International Audit of Research, Development & Innovation in the Czech Republic

Final Report - 3

The Quality of Research, Institutional Funding and Research Evaluation in the Czech Republic and abroad

Technopolis Group

Erik Arnold, Barbara Good, Fritz Ohler, Brigitte Tiefenthaler, Niki Vermeulen

In collaboration with Thed van Leeuwen and Rodrigo Costas Comesana, CWTS (bibliometric analyis)

esf evropský sociální fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Table of Contents

Table of Figures

List of Abbreviations

AERES	French Evaluation Agency for Research and Higher Education
ASCR	Academy of Sciences of the Czech Republic
BERD	Business expenditure on research and development
BMWF	Federal Ministry for Science and Research (Austria)
CERN	European Organisation for Nuclear Research
CWTS	Centre for Science and Technology Studies, Leiden University, Netherlands
CZK	Czech crowns
DEL	Department for Employment and Learning, Northern Ireland
ECTS	European Credit Transfer System
EFQA	European Forum for Quality Assurance
EGS	Economic Growth Strategy
EM	Evaluation Methodology (the “coffee grinder”)
EPO	European Patent Office
EriC	Evaluating Research in Context (Netherlands)
ERIH	European Reference Index for the Humanities
ESA	European Space Agency
FFG	Austrian Research Promotion Agency
FINHEEC	Finnish Higher Education Evaluation Council
FWF	Austrian Science Fund
HEFCE	Higher Education Funding Council for England
HEFCW	Higher Education Funding Council for Wales
HEI	Higher Education Institution
HESA	Higher Education Statistical Agency (UK)
HES	Higher Education Sector
HR	Human resources
FTE	Full-time equivalents
GDP	Gross domestic product
IPR	Intellectual property rights
ISI	Institute for Scientific Information (now Thomson Reuters)
ISO	International Organisation for Standardisation
KNAW	Royal Netherlands Academy of Arts and Sciences
KTH	Royal Institute of Technology, Stockholm
MEYS	Ministry of Education, Youth and Sports
NDP	National Development Plan
NERR	National Excellence Reference Framework

NIP	National Innovation Policy
NPM	New Public Management
NR&DP	National Research and Development Policy
NSRF	National Strategic Reference Framework
NWO	Netherlands Organisation for Scientific Research
OECD	Organisation for Economic Cooperation and Development
OP R&DI	Operational Programme Research and Development for Innovation (European Structural Funds)
PA	Priority Axis (European Structural Funds)
PNAS	Proceedings of the National Academy of Sciences
PRFS	Performance-based research funding system
QA	Quality assurance
QR	Quality-related (funding) (UK)
RAE	Research Assessment Exercise (UK)
R&D	Research and development
RDI	Research, development and innovation
R&D&I	Research, development and innovation
REF	Research Excellence Framework (UK)
RIV	R&D Information System
RVVI	R&D&I Council
SCI	Science Citation Index
SEP	Standard Evaluation Protocol (Netherlands)
SHEFC	Scottish Higher Education Funding Council
SLU	Swedish University of Agriculture Sciences
SME	Small and medium-sized company
SSH	Social Sciences and Humanities
S&T	Science and Technology
UAS	University of Applied Sciences
USPTO	United States Patent and Trademark Office
SWOT	Strengths, Weaknesses, Opportunities, Threats (usually in SWOT analysis)
STEM	Science, technology, engineering and mathematics
VSNU	Association of Dutch Universities
WIPO	World Intellectual Property Office
WoS	Web of Science
WP	Work Package

Executive summary

This Work Package investigated the quality of research in the Czech Republic and analysed the Evaluation Methodology (Metodika Hodnocení) in place to evaluate research and allocate institutional funding for R&D.

Bibliometric analysis

We analysed the performance of the Czech research system compared to international developments using bibliometric data. We analysed the total output of 85,572 indexed publications from the CR in the period 1993-2000. The number of articles produced per year more than doubled in the period while the number of times the average paper was cited (excluding self-citations) went up eight times, indicating a dramatic improvement in quality. Productivity improved: Czech researchers increased their rate of production of indexed journal articles per year and are now about 50% more productive than they were in 1993. At the same time, the impact factors of the journals chosen rose steadily, indicating growing confidence and ambition as well as quality. At the start of the period, over 70% of the papers were never cited but by the end this proportion had fallen to 50%, indicating a great improvement in the international visibility of Czech research. The quality of the publications did not change much until around 1996, when an upward trend began that carries on to this day. However, on any impact measure average quality is still below the world average.

In the period 1993-2009, the largest field was physics & materials science, providing over 18% of the total Czech output, followed by chemistry & chemical engineering (over 16%). Thus the Czech Republic has a traditional continental European research profile, similar to the profiles of Germany, France and Italy. However, the country's publications profile is now slowly changing so that (bio)medicine and health sciences play a somewhat stronger role. Only a few disciplines reach average international quality levels: mathematics and statistics; environmental science and technology. Impact factors have increased particularly strongly in physics and materials science and in clinical medicine.

Over time, the CR's research community has both become more coherent – in the sense that authors from different Czech organisations are increasingly publishing joint papers – and better embedded in the international research community, indicated by a growing proportion of articles co-written with authors abroad. As is the case elsewhere, joint papers tend to be of better quality than those written by a single author, so collaboration – in particular international collaboration – seems to be one of the ways in which the Czech research community is learning how to perform better.

Universities and ASCR institutes account for the lion's share of Czech publications. Quality is improving among all groups of institutions but the ASCR institutes tend to perform the best. Scientific production has been growing in all the institutional sectors of the Czech Republic. But universities' production has grown more rapidly than that of the pure research organisations, outperforming their production from the year 2000 onwards. This may be a 'catching up' effect, with universities building up research capacities only recently after their role changed.

Czech publication performance is developing in a very positive manner – rather similar to developments in Slovenia and Hungary – and looks set over time to converge with international performance levels. Of course, if we compare with the leading small research-performing countries such as Sweden, Finland and Denmark there is clearly still a long way to go but Czech output and quality are growing faster

than is the case in these more established research-focused countries in Western and Northern Europe.[1]

Quality of research

Although the bibliometric analysis gives a good idea of how Czech publication outputs are situated in an international context, the quality of research cannot only be measured through publications. In order to get a more comprehensive view of quality and of where and how to improve it, we performed peer reviews of 16 research groups/institutes. We selected high-performing groups mainly based on bibliometric data, while at the same time paying attention to diversity in discipline, type of organisation, and region. Selecting high-performing research groups allowed us to identify systematic quality issues in the Czech Republic. Research groups came from the ASCR, the university system and the applied research institute sector and represented the natural sciences, technical sciences, medical sciences, social sciences and humanities, economics and the applied sphere.

Each research group wrote a short structured self-assessment report and obtained feedback from the international peers following a site visit. We chose all the international peers ourselves, thus making sure that the reviews were completely independent.

The criteria on which the peers made their judgement included the existence and number of high-level academic publications, the scientific standing of key staff in the more academic types of research; a clear understanding of users and access to users in more applied environments. Coherence and compactness of the research portfolio was another important criterion, as were management and human resource policy, research funding, and perceived barriers to development. Many of the institutional environments in which the groups worked were challenging, and the peers were generally impressed that they were able to produce work of a high standard under such circumstances.

The majority of the research groups evaluated were of good to excellent quality measuring up to international standards. Especially in the fields of natural sciences, technical sciences and economics, we encountered groups that are very well embedded in the international research community. They have achieved their international profile either through the training of research group leaders in the United States and/or Europe, or through the development of extensive collaboration with international universities or companies. International experience and contacts seem to have helped them understand and reach international standards and also inspired the organisation and management of the research groups.

Other research groups have come a long way. While the Czech Institute of Egyptology is a good example of a research institute from the social sciences and humanities that is internationally recognised, the other research groups we reviewed in this field are not at an international level yet. This is not surprising as several disciplines needed to be completely (re)established after the Velvet Revolution.

A final category of research groups still needs to improve considerably in order to reach international standards. In particular, the applied research institutes visited were mainly nationally oriented. Although a national focus is in line with their mission, their R&D does not live up to international standards. Also, rather than being systematic, their interaction with users is driven by opportunities.

A number of general but important research management issues emerged from the peer reviews.

[1] Extrapolating research output to calculate when the CR will reach these benchmark countries does not make sense because publication activity does not behave in a linear way. In other words, we cannot say how long it takes before the CR reaches the benchmark countries because publication behaviour is not predictable, being influenced by many factors

- *Fragmentation*. In the natural sciences, medicine, and technical sciences research groups tended to be small, especially when they were located within larger institutions. This fragmentation was caused by the high level of autonomy enjoyed by many professors and the shortage of the kind of incentives increasingly seen in other countries for working in larger centres.
- *Lock-ins*. Fragmentation also meant that it was difficult to alter the research trajectory by doing interdisciplinary research, orientating towards applications, doing long-term development or establishing strategic partnerships. These things require firmly established competences in core activities as well as resources available for exploring new topics. Small units typically do not have enough resources to do both, impeding innovation.
- *Reluctance to do interdisciplinary or application-orientated research*. The peers were surprised often to find that groups were reluctant to make contact or collaborate with others outside their discipline. As a result, they fail to benefit from the opportunities provided by interdisciplinarity to make breakthroughs (which often occur at disciplinary boundaries) or to benefit from the supply of new research challenges provided by tackling applications.
- *Little awareness or room for manoeuvre in human resource issues*. A number of units evaluated suffered from an unhealthy age and gender structure, problems in recruitment, and lack of mobility.
  - Several groups were dependent upon a single leader, who was ageing but was also critical to the group in terms of leadership, reputation and external networks. Also, women were clearly underrepresented among research staff, especially at higher hierarchal levels.
  - Many groups were in-bred, with careers characterised by a Master thesis, PhD thesis, research position and professor position within the same group. This limits the perspectives of the researchers and reduces their options for participation in international projects, which are often based on personal networks.
  - Small group size also means that there is not the flexibility required for people to be internationally mobile. Stays abroad are mostly restricted to a couple of weeks and at best months, limiting the opportunities for learning.
- *Funding policy as a major cause for short-termism*. Groups struggled with aspects of both the national and the international funding regime. There is a lack of continuity, owing to continuous rule changes in institutional funding for R&D, which is exacerbated by poor financial planning within research performing organisations. While the need to allocate money using evidence and transparent criteria was universally recognised, the Evaluation Methodology was widely seen as promoting short-termism, backward orientation and encouraging perverse behaviour. Project funding (targeted funding) tends to be small-scale and short term, further impeding the development and deployment of research strategy. Finally, participation in EU programmes was generally weak, partly because the problems of small size mean groups have weak international networks.

