in the case of the public HEI, a Faculty or Institute or any other organisational unit at that level such as Centres
- The evaluation covers **all research organisations of a minimum size**
- The evaluation is a process of **informed peer review.** The expert panels will draw on a draw on a **mix** of appropriate quantitative and qualitative data to support their professional judgement
- The evaluation is a **fair and egalitarian** system. It will use a single framework for assessment across all disciplines and types of research organisation while allowing for a reasonable level of field- and RO type-specific variations
- The **cost and burden** of the evaluation will be the minimum possible to deliver a robust and defensible process

4.2 The evaluation structure

We have defined the evaluation structure as follows:

- An **Evaluated Unit** (EvU) is a research organisation, except for the public HEIs where the Evaluated Unit is a Faculty or Institute or any other organisational unit at that level such as Centres
- A **Research Unit** (RU) includes all individual researchers in an EvU (across the organisation structure) that conduct research in a single scientific field. Researchers need to be assigned to research units in their major field of research; each researcher can be assigned only to one research unit in an evaluated unit.

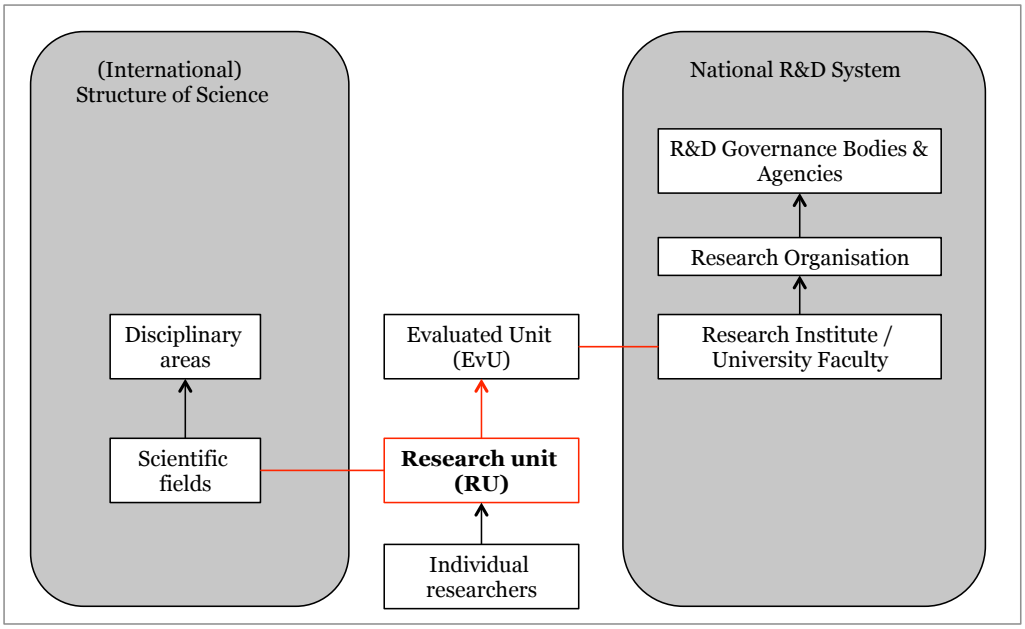
In other words, a Research Unit is a sub-set of an EvU does not necessarily represent a coordinated or collaborating research group. All researchers in an EvU have to be allocated to an RU.

The scientific fields are based on the OECD field classification (see Exhibit 28, below).

We have defined the primary unit of evaluation, i.e. the Research Unit, at the intersection of the 'natural' dimension for evaluation of research - the scientific field - and the dimension determined by the need for information at the level of institutions (Exhibit 22).

This will allow for the assessment of the RUs' role, positioning, and competitive value in the **national** R&D and innovation system as well as in the **international** R&D landscape. As research is becoming more and more international, and the competition - as well as collaboration in research - is at a global level, a view on the performance and position of the research actors in the Czech Republic relative to the international landscape is a critical piece of strategic information, for any actor in the R&D system.

Exhibit 22 The Research Unit as primary unit of evaluation



The evaluation is **not** conducted at the level of individual researchers. International experience (in particular in PRFS) shows that any inclusion of even a single component of the evaluation methodology at the level of individual researcher risks having considerable negative unintended effects (see Section 3.5.2, above). Any use of the evaluation results - or part of the evaluation results - at the individual researcher level is therefore inappropriate.

We propose that research organisations will participate to the evaluation on a **voluntary** basis. Costs for participation in a national evaluation can be high, requiring considerable investments in terms of time and resources and it would be unfair to oblige all research organisations - and in particular the small ones - to bear these costs. By deciding not to participate, however, the research organisation will not be entitled to the funding related to the performance-based research funding component in the new institutional funding system. We explain this in more detail in the 2nd Interim report.

This 'opting-out' possibility applies only at the level of research organisation. In other words, a research organisation cannot decide that only some of its EvU that are active in research will be subject to evaluation.

4.3 The scope of evaluation

4.3.1 The typologies of research organisations

In Section 3.3.2, above, we explained that in international practice, a basic concept for evaluation is the focus on the attainment of specific goals and objectives, based on the functions in the National Innovation System of the different actors involved.

This is even more the case for an evaluation that is expected to assess the performance of these organisations in terms of their value for research as well as for society. Inevitably it needs to develop a categorisation that takes into account the mission of these organisations' activities within the National Innovation System, based on the profile of the primary beneficiaries of their research activities.

The research organisations in the Czech Republic have a variety of legal forms, based on the historical background of the RD&I system. While the mission of public HEIs,

universities and the Academy of Sciences research institutes are clear, for other research organisations in the Czech system it cannot be identified in a straightforward manner based on their legal form.

We therefore propose the following four categories of research organisations for this Evaluation Methodology, taking into account missions rather than legal forms:

- **Scientific Research Organisations** include the HEIs, university hospitals and the Academy of Sciences research institutes - and eventually other research institutes. They are research organisations that have as primary function to conduct scientific research and/or teach future researchers, to the benefit of the *research community*
- **Research and Technology Organisations (RTOs)** are research organisations that have as primary function to provide knowledge and technology transfer services, to the benefit of the *industry sector*. Examples are the Research Centre Rez, the Aerospace Research and Test Establishment, and the Research Institute of Building Materials
- **Public Service Research Organisations** are research organisations that have as primary function to deliver knowledge transfer services, to the benefit of the *public sector or society*. Examples are the Research Institute for Labour and Social Affaires, the Centre for Higher Education Studies, the Institute for International Relations, the Czech Metrology Institute
- **National Resource/Infrastructure Research Organisations** provide infrastructure, to the benefit of the *research community*. Examples are CESNET, the National Library of the Czech Republic

This typology of research organisations is based on international practice and is universally understood. It will allow the (international) evaluation panels properly to assess the performance of the evaluated research organisations against the different assessment criteria, within the context of the different fields.

Apart of informing the evaluation panels, this typology will also form a basis for funding allocation in the PRFS in the new funding principles. We cover this in more detail in the Second Interim Report.

Upon registration for the evaluation, the research organisations will **apply** for the category of research organisations they consider to belong to. The Evaluation Management Team will take the final decision, wherever relevant in consultation with the Evaluation Management Board.

4.3.2 The use of thresholds

The minimum threshold for the registration of a **Research Unit** is **50 research outputs** within 1 field of research over the evaluated period (currently envisaged as 5 to 6 years).

This implies that an EvU or RO that has *in total* less than 50 research outputs within 1 field of research over the evaluated period cannot register for the evaluation.

An EvU can register more than 1 Research Unit only if each of these Research Units have produced a minimum of 50 research outputs in the field during the evaluation period.

There is no maximum threshold. This implies that an EvU where all of the research is conducted in one field of research can register only one RU.

The type of research outputs that are eligible for this count-out are the scholarly outputs, the non-traditional scholarly outputs and the IPR-related outputs as listed in Exhibit 27, below.

The field of research that will apply for this count-out is the one indicated upon registration of the research output in the RD&I IS.

In exceptional cases, the EvU can apply for the registration of an Interdisciplinary Research Unit. We describe this in more detail in Section 3.7.5.

The introduction of the minimum threshold is to be set in the context of the Czech R&D system, characterised by a considerable fragmentation in the R&D system and a relatively high number of research organisations that are active only to a limited extent in research (see Section 2.2.2, above). A minimum level of research activities and research outputs are required to guarantee the robustness of bibliometric and statistical data analyses, and for the assessment as such.

Non-participation in the evaluation exercise does **not** imply a total loss of institutional funding of research organisations. Similar to the case of research organisations that opt out of the evaluation, the research organisation will not be entitled to the funding related to the performance-based research funding component in the new institutional funding system. We explain this in more detail in the 2nd Interim report.

Exhibit 23, below, gives only an **indicative** view on the potential effects of the minimum threshold as it considers only the production of scholarly outputs in a 4-year period (2008-2012). It shows that even with this more restrictive approach, the effects are relatively limited. Exhibit 24 makes it even clearer, listing the indicatively excluded RO based on the production of scholarly outputs in a 4-year period.

Exhibit 23 Indicative effect of the minimum threshold - scholarly outputs only in a 4-year period (Number of RO below/above threshold)

RO grouping	Below threshold		Above threshold	Total
	Excluded	Eventually excluded		
ASCR	1		52	53
HEI - private	3	1	3	7
HEI - public	1		25	26
Hospital			13	13
Ministry Interior/Defence institutes/university	3	1	2	6
Museum/Library	11	3	8	22
Other RO	10	5	54	69
Grand Total	**29**	**10**	**157**	**196**

Notes: "Eventually excluded" are research organisations that have a total of scholarly outputs above threshold but they are spread over different fields

Exhibit 24 Indicative list of RO excluded - scholarly outputs only in a 4-year period

	Research Organisation	Total publications 2008-2012
ASCR	Library of the AS CR, v.v.i.	29
HEI - private	College of European and Regional Studies, o.p.s.	17
	Institute of Hospitality Management in Prague 8, s.r.o.	44
	ŠKODA AUTO University	42
HEI - public	College of Polytechnics Jihlava	37
Min. Interior/ Defence	Fire Headquarters of the Czech Republic - Fire Technical Institute	28
	Institute of Criminalistics Prague	33
	Sports Research Institute of Czech Armed Forces	37
Museum/ Library	Hussite Museum in Tábor	4
	Jewish Museum in Prague	3
	Museum of Czech Literature	28

	Research Organisation	Total publications 2008-2012
	Museum of Eastern Bohemia in Hradec Králové	2
	Museum of Highlands Jihlava	21
	National Film Archive	4
	National Technical Library	25
	Technical Museum in Brno	27
	Wallachian Open Air Museum in Rožnov pod Radhoštěm	8
Other RO	Akademie, o.p.s.	2
	BIC Brno, spol. s r. o.	1
	CENIA, Czech Environmental Information Agency	12
	Center for Organic Chemistry s.r.o.	13
	GaREP, spol. s r.o.	44
	Institute for Postgraduate Medical Education	37
	Institute for the Care of Mother and Child	47
	Institute of Archaeological Heritage Brno, v.v.i.	45
	MemBrain s.r.o.	13
	Mikropur, s.r.o.	6
	National Information and Consulting Centre for Culture	3
	SELTON Research Centre, s.r.o.	41

4.4 The evaluation method

The Evaluation Methodology will be implemented through a process of **informed peer review**, i.e. expert evaluation panels will base their assessment on the information provided by the evaluated RUs and bibliometric data.