Evaluation Methodology

The Czech Republic has developed an intricate Evaluation Methodology to evaluate research and to allocate funding. The Evaluation Methodology, aptly called 'coffee grinder', counts various sorts of research outputs and assigns a certain number of points to them (this is the evaluation part). The points, each representing a certain amount of CZK, are then translated into money (this is the allocation part). Hence, the Evaluation Methodology 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. As such, the Evaluation Methodology is a performance-based research funding system (PBRF).

A growing number of countries are using PBRF that allocate some or (rarely) all of the institutional resources, based on performance so that institutional funding also becomes competitive. Historically, the trend has been for countries to move from 100% institutional funding to a mix of institutional and project funding, with project funding acting as a quality check for the institution to make it more competitive. Later also part of the institutional funding became performance-based.

The earlier PBRF tend to be heavily based upon peer review – the most well-known example is the UK Research Assessment Exercise (RAE). But many systems nowadays are indicator-based. While the use of indicators in the Czech Republic is driven by a desire to de-politicise and de-personalise the evaluation and funding process, elsewhere it reflects a desire to simplify and reduce the cost of the assessment.

Most indicator-based PBRF use a variety of indicators – not only scientific publications and research output like the Czech. If they use scientific publications and other research outputs at all, these typically make up only a small part of the funding formula (e.g. 5% in Finland). Also, PRFS typically move small amounts of money around each time they are performed (e.g. 2% in Norway) – in strong contrast to the situation in the Czech Republic, where all institutional funding for R&D is going to be allocated based on the Evaluation Methodology. What is more, most PRFS are for one type of research organisation only, mostly universities, while other types of research organisations have their own funding systems in line with *their* missions and objectives – again in contrast to the Czech PRFS, which is applied to all sorts of research organisations. Finally, not all systems automatically translate points into funding (e.g. Poland). Hence, it is safe to say the Czech Evaluation Methodology is *uniquely radical* in its sole focus on scientific publications and other research outputs and their automatic translation into money, applied identically to all research-performing organisations in the country.

It will come as no surprise then that the Evaluation Methodology – despite regular attempts to correct the most obvious imbalances – suffers from various problems.

- *Not addressing differences between disciplines sufficiently:* The present system ignores important and legitimate differences between disciplines. They differ significantly in their publication patterns both in terms of publication type and frequency, and not all types of results defined in the Evaluation Methodology are equally achievable for all disciplines (e.g. patents)[2]. The Evaluation Methodology has repeatedly tried to take into account differences between disciplines but has failed to address them adequately.
- *Not differentiating among different types of organisations:* The Evaluation Methodology treats all research organisations in the same way, regardless of their different missions and goals. While the intention is to take account of this by using a long list of differently weighted output categories, the weights are in practice arbitrary. Also, this number is the same for all institutions without regard to the importance this particular output has (or should have) for a given institution.
- *Exclusive focus on outputs:* As an evaluation tool, the Evaluation Methodology is not in line with international good practice because it only counts outputs. It ignores key factors affecting research quality, such as research management, research strategy, and human resource policy. Most importantly, however, it fails to deliver the kind of information needed for the advancement of Czech research organisations.
- *No room for improvement and building up of capacity:* Because it is purely retrospective, looking at past performance only, the Evaluation Methodology does

[2] Because the bias inherent in bibliometric data is the same for all countries, it is standard to use bibliometric indicators at the macro level to compare publication output internationally. However, it is highly problematic to use bibliometrics at lower levels of aggregation

not leave room for improvement and building up of capacity – two of the main purposes of both evaluation and institutional funding.

The Evaluation Methodology is affecting the research and innovation system in a way that threatens to harm it in the medium to long term:

- *Widespread opportunistic behaviour:* The Evaluation Methodology provides incentives for opportunistic behaviour and encourages cheating. As a result of widespread opportunistic behaviour – together with the arbitrary assignment of points to various (non quality-checked) research outputs –, the Evaluation Methodology is unable to distinguish between weak and strong research groups.
- *Unpredictability and instability:* While the role of institutional funding is to provide stability, the Evaluation Methodology can cause large – and unpredictable – changes in institutional funding, making institutional funding unreliable and planning for an institution difficult. Moreover, at a political level, the Evaluation Methodology fails to take into account other relevant (thematic) policy priorities or in effect even contradicts them.
- *Short-termism:* Because of the instability and unpredictability of institutional funding, institutes start to focus on short-term fixes to solve immediate problems. However, long-term planning is important to create high-quality research results, especially in a research environment where time horizons are inherently long.
- *Weak ownership:* The Evaluation Methodology has little support among researchers and research leaders as well as policy-makers. This can lead to frustration and low morale among researchers, which is not conducive to research quality and productivity. Ultimately, it can also lead to brain drain – researchers either leaving research or leaving the country.

Recommendations

Our conclusion is that the existing Evaluation Methodology is inappropriate for both the evaluation of research quality and the allocation of institutional funding. For this reason, we recommend *discontinuing* it.

To fill in the evaluation function of the Evaluation Methodology, we recommend introducing *informed* (indicator-based) *peer review* with independent selection of (international) peers. We strongly recommend *not* linking the evaluation of research to allocation of institutional funding for R&D directly. Not doing so will permit the Czech RDI system to build up an evaluation culture and allow policy-makers and research organisations to come to see evaluation as a tool for learning and improving research. Evaluation results can inform decision making but must not determine it.

For the allocation function of the Evaluation Methodology, we recommend introducing a new *funding formula* and *performance contracts*. The funding formula takes up the idea of the Evaluation Methodology by basing the allocation of institutional funding on past performance, while performance contracts, as contracts between a principal (e.g. a ministry) and an agent (e.g. the research organisation), represent the prospective element. They define objectives that the research organisation needs to attain and are embedded in proper procedural steps to prevent nepotism. However, performance contracts are not purely prospective: past performance is the basis for credible and achievable, yet ambitious objectives. As such, they differ radically from the research intentions.

For the research organisations that receive small amounts of institutional funding as a share of their budget (applied research institutes, universities), it may be an option to use either a formula budget or a performance contract.

Most importantly, bespoke systems for different types of research organisations are needed, *in line with their missions and roles in the Czech research and innovation system*. Research evaluation basically has to answer the question of whether or not an organisation is fulfilling its role and research funding needs to be such that research organisations can fulfil their roles in the Czech RDI system. The new funding and

evaluation systems we are proposing are able to accommodate different missions and roles. For example, formulas with different indicators can be set up for different types of research organisations.

In order to build up new evaluation and funding systems, institutional capacity will need to be built up. Ministries and research organisations alike will have to learn how to deal with evaluation and how to negotiate a performance contract. However, there already is a good basis for learning:

- The Academy of Sciences has established its own internal evaluation system and this can be used as a starting point for learning about evaluation, although some elements of the ASCR evaluation, in particular the peer selection process, need to be brought in line with international good practice. In addition, the reviews of institutes we did as part of this audit can be used as a basis for developing an evaluation culture. These reviews, being completely independent and international, were typically welcomed as useful exercises and an opportunity for learning. Incidentally, a comparison showed that the results of the ASCR evaluations are largely in line with the results of our own reviews.
- The Centres of Excellence and the Regional Research Centres within the Operational Programme 'R&D for Innovation' base funding on performance contracts, the related negotiations have recently been concluded. In other words, there already is some knowledge about performance contracts in the Czech RDI system and there will be more as these Centres move on. We will recommend actively sharing these – and international – experiences.

1. Introduction

This report presents the findings of two work packages: 'Assessment of the 2004 Guidelines for Evaluating R&D Results' (WP d, i) and 'Evaluation of R&D carried out in all types of institutes eligible for public research support' (WP d, ii).

According to the Terms of Reference for this Audit, assessing the quality of research (as foreseen in WP d, ii) and the 2004 Guidelines for Evaluating R&D results (as foreseen in WP d, i) as well as making recommendations for their improvement have been key tasks of this Audit. We are presenting the results of the two Work Packages in one report as the improvement of the quality of research in the Czech Republic is of course closely interlinked with the evaluation of research. More specifically, the Guidelines for Evaluating R&D Results (henceforth called Evaluation Methodology) – the focus of attention of WP d, ii – is seen as a tool for assessing the quality of research and increasing it.

In our study we combined several methodological approaches to answer the study questions and come up with conclusions and recommendations:

- A bibliometric analysis of the Czech Republic research output in an international context performed by the Centre for Science and Technology Studies (CWTS) of Leiden University, the Netherlands.
- An international peer-review (panel-based assessment) of 16 research institutes from different fields and of different organisation type. The institutes were selected on the basis of bibliometric analyses and the expertise of the Czech mirror group for the study.
- Analysis of both the relevant Czech policy documents, in particular the documents describing the 2004, 2009 and 2010 Evaluation Methodologies, and the international scientific literature about institutional research funding. The analysis benefited from the experience of the experts involved in the project team. Our analysis of the Czech documents was based on English translations.
- International case studies describing and analysing evaluation and funding systems and practices in other countries, comparing them with those in the Czech Republic. Based on the candidates identified in the proposal, we selected four countries: Austria, Finland, the Netherlands, and the UK. We also took experience from other countries into account in our literature analysis (e.g. Australia, Norway, Sweden).
- Czech case studies describing and analysing evaluation and funding systems in the Czech Republic *other* than the Evaluation Methodology.
  - Comparison with the evaluation guidelines developed and applied in the allocation of funds in the framework of the Operational Programme Research and Development for Innovation (OP R&DI) in the priority axes 1 and 2, i.e. for Centres of Excellence and Regional R&D Centres.
  - As an institutional case study we have chosen the internal research evaluation system at the Czech Academy of Sciences.
- Empirical testing and assessment of the use and usefulness of the existing Czech Evaluation Methodology through questionnaires and interviews. We also tested our fundamental ideas for improvements and our recommendations in interviews with Czech experts and during a workshop.