Peer review and panel evaluations have their roots in the culture of science itself, where the use of peers for judgment on quality is a long-standing practice. Peers conduct research project appraisals (both ex ante and ex post) across basic and applied research and development as well as for the assessment of scientific quality in connection with the publication of articles and scientific papers.

Peer review, as any other method, has its weaknesses and we have established the following to compensate for them:

- The use of bibliometrics to inform the panels means that they have access to a greater body of evidence than would otherwise be available to them and can also mitigate some of the sources of potential bias that can occur in panel reviews, as mentioned in Section 3.4.3, above
- Based on international practice in peer and panel review, and taking into consideration the context in the Czech Republic, we have articulated a set of rules and processes intended to ensure the proper functioning of the panels. We cover this more in detail in Section **Error! Reference source not found.**, below

The reviews of the submitted research outputs will be done **remote**; in a first instance, also the panel members' assessment of the provided information will be done remote, complemented in a second instance by panel meetings. This is increasingly common practice in the international environment; it reduces costs and sets a lower level of time required from the evaluation panel members, which is a typical barrier for the involvement of potential experts in panel evaluations.

In order to limit the costs of the evaluation exercise as well as the burden on the evaluated RUs, the current EM does not foresee the inclusion of **on-site visits**. Site

visits are a regular feature in *institutional* evaluations where the in-depth assessment requires closer contact with the evaluated subjects and an understanding of the concrete research conditions in the institutional environment. In national evaluations, however, they are extremely rare because of the scale of these exercises and the significant costs they imply.

Site visits enable a direct contact between evaluators and evaluated subjects, which enables for a smoother communication between the two parties, focused on the factors that enable or hinder good research performance in the specific institution. In this context, implementation of an (additional) institutional evaluation including site visits may constitute an appropriate follow-up of the national evaluation, focused on research organisations that showed a poor performance in the national evaluation.

The importance of a direct contact with the evaluated research organisations is recognised in the evaluation community and alternatives to onsite visits have been developed to allow for such contacts to occur, at the least in those cases where evaluation panels feel the need for additional information. Examples of such alternatives are Question & Answer (Q&A) sessions using videoconference facilities and the invitation of the evaluated organisations' representatives to a central location for interviews.

4.5 The assessment criteria

The EM uses a single framework for the assessment of all scientific disciplines and RO types, no matter their size, thereby ensuring full comparability of the evaluation results across all dimensions. Comparability is a fundamental condition for the use of the evaluation results in the performance-based research funding system as it is currently designed. A starred quality level should have the **same** value for all RUs evaluated and the assessment should base itself on the same set of indicators and information.

The EM uses assessment criteria that are relevant for **all** types of RO, no matter the type of research they perform. Indicators included cover the conditions that enable for quality research to occur in any type of research organisation (research management, strategy, collaboration for research at international and national level), the key factors that can indicate the quality of the overall research performance (research output and competitiveness in research), and the activities that constitute pathways to impact – on research and the society at large. A potential exception is the criterion 'scientific research excellence', which may be of little relevance for some non-scientific research organisations – but surely not for all.

The assessment criteria and their formulation are inspired by international practice, but have been adapted to the realities of the Czech R&D system, with an attention to guarantee a sufficient spread of the quality over the starred quality levels. An exception is the criteria used for the assessment of scientific research excellence.

At the very first stage of the evaluation process, the evaluation panels will implement a **calibration** exercise in order to ensure that all members of the panel have a similar understanding of the assessment criteria. They will decide on the field-specific interpretation of key terms used in the assessment criteria and on the importance of the main sub-criteria for the different types of research organisations. This will be expressed in the form of % weights for the sake of simplicity, but the overall judgment will **not** be based on arithmetic. The panel members can also decide that all sub-criteria are equally important.

It should be noted that in this study we use the term 'research' in the broad sense, i.e. encompassing research, development and innovation. We use the term 'scientific research' when referring to research in the narrow sense of the word.

4.5.1 The context for the assessment criteria: the policy objectives

International practice takes a fairly standard approach to the definition of assessment criteria: they reflect the policy objectives and the panels are requested to define a score of 1 to 4 or 5 (see Chapter 3.8.5, above).

We defined **5 main assessment criteria**, i.e. the research environment, membership of the national and global research community, scientific research excellence, overall research performance, and societal relevance.

These criteria were defined in order to reflect the policy objectives (Exhibit 25, below), bearing in mind also that the evaluation results need to inform the PRFS. For this reason, we paid particular attention to the (policy) need of reaching an understanding of the overall quality while creating incentives for the ROs to become better as well as to award excellence.

Exhibit 25 Assessment criteria in response to the policy objectives

Objective category	Objectives	Assessment criteria	Sub-criteria
R&D capacity	To improve research and development management, at all levels	Research environment	The quality of the research management (including HR management)
	To improve human resource development, reflecting the needs of the knowledge economy of the CR		The adequacy of the research strategy
	To strengthen cooperation between the RD&I actors at the national level	Membership of the global and national research community	International research presence and collaboration
	To strengthen international cooperation		National research presence and collaboration
Excellence in R&D	To motivate research organisations (ROs) to excellence	Scientific research excellence	Peak quality, i.e. the quality of the selected outputs
		Overall research performance	Research output
			Competitiveness in research
Societal relevance	To motivate ROs for collaboration with industry	Societal relevance	Knowledge & technology transfer activities
	To motivate ROs for the transfer of knowledge to practice		
	To stimulate ROs to research corresponding to the needs of society and the business sector		

Note: the term 'societal' refers to all sectors in society, including industry, education, and the society at large

4.5.2 The assessment criteria in detail

In this section we describe each assessment criterion, the values for the attribution of starred quality level, the main topics of investigation, and the factors and information to take into account.

The careful reader will note some significant changes in the assessment criteria compared to the draft version of the 1st Interim report. These changes were triggered by comments of the panel members in the Small Pilot Evaluation as well as feedback from the research community.

The most important changes are

- The use of 5-point starred quality levels, which is a scale that is more internationally recognised than the 4-point one. It allows for a wider spread of scores in the medium range, and facilitates a more adequate interpretation

- The change in the use of starred quality levels led to a revision of the assessment criteria, aimed at improving the descriptions
- Panellists asked for the documentation to be designed so that adequate information would be available also without site visits and the research organisations would better understand the importance and value of self-assessment
- Request by panellists and the research community for a higher minimal number of submitted outputs in order to avoid arbitrariness if one publication only can be selected

Research environment

For this criterion, the focus of analysis is on those conditions in the institutional environment that set the basis for the conduct of quality R&D, now and in the future, i.e.

- The quality of the research management
- The adequacy of the research strategy, its feasibility and sustainability

In the course of the calibration exercise, the panel members will discuss and define the importance to be attributed to the 2 sub-criteria, in *the field and for the types of RO*.

Starred quality level	Definition	Description
5	Outstanding	**The RU is a Global Leader** In terms of the quality of the research strategy and management, the Unit's research environment is fully comparable to that of global leaders in the field. It can attract the highest quality international researchers
4	Very good	**The RU is a Strong International Player** The Unit is able to provide an internationally comparable excellent research environment to high-level international researchers in the given field
3	Good level	**The RU is a Strong National Player** The Unit is able to provide a research environment that is comparable with internationally recognised research organisations in the field
2	Adequate	**The RU is a Satisfactory National Player** The Unit's research environment is still evolving to achieve a level that is expected in the international research community of a respected research organisation in the field
1	Poor	**The RU is a Poor National Player** The Unit is still only in the process of creating an internationally comparable research environment
Unclassified		N/A

In this criterion, 'global', 'international' and 'national' refer to quality standards. They do not refer to the geographical scope of the strategy or management activities.

The panel members will base their judgment against this criterion on quantitative and qualitative data related to

- For the assessment of the research management:
  - The *research capacity*, including the longer-term financial resources stability, the size of the institution (does it have critical mass), the age of researchers, the ratio of PhD students involved in research versus FTE researchers

  - Quality of *Human Resources (HR) management*: the processes for career development (appraisal and monitoring systems, competency framework, frequency of performance reviews and the employees' feedback, promotion criteria and individual targets), the level of inbreeding as an expression of the 'openness' of the RU
  - (Only for the RU that in practice teach and train PhD students and the HEIs) The *career development of postdocs and PhDs:* the support to early-career researchers, objectives and outcomes of the PhD programme, approach to PhD supervision, educational components for the training of the PhD students, the attention for gender
  - The capability of the research organisation adequately to *support research* in the field. This includes the availability of technical staff, the nature and quality of the research infrastructure, the level of investments in its renewal, the RU's shared or collaborative use of research infrastructure, the RU's capacity to gain competitive access to major research infrastructure and facilities
- For the quality of the research strategy:
  - The description of the RU research plan, its main objectives and activities
  - The competitive positioning of the RU in terms of focus and field(s) of activity and how it intends to improve it
  - The intended use of resources (human, financial, equipment) and how the RU intends to combat the RU weaknesses and exploit the strengths
  - The alignment of the RU strategy with the strategies and research priorities at the national level

Membership of the global and national research community

The assessment will focus on

- Level of participation and recognition of the RU at the international level, and
- Level of participation and recognition of the RU at the national level

In the course of the calibration exercise, the panel members will discuss and define the importance of these 2 sub-criteria, in *the field and for the types of RO.*

Starred quality level	Definition	Description
5	Outstanding	**The RU is a Global Leader** The Unit participates and is recognised in excellent international networks involving global leaders in the field.
4	Very good	**The RU is a Strong International Player** The Unit participates and is recognised in international networks in the field.
3	Good level	**The RU is a Strong National Player** The Unit participates and is recognised in excellent national networks involving national leaders in the field.
2	Adequate	**The RU is a Satisfactory National Player** The Unit participates and is recognised in national networks in the field.
1	Poor	**The RU is a Poor National Player** The Unit has little to no substantive collaboration.
Unclassified		N/A

The panel members will base their judgment on quantitative and qualitative data related to

- The *positioning of the RU at the international level*, including
  - The *intensity and quality* of the collaborations: the profile and prestige of the partner organisations, incoming and outgoing study visits, the geographical distribution of the partner organisations, the level of co-publications
  - *International competitiveness*: the capability successfully to participate in international competitive projects, the competitive use of major international research infrastructure
  - *Esteem*: the interest of the international conferences organised by the RU for an international audience, membership of international editorial boards of journals
- The *positioning of the RU at the national level*, including
  - The *intensity and quality* of the collaborations: the RU involvement in centres and research infrastructure in the country, the focus and diversity of research collaborations and partnerships, national co-publications, the shared or collaborative use of RI
  - *National competitiveness*: the capability successfully to participate in national competitive projects
  - *Esteem*: scientific awards and memberships of scientific advisory boards in academia

Scientific research excellence

This criterion assesses scientific research **excellence**, i.e. the quality of the RU research output against the **highest** international standards. It aims to reveal "pockets of excellence" in the Czech research community by assessing to what extent the RUs have the capacity to deliver outstanding research outputs. It also serves as an indicator of the RUs' potential level of competitiveness on the international scene of scientific research.