Based on all these diverse inputs we will first give an overview of the quality of research in the Czech Republic (section 2), looking at the outputs of research through bibliometric analysis, but also at important factors that influence the quality of research: the organisation of research, research strategy, research budgets, human resource management, international collaboration, etc. This analysis of the quality of

research is followed by a thorough analysis of the Evaluation Methodology: After addressing the development of the Evaluation Methodology in the Czech Republic (section 3) and presenting a first analysis of it (section 4), we describe Czech researchers' and research managers' view of the Evaluation Methodology and analyse its effects on the Czech research community (section 5). Based on four international case studies, we assess the Evaluation Methodology against international practice (section 6). We also describe alternative approaches to the Evaluation Methodology found in the Czech Republic (section 7). All this information results in a SWOT analysis of the Evaluation Methodology (section 8) and recommendations for its reform (section 9), or rather for the two main issues it addresses: evaluating and developing the quality of research and allocating institutional funding on the basis of performance.

An important evolution of the Methodology for the Evaluation of R&D results is its use as a tool for decision-making on the allocation of institutional funding (at the funding providers' level), which was announced in the 2008 Reform Plan and implemented in the National R&D&I Policy 2009–2015. In other words, the Evaluation Methodology was given the additional role of guiding a Performance-based Research Funding System (PRFS). Therefore, we did not only analyse the Evaluation Methodology of 2004, as requested, but also the more recent versions that take into account this new role, namely those of 2009 and 2010.

In other words, the Czech research evaluation system tries to combine two activities: 1) evaluation of research and 2) decision-making about allocating institutional funding. Inevitably our assessment needed to take into account the diverse aspects of the application of the methodology in the *two* systems, i.e. the evaluation system and the PRFS. Moreover, the dual function of the methodology results in recommendation on two fronts: at the end of this report we suggest the implementation of a new way to allocate funding as well as a new way to evaluate the quality of research, both in line with international practice and specifically adapted to the circumstances and needs in the Czech research system.

2. Quality of Research in the Czech Republic

2.1 Introduction

This section starts with the presentation of the main results of the bibliometric analysis of the Czech Republic's research output in international context (section 2.2). The objective of this bibliometric analysis is to provide a good overall comparative picture of the quality of research in the Czech Republic, identifying strong and weak areas of research, covering the time period 1993-2009. The analysis comes in two parts; an overview on a national level is followed by an institutional analysis.

Although the bibliometric analysis gives a good idea of how Czech publication outputs are situated in an international context, the quality of research cannot only be measured through publications. When evaluating the quality of research, also other factors have to be taken into account, like the organisation of research, research strategy, research budgets, human resource management, collaboration, etc. These factors that constitute the environment in which research is performed are an important part of the quality of research as they fundamentally influence the performance of research, and as such are crucial for a good research output. Moreover, these conditions are important to take into account because they can be changed and improved in order to develop the quality of the research and its results. In short, they are decisive factors in – the improvement of – research quality.

After the conclusions of the bibliometric study, this section therefore gives an overview of the international peer review (panel assessments) of research institutes that we performed in various types research organisations in the Czech Republic, giving an overview of the quality of research in the five main different fields of research as well as the applied research centres (section 2.3). In addition, survey and interview results are highlighted, which cover feedback on important factors of research quality from research directors and researchers in the Czech Republic (section 2.4). Finally, based on the bibliometric analysis, the international peer reviews, the survey as well as the interviews we conducted, the last section draws conclusions on the quality of research in the Czech Republic and gives recommendation about how to improve the quality of research (section 9).

2.2 Bibliometric analysis of the Czech Republic research output in an international context (1993-2009)

2.2.1 Introduction

The bibliometric analysis of the Czech Republic research output in an international context is performed by Thed van Leeuwen and Rodrigo Costas Comesana of the Centre for Science and Technology Studies (CWTS) of Leiden University in the Netherlands. The analysis of the national and institutional levels are summarised below while the complete reports, including important clarifications in terms of instruments and methodologies used for this analysis as well as data tables, can be found in the Final Report, 8 – Bibliometric analysis.

2.2.2 Conclusions of the bibliometric analysis of the Czech Republic research output in an international context – overall performance

The first study of the research output on a national level clearly shows that the Czech Republic's research system has gone through drastic changes. Accession to the European Union has created large opportunities for the Czech Republic.

Important findings of the analysis include:

- The increasing international visibility of the output of researchers from the Czech Republic and a remarkable improvement in the quality of the scientific publication output of the Czech Republic over time – especially in the last decade.
- The increasing trend in international and especially national cooperation – the latter indicating an improvement of the internal cohesion in the Czech research system.

Although a comparison with a number of benchmark countries made clear that the Czech Republic still has a long way to go when compared with some small European countries with a strong scientific performance (e.g. Denmark, Finland, the Netherlands, and Sweden), it is bridging the gap, especially in terms of output development.

The research output and impacts

During the time period 1993-2009, the output of the Czech Republic in the journals processed for the Web of Science[3] has nearly tripled. When comparing data on citations received by researchers from the Czech Republic in 2005-2009 with those received in 1993-1997, one sees that in the most recent time period, the number of citations received (which are indicators of impacts) is more than six times higher.[4]

The impact of the publications from the Czech Republic has increased strongly, and although still somewhat below worldwide average impact level, the impact was at such an initially low level that an improvement to a worldwide average level was not to be expected in such a short period.

In general, we therefore observe an increasing international visibility of the output of researchers from the Czech Republic. The sustained increase in the impact, developing at a much faster pace than the number of publications, indicates a remarkable improvement in the quality of the scientific publication output of the Czech Republic over time – especially in the last decade.

Research Profile of the Czech Republic

The composition of the research conducted in the Czech Republic shows a slow change in scope. As a country with a traditionally strong focus on the natural sciences and mathematics, the country's research profile is now slowly changing into a profile in which (bio)medicine and health sciences play a somewhat stronger role.

In the period 1993-2009, the largest field is physics & materials science, covering over 18% of the total Czech output, followed by chemistry & chemical engineering (with over 16% of the national output of the Czech Republic). The 35% of the output in these two fields of the natural sciences gives the Czech Republic a rather traditional continental European research profile, similar to the profiles of particularly Germany, France and Italy (contrary to a more Anglo-Saxon profile, where the life and medical sciences play a more dominant role).

---

[3] The CWTS/WoS database is an upgraded and dedicated 'bibliometric' version of the widely available online/offline 'bibliographic' versions of the database provided by Thomson Reuters Scientific to its customers. The CWTS bibliometric information system integrates the CWTS/WoS database and a series of software routines and research performance indicators based on publication output and citation impact statistics

[4] Within the CWTS citation analysis, field normalization is applied. This means that every paper, and particularly its impact, is compared within its own environment first before it is compared with others. As citation practices differ among fields, it is necessary to create benchmark values for citation data, in order to do right to the specific character of a country's or unit's output profile. Within this field-normalized impact measurement, we take into consideration the type of document (as various types of documents have different citation characteristics), and the age of the publications (older publications have had more time to collect citation impact).

Clinical medicine provides only 10% of the national publication output, underlining the observed preference for the natural sciences. Other disciplines covering more than 5% of the national output are basic life sciences, biomedical sciences, and biological sciences. In all of these disciplines, however, we observe low impact scores.

We observe only a few disciplines with average impact levels, namely mathematics & statistical sciences, environmental sciences & technology, and a few with a much smaller output, i.e. literature, and management & planning.

Looking into the research profile of only the most recent time period (2005-2009), we see a few remarkable differences:

- The somewhat more modest volume of publications in the hard natural sciences related to physics and chemistry (physics & materials science, and chemistry & chemical engineering), compared to the life and medical sciences related disciplines.
- An improvement of the impact in the most important fields in the national research profile. This was most marked for two of the six largest disciplines in which Czech scientists are active, i.e. physics & materials science, and clinical medicine, but the other top ranking disciplines in the profile also show an increasing impact for Czech science. Other fields in which the impact increased were mathematics & statistical sciences, environmental sciences & technology, instruments & instrumentation, and notably, social and behavioural sciences, all of which have reached internationally average impact levels. In multidisciplinary sciences, a field that covers multidisciplinary top journals such as Nature, Science, and the Proceedings of the National Academy of Sciences of the USA, we note a high impact score for Czech publications.

Scientific Collaboration

Czech publications show an increasing trend in scientific collaboration. Over the whole period, some 45% of all publications are the result of international collaborations. The part of the Czech Republic's output involving international collaboration has remained relatively stable in the last decade, while elsewhere it has tended to grow.

The largest change in the Czech research system was the huge shift from publications with no collaboration at all, towards publications resulting from national cooperation, thereby showing that internal cohesion has improved in the Czech research system. In that light it is important to stress that these types of scientific activity also show increased impacts. This is an important observation, as in other countries increased impact is normally generated by international cooperation. Of course – and this is partially the case for the Czech Republic – the publications resulting from national cooperation follow a similar pattern, thereby contributing to the strength of the system.

Benchmarking analysis of the Czech output in an international context

We compared the research performance of the Czech Republic with a number of benchmark countries including neighbouring countries (Austria, Germany, Hungary) and countries of similar size (such as Slovenia), as well as some smaller European countries with a strong research performance (such as Denmark, Finland, the Netherlands, and Sweden). We analysed the research performance of this set of countries over a longer period to be able to identify significant trends in the development of the Czech Republic in this international context.

Overall benchmarking

The comparison with the benchmarking countries makes it clear that the Czech Republic still has a long way to go. In the period 1993-2009, the Czech Republic's output was among the smallest and of a similar volume to those in Hungary and Slovenia. The number of citations received was small. Only Slovenia received fewer citations than the Czech Republic.

With respect to normalised impact scores, we find the Czech Republic again next to Hungary and Slovenia, two countries with a similar impact level, although the latter two countries tended to publish in journals with somewhat higher impact factors in the field to which the journals belong. Countries with high impact scores were Finland, Sweden, Denmark, and the Netherlands (with impact levels of 19%, 20%, 27%, and 30% above worldwide average field impact level, respectively).

Trends in outputs & impacts

The development of the Czech Republic is comparable to that of Hungary and Slovenia. One also observes a distance, particularly with respect to scientific impact, with the smaller countries in the study: countries as Denmark and the Netherlands, well-known scientific European powerhouses, outperform the Czech Republic.

It is important to stress that the growth in *output development* was more marked for the Czech Republic than for the other benchmark countries, thereby bridging the gap. Also in terms of impact we observe an upwards trend for the Czech Republic, albeit not as pronounced as for the volume of production.

- When we look at the *productivity of researchers* in the Czech Republic, we clearly observe a positive trend. When we compare the full period 1993-2009 with the most recent period 2005 - 2009, the researchers in the Czech Republic are now catching up strongly, together with their colleagues in Hungary and Slovenia. The improvement in output per inhabitant increased by roughly 50% for the researchers in these three countries.
- When we compare the impact development over the two periods, we find an even more important development, namely an increase of the Czech impact of nearly 25%, making this the strongest increase observed among all benchmarking countries.

Research fields

The analysis of the research fields across the set of benchmark countries clearly shows the growing strength of the Czech Republic. The country is more active in a number of fields traditionally belonging to its research profile, such as chemistry and chemical engineering, physics and materials science, biological sciences, and to a lesser extent, mathematics and statistics.

A field in which the Czech Republic is underperforming in terms of output is clinical medicine. This is partially due to the fact that the Scandinavian countries and the Netherlands have been moving towards an Anglo-Saxon research profile for a longer period of time, i.e. a profile in which biomedicine takes a more prominent position compared to a more classical continental European research profile.

There have been improvements in impacts in many fields. The most striking is the strong increase in the impact of social scientists from the Czech republic, a clear sign that these researchers are catching up with international performance levels. This also indicates that the positive changes in research performance of the Czech Republic's resyearch system are not limited to the natural, life and medical sciences, but are occurring across the whole system.

2.2.3 Conclusions of the bibliometric analysis of the Czech Republic research output in an international context –institutional analysis

In this second part we present the findings related to research performance in the international context at the *institutional* level (meso-level) during the years 1993-2009.[5]

In this meso-level analysis, the focus was on the main scientific institutions and organisations in the Czech Republic with production in the Web of Science.

Overall results from the bibliometric analysis of Czech research organisations

Universities and research organisations are the main players in the research activities of the country, participating in more than 50% of the production of the country[6]. Hospitals also play an important role, but we have to keep in mind that part of their production is also included in the sector of universities. Finally, it is also interesting to note the role of private companies that participate in around the 4% of the production of the country.

There are two major actors on the international scene in the Czech Republic, the Academy of Sciences as a whole, participating in more than 45% of all the scientific publications of the country, followed by Charles University in Prague. These, of course, are also the largest research organisations in the country. The picture is strongly dominated by the institutes of the Academy of Sciences (the Institute of Physics, the Biology Centre, the Institutes of Macromolecular Chemistry, Microbiology, etc.). Besides them, other major institutes and universities that publish a great deal are the Masaryk University, the Institute of Chemical Technology in Prague, the Czech Technical University in Prague, and the Palacky University in Olomouc (all of them with more than 3000 publications during the period 1993-2009) among others.

None of the main research organisations in the Czech Republic presents a field-normalized impact higher than the international level (indicated by the value 1). Most of the institutions with the highest field normalized impact values are institutes of the Academy of Sciences (the J. Heyrovsky Institute of Physical Chemistry – very close to the international value of 1 –, the Institute of Physics or the Institute of Macromolecular Chemistry). The University of South Bohemia also presents a relatively high normalised-impact as compared to the other institutions.

Evolution over time of scientific publications

Scientific production has been growing in all the institutional sectors of the Czech Republic. Universities' production has grown more rapidly than that of the pure research organisations, outperforming their production from the year 2000 onwards.

Publications impacts have also risen over time for all the main institutional sectors in the Czech Republic. There has been remarkable growth in average publication impact, although this *is still below the international level.* Only the Institute of Physics of the Academy of Sciences has occasionally exceeded the international level.

Co-publications

To study collaboration patterns, we analysed three main types of cooperation: international cooperation, national cooperation (where only Czech institutions are

[5] The full analysis as well as all bibliometric data are reproduced in the Final Report, 8 – Bibliometric Analysis

[6] The total production of the Czech Republic recorded for the period 1993-2009 amounts a total of 85575 publications (see report at country level). This is the value we use for the calculation of the percentages.

involved) and single institute cooperation (when the publication is authored by one or more researchers from just one single institution).

In all sectors, the field-normalized impact of all sectors is higher for publications in international collaboration, while the lowest scores are found for the production without collaboration.

Co-publication among Czech institutions and between Czech institutions and others abroad has increased; the number of publications from single institutions has stagnated. Other important findings were:

- Publications involving international collaboration tended to have a normalised impact factor of about '1'. *Hospitals* is the only sector with a normalized impact well above that, reaching the international level.
- While the output of most institutions focused on international collaborations, certain of the regional universities were more focused on national collaborations.
- The growth of international collaboration as well as the level of production in international collaboration is higher for universities and research organizations (mainly the Academy of Sciences), while national collaboration has a major importance for hospitals, governmental institutions and especially companies.
- Some universities and institutes have for a long time been well linked into international R&D networks, so their share of international collaboration does not change much over time (this is the case of the Academy of Sciences as a whole, the Charles University in Prague or the Institute of Physics of the Academy of Sciences). Other institutions, instead, show a remarkable growth in the share of international collaborations during the period of analysis, such as for example the Czech Technical University in Prague. This university saw a large increase in the impact of its papers in international collaboration during the last years of the period.

2.3 Results of the international peer-review of research institutes

2.3.1 Introduction on the methodological approach

In order to get a good impression of the quality and the organisation of research in the Czech Republic we have performed an international informed peer-review of good research groups in the Czech Republic. We made a selection of high-quality groups based on bibliometric data, while at the same time paying attention to diversity in disciplines, types of institutes and region. Selected groups wrote a self-assessment report, which was read by a panel of international experts who then also visited the groups on their premises. Subsequently, the panels wrote a review report with conclusions on the research and the organisation of the specific groups, as well as general conclusions on the opportunities and challenges of institutes and researchers in the Czech research system.

Selection of star research groups

For the evaluation of R&D we selected 18 good research groups that were subject to a thorough qualitative evaluation. The selection of these groups was performed taking specific requirements into account, and we therefore developed different methods to select these research groups. This section describes the requirements, selection process and the results of our selection.

Requirements of the selected groups

For the selection of groups it is important to note that the 18 star research groups not only needed to be of good quality but they also needed to be diversified, most importantly across disciplines. So we decided that we required a selection of three groups from each of the following fields:

- Technical sciences

- Natural sciences
- Medical sciences
- Humanities & social sciences
- Economics
- Applied research centres

However, in order to get a good overview of the quality of research in the Czech Republic we also wanted to have a distribution over different types of institutes (more specifically the Academy of Sciences, universities and applied research institutes) as well as a regional distribution.

Methods of selection

As the requirements for the group selection were quite complex, different methods were developed to accomplish the selection of appropriate groups for our evaluation of R&D in public research institutions in the Czech Republic. Consequently, we have used the following sources for selection:

- An analysis of the Centre of Science and Technology Studies (CWTS) of the University of Leiden, The Netherlands. For the selection of star research groups based on international recognized bibliometric databases, CWTS built a special methodology that delivered a short list of the best research groups. The method focused on highly cited publications, and the writers of these publications. This means that for all Czech publications, it was determined which publications belong to the top 20% of the most highly cited publications of the worldwide (journal) literature. From this set of data, CWTS selected those names that appear frequently in the total set. Among the publications of the researchers with the largest volume of papers, they distinguished between output produced alone, or in combination with other researchers appearing on the list. Then they added field labels distinguishing between technical sciences, natural sciences, medical sciences, humanities & social sciences and economics. For the selection of applied research institutes the data for the technical sciences were used as applied research is not a category in citation databases.
- An analysis of the Technology Centre. The Technology Centre composed a list of good groups sorted according to their citation index, based on the citation analysis of Czech publications (2003 – 2007) in the database Web of Science. The list contained 47 names of which the 18 best were selected, taking distribution into account. The selection was made at the level of scientific field in order to cover all the scientific fields (except for social sciences and humanities). For the social sciences and humanities, the well-recognized individuals were selected. For the selection of applied research groups they additionally performed an analyses of patents applied at the EPO and US patent office.
- The 2009 results of the Evaluation Methodology also deliver a selection of research groups for the mentioned disciplines on the basis of the points collected. We have first selected the top 152 institutions, which we then analysed according to discipline to select the groups.
- Recommendations of Mirror Group members of this study on the selection of star research groups. These recommendations varied and included the mention of specific top researchers or research groups, as well as the recommendations of existing national rankings or prizes. In addition, the Mirror Group recommended knowledgeable individuals whom we asked about top research groups within an organization they know very well.

Primarily we used the selection based on bibliometrics made by the Centre of Science and Technology Studies (CWTS) of Leiden University in The Netherlands, which was then complemented by a bibliometric selection made by the Technology Centre, the results of the Evaluation Methodology, and expertise of the Mirror Group members.

As all these different selection methods have their own bias, we found that only the combination of these methods delivered us a selection that fulfilled all our criteria. By way of example, relying only on bibliometric analysis would have provided us with a sample with a strong bias towards the Academy of Sciences, basic (natural) sciences, and the Prague region. Only through the combination of different methodologies could we come to a balanced selection of groups that was approved by the members of the Mirror Group.