During the calibration exercise, the panel members will define *the field-specific interpretation* of the terms originality, significance and rigour prior to the review by the referees.

Starred quality level	Definition	Description
5	Outstanding	**The RU is a Global Leader** In terms of originality, significance and rigour, the Unit's research output is comparable with outstanding work internationally in the field. The research possesses the requisite quality to meet the highest international standards of excellence. Work at this level can be a key international reference point in the field. The RU output profile is comparable to the one of the best international research organisations in the field.
4	Very good	**The RU is a Strong International Player** In terms of originality, significance and rigour, the Unit's research output is comparable with excellent work internationally. The research nonetheless does not yet meet the highest standards of excellence. Work at this level can arouse serious interest in the international academic community. The RU output profile is comparable to the one of very good international research organisations in the field.

3	Good level	**The RU is a Good International Player** In terms of originality, significance and rigour, the Unit's research output is comparable with the best work internationally. The research possesses the requisite quality to meet high international standards. Internationally recognized publishers or journals could publish work of this level. The RU output profile is comparable to the one of good international research organisations in the field.
2	Adequate	**The RU is a Good National Player with Some International Recognition** In terms of originality, significance and rigour, the Unit's research output is comparable with good work internationally. The research possesses the requisite quality to meet international standards only to a certain extent. The RU output profile is comparable to the one of modest international research organisations in the field.
1	Poor	**The RU is a Poor National Player** In terms of originality, significance and rigour, the Unit's research output falls below the international quality standards. The RU output profile is not comparable to the one of modest international research organisations in the field.
Unclassified		N/A

In this criterion, 'Global', 'International' and 'National' refer to quality standards. They do not refer to the geographical scope of the research outputs and/or publication channels.

We designed a three-stage process with a clear division of roles for the panels versus referees:

- The two referees assess the research outputs and assign to each submitted scholarly output a starred quality level, accompanied by an explanatory statement
- The panel member(s) expert(s) in the field assign(s) the final starred quality level for each submitted scholarly output
- Based on the **average** scores for all submitted scholarly outputs, the subject panel decides on the final starred quality level for the RU

Overall research performance

The panels will assess the RU overall research performance looking into

- The research output, including productivity
- The RU competitiveness in research

In the course of the calibration exercise, the panel members will discuss and define the importance of these 2 sub-criteria, in *the field and for the types of RO*.

Starred quality level	Definition	Description
5	Outstanding	**The RU is a Global Leader** In terms of research output and competitiveness, the Unit's overall research performance is internationally excellent, i.e. at the level of the best international research organisations in the field.
4	Very good	**The RU is a Strong International Player** In terms of research output and competitiveness, the Unit's overall research performance is optimal, i.e. at the level of very good international research organisations in the field.

3	Good level	**The RU is a Strong National Player** In terms of research output and competitiveness, the Unit's overall research performance is at a good standard.
2	Adequate	**The RU is a Satisfactory National Player** In terms of research output and competitiveness, the Unit's overall research performance is at an acceptable standard.
1	Poor	**The RU is a Poor National Player** In terms of research output and competitiveness, the Unit's overall research performance is poor.
Unclassified		N/A

In this criterion, 'global', 'international' and 'national' refer to quality standards. They do not refer to the geographical scope of the research activities.

The panel members will base their judgment on quantitative and qualitative data related to

- For the assessment of the research output:
  - *Research productivity*: the production of research outputs versus the size of the research unit in terms of FTE researchers (including PhD students in case their publications are assigned to the RU)
  - The *publication profile* of the RU, including the trends in publication and types of research outputs, the use of national/international journals and type of publication channels, the citation impacts. These bibliometric data will be considered in absolute and relative terms, i.e. set against the field total in the CR
  - The value of the RU activities for the *advancement of research* (self-assessment)
- For the assessment of the RU competitiveness in research:
  - The capacity to gain *external funding, i.e.* competitive funding and contract research - from international and/or national sources
  - The reputation of the RU in the research community, expressed in its ability to *attract PhD students*
    - For the RU that in practice train PhD students: the number and trends of PhD students trained and the level of investment in PhD training (PhD students versus FTE researchers)
    - For the HEIs: the number and trends of PhD students enrolled, level of investment in PhD training (PhD students versus FTE researchers), and the effectiveness of the PhD education and trend (ratio PhDs awarded/PhD students enrolled)
  - The national and international competitive positioning of the RU (self-assessment)

Societal relevance

The panels will assess the societal relevance of the RU activities in terms of their reach and significance, considering the RU knowledge and technology transfer activities and impacts - to the benefit of industry and/or the public sector and other societal actors.

Prior to the assessment, the panel members will define the *field-specific* interpretation of the terms reach and significance.

Starred quality level	Definition	Description
5	Outstanding	**Work in the RU has a Very High Potential for Societal Impacts** In terms of reach and significance, the RU is an important driver of societal development. The RU's collaborations and/or interactions with non-academics (i.e. business, policy-makers, the public) stand out in terms of their extensive and dynamic nature.
4	Very good	**Work in the RU has a High Potential for Societal Impacts** In terms of reach and significance, the RU strongly contributes to societal development. The RU's collaborations and/or interactions with non-academics (i.e. business, policy-makers, the public) are at a very high level.
3	Good level	**Work in the RU has a Good Potential for Societal Impacts** In terms of reach and significance, the RU contributes well to societal development. The RU's collaborations and/or interactions with non-academics (i.e. business, policy-makers, the public) are at a good level.
2	Adequate	**Work in the RU has a Low Potential for Societal Impacts** In terms of reach and significance, the RU contributes to societal development. The RU has some collaborations and/or interactions with non-academics (i.e. business, policy-makers, the public).
1	Poor	**Work in the RU has Little to No Potential for Societal Impacts** In terms of reach and significance, the RU makes little to no contributions to societal development. The RU does not collaborate and/or interact with non-academics (i.e. business, policy-makers, the public).
Unclassified		N/A

'Societal' impacts refer to impacts on the economy and social welfare, the latter including health, environment, culture, social inclusion, education and gender.

The panel members will base their judgment on quantitative and qualitative data related to

- The **intensity and quality** of the knowledge and technology transfer
  - To the benefit of industry: the collaboration with industry, membership of advisory boards, the volume of competitive & contract research with/for industry, income from the commercialisation of research outputs, IPR-related outputs (patents) and the geographical distribution of the patent offices, participation in incubators or clusters, the profile of the industry partners and/or clients, the use of research outputs in the industry/business environment
  - To the benefit of the public sector or other societal actors: the volume of competitive & contract research with/for the public sector or other societal actors, the publication of non-traditional scholarly outputs, the profile of the societal actors supported, the use of media channels, the implementation of other outreach activities (activities to the benefit of schools, NGOs, amateur associations like arts, reading or writing groups, local history societies, museum-goers, tourists, etc)
- The creation of **spin-off companies**
- The **value of the RU research activities for society** (self-assessment)

4.5.3 Tools for the data collection

The evaluation builds on information that is available in the national RD&I IS (research outputs and national competitive funding) as well as on information provided by the evaluated RUs themselves.

The evaluated RUs are requested to provide quantitative and especially qualitative information. Quantitative data are related to their research personnel, PhD students enrolled or trained, external funding and income gained from the commercialisation of research outputs.

The **qualitative information** is provided through

- Background information that highlights certain conditions for the research that is evaluated. This information will not be appraised, but it provides context and understanding to the other evaluation material
- The Unit's own description of its research
- The RU self-assessment
- The RU SWOT analysis

Compared to the draft version of the EM, we have set a stronger focus on qualitative rather than quantitative data and introduced a stronger self-assessment focus in the EM.

This follows the feedback from the panel members in the Small Pilot Evaluation. On the one hand, they underlined the need for such more qualitative information that would enable them setting data in their context. This will compensate better also for the lack of site visits.

On the other hand, a major remark of the panel members was the apparent general lack of attention for an appropriate transfer of *qualitative* information by the evaluated RU. No doubt the time factor will have played a role there (the evaluated RUs had relatively little time available for filling in their submission forms). However, there is little doubt that also the understanding of evaluation as a somewhat arithmetic exercise has played a role.

An important value of evaluations such as the one designed in this study is that it constitutes a major opportunity for a collective self-assessment of research performance by the researchers and their management structure, leading to an improved understanding of strengths and weaknesses, and ultimately – an improved research strategy. By increasing the importance of these activities in the current EM, we hope to foster this type of approach to research management in all research organisations.

The topics for self-assessment included in the EM are:

- The adequacy of the research infrastructure and facilities
- The RU research strategy
- The value of the RU activities for the advancement of research
- The competitive positioning of the RU in the national and the international context
- The societal relevance of the RU activities
- The RU SWOT analysis

The background report on the Guidelines for the evaluated RO provides more detailed information on the description of the information expected in these narratives.

4.5.4 The indicators used

We selected the indicators to be used in the EM with a set of simple guidelines in mind. The set of indicators

- Reflects the strategic priorities
- Is based on our knowledge of international practice and on the value of specific input, output, process and system indicators as outlined in Section 3.8, above
- Takes into account data availability, in particular in the Czech RD&I IS, in order to promote cost efficiency
- Constrains to the degree of complexity, fairly simple and understandable to the research community and the broader public
- Focuses on indicators that are reliable and data that is verifiable
- Is fair in its coverage of indicators that are of relevance to the different scientific disciplines and research organisation typologies
- Takes into consideration potential gaming - to the extent possible
- Considers the potential effects of the indicators on the Czech R&D system
- Takes a comprehensive approach: it covers research quality from a scientific perspective as well as in terms of societal relevance, and considers the quality of the research environment
- Last but not least: is composed of a mix of quantitative and qualitative data allowing for triangulation

The **variables** used for the quantitative data and their analysis will give the panels a view on

- Trends over the years in the evaluated period
- The positioning of the RU in the EvU and the field (share of the RU in the EvU; share of the RU in the field)

In the remainder of this section we cover some specific aspects related to the indicators used, i.e. the use of an indicator mix allowing for triangulation, the calculation of FTE researchers and the challenges it presents, thresholds and rulings for the research outputs, bibliometric indicators, and some final considerations on how to deal with the limits in bibliometric data availability for some fields of science.

A mix of quantitative & qualitative data allowing for triangulation

In the description of the assessment criteria above, we gave an overview of the main indicators that will provide the information for the assessment against each criterion.

They cover the broad range of indicator typologies and encompass a **mix** of quantitative and qualitative data. This approach reduces the potential effects of gaming – and therefore indirectly the phenomenon itself.

The robustness of the assessment is ensured also by the use of **multiple** data to inform the panels against a specific criterion. Many of these indicators provide information against various criteria and enable for **triangulation** of the information, i.e. crosschecks of the data provided. We illustrate this concept in Exhibit 26, below.