Results of the selection

Through the different selection methods we identified five clusters of three research groups in the fields of natural science, medical science, technical science, economics, socials sciences and humanities. The different groups were embedded in different organisational contexts, varying from the Academy of Sciences to universities. In addition, we selected three institutes for applied research. While the majority of these institutes was located in Prague, we also included groups in other parts of the country in order to achieve regional distribution. We invited the groups to participate in the evaluation, and although many responded positively it also turned out that some were reluctant or did not participate. In case of non-participation we were often able to find a substitute research group, except for two cases: a group in the realm of economics and an applied research centre. As a result we have selected 18 research groups, but performed only 16 peer reviews.

Self-assessment of the selected research groups

A self-assessment report is often used in the assessment of research groups (see for instance the Dutch Standard Evaluation Protocol). This report has a dual function. On the one hand, it makes the research group or institutes under review reflect on their own performance, organisation, and research plans; on the other hand, it is also a good way to inform international peers in preparation of the site-visit.

For the self-assessment that the selected groups performed we composed a template. This template included questions concerning:

- The description of the research organisation
- The research programme
- Teaming and strategic partners
- Key staff
- Achievements and users of your research results
- Management and human resource policy
- Research funding
- perceived barriers in the context of the organisation
- allocating institutional funding

The self-assessment template was based on international evaluation practices in which self-assessment often plays an important role. The selected groups filled in the reports and these were used as a basis for the international evaluation and site visit.

2.3.2 International reviewers

The review process required an international panel of three experts per selected field of study. This meant that for instance in the field of natural sciences there were three research groups selected with a different research focus: mathematics, chemistry and physics. Hence, for every institute we selected one expert (e.g. a mathematician for the mathematics institute). But since all experts in a panel were natural scientists, they were of course also able to assess the quality of research and the research environment of the other institute. In other words, every group was visited by three experts in the field, of which one an expert in the focus of study of the specific group.

The international review process was led by a so-called guarantor who has expertise in the scientific field that was reviewed. These guarantors were also in charge of the selection of the international reviewers. Thereby they combined their knowledge of the field and desk research to select a group of reviewers with different expertise and various national backgrounds that were invited to participate in the study.

The international reviewers received a reviewers' guideline that explained the reason and the procedure of the evaluations as well as their own role. They read the self-assessment reports that were produced by the research groups and together with the guarantor they went to the Czech Republic to perform the site-visits. Finally, they wrote the final evaluation report together with the guarantor.

Site-visits

After reading the self-assessment reports and preparing discussion points and questions for the representatives of the research groups and institutes, the international reviewers and the guarantor visited the groups in the Czech Republic. After arriving at the institute, the leader of the research group gave a presentation on the research and the research group. Afterwards the self-assessment report was discussed with the representatives of the research group. In addition, there was a tour of the institute or the laboratory to see the research facilities and assess the working conditions of the researchers. The international experts also spoke to junior and senior researchers of the group in order to get a more complete impression of the composition of the research group and the organisation of research. Finally, the international experts discussed their first impressions with the representatives of the research group or institute. After the site-visits they wrote a review report with conclusions for the specific research groups as well as general recommendations for the improvement of the organisation of research in the Czech Republic. The conclusions for the groups have been sent to the group leaders to give them an opportunity to give feedback before the review reports were finalised.

2.3.3 Results of the evaluations of the Czech research groups

In this report we will not include the specific conclusions for the selected groups[7], but we will present the general conclusions of the six international reviews.

1. Overall, the participants in the peer reviews (the team and department leaders, the staff members, and sometimes the head of the super-ordinate organisational unit) indicated that the participation in the international review and the conclusions of the reviewers were very useful for them. Despite the fact that it was sometimes quite burdensome to convince and win over the pre-selected research teams to participate in the review exercise, in the end each single participant demonstrated a cooperative and open-minded attitude. Many of them came to see the exercise as a serious and challenging experiment and a learning opportunity. They particularly appreciated the feedback that came from the foreign reviewers.

[7] In the Czech circumstances, where the evaluation of research is highly politicised, we did not think the publication of review reports of specific institutes would be very wise, all the more so as this would probably have deterred them from participating even more than was already the case. There is no general rule or practice as to whether review reports should be published or not. In some countries they are, in others they are not. The Dutch have found a solution in-between, which could serve as a role model for the Czech Republic. In the Dutch case, a part of the report is made public and a part with more sensitive information stays inside the walls of the institute: "The objective of accountability can only be met by producing a transparent and informative public report of the evaluation's outcomes. On the other hand, to meet the objective of improvement and advice to the research management and the board of the institute, the evaluation committee should feel free to discuss the future of the research and of the institute. For this second objective, the evaluation committee can organise discussions with the institute's scientific leaders during their site visit and draw up a management letter to the board. Matters of personnel policy and sensitive decisions are generally treated in the confidential management letter to the board and do not form part of the public report." (SEP 2003-2009, p. 16)

2. In many cases the reviewed units appreciated the review exercise as it 'forced' the leaders to think about their performance and organization. The process itself helped the leaders to sharpen their focus and in the longer run, improve their performance.
3. At the same time, for many of the reviewers the exercise was enhanced by an extra curiosity-driven motive to observe a system in transition (with a strong research tradition) 'in situ'.

General conclusions of the technical sciences review

The first conclusion of the peer review committee is that running a research group or institute in the Czech Republic has large similarities with running a research group or institute in Western Europe (e.g. strive for scientific excellence, attracting and stimulating young researchers, attracting funding, finding partners, international contacts, importance of publishing in peer-reviewed journals) but also remarkable differences (e.g. extremely large possibilities of discontinuity in funding, very large job and remuneration insecurities for personnel, large divide between universities and research institutes, limited national industrial demand) that do not make running a research group easier.

The three research groups visited show that the research leaders, despite the difficulties, have been able to deliver good to excellent quality research at an international level. Leadership of the Principal Investigator seems to be a critical success factor here, in combination with management skills (including human resources management). The peer review committee is of the opinion that the (research system in the) Czech Republic could benefit from addressing these difficulties at a higher level than the research group (university level or system level). The system can learn from examples of good practices in research funding and research evaluation in other countries.

The first issue to address is the *position of research in relation to teaching*. Although the review committee does not have a profound knowledge of the Czech university system, the impression has taken root with the committee that in Czech universities teaching and research are not considered of equal importance, and therefore improvement of research is not very strategically targeted.

Research and teaching seem to be rather separated in the education system. Good researchers should however teach, and (almost all) teachers should be involved in the research in order to better train students and prepare them for a research career. To reduce the disparity of having groups almost solely involved in teaching and others employing a major part of their efforts on research, a more distinct separation between 'research universities' (that educate researchers) and 'teaching universities' (that educate for non-research functions) might also be considered. A stronger link between research and teaching also gives good research groups better access to students to train and select the next generation of researchers. In this context, it is also worth developing more relationships between Academy researchers and universities, for example by means of joint positions at the Academy and in a university similar to the situation in Germany where many researchers at Max Planck Institutes hold a formal (zero-salary) professor position at a nearby university.

The second issue to be addressed is *continuity:* continuity in financing of (good) research and continuity in good research personnel. We have been impressed by the research talent in these groups. The country would benefit from a more balanced financing for these good groups in a mix of long term and project funding.

In order to increase continuity in personnel, permanent positions for senior research staff should be created, with good, stable salaries but regular (once every 4-5 years) stringent evaluation.

Continuity of personnel should also be promoted by increasing the percentage of PhD students that actually finishes his/her PhD study. Better supervision, more room to develop their own research lines, better facilities and higher and more stable salaries

including stimuli to actually finish the PhD should be realized. The present situation where many PhD students are formally employed by the university for a minimal remuneration, but are mostly (or solely) involved with a second main job in industry is undesirable and should be replaced by a system with fewer PhD students with better and stable salaries.

A third issue that, according to the review committee, should be addressed is *participation in EU programmes*. Czech participation in EU projects is limited. EU research projects however offer substantial funds and cooperation with international research groups at the forefront of knowledge. These projects open up new perspectives for research and researchers. Without participation in EU projects keeping up with the international pace is almost impossible.

Application for EU grants is a difficult and often frustrating process due to the complicated bureaucratic rules and procedures. Better support for researchers willing and able to participate in EU projects should be organized (e.g. application support office(s); training courses; support in financial administration; showcasing successful Czech examples,…).

A fourth major issue is the *lack of mobility*. Even in groups with a strong international reputation, many careers are characterized by a Master thesis, PhD thesis, research position and (in one case) professor position within the same group. These internal careers do not only limit the (research) perspective of researchers (and therefore their options to follow new research paths) but also limit the number of solid external contacts of a group and therefore their options for participation in international projects (as these are most often based on personal contacts). The present recruitment policy aimed at hiring the best PhD students from their own group for staff positions should be discouraged and open recruitment should be encouraged.

(International) mobility should be promoted by bringing students in contact with foreign researchers earlier, by awarding foreign exchange scholarships, promoting participation of Czech researchers in EU mobility programmes and by good examples from senior researchers. If this is insufficient, it may be worth considering punishing immobility, e.g. by taking the international orientation of researchers into account when evaluating research proposals and research positions.

The number of foreign researchers in the Czech Republic should also be promoted. At present the system is not prepared to accept foreign people. It costs a lot of time to arrange everything. Active international advertising of positions (in English), more attractive salaries, fewer bureaucratic procedures, lower tuition fees for foreign students and other measures are needed to attract foreign researchers.

A fifth major issue is the *use of research results by industry and/or society*. Although the groups we visited were in general paying attention to the applicability of their results in economic or societal practice, mechanisms for smooth technology transfer are not well developed.

Universities and the Academy should be stimulated to promote tech transfer. An entrepreneurial spirit should be created and support mechanisms (e.g. tech transfer offices handling IPR issues and other contracts, VC funds for spin-offs) developed.

Industrial demand for research results, however, is not well developed in the Czech Republic and needs attention. Support for SMEs hiring researchers with an academic background might help in bringing research in the strategic portfolio of these businesses and might increase demand for research results as well as valorisation of research result.