Exhibit 26 Mix of qualitative and quantitative indicators informing the assessment criteria

Indicator category	Quantitative	Qualitative	Main Indicators	Research environment		Membership of the global & national research community		Scientific research excellence	Research performance		Societal relevance
				Research management	Research strategy	International research presence and collaboration	National research presence and collaboration	Peak quality of the scholarly outputs	Research output	Competitiveness in research	Knowledge & technology transfer
Input	X		External funding			**XX**	**XX**			**XX**	**XX**
	X		Research capacity	**XX**						X	
	X	X	Infrastructure for research	**XX**		X	X			X	X
		X	Research strategy		**XX**						
	X	X	Human resource development	**XX**							
Process		X	Reputation and esteem			**XX**	**XX**				**XX**
	X	X	PhD education & training	X			X			**XX**	
	X		International citations						**XX**		
Systemic	X	X	International mobility			**XX**				X	
	X	X	International collaborations & partnerships			**XX**				**XX**	
	X		International co-publications			**XX**				X	
	X	X	National collaborations & partnerships				**XX**			X	**XX**
	X		National co-publications				**XX**			X	
Research outputs	X	X	Research outputs					**XX**	**XX**		**XX**
	X		Use of national/ international journals						**XX**		
Impacts	X		Licence income								**XX**
			Spin-off companies								**XX**

FTE researchers

Somewhat surprisingly for an international observer, the definition of the size of an RO in terms of FTE researchers is not a straightforward matter in the Czech RD&I system.

Factors that make this exercise particularly cumbersome are

- There is no central registry of researchers in public organisations
- The data on FTE researchers in a single research organisation differ substantially depending on the sources used
- By law, employers are not allowed to investigate double or even triple full-time occupations of their employees. As a result, there are no official data on this common phenomenon
- Some PhD students may have a contract to conduct research (n addition to their student grant) and in the HEI they are at that point officially counted into the FTE figures
- Some HEI count in also employees that would normally be categorised as technicians
- There is no consensus in the research community on how much time a university professor spends on average in teaching versus research, nor is this topic a component of the employment contracts

As a result, this is an indicator at high risk for gaming.

We managed this risk as follows:

- Based on the experience gained in the Small Pilot Evaluation, in the Guidelines for the evaluated ROs (see the Background report) we give precise definitions for the different job titles and functions in the research organisations, based on existing definitions in the different research organisations and/or used by the Czech Statistical Office
- We ask for Headcounts and FTE numbers on researchers – at the overall EvU level and for each RU. Because each researcher in an RU has to be registered in an RU and can be registered only once, the sum of the data for the RUs constitute 100% of the data for the EvU
- We underline that the number FTE researchers should be calculated in line with the indications from the Czech Statistical Office, i.e. taking into account only the 'real' working time that is dedicated to research
- We ask for Headcounts and FTE numbers of PhD students that are included in the 'official' number on researchers. The concept is that technicians and PhD students should not be considered 'researchers'
- The data will be checked on consistency by the evaluation management team

The complexity and cost for this exercise will be tested during the 2nd Pilot Evaluation.

Quite obviously, these risk management measures will increase the burden on the evaluated RO, in particular the Higher Education Institutions. We therefore invite the Czech community to consider the following longer-term solutions:

- Reach a consensus for the average time spent for research versus other activities such as teaching in the different types of research organisations or introduce the concept of teaching versus research time indications in the staff contracts
- In relation to the double employments, re-install a Central Registry including the names of researchers employed by all RO receiving institutional funding or use the RD&I IS to identify the (most troublesome) cases

Research outputs, thresholds and rulings

The research outputs that are taken into account for the calculation of the threshold for an RU registration to the evaluation and for the assessment of the RU scientific research excellence and research output are listed in Exhibit 27, below.

We have established upper and lower limits for the number of publications to be submitted for the assessment of **scientific research excellence:**

Each Research Unit will submit for review a number of research outputs that accounts for ***minimum 1% and maximum 2%*** *of the total number of scholarly outputs by the researchers in the Research Unit over the evaluation period* ***- but however no less than 3 and no more than 20.***

For example,

- A research unit that has published in total 1000 scholarly outputs from 2009 until 2014 can submit between 10 to 20 research outputs for review
- A research unit that has published 100 scholarly outputs should submit 3 outputs for review
- A research unit that has published 3000 scholarly outputs should submit 20 outputs for review

Rulings related to **research output** and **threshold** calculations are:

- Books will count as 4
- Co-publications of researchers active in the same RU will be de-duplicated (i.e. counted once)
- Co-publications of researchers active in the same EvU but registered in different RUs will be counted as one each, on condition that they are of a clear *interdisciplinary* nature. This will be subject for audit by the evaluation management team (see Section 5.3.2)
- Co-publications of researchers active in different EvUs will be counted as one each, thus keeping the incentive for collaborations among the research organisations

Exhibit 27 Research outputs eligible for the threshold and the research excellence and productivity assessment

	Threshold for RU registration	**Scientific research excellence**	**Research output**
Scholarly outputs			
Papers in peer-reviewed journals (J)	X	X	X
Conference proceedings (D)	X	X	X
Monographs, books and book chapters (B), provided they are identified with an ISBN number	X	X	X
Non-traditional scholarly outputs			
Results used by the funding provider, i.e. projected into legislation or norm, projected into non-legislative or strategic documents (H)	X		X
Research report containing classified information (V)	X		X
Certified methodologies, art conservation methodologies, specialized map works (N)	X		X
Patents and other IP			

	Threshold for RU registration	Scientific research excellence	Research output
Patents and patent applications (P)	X		X
Plant/ breeders rights (Zodry & Zplem)	X		X

Notes: the definition of the research outputs is as listed in the Methodology of Evaluation of Research Organisations and Evaluation of Finished Programmes (valid for years 2013-15, Office of the Government of the Czech republic, File No.: 1417/2013-RVV

The selection of the research outputs listed above has been subject of major critique by the research community, in particular the actors active in applied research.

The expectation by research communities to have a broader range of research outputs included in the count-out for the assessment of their research performance is of course legitimate. There are, however, a number of reasons why - under current circumstances - a broader range of research outputs seems not desirable.

First, trust of the research community in the EM depends first of all on the reliability of the data collected. The Metodika 2013-2015 considerably restricted the types of outputs that previously were eligible in response to critique in the research community that many of these outputs were subject to a significant level of gaming. There was a general lack of trust that the registered research outputs were 'real' and effectively constituted research, despite various control mechanisms. In the context of a PRFS where research outputs constitute the only criterion, the issue is quite obviously serious. The approach taken was to include only verifiable (and verified) research outputs.

In the context of the new EM, where the number of research outputs is only one of many indicators and where metrics are only one of the elements informing panel assessment, the topic is obviously less problematic – or at least, so would one expect.

Nevertheless, we felt that inclusion of types of research outputs beyond the ones included in the Metodika 2013-2015 needed to provide sufficient guarantees that the risk for a repetition of the 'old problem', i.e. gaming, was minimal or however acceptable.

In order to ensure a fair coverage of research outputs also for research organisations beyond academia, also a small number of non-traditional scholarly outputs were included. These outputs were selected on the basis of the reliability of their verification process, which is linked to their registration in external databases. The inclusion of the V-type non-traditional output was needed to cover security research.

Second, it should be considered that while evaluation panels can assess the quality of scientific research outputs on the basis of bibliometric data, the quality assessment of research outputs in the applied research sphere requires a view on the value of the research results for future development and most importantly, for the *users* of the research results. A simple number of prototypes or softwares produced by the RU does not provide any useful information to the evaluator.

In the EM, the evaluated RUs are provided ample opportunity to inform the expert panels on the value of their research activities from these two perspectives.

Bibliometric indicators

The selection of the bibliometric indicators is based on:

- The purpose of informing field specific panels in the best possible way
- Relevance for the institutional level of evaluation
- Well-established international practice in the field
- Availability, compatibility and transparency in relation to chosen data sources

The indicators are in three main categories: Publishing Profile, Citation Impact, and Collaboration.

Indicators are calculated and presented per OECD field. Indicators can be aggregated (weighted for average citation rates) up to major area, research organizations, country. A measure of interdisciplinarity will be number of articles per field at RO level.

Publishing profile

1. Number of publications in RD&I IS categories: articles in peer-reviewed journals (type J); monographs (type B); book chapters (type C - includes Articles in Books and Chapters in Books); Proceedings papers (type D)
2. Percentage Web of Science/Scopus publications among all peer-reviewed RD&I IS publications
3. Number and percentage publications in each field as an indicator of the research profile of the RU
4. Number of Web of Science publications in the document types: Article (incl. combined with other types); Review; Letter.
5. Mean and median number of authors and addresses per WoS publication
6. Percentage WoS/Scopus publications in the most cited 10 per cent and 25 per cent of the journals in the field (counted from the top by the number of articles in the field), based on Journal Impact Factor in the latest edition of Journal Citation Reports

Citation impact

a. Field Normalized Citation Impact (world, EU28)

b. Number and percentage publications among the top 10%, and 25% most cited publications (world, EU28).

Purpose: To present the panel with information about the international influence or impact of the research, as seen in the frequency of received citations per publication

Collaboration

- Percentage WoS/Scopus publications exhibiting international collaboration in the addresses
- Five most frequent collaborating countries in the field, and their shares of the publications in field
- Field Normalized Citation Impact of articles in each country relation
- Percentage WoS/Scopus publications exhibiting national collaboration among Czech institutions in the addresses
- Five most frequent collaborating institutions in the field, and their shares of the publications in the field

Solutions for fields with insufficient coverage in WoS and Scopus

As demonstrated in Section 3.6, above, the social sciences and humanities (SSH), partly also the engineering sciences, are so far insufficiently covered in commercial citation indexes Web of Science and Scopus. The situation has been improving during the last years (after the introduction of Scopus as a competitor), but both data sources are still far from covering scholarly publications on the national level, especially publications in books, to the extent that would be acceptable for evaluation purposes in these fields. Even international journals are still insufficiently covered in some the fields in the SSH.

With insufficient coverage in the commercial international data sources, we trust that the Czech RD&I IS will cover the outputs from these fields comprehensively. This is one of the main reasons for including the RD&I IS as a data source.

But even with this solution, some problems remain. We will list the problems here and explain how or to what extent they can be solved. The focus here will mainly be on SSH fields.

Partial coverage creating tensions among fields

As seen in Section 3.6, above, there are large variations in the coverage of the SSH fields in WoS. Scopus follows the same pattern of insufficiencies, although the coverage of each field there is broader.

The result is that some SSH fields may "feel at home" with the way their output is represented in citation data, and they may even regard citation indicators as useful and valid because it corresponds to their publishing and citing practices. Common examples are economics in the social sciences and general linguistics in the humanities. At the other extreme are fields, such as law or philology (except English philology), that rightfully can see themselves as misrepresented, and where citation indicators would be completely invalid. Even within a field, there may be differences. An example within political science is International Relations, which is well covered, and Local Government Studies, which is not as well covered.

It is important to see these differences as rooted in different publishing and citing practices, which again are more deeply rooted in the missions and methodologies of the research in the fields. Law or philology cannot be blamed for the insufficient coverage, neither can the companies that produce the citation data be blamed. This is important to note, because these data sources are often regarded as selecting the highest quality and most important journals for coverage. The producers also state that they follow certain quality when including new journals, but we also observe that expansion the journal list mainly happens in countries in parts of the world that represent new markets for the producers. From a European perspective, the impression is rather that the journal coverage in the SSH is a lottery.

To avoid creating tensions among fields in the SSH, it is important not to regard coverage in the commercial databases as an indication of research quality in a given field. Instead, less covered fields should be respected for their specific publishing pattern that is insufficiently covered. More sufficiently covered fields should on the other hand be respected for their increasing interest in citation indicators.

Accordingly, the indicator *Web of Science coverage* is not an indication of research quality in our methodology, but on the contrary, an indication that awareness should be taken if numbers are low.

Partial coverage – but the indicators are still invalid

Coverage in a citation database is not sufficient for the validity of some of the indicators we use:

- For the citation indicators, the references in typical citing documents must mainly relate to other recently publications that are also covered in the database. If this is not the case, the number of citations to a given publication, even if it is covered, will be very low and invalid as an indication of influence on further research
- The indicator for the publishing profile that is based on journal impact factors does not give much meaning if citations can seldom be measured in the field (see above) and a large proportion of the publications in the field are not in journals covered by the data source
- Collaboration indicators with data from co-authorship will not be valid in fields where there is usually one author per publication. The lack of co-authors abroad is in these cases not an indication of lack of collaboration

In our methodology, we will not use the three mentioned (groups of) indicators in fields where we see them as invalid, even if the field is partially covered in the citation data.

More weight on qualitative judgment by peers

In fields with insufficient coverage in the international databases, and where some of the indicators will be invalid, it is indeed possible for experts in the field to judge the same aspects by reading publications and using the contextual information they are provided with. It is, as an example, possible in typical book publishing fields to evaluate the originality, significance and impact of a monograph from an international perspective without looking at citation indicators. In countries using bibliometrics to inform research evaluation for institutional funding (e.g. Australia, Italy, UK), it is normal practice to give more weight to qualitative judgment by peers in fields where bibliometric indicators have limitations. Thereby, one opens up also for the evaluation of a larger multitude of aspects of research quality that SSH scholars are used to judge.

In some countries, these considerations led to an additional set of submitted research outputs assigned for peer review in these fields; in other countries, the evaluators compensated with a more extensive analysis of the valid information at hand, such as the types of journals used, the profile of the conferences for which proceedings were registered etc. Finally, some countries tried to solve the issue by defining quality categories for journals published in the national language (and not indexed in the international databases) or for the conference proceedings, categories of the conferences (with our without peer review).

In the EM, we opted for the second approach, i.e. a more extensive use of the information at hand by the evaluators. In a longer-term perspective, the Czech research community should consider solutions listed above as the third approach. WE will cover this in further detail in the Third Interim report.

4.6 Taking diversities into account

4.6.1 Disciplinary cultures and missions

Field specifics are at the core of the evaluation methodology that we propose in this report. Subject panels made up of field experts are best placed to take account of the particularities of the scientific fields and differences among sub-fields as well as the characteristics of different types of research organisation, ensuring fair assessment (see Section 3.6, above). The fact that experts only assess research in their own field means that they can apply their field-specific understanding in order to translate particular patterns of evidence into judgements about performance that are universally recognisable within the overall research community.

Consideration of the field specifics is nonetheless sensitive in research assessments, and in particular in PRFS where the results drive a considerable proportion of institutional funding. The UK RAE/REF is an example of such a PRFS. Over time it has developed a highly nuanced approach by allowing the panels to take account of the field-specific characteristics of the generic indicators used for the assessments, as well as in certain cases admitting specific types of evidence for consideration in some fields but not in others (see Section 3.7.4, above).

While it seems appropriate **not** to adopt an overly detailed and resource-intensive approach, one aspect of the RAE/REF methodology of relevance for this study is the extend to which panels are able to assign different amounts of importance to different types of indicators, based on their relevance to each particular field. The link from evidence to assigning assessment scores is judgemental and not based on arithmetic.

A reasonable level of freedom for the expert panels to define field-specific variations of the common generic criteria seems appropriate. Our proposal is therefore for the subject panels to implement a calibration exercise at the very start of the evaluation process. As mentioned in Section XXX, above, this discussion during the first meeting

of the subject panel should generate a common understanding of the indicators among the panel members and agreement about how these should be interpreted in the context of specific fields and research organisation types.

It will be the task of the main panel chairs to ensure coherence in the interpretation of the assessment criteria and their key words among the panels, so that, for example, a score 4 against the criterion 'research performance' has an **equal** value in physics as in social sciences, for any type of organisation.

While the scores against the different assessment criteria will therefore be 'field-neutral', the different missions of the research organisations will be taken into account in the second stage of the PRFS process, i.e. for the allocation of the funding. We explain this in further detail in the 2nd Interim Report.

4.6.2 Inter-disciplinary research

We handled the issue of assessing inter-disciplinary research by

- Defining a limited set of 'subject' panels, covering broad fields of research
- Allowing for research units to flag work across fields within a single disciplinary area, and foreseeing cross-referrals among panels
- Exceptionally allowing for the registration of Interdisciplinary Research Units

The considerable fragmentation of research in the Czech Republic and the allocation of research outputs in the scientific fields by the researchers themselves upon registration imply a potential abuse of these opportunities offered. Clear boundaries and rulings are therefore required. These are:

- The evaluation is not conducted at the level of individual researchers but research units, so requests for cross-referrals among panels will be considered only for multi-disciplinary work of a research unit as a whole
- Both cross-referrals and the status of Interdisciplinary Research Unit will be subject to specific thresholds.
- A Research Unit can apply for the status of Inter-disciplinary Research Unit in case it conducts at least 30% of its research *across* disciplinary areas. In order to qualify for this status, it will need to demonstrate true inter-disciplinarity, by means of research outputs/bibliometrics, the scientific background of the researchers, and a statement of research strategy
- Research Units can indicate their interdisciplinary activities *within* a disciplinary area. They can recommend cross-referrals for their assessment among subject panels in the disciplinary area only if at least 30% of their activities take place across fields. This is to be demonstrated in terms of research outputs/ bibliometrics or the scientific background of the researchers in the unit

Rulings on who will decide on the approval of an Interdisciplinary Research Unit and the cross-referrals are in Section 5.2.4, below.

Further measures to facilitate the assessment of interdisciplinary research are foreseen in the panel working methods, described in Section 5.3.1, below.

4.7 The evaluation results

The panel evaluations will result in the following reports:

- A panel report per RU
- An overview panel report for EvUs with more than 1 RU, prepared by the assigned subject panel chair
- An analytical report per field, prepared by the subject panel chair

- An analytical report per disciplinary area, prepared by the main panel chair

These reports will become public when the evaluation process is finalised.

The panel reports will entail the final score for the RU and an explanatory statement arguing the evaluation result against each assessment criterion, and conclusions and recommendations for the future development of the RUs in the context of their area of research and the national science and innovation system. This may include:

- The potential evolution of the research environment and infrastructure, including strategic management and operational issues and the composition of research staff
- Opinions regarding the potential for collaboration with other institutions and for interdisciplinary research

The panels will **not** combine the scores for the 5 criteria into a single score for each RU as an overall assessment result.

The reflections that led to this decision are the following:

- The **risk for bias**: the Evaluation Methodology covers all research organisations, with activities ranging from basic research to development. To a certain extent, these research organisations are set in competition with each other. In the Evaluation Methodology that we propose, this competition takes place against each of the 5 criteria separately. At this level, the risk for a bias in the judgment of the panel members due to their research profile is limited. Nobody will have difficulties in accepting that an RTO performs better in reaching societal relevance than a basic research institute.

  The situation changes when an overall judgment of performance needs to be made. At that point, research organisations 'compete' against each other in terms of 'quality in overall performance'. The risk for scholarly bias in favour of research organisations is high in those cases.

- The **different contexts for the assessment criteria**: the criterion on research environment, and in particular the research management sub-criterion, reflects the desire for performance assessment at the institutional level. Information on these 'institutional conditions' also gives the expert panels a view on the context in which the research is conducted.

  However, decision-making on the research management aspects takes place at different level in the various organisations involved in the evaluation. For the small non-university research organisations, it is at the level of the RU; for the larger organisations and the public HEI, the responsibility is at the level of the EvU.

  As a consequence, the assessment of the research management is effectively an assessment of the RUs' performance in the smaller research organisations, similar to the other criteria. In the other cases, instead, it is an assessment of the management by the EvU, in contrast to the other criteria. In those cases, the aggregation of the scores against the different criteria will combine and compare apples and pears.

A maximal use of the evaluation results from a **formative** perspective consists in providing information on performance against the different assessment criteria, rather than an aggregate score. Aggregate scores are used to enable rankings of units or research organisations, indirectly creating competition amongst them. That was not the purpose of this evaluation.

As can be noted from the structure and content indications for the Panel Reports, the emphasis is on the delivery of **qualitative** information, aimed at supporting R&D governance at the institutional and national level.

An appropriate use of IT tools will allow for the grouping of the panels' reports on the RUs level at EvU level, and the collection of the chapters on field performance to the benefit of the national policy makers.

The benefit and value for the **Evaluated Unit** will consist in:

- A view on the Research Units' performance against each of the 5 assessment criteria, allowing for identification of the weaknesses and strengths in the EvU's overall performance, as well as eventual pockets of excellence and areas of lower-than-average performance
- The panel conclusions and recommendations for future development of the RUs, providing a valuable input and suggestions for future strategy development, including
  - A feedback on the effects of the EvU research and HR management policies on research groups active in specific fields as well as an assessment as such of these policies
  - Input for the development of interdisciplinary collaborations
- A comprehensive view on the EvU international and national level of competitiveness in R&D

The benefit and value for the national policy makers will consist in:

- A comprehensive view of research performance in the country at the level of fields
- Identification of areas of weaknesses and strengths at the level of fields
- An overview of performance against each of the five assessment criteria, suggesting the areas of major failure in the R&D system

The Third Interim Report, containing the 'lessons learned' from the Small Pilot Evaluation, will contain examples of Panel Reports.

4.8 Data sources

4.8.1 Categorisation of scientific fields & areas

In order to facilitate comparisons and linkages among data stored in the Czech RD&I IS and other international databases (such as bibliometric ones), we have decided to adopt the OECD FOS 2007 classification, the latest update of the Frascati manual classification, for the structuring of data in the RD&I IS. It consists of 6 major areas with 36 fields and 6 "other" categories. There is a third level of the classification that has approximately 190 sub-fields.

A first analysis of how these fields match with the Czech RD&I system led to the following modifications that are included in table 3 below:

- Computer and information sciences (1.2) is moved from major area 1 (Natural Sciences) to area 2 (Engineering and Technology). Reason: This is the normal classification of the field, also according to publication patterns.
- Biological sciences (1.6) is moved from area 1 *(Natural sciences)* and united with Agricultural sciences (area 4). Area 4 is renamed *Biological and Agricultural Sciences*. Reason: The other natural sciences (physical sciences) have a particularly large activity in the research profile of the Czech Republic, and so do the Biological Sciences if united with the Agricultural Sciences.
- Area 1 (Natural Sciences) is renamed *Physical Sciences*.

The current categorisation is shown in Exhibit 28, below. A transition period of approximately 5 years will be needed before a *full-fledged* use of this new classification system will be possible.