The 'Methodology of Evaluation of Research and Development Results in the Czech Republic' is received with mixed feelings by the research groups visited. The peer review committee shares this mixed view. On the one hand it is good that the system tries to allocate funds for research on evidence-based transparent criteria, on the other hand the present system has too short a horizon (results in one year means budget

changes in the next) and is too crude in its indicators (especially not taking quality of the outputs into account sufficiently).

The use of a past performance based system alone may discourage renewal of the research portfolio and the short time horizon may hamper development and execution of coherent strategic research agendas.

Although the committee has not studied the details of the 'coffee mill', all research groups underlined that there should be more attention paid to the relation between the number of points obtained for a certain output and the effort required to achieve this output.

Improvements suggested are taking the quality of papers into account (e.g. by higher rating of high-impact journals; taking citation scores into account; by assessing the quality of industrial models, prototypes, software, etc.; and by assessing the value of a patent (in one way or another).

Checks on the spending of project subsidies is considered necessary by the peer review committee, but this control should not be too bureaucratic and should take the unpredictable nature of research into account. There should be (some) freedom for researchers to move funds between different budget items as well as some freedom to move budgets between years of the project.

The final conclusion relates to the peer review itself. Based on the review results as well as the reactions from the research leaders the peer review committee considers the site visits successful.

The peer review process 'forces' the group leaders to think about their performance and organization, and the independent international quality assessment, adds to the power of reflection. This helps the group leaders to sharpen their focus and in the longer run, improve their performance.

A balanced quality system taking into account quantitative data on output and impact citations as well as independent (international and critical) quality assessment also focusing on soft/qualitative issues would be of benefit for the Czech research system.

General conclusions of the natural sciences review

From an international perspective Czech researchers receive relatively low basic salaries. Although these salaries may be augmented with bonuses for above-average performance, this is a barrier for full integration in the international world of science and technology as it is financially not attractive for foreign scientists to come to work in the Czech Republic.

A limitation (or drawback) is concerned with the present structure, with on the one hand universities (degree-granting, research and teaching) and on the other hand the Academy of Sciences (fundamental research). The two entities profit not yet as much from the synergetic possibilities of their co-existence as in many foreign countries.

For two of the three Academy of Sciences institutes that we visited, part-time full-professor appointments appear to be hard. Requirements for full professorships are in some cases heavily based on education related factors. For the Czech research system to get more in tune with the rest of the scientific world and to profit more from the research in other institutions such as the Academy of Sciences or companies it is recommended that also qualified scientists from outside the university can be appointed as part-time full professor. When, vice versa, more university professors hold a part-time research position outside university (in industry for instance), cross-fertilization may be the result.

In order to have a more complete evaluation of a research team it is recommended that the context within which the team operates (the Institute) should also be taken into consideration.

In order for the Czech Republic to profit more from the knowledge created within its research system, researchers should be made better aware of the societal and

commercial value of their work and – if appropriate – be trained in protecting and commercialising their intellectual properties.

In order to spread knowledge on importance and impact of scientific research, outreach events should be developed for a broader public, including children and opinion leaders.

General conclusions of the medical sciences review

The panel states that medical research suffers from a lack of national institutional funding for research, which is a generic problem in Czech research. In general, the panel thinks that the research budgets are extremely modest, compared to international top laboratories, which spend these amounts of money in a month. The infrastructure of research institutes is often not of international quality, and there are problems with space: limited laboratory and office space.

Nevertheless, the visited institutes perform well and even in one case internationally up to par. However, while the current budgets may be sufficient for the current research, it is difficult to expand.

Overall, the results of research originate from basic exploratory research and are mainly used by the scientific community.

When looking at the research programmes it is important for research groups and institutes to make a long-term mission that is focusing on solving a specific medical, environmental, or biological problem. This will make research more targeted and functionally coherent, it will enable outreach and make research truly multidisciplinary.

Better papers should not be a goal in itself, it is a reward for a broader vision, which also relates to outreach. Basic science, above all benefits from being published in high impact journals. However, also publications in high impact journals should not be a goal in itself, but the reward for a long-term vision.

A more focussed mission could also enhance collaboration, as it enables the attraction of new (inter)national strategic partners, also outside of the own research communities.

The panel recommends advocating transregional grants between groups working at different universities/institutes. Other countries have had very good experience with funding proposals brought forward by a group of labs that focus on a particular topic. This fosters collaboration/translation of research; communication between post-docs and researchers of different laboratories and it makes such a group more competitive on a national, but also international basis. Obviously, transregional funding programmes require policy involvement as well.

International collaboration is present, but could be further explored, for instance in the context of the European Framework Programme.

Human resource policy is crucial:

- A brain drain in medicine is a not so unlikely a threat. The panel is worried that much talent is lost as it is hard to follow a scientific career in medicine, because of the insecure funding situation and the sometimes high teaching obligations, but also because the incentives to stay in science are not appealing: even dedicated young scientists have to secure exit strategies, which in medicine are readily available.
- The possibility to attract qualified academic personnel, post-docs and researchers from abroad remains difficult because of low salaries in science in the Czech Republic, for young as well as senior scientists.
- There is an absence of stable long-term sources (regular institutional funding) for PhD students, which is a problem for educating talent and structured planning of developing talent

- Institutes would benefit from more (competitive) post-doc positions. Czech salaries seem not to be competitive to be able to hire post-docs from abroad. The return from experienced Czech post-docs from abroad would be beneficial but the panel acknowledges that it is mainly for social reasons that Czech researchers return. The panel acknowledges that the lack of post-docs could at least be partially solved by clearly defined and reliable management of institutional funding of science by the Ministry of Education.
- It is recommended to improve international mobility of researchers. For example, make post-docs fit for international grants, but also already start with international mobility in the PhD phase.

The review panel recognises that the Evaluation Methodology favours short-term results. The system enforces rapid publication in lower impact factor journals, while good research benefits from comprehensive results, the obtaining of which usually takes several years, presented in high impact journals.

The review panel endorses the vision that it is better to produce a few high-quality papers than many mediocre ones. The high-quality papers, however, will follow from a transition into the international current top standards, which will include translational, problem oriented studies.

General conclusions of the social sciences and humanities review

The first observation is that all three institutes heavily criticize the current *Evaluation Methodology*. The criticism focuses on the following points:

- All three institutes underlined that the Evaluation Methodology does not take the differences between the disciplines into account. It is discriminative towards the social sciences and the humanities, as the system has a bias to the natural sciences. The scientometric criteria used for judging outputs are not applicable to the social sciences and the humanities. A good example of this is patents (which are worth a lot of points), which are very rare in these disciplines. Furthermore, many disciplines in the humanities focus on national issues like literature, culture and language. Therefore, it is not common to publish in international peer reviewed journals. Equally rewarded activities in these disciplines, like scientific conferences and exhibitions, are not taken into account in the Methodology.
- The Evaluation Methodology is too rigid. It gives points for outputs, but does not do justice to the differences in quality of that output.
- The Evaluation Methodology encourages competition between research institutes of the Academy of Sciences and universities instead of closer cooperation. In addition, for the viability of the research institutes of the ASCR it is important to cooperate closely with the universities. Although there are many ties between the institutes and the universities, in order to get funding, the institutes and the universities have to compete for output-points.
- The Evaluation Methodology gives rise to strategic behaviour. In order to maintain the same level of funding institutes have to acquire as many points as possible, which encourages a very short term strategy, while quality of research requires a long term strategy. As a result, the big threat is that the Evaluation Methodology determines the research efforts, instead of an evaluation system that is suited to assess the quality of the research.

The committee strongly recommends a more comprehensive Evaluation Methodology based on qualitative indicators and some kind of peer review. Secondly, the indicators used in the Methodology should be made more appropriate. A more diversified set of indicators should be incorporated in the Methodology.

The second observation is that the social sciences and the humanities are under pressure due to *budget constraints*. Like in many other countries the 'economic relevance' of research is increasingly emphasised. As a result, the social sciences and humanities face dramatic cut backs in budget. The Council of Research, Science and

Innovation does not seem to reward the 'societal relevance' of the social sciences and the humanities and the richness of their research but to solely focus on more applied technological research.

The third observation is that there is a *lack of continuity in research policy*. A long-term vision on the development of the research and innovation system seems to be missing, in particular regarding the contribution and role of the social sciences and humanities. In addition, the criteria of the Evaluation Methodology are often subject to change, sometimes even retrospectively. As funds are allocated on a year-to-year basis, the Methodology causes a lot of instability and turbulence.

A fourth observation is that from an international perspective Czech researchers receive relatively *low basic salaries*. Some researchers have two jobs in order to make a living. This could well be at the expense of the quality of research and teaching. Furthermore, the level of the salaries is a barrier to attract the best and brightest researchers or to attract foreign scientists to come to work in the Czech Republic.

The final observation is that there are some barriers for the researchers to fully *integrate in the international world of science*. There seems to be too little budget for travelling and attending international conferences. Furthermore, there is too much red tape. The procedure for going abroad should be lean and mean and the committee recommends reducing the administrative burdens. Finally, the supportive staff for international projects (for the application of grants) could be extended. Especially the EU procedures require special skills and knowledge of the administrative system. Support staff that helps researchers with the application process could be of great value.

General conclusions of the economics review

In the economic sciences, two economics institutes participated in the review. The research output of both institutes appears to be considerably above the national average, with one of the institutes displaying international research excellence, and the other national excellence. Hence, in the economic science it was clearly the 'star groups' that underwent review.

Economics had to be built up from scratch after the 'velvet revolution' of 1989. In this light, the achievements of both institutes are doubly impressive. Both research institutes attempt to deviate as much as possible from national research standards and to develop international academic recognition. However, both of them are facing organisational and funding pressures that threaten past achievements and future ambitions. This is mainly due two to factors:

- Local promotion policies constitute a major obstacle to building and maintaining a high-quality, research-orientated research institute because they emphasise the quantity of publication rather than quality.
- As in the social sciences and humanities review, the Evaluation Methodology was heavily criticised as unsuitable to the economic sciences. Both institutes were critical of the Evaluation Methodology, which in their views does not differentiate enough between different types of outputs, especially in terms of quality, and is subject to widespread gaming. Reviewers agreed that the Evaluation Methodology needs to be re-thought and redesigned to take into account the nature of economic research and specifically commented on the Evaluation Methodology.