In Exhibit 28 we show the matching of the fields currently in the RD&I IS to the OECD FOS structure.

Exhibit 28 Structure of disciplinary areas and fields

Major area	Field
1. Physical Sciences	1.1 Mathematics
	1.3 Physical sciences
	1.4 Chemical sciences
	1.5 Earth and related environmental sciences
	1.7 Other natural sciences
2. Engineering and Technology	2.1 Civil engineering
	2.2 Electrical engineering, electronic engineering, information engineering
	1.2 Computer and information sciences
	2.3 Mechanical engineering
	2.4 Chemical engineering
	2.5 Materials engineering
	2.6 Medical engineering
	2.7 Environmental engineering
	2.8 Environmental biotechnology
	2.9 Industrial Biotechnology
	2.10 Nano-technology
	2.11 Other engineering and technologies
3. Medical and Health Sciences	3.1 Basic medicine
	3.2 Clinical medicine
	3.3 Health sciences
	3.4 Health biotechnology
	3.5 Other medical sciences
4. Biological and Agricultural Sciences	1.6 Biological sciences
	4.1 Agriculture, forestry, and fisheries
	4.2 Animal and dairy science
	4.3 Veterinary science
	4.4 Agricultural biotechnology
	4.5 Other agricultural sciences
5. Social Sciences	5.1 Psychology
	5.2 Economics and business
	5.3 Educational sciences
	5.4 Sociology
	5.5 Law
	5.6 Political Science
	5.7 Social and economic geography
	5.8 Media and communications
	5.9 Other social sciences
6. Humanities	6.1 History and archaeology
	6.2 Languages and literature
	6.3 Philosophy, ethics and religion
	6.4 Art (arts, history of arts, performing arts, music)
	6.5 Other humanities

Based on and elaborated from OECD. REVISED FIELD OF SCIENCE AND TECHNOLOGY (FOS) CLASSIFICATION IN THE FRASCATI MANUAL, version 26-Feb-2007, DSTI/EAS/STP/NESTI (2006)19/FINAL.

The structure of the OECD disciplinary areas and fields will define the work and coverage of the main panels and panels in the evaluation, respectively.

However, this does not imply that 36 panels will be established. In the preparatory phase of the performance assessment, **subject panels** will need to be defined, taking into consideration the volume of the research conducted in the CR in the specific fields, in terms of number of research units and research outputs produced over the evaluation period. This will be based upon input from the research organisations on the fields they wish to be assessed against, i.e. at the moment of registration of the research unit(s). The intent is to spread the assessment workload over the different panels as much as possible on an equal basis as well as reach the maximum level of efficiency.

The identification of the subject panels will

- Be in line with the OECD categorisation of scientific fields in disciplinary areas and sub-fields into fields
- Spread the assessment work for scientific fields with exceptionally high research volume over two or more subject panels (i.e. aggregating the relevant subfields into 2 rather than 1 field)
- Concentrate the assessment work for scientific fields with exceptionally low research volume into one subject panel

4.8.2 The use of the national RD&I Information System (IS)

The RD&I IS constitutes an important tool for the design and implementation of the evaluation exercise. It has the capacity to support the evaluated RU with data that is stored in the national information system, contributing to a significant reduction of the evaluation burden for the research organisations. The RD&I IS also is an important tool for the implementation of eligibility and data quality checks during the evaluation, as well as for the validation of data submitted by the RU.

No doubt the RD&I IS can be improved so that it can be better exploited for evaluation purposes. This regards for example

- Increased data quality in the R&D IS. This concerns both the quality assurance of the input and how the data are structured, e.g. with standardized registers of journals, foreign institutions, etc.
- A further development of output categories for the purpose of evaluation

We will provide suggestions from this perspective in the Third Interim Report.

5. Implementation of the Evaluation Methodology

The entities involved in the implementation of the Evaluation Methodology are

- The Evaluation Management Board, acting as overall governance and supervisory body
- The Evaluation Management Team, responsible for the operational management of the evaluation
- The structure of Main Panels, Subject Panels and Referees, eventually supported by Specialist Advisors

The language for the evaluation exercise will be English.

Based on our current cost estimates, we propose the evaluation to take place every 5 to 6 years.

In this Chapter we describe the structure for the evaluation governance and management (Section 5.1), the evaluation panels and referees (Section 5.2) and the measures developed to ensure the fairness and integrity of the evaluation process (Section 5.3).

5.1 Evaluation governance and management

This Chapter describes the governance structure for the evaluation and the roles and tasks of the central Evaluation Management Team.

5.1.1 The governance and management structure

The **Evaluation Management Board** is the supervisory and ultimate governance body for the evaluation exercise. In order to reach the maximum level of legitimisation, this body should be composed of representatives of the R&D governing bodies (funding ministries and ministries with responsibilities for research institutes, the Academy and the agencies), chaired by a representative of the Government Office. The Government Office installs the Evaluation Management Board.

The Evaluation Management Board oversees the entire evaluation process and selects the members of, and installs, the Evaluation Management Team.

The **Evaluation Management Team** is the body responsible for the operational implementation of the evaluation. It consists of

- An *Evaluation Directorate*, taking the operational decisions
- An *Evaluation Secretariat*, responsible for the operational implementation. This will include evaluation experts, statistical data analysis and bibliometrics experts, and IT staff
- The *Panel Secretariats*, supporting the main and subject panels. Members of these Secretariats should be acknowledged on research in the specific fields and disciplinary areas.

The Evaluation Management Directorate selects the staff of the Evaluation Secretariat and members of the Panel Secretariats.

The research communities, represented by the RD&I Council, provide scientific advice as well as support for the evaluation implementation to both the Evaluation Management Board and the Evaluation Management Team.

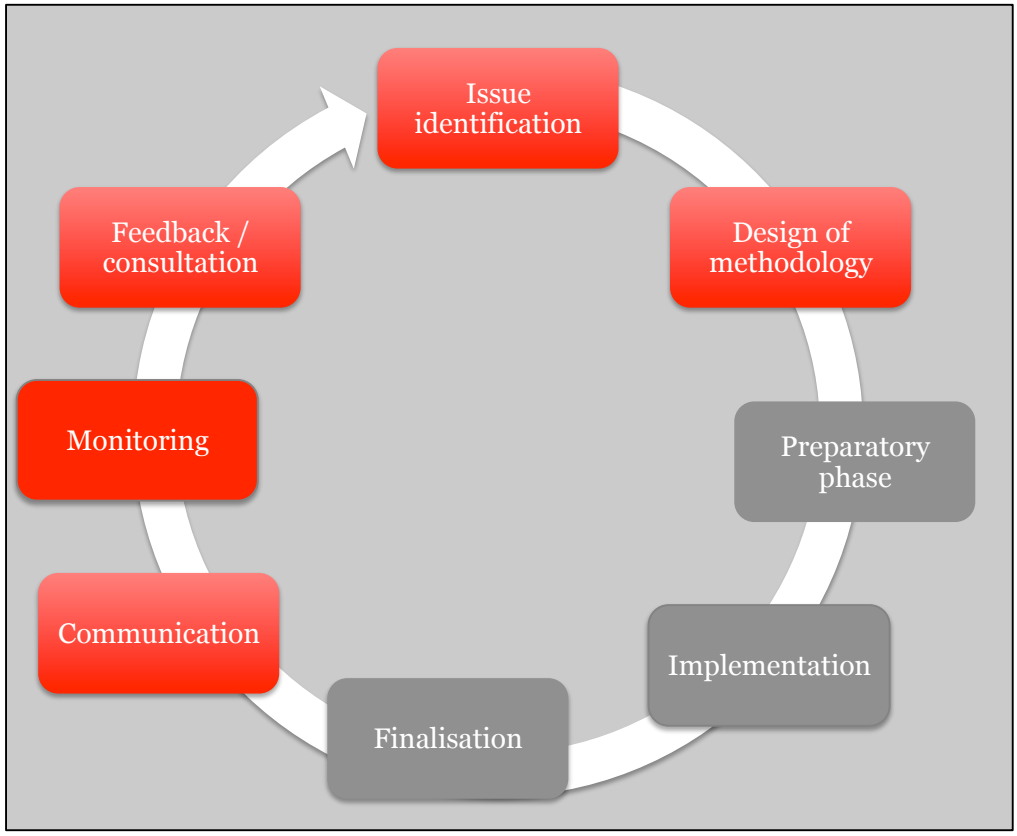
The implementation of the evaluation should **not** be considered as an alone-standing ad-hoc exercise. Instead, it is (only) a component in the evaluation cycle as shown in

Exhibit 29, below. Especially the activities of the Evaluation Directorate and the Evaluation Secretariat will be part of a continuous process of evaluation design, implementation, monitoring of its (intended and un-intended) effects, and design of an updated methodology, taking 'lessons learned' into consideration and eventually responding to new policy priorities and needs.

In this context, the establishment of a **stable** structure, allowing for the build-up and continuous use of skills and expertise, seems an appropriate choice. The IPN project team, which is responsible for this study, has among its tasks to analyse the feasibility of establishing such a structure and its possible variants.

In this report, however, we will describe

Exhibit 29 The evaluation cycle



5.1.2 Roles and tasks of the Evaluation Management Team

The Evaluation Management Team is responsible for the organisation and implementation of the whole evaluation process. This process can be sub-divided into 3 phases as shown in Exhibit 30, below, i.e. the Preparatory Phase, the Assessment Phase, and the Finalisation Phase.

The main tasks of the Evaluation Management Team are

In the Preparatory Phase

- **Designing and testing information systems.** For the evaluation process several information systems are used, such as an online tool for the registration of the EvUs, submissions of research output, and the national RD&I Information System. In the preparatory phase all these systems have to be designed, optimised and tested.

- **Analysis of the data in the RD&I Information System.** During the preparatory phase, the Management Team and Evaluation Secretariat will conduct a preliminary analysis of the information that is available in the national RD&I Information System. This analysis allows for updating the overview of relevant research in the Czech Republic and anticipating the panels' workload during the submission phase (such as the volume of research in specific scientific disciplines). This will inform the decision making on the focus of the subject panels for the evaluation
- **The design and publication of the Evaluation Protocol and Guidelines**.

  The Evaluation Protocol explains the approach to the evaluation. It lists the objectives of the evaluation, the assessment criteria and indicators used, the subject panels and their focus in terms of fields and sub-fields, a description of the evaluation process, the time line for the evaluation, the deadlines for registration and submission of information, etc.

  The Guidelines for the Evaluation Panels set out the tasks and responsibilities of the different panels and panel members, describe in detail the assessment criteria and the indicators used, explain the work process and the support provided by the Evaluation Management Team, and contain the templates for the remote review reports, the remote assessment reports, and the panel reports.

  The Guidelines for the Evaluated Research Organisations set out the process and criteria for the registration to the evaluation and the submission of the requested information, indicate the time line, provide detailed descriptions and instructions for the information to submit and the self-assessment.

  These documents will be published on the Evaluation Secretariat website and a helpline will be set up to provide further support and information.

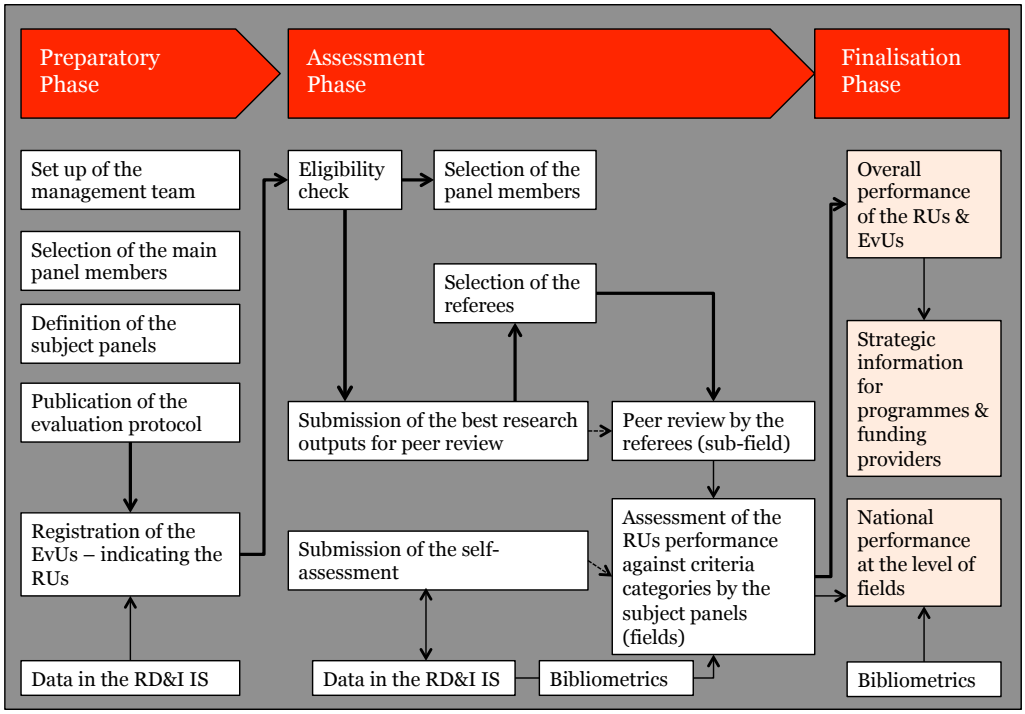
During the Assessment Phase

- The delivery of support for the selection and nomination of the panel members and referees
- The implementation of eligibility checks for the registered RU and the quality check of the data submitted
- The delivery of data support to the evaluated research organisations, transferring information from the RD&I Information System
- The delivery to the subject panels of the data reports for each RU, based on the processing of the data submitted by the RU and the analysis of the bibliometric data
- The support to the subject panels by the Panel Secretariats

During the Finalisation Phase

- The coordination of the panels' drafting of the Panel Reports
- The publication of the evaluation results and the transfer of structured information to the national R&D governance bodies

Exhibit 30 Workflow in the evaluation process



5.2 The evaluation panels and referees

In this section we describe the different entities involved in the evaluation process itself, their roles and tasks, the processes for the staffing, and the working methods.

5.2.1 The entities involved

The evaluation implementation is entrusted to a core structure of **main panels**, **subject panels** and **referees**.

- There will be 6 main panels, organised at the level of disciplinary area, and approximately 24 subject panels, organised at the level of field. Referees will assess submitted research outputs and will work at the level of sub-field
- The main panel will have a chairman and 3 additional members
- Panels need to be small and high-level. Recommendation is to keep the number to 5 max 6 members per panel
- Main panels and subject panels convene for three to five physical meetings

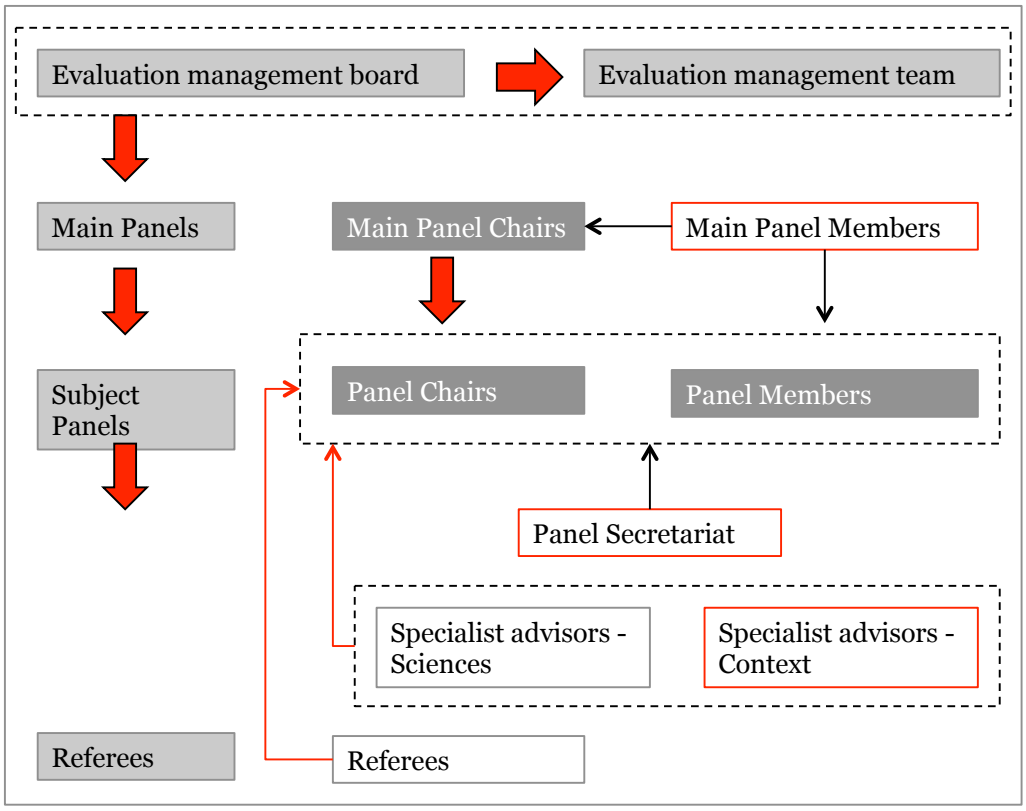
A **Panel Secretariat** assists each subject panel and main panel. The task of the panel secretariat is to support the panels in their work, in particular during the subject panel meetings, taking notes on the panel discussions, registering the scores decided, and acting as a bridge between the Evaluation Management Team and the panels whenever problems occur. They are also to ensure that the rules are followed, supporting consistency in the approach among different panels. The notes taken during the panel discussions constitute an important source for the drafting of the panel reports. Each subject panel will have 2 secretariat staff members.

The subject panels will have the possibility to call in the support by a number of **specialist advisors** (max 4). These can be international field experts, providing additional field expertise, or national experts providing context information, depending on the needs of the panel members.

In the context of an evaluation making use of panels exclusively made up by international experts, it is important to ensure mechanisms and processes that can provide the panel members with context information, e.g. on the research field in the national context and/or on practices in the national R&D system.

In Exhibit 31, below, illustrating the structure for the evaluation implementation, the 'local' actors involved that can provide the international panels with such information are indicated in red.

Exhibit 31 The main panels, panels and referees



5.2.2 Staffing of the panels

Process

The staffing of the panels will take place in a cascade process:

- The RD&I Council proposes the main panel chairs and members, based upon a consultation process, to the Evaluation Directorate
- The Evaluation Directorate considers the suggestions and presents the list of preferred main panel chairs and members, and/or alternative candidates, to the Evaluation Management Board, responsible for the nomination
- The main panel chairs and members assign the subject panel chairs and members, which will be nominated by the Evaluation Directorate. The Evaluation Management Team will support the selection process, collecting names of candidates and their profiles, and suggesting them to the main panels. During their registration process, the Research Units have the possibility to suggest and not-suggest names for panel members
- The subject panel chairs and members assign the referees (they need to have confidence in their skills/expertise)

In their selection of chairs and other members of the subject panels, the main panels will prioritise breadth over depth in relation to the candidates' field expertise. The main panels will consider also variety in terms of scientific disciplines, inter-disciplinary expertise, organisational background, nationality, gender and other relevant criteria.

Profiles

Each main panel will have 4 members

- The Chairs of the main panels will be international experts with a strong reputation in the field of discipline and (preferably) experience in industry or other user communities.
- The three other members will be Czech nationals. One member originates from the national research community and two members are 'outsiders' to the research communities, e.g. one member of the user communities (such as industry) and one member from a relevant funding agency or ministry
- At least one member has expertise in inter-disciplinary research

All members of the subject panels will be international experts, to minimise conflicts of interest.

Subject panel members should have a broad view of fields and major areas. Breadth should be prioritised over depth, especially for the panel chairmen. They should be respectable and have the capacity to consider societal relevance. Faculty deans of universities and researchers with strong collaboration with industry are good profiles In each subject panel, some members should have the expertise and experience that is required to assess the societal relevance of research.

Referees will have a fine-grained expertise. They will work in remote to keep costs down. Also the referees will be international experts only.

5.2.3 Roles and tasks of the panels

The role of the **main panel** is to moderate. It has an auditing function and provides a bridge between the Evaluation Management Team and the panels.

- Part of the main panels' tasks in auditing the work of the subject panels is to quality review the draft subject panel reports and sign off the assessment outcomes.
- Part of the main panels' task of being the liaison between the subject panels and the Management Team is to enquire, on a quarterly basis, about the progress of subject panels and to report back to the Management Team.
- Part of coordinating the work of the subject panels is to advise the subject panels on the interpretation and use of the evaluation protocol and panel guidelines; to advise on matters relating to insufficient or unclear information in the RUs' submissions; to discuss and take decisions in matters of conflicts of interest and other sensitivities; to discuss and take decisions in eventual cases of gaming or insufficient fairness during the process; and to handle cases of inter-disciplinary work for which attribution to a certain subject panel is contested, looking into the issue and take the decisions.

In the event that a panel chair is unable to perform a task, because of illness or other reason, the Management Team will appoint an acting chair from the other members of the main panel.

The main panel chair will participate in the meetings of the subject panels. The three national members in the main panels constitute an additional 'auditing' element to the panel review and can provide the main panel chair with context information, to be fed back into the panel meetings.

Specific tasks of the main panel chair are to closely monitor matters related to timing and progress; chair the physical and other meetings of the main panel; ensure that all members of the main panel contribute substantially in reaching a collective judgment; address conflicts of interest and confidentiality; and sign off the assessment outcomes by the subject panels.

The **subject panels** will have the primary function of conducting the performance assessment. The structure of the panel review (panels broad view, referees specialised) facilitates the management of the inter-disciplinary research issue. There will also be rulings for collaboration among panels and reporting to each other.

The subject panels will examine the submission of selected outputs for peer review and identify where and how many referees are required. They will assign outputs to the referees (first and second reader).

Subject panel chairs are to report to the main panels on the progress and on how the working methods are implemented. They will also be in charge of preparing an analytical report on the state of research in their field of discipline, based on the assessment outcomes for the RUs and supported by bibliometric data at the national field level. The analytical report contains an overview of the outcomes and the scores (on each of the five criteria) across the relevant RUs and draws conclusions and recommendations.