  "During the analysis of the performance of both institutes, it was interesting to learn about the funding distribution mechanism used by the Czech authorities. […] Across the globe, governments are now designing rules for allocating funding for higher education teaching and research on the basis of parametric models. These models should have the advantage of being transparent and neutral. However, to this end, they should be well balanced and they should aim at strengthening quality and excellence, and not just quantity. More specifically, the weights attributed to various parameters and countable items that provide the parameter values, should reflect an equitable balance between quantity and

quality. It is clear from the various analyses reported during the interviews and performance overviews that this balance may at presence be out of balance. This situation has quite a negative impact on the funding performance of centres of excellence that operate in disciplines like economics, as opposed to the natural sciences. The inter-disciplinary differences in publication performance require careful attention when designing parametric allocation models.[8][9] In addition, it is important not just to take into account the inter-disciplinary issues as just mentioned. Also intra-disciplinary parameter weights deserve careful attention and scrutiny. More specifically, the weight attached to books versus international publications in top-ranked high-impact journals. The analysis of the information provided during the visit demonstrates that also in that respect, remediation may both be warranted and desirable."[10]

General conclusions of the review of applied research centres

Two applied research centres were evaluated, and their specific profiles exemplify that even within the seemingly homogeneous group of 'applied research centres' R&D practices differ widely: the different institutes work in different thematic fields and for different user groups, ranging from small manufacturing companies in the region or multinational companies through sectoral user groups to municipalities, regional and national authorities or even citizens in general. This, of course, shapes the interaction between the research centre and its users. While in one field, patents and licences are important, in another field, publication in regional or national journals and direct communication with the customers are more appropriate. Moreover, other activities apart from research can be important and relevant in certain areas (e.g. services and consulting such as preparatory work for regulations or high-tech development services for local companies). Any evaluation of an applied research centre and – even more – related R&D policies have to take these specificities into account.

Both institutes evidently benefit from good working relations with other research organisations and with customers. Partnerships with universities are not only crucial for research quality, they are also a major way of recruiting young researchers. The relations with customers often seem to be opportunity driven rather than strategic which can be risky on the long-term because for an applied research centre, strategic partnerships with users are the basis for understanding their (changing) needs which in turn is the basis for staying a relevant and competent research partner.

Over a long period, both institutes have been clearly focusing on applied research in the national market and within these markets they had and still have a good standing. In the recent past, they have experienced substantial changes of both their markets and the national R&D system. Nonetheless, they have managed to survive in unstable conditions and to remain competitive at national and (partly) the international level.

However, due to the external changes, the institutes have expanded their portfolio of R&D activities and services, but without closing down outdated areas of activities. The need to grasp each funding opportunity has thinned out competences in terms of team size, critical mass, and keeping up-to-date with new developments in the field. Research is very much driven by the search for funding opportunities, resulting in long lists of activities rather than a coherent programme. Consequently, both centres miss a stringent research strategy and clear understanding of their core competences.

---

[8] Glänzel, Thijs, Schubert & Debackere, 'Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance', *Scientometrics*, Vol. 78, No. 1 (2009) 165–188

[9] Glänzel, Thijs, Schubert & Debackere, 'A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking', forthcoming in Scientometrics 2011

[10] Panel Report Economics, p. 10-11

The changes in funding policy of the government which has reduced or taken away research budgets from some sectoral ministries has caused difficulties for some applied research institutes: they have literally lost main contractors of research and users of results. Some of these funds have been transferred to competitive thematic research programmes where institutes receive funding for mainly self-determined applications. These are important tools of R&D policy but they cannot substitute specific contract research where a ministry acts as the very user of results. Therefore, a debate about the role of Czech ministries as research policy makers on the one hand and as buyers and users of research results is required.

Management is a recurring issue in the Czech Republic (see also the case study about the OP R&D for Innovation). What's more, R&D policy makers obviously have not yet identified it as a key factor influencing research quality to the good or the bad. Above all, also on the applied centres, management systems suffer from unclear divisions of tasks and responsibilities between executive, supervisory and advisory bodies. They often lack effective HR policies, and clear career paths for (young) researchers and other staff. Similarly, bespoke mobility activities and gender programmes are rare. This is exacerbated by the tight financial situation which makes it difficult to pay competitive wages.

Both centres evaluated (and many others in the Czech Republic) have succeeded in the international peer evaluation of the Operational Programme R&D for Innovation and they will set up Regional Research Centres. The centres will not only receive (dearly needed) new buildings and advanced equipment, they will also implement stringent research programmes, more effective management models and a more explicit human resource policy.

2.3.4 Lessons learned from the international peer-review

In this section we draw the major lessons on the quality of research based on the international informed peer-review.

1. **The quality of research in terms of international visibility (publication, strategic partners) was generally appreciated.** There was no doubt that Czech research teams are well recognised members in the international research community. The units under consideration have been labelled ranging from 'good' to 'competitive'. In the context of what can sometimes be quite unstable and even adverse institutional environments research quality came to be highly appreciated.
2. **Using a limited number of solid criteria for testing the quality of research.** The existence and number of high level academic publications, a good scientific standing of the key staff in the more academic types of research; a clear understanding of users and the access to users in more applied environments, were all important elements. Coherence and compactness of the research portfolio was another important criterion. From a procedural point of view the question of whether or not a unit displayed good research quality was settled within a fairly short time and in all cases without much dispute.
3. **A tendency towards small units.** Mainly in the field of natural sciences, medicine, and technical sciences a tendency towards small units has been observed, particularly in larger institutes. This tendency is to a large extent determined by a number of rather persistent behavioural patterns, mainly rooted in the institutional and political trajectories of the research institutions. Regarding the institutional background it is mainly cultural, as academic institutions delegate a high degree of autonomy to their key staff, the professors. The selection of research topics is accordingly mainly bottom-up and de facto determined by personal preferences. Political trajectories, particularly from the communist period, also support the tendency to create small research niches (which, in communist times, were perceived as niches of relative freedom to work, to exchange ideas, to travel, etc.).

4. **Missed opportunities from limited capability of the research units to re-orient themselves.** The inherent tendency towards small entities mainly in the field of natural sciences, medicine, and technical sciences is a cause for missed opportunities as small units typically face severe problems when it comes to entering into interdisciplinary research, orientation towards application, long-term development, or strategic partnering. All these activities require firmly established competences in the core fields as well as free resources for exploring new topics. Small units typically do not have enough resources for both, which systematically prevents innovation (particularly in the absences of leadership at higher hierarchical levels).

5. **Limited awareness and incentives for entering into interdisciplinary or application oriented research.** In the reviews a rather peculiar attitude has been observed, namely the low level of awareness of, a relative absence of and sometimes reluctance to enter into inter-disciplinary research and application orientation within academic research. This observation is not about a sometimes politically motivated claim for commercialising academic research; rather it is about systematic interaction and exchange between different fields of academic research. A good example at hand is modelling. On the one hand, modelling is a core topic in mathematics. On the other hand, modelling is applied in fields like advanced materials, complex energy systems, or finance. It is indispensable for mathematicians and researchers in the fields of application to work together systematically. This type of collaboration typically requires a minimum number of staff (> 50), long-term orientation and stability (up to ten years), adequate organisational settings, and a strong, visionary leadership. In institutional terms it requires interaction between, say, two Institutes of the Academy of Sciences or a faculty of a university and an institute of the Academy of Sciences. Actually, there is not enough awareness of such opportunities, nor are there incentives, not to mention adequate funding schemes supporting interdisciplinary research . As research at the interface of established research fields often turns out to be a source of breakthroughs, the lack of interdisciplinary research is a missed opportunity of considerable long-term significance.

6. **Little awareness and room for manoeuvre in human resource issues.** While the some of units evaluated had staff well balanced in terms of age and experience as well as access to young researchers, quite a high share suffered from an unhealthy age structure, problems in recruitment, and mobility. A closer look reveals a number of specific issues. Most of them are missed opportunities rather than institutional barriers. Experience shows that there is always at least one unit that has found a proper solution to a problem with which others are still struggling.

   - Age structure. If a stone were to fall on the head of the typical director, his team would face serious problems in terms of leadership and relationships to the (outside) scientific community. This structural problem is systematically related to small size and hence to limited room for division of tasks, responsibilities and risks.

   - Recruitment by self-breeding. Even in groups with a strong international reputation, many careers are characterized by a Master thesis, PhD thesis, research position and professor position within the same group. These internal careers do not only limit the perspectives of researchers but also limit the number of solid external contacts of a group and therefore their options for participation in international projects (which are often based on personal networks). The present recruitment policy aimed at hiring the best PhD students from one's own group for staff positions should thus be linked with open recruitment policies.

   - (International) mobility. The 'closed loop' recruitment policy prevents an explicit mobility policy. Moreover, the small size of many units typically causes another systematic problem: they have difficulties managing a 'portfolio' of

projects in which both young and more experienced staff are involved. Accordingly, stays abroad are mostly restricted to a couple of weeks and at best months, limiting the opportunities for learning and discovering new horizons. Thus it should be a high level goal to promote young researchers earlier and make more intensive contact with foreign research teams, funded e.g. by foreign exchange scholarships and the participation of Czech researchers in EU mobility programmes. However, senior people have to play their role in establishing these international relationships. Mobility has to be rewarded in the appraisal and career development schemes.

- Limited collaboration between universities and Academy of Sciences. Universities do have a monopoly in awarding academic degrees. Some (mainly Academy of Sciences institutes) perceive this division of labour as a massive barrier, while others, in the same type of institutes, consider it a cheap research workforce. However, it is hardly ever resolved in a creative way, e.g. by implementing joint programmes. The implementation of 'research schools' could be such an opportunity and a giant step in the training of young researchers in the Czech Republic.