The **referees** will have the exclusive role of assessing the excellence of a limited set of submitted research outputs. Referees will work in remote to keep costs down. They will be informed on the outcomes of the subject panel calibration exercise. They will produce a written assessment about the research output submitted by the RUs and send it to the subject panel. The Evaluation Team will provide referees with the standard template for their assessment report.

5.2.4 Working methods

Main Panel working methods

- *Consensus and decision-making*. The main panels are to review assessments at RU level from their subject panels as they emerge throughout the assessment phase. They are to confirm that the published procedures and criteria have been implemented, and that overall standards of assessment have been applied consistently. The main panels will ensure that any inconsistencies in the assessment standards are investigated and explained, before accepting the evaluation results of the subject panels. Upon approval, the main panels will report the assessment to the Management Team
- *Meetings*. The different main panels will meet regularly throughout the preparatory and assessment phases with the purpose of ensuring close communication and collaboration, addressing any issues, and ensuring that procedures are followed
- *Cross-panel collaboration*. The main panel may identify boundary areas where the fields of multiple panels overlap, and may decide to assign individual panel members to more than one of these overlapping panels. As such, research within these boundary areas can be assessed with greater consistency
- *Ad-hoc panels for interdisciplinary research units*. The main panel will also decide on the applications for the registration of Interdisciplinary Research Units. If accepted, the main panel will install an interdisciplinary ah-hoc panel. The main panel chair of the major research area of the RU will act as the chair of this ad-hoc panel and suggests the members for this interdisciplinary panel to the Evaluation Team
- *Evaluation reports at the EvU level*. In those cases where an EvU has more than 1 RU, the main panel chair(s) will decide on the subject panel chair that will be in charge of drafting the EvU panel report. The responsible subject panel chair will

draft an overview of the evaluation results for each RU and draw conclusions and recommendations with a specific focus on the research environment criterion

- *Specific issues for the operation of panels*. Important issues for the operation of panels (main panels and subject panels) are:
  - *Conflicts of interest*. All main and subject panel chairs, members, referees, secretaries and special advisors are to observe the arrangements for managing potential conflicts of interest. As such, they are obliged to record declarations of interest and avoid potential conflicts of interest.
  - *Dealing with absences of the chair*. A deputy chair is elected for each main and subject panel, to chair the panel in case of planned and unplanned absences of the chair
  - *Confidentiality arrangements*. All main and subject panel chairs, members, referees, secretaries and special advisors are bound by the terms of confidentiality arrangements. These arrangements must ensure an effective management and operation of the evaluation process.
- *Calibration exercises*. Calibration exercises will be undertaken at an early stage in the assessment by all subject panels, in order to establish a common view on the implementation of assessment standards and quality levels. The main panel chair will attend these calibration exercises of the subject panels. In addition, the main panels are to receive and review reports on these calibration exercises from the subject panels.

Subject panel working methods

- *The subject panel's expertise*. The subject panels are to make collective judgements about the submissions. In the early assessment phase, they will consider the breadth of work in the submissions in order to ensure they have the appropriate expertise at their disposal to assess the submissions. Where needed, parts of submissions are cross-referred to another subject panel
- *Decision-making*. Decisions must be taken collectively by the subject panels within the framework of the exercise, following the published procedures and criteria. The subject panels will develop quality scores for each of the five criteria, and will debate the reasoning behind the quality profiles in order to reach consensus. If consensus cannot be reached, decisions will be taken by majority vote, in which the chair holds a casting vote.
- Specific working methods for the subject panels include:
  - *Allocating work*. The subject panel chair will allocate work to panel members, taking into account their expertise. Each member and referee must be allocated a significant volume of material to assess, to ensure that each makes a significant contribution to the panel's assessment work
  - *Calibration of assessment standards*. Subject panels will undertake early calibration exercises to ensure that subject panel members (and referees) develop a common understanding of the quality levels. The subject panels will continue to discuss the application of the quality levels and will keep under review the scoring patterns of panel members throughout the assessment process to ensure consistency in the subject panel's standards of assessment.
  - *Clarifications*. Where the subject panel feels that essential information is missing or unclear, they can ask the EvUs for clarification, upon which the panel secretariat will collect the additional information by email or phone. The additional information will be distributed to the subject panel members.

5.3 Risks and risk management

In this Chapter we set out the main measures developed in order to ensure the fairness and integrity of the evaluation process.

5.3.1 Integrity of the panel review process

Confidentiality

- The Management Team is responsible for preparing a confidentiality agreement and may prepare detailed guidelines for ensuring confidentiality of the evaluation process. In doing so, the Management Team will consult with the Evaluation Management Board and the chairs of the main panels. The confidentiality agreement should address, at least, the topics described below.
- The confidentiality agreement protects the rights (including Intellectual Property Rights) of the research organisations, EvUs, RUs and researchers, as well as the rights of panel members and other persons that contribute to the evaluation process (for example, to minimise the risk of claims for a breach of Intellectual Property Rights or disclosing confidential information).
- All persons involved in the evaluation process, including the Evaluation Management Board, the Management Team, the main panels, subject panels, referees and special advisers, are obliged to sign the confidentiality agreement.
- All meetings of the Evaluation Management Board, the Management Team, the main panels and subject panels as well as information and materials used in the evaluation process are confidential. The panel members are obliged not to disclose any information and materials related to the evaluation to third parties either in oral, written or electronic form, for the whole period of work as a panel member and after that.
- Panel members and all other persons involved in the evaluation process only use the confidential information for the purpose of the evaluation.
- On behalf of the main panels, the Management Team will assess the submissions by the EvU for any information that is highly confidential and that may not be shared with all members of the relevant main panels and/or subject panels. The Management Team will take the necessary actions (such as only disclosing specific information to selected panel members) after consultation with the chairs of the main panels.
- Panel members and all other persons involved in the evaluation process have to take appropriate measures to address the security risks related to making physical copies for personal use or using electronic communication (such as email) for communication with panel chairs and other persons involved in the evaluation process.
- Panel members are not allowed to contact research organisations or researchers in relation to information that is shared on a confidential basis, in the context of the evaluation process.
- In case of doubts about confidentiality issues, panel members and other persons involved in the evaluation process immediately contact the Management Team or a main panel chair.

Transparency

- The evaluation protocol, the text of the confidentiality agreement and any additional guidelines are publicly available.
- The names of the main panel chair and members are public. The names of the subject panel chairs and members will be made public after finalisation of the evaluation process. The names of referees are not public.

- Minutes must be taken of all meetings of the Evaluation Management Board, the main panels and subject panels. These are made available to the Evaluation Management Board, the Management Team and the main panels. Members of subject panels will have access to the minutes of the main panel to which they are accountable. The minutes of meetings will not be public.
- The panel report per RU, the overview report for EvUs with more than 1 RU (prepared by the subject panel chair) and the analytical report per field (prepared by the subject panel chair) and per disciplinary area (prepared by the main panel chair) will become public when the evaluation process is finalised.

Conflicts of Interest

- The Management Team prepares a statement about Conflict of Interest, in consultation with the Evaluation Management Board and the chairs of the main panels. The statement about Conflicts of Interest should address, at least, the topics described below.
- The statement about Conflicts of Interest intends to protect the rights of the research organisations, EvUs, RUs and researchers, as well as the rights of panel members and other persons that contribute to the evaluation process.
- All persons involved in the evaluation process, including the Governing Body, the Management Team, the secretariat, the main panels, subject panels, referees and special advisers, are obliged to fill in and sign the statement about Conflicts of Interest.
- All members of the main panels and subject panels must indicate any close personal or professional relations with the research organisations, EvUs and RUs in the field of discipline or the sub-field for which they will contribute to the evaluation process. Examples include planned, recently ended or honorary positions in a research organisation, more than three co-publications with researchers from one research organisation, or collaboration in applied research and commercialization.
- Statements about Conflicts of Interest will be analysed by the panel secretariat and discussed by the Management Team. The Management Team will suggest corrective measures to the chairs of the main panels and subject panels.
- A conflict of interest exists if an expert:
  - Stands to benefit directly or indirectly from the evaluation
  - Has a close family or personal relationship with any person employed in the evaluated organisation
  - Has been employed or contracted by an evaluated organisation
  - Has or has had during the last five years, a scientific collaboration with the evaluated organisation.
  - Has or has had in the past, a mentor/mentee relationship with staff of the evaluated organisation

Dealing with inter-disciplinary research

- To acknowledge the importance of inter-disciplinary research, there is a small number of main panels (around six) that cover a broadly defined field of discipline and that include panel members with expertise in inter-disciplinary research or in application areas (for which different fields of disciplines are often combined or integrated).
- In selecting the members of the subject panels, expertise in inter-disciplinary research is an explicit criterion.

- At least three out of five criteria for assessing RUs allow for appraising inter-disciplinary research: scientific research excellence (that includes originality of the ideas and the research approach), overall research performance (with research productivity that takes into account a variety of outputs) and societal relevance (that often requires the combination or integration of different disciplinary fields or sub-fields).
- In the submission document, RUs can indicate and explain which part of their research is inter-disciplinary. In addition to suggesting the disciplinary field or sub-field that is most relevant for the RU (the host field or sub-field), RU can mention two additional disciplinary fields or sub-fields and ask for cross-referral.
- The chairs of the relevant main panels and sub-panels will use this information to allocate the EvU to one of the main panels and subject panels, and to ensure that the subject panel members (supported by referees) have sufficient expertise to assess the EvU and that panel members of a second or third subject panel are involved in the assessment of the RU (cross-referral). Involvement of members from other panels will concern the assessment of the EvU on one or several criteria.
- When several RUs, allocated to the same subject panel, submit similar types of inter-disciplinary research, the subject panel may decide to identify and appoint a special adviser for multi-disciplinary research.
- In special cases where the volume and nature of inter-disciplinary research of an RU does not allow for allocating the RU to one (host) sub-field and subject panel, an ad-hoc inter-disciplinary subject panel is installed. This subject panel is chaired by one of the main panel chairs and includes members from two or more subject panels, from two or more disciplinary fields.

5.3.2 Audits and verifications

- The submission document that RUs use to submit information includes a statement about the accurateness of the information and the willingness to deliver any proof, upon request.
- The panel secretariat, in close collaboration with the Management Team, performs random audits of the submissions by RUs. As much as possible, random audits are spread across different research organisations
- The first type of audit concerns a request for proof regarding, for example, the number of researchers, PhD graduates, and strategic partnerships and the volume of grants and contract research. In the interest of proportionality, the first type of audits is done for around 5% of EvU submissions.
- A second type of audit is the confrontation of submitted information with information in databases about, for example, staff and revenues of research organisations, dissertations, grants and service contracts. In the interest of proportionality, the second type of audits is done for around 10% of EvU submissions.
- The panel secretariat, in close collaboration with the Management Team, performs targeted audits in response to concerns raised by panel members and referees.
- Specific verifications in using the R&D Information System and the bibliometric analyses will take place.

The assessment criteria are accompanied with a clause declaring: If fraud or dishonesty is detected, the panels will assign the lowest starred quality levels for the RU against **all** assessment criteria.

technopolis |group|

In collaboration with

NIFU

Nordic Institute for Studies in
Innovation, Research and Education

TECHNOLOGY
CENTRE ASCR