7. **Funding policy as a major cause for short-termism.** Many research institutes suffer from the 'funding regime'. It is not primarily the limited availability of funding in general – there should always be certain level of complaints about limited resources as it is an indicator of a productive research environment. The major problems arise from two main sources: national public funding, both institutional funding and targeted funding, as well as the handling of funding within the research institutions.

   - Missing continuity. This has essentially two aspects. One is the missing continuity due to fast changing rules for institutional funding at the level of national policy institutions. The other one is weak leadership at the level of the research institutions in terms of internal funding allocation. In practice this means that annual budgeting at national level is often more or less passed on to the sub-units.

   - Small size and short term. Obviously, the dominance of small-scale project funding hampers the running of bigger research projects or programmes over longer periods. Likewise, the allocation of institutional funding based on the annual measurement of past achievements, as does the Evaluation Methodology, systematically drives out long-term large-scale considerations. This type of short-termism is caused not only by the funding policy at national level (institutional funding, grant agencies) but also within research institutions, where annual appraisals are mainly based on last years' achievements rather than on the assessment of future plans, with past achievements considered an indicator not of only performance but also of credibility.

   - Not surprising, the Evaluation Methodology has been perceived with mainly negative feelings both by the research groups visited as well as by the reviewers. While it was acknowledged there was a need to allocate funds on evidence based transparent criteria, the practical implication of this was criticised for the following main reasons: short-termism, backward orientation, negation of disciplinary differences and differences between types of institutions, and encouragement of perverse behaviour.

   - Poor participation in EU programmes. The reason for this is rather obvious: weak linkages due to poor experience. Putting participation in international programmes on the top of the agenda – at the level of institutions, not necessarily of the teams – is thus an immediate advice, including the provision of support and training.

2.4 Survey results

In addition to the informed peer-review we performed a survey to ask researchers and those in charge of research management about the ways in which research is organised in the Czech Republic. In both the survey of rectors, deans of faculties, and directors of research institutes and the survey of researchers we asked a number of questions about the research environment they operate in.

In the survey of rectors, deans, and directors we asked them about the existence of various management tools. Of course, the existence of a tool does not tell us anything about its content, let alone its quality, its effects and the priority it has with the research organisation. Nonetheless, a first step is to *have* a certain tool.

Figure 1 Management tools at research organisations

Our organization ...	Yes (in %)
Publishes an annual report	97%
Has a gender policy	96%
Has an advisory board that includes internal members	93%
Has an advisory board that includes external national members	92%
Has a long-term research strategy	91%
Has evaluation procedures and criteria for quality of research	91%
Implemented a multi-annual plan addressing major issues (research programmes, investment, human resources ...)	88%
Has a career policy	73%
Has an IPR policy	73%
Requires a minimum number of publications to obtain a PhD	69%
Has a recruitment policy	57%
Has evaluation procedures and criteria for quality of teaching	51%
Has evaluation procedures and criteria for quality of administration and internal services	51%
Conforms to quality standards of ISO 9001, EFQA, or others	45%
Has an advisory board that includes external international members	39%
Has a technology transfer facility	31%

Source: survey of rectors, deans, directors, n=74

As can be seen in Figure 1, many formal tools are in place: most research organisations publish an annual report, have a gender policy and an advisory board (although not very often with international experts), have a long-term strategy and evaluation procedures and criteria in place for evaluating the quality of research (although in many cases presumably the Evaluation Methodology).

However, the most important management tools in a research organisations which every research organisation should have – career policy and recruitment policy – score much lower (73% and 57% respectively).

Somewhat oddly, while relatively few have career and recruitment policies, almost all have a gender policy. How does this go together? Gender policy is by necessity an integral part of recruitment and career policy, as a gender policy typically encompasses strategies for female recruiting and female research careers. Indeed, based on the various interviews we conducted, based on our experience with the international peer-reviews and the Operational Programme R&D for Innovation we had the distinct impression that if gender policies exist, they are superficial ("We do not discriminate against women!") and are not based on an analysis of the

organisation's situation[11]. In fact, in the international peer-reviews we hardly saw any women researchers and directors at all, and gender was hardly ever mentioned in any of the self-assessment reports, where, within the framework given to them, research organisations were invited talk about what they considered the most important points.

The gender policy example nicely illustrates the limitations of Figure 1 – that there may be discrepancies between formal management tools in place and the priority they have for the research organisation and the quality of the management tool. Ticking the box and thus confirming that a certain policy is in place, does not mean that the policy is developed and implemented in a good way. In the case of gender policy this is particularly problematic as the situation of women in research in the Czech Republic is far from ideal[12]. At the same time research groups are complaining about limited capability to attract senior researchers (48% of research groups are 'not at all' or 'to a limited extent' able to attract qualified senior researchers), technical-administrative staff (38%), postdoctoral researchers (29%) and PhD students (16%). Given these staff shortages, it would be particularly necessary to have good strategies to better recognise the female potential.

We also asked research organisation leaders about working conditions at their organisations. In this instance, we also asked the same question to researchers (Figure 2). Most organisations (over 90%: 20% 'to a moderate extent', 33% 'to a large extent', 38% 'to a very large extent'[13]) have incentives for individual researchers in place to apply for R&D projects, but fewer (60%) provide support and training in preparing funding applications. Most organisations (over 90%) also encourage researchers to engage in academic research and related publication activities. However, only about half of organisations (49%) offer some sort of career development plans for researchers or encourage researchers (26%) to adopt entrepreneurial activities outside our institution.

Interestingly, research organisation leaders consistently rate working conditions at their organisations higher than researchers themselves. For example, while 49% of researchers agree that their organisations have career development plans for researchers, it is 65% for rectors, deans, and directors[14]. This either points to a communication problem (researchers do not know that there are such plans) or a quality problem (researchers do not consider the existing career development plans good enough to agree to the question).

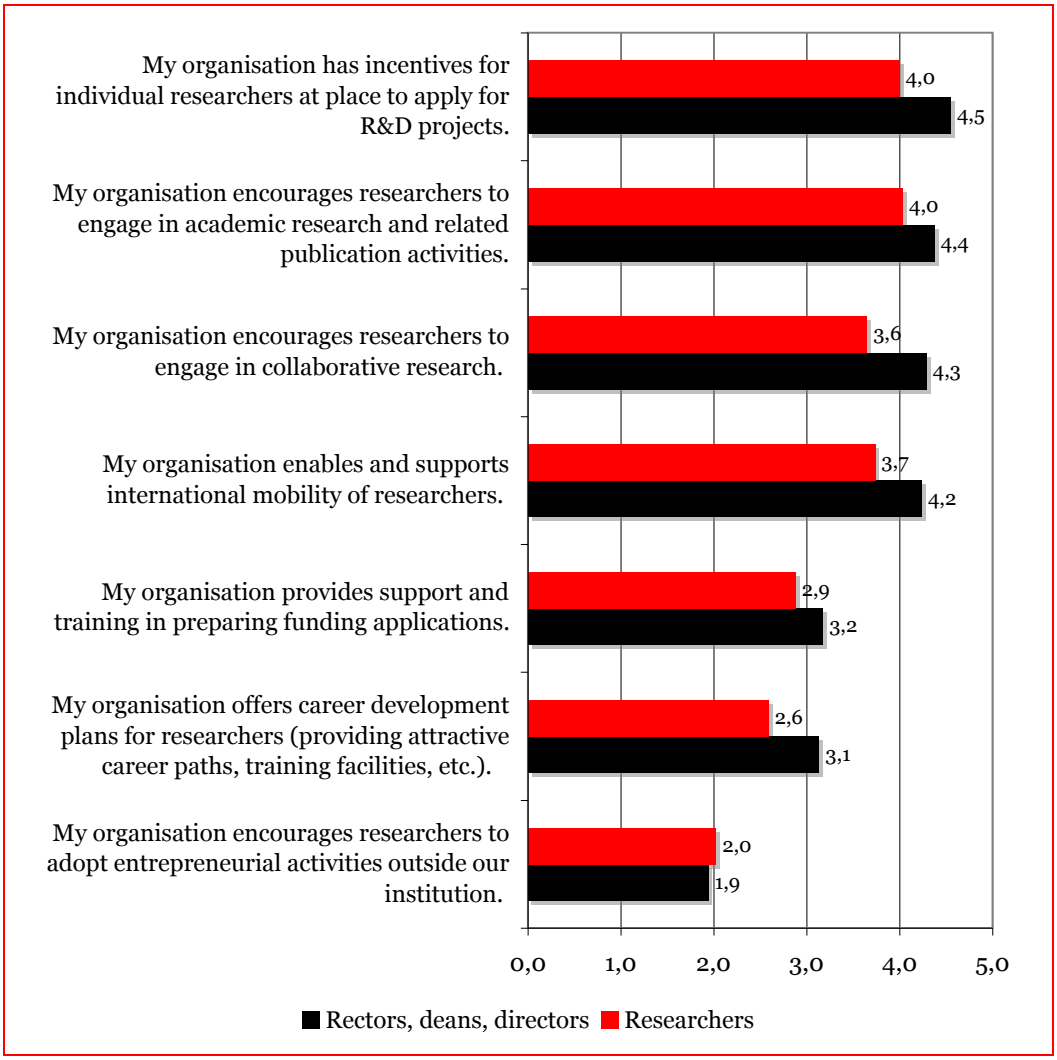
[11] An analysis would answer questions such as: How many women do we have in our research organisation? At what hierarchical levels? What kind of contract are they on compared to their male counterparts? How much do they earn in comparison with their male counterparts? How high has female turnover been in comparison with male turnover? Etc.

[12] Alena Krizkova, Czech Republic Country report, Meta-Analysis of gender and science research, FP7 RTD-PP-L4_2007-1, March 2009

[13] The percentages refer to the researcher survey

[14] See also Figure 1, where 73% of rectors, deans, and directors say they have a career policy in place

Figure 2 Working conditions at Czech research organisations, as perceived by research organisation leaders and researchers*



*Arithmetic mean on a scale from 1=not at all to 5=to a very large extent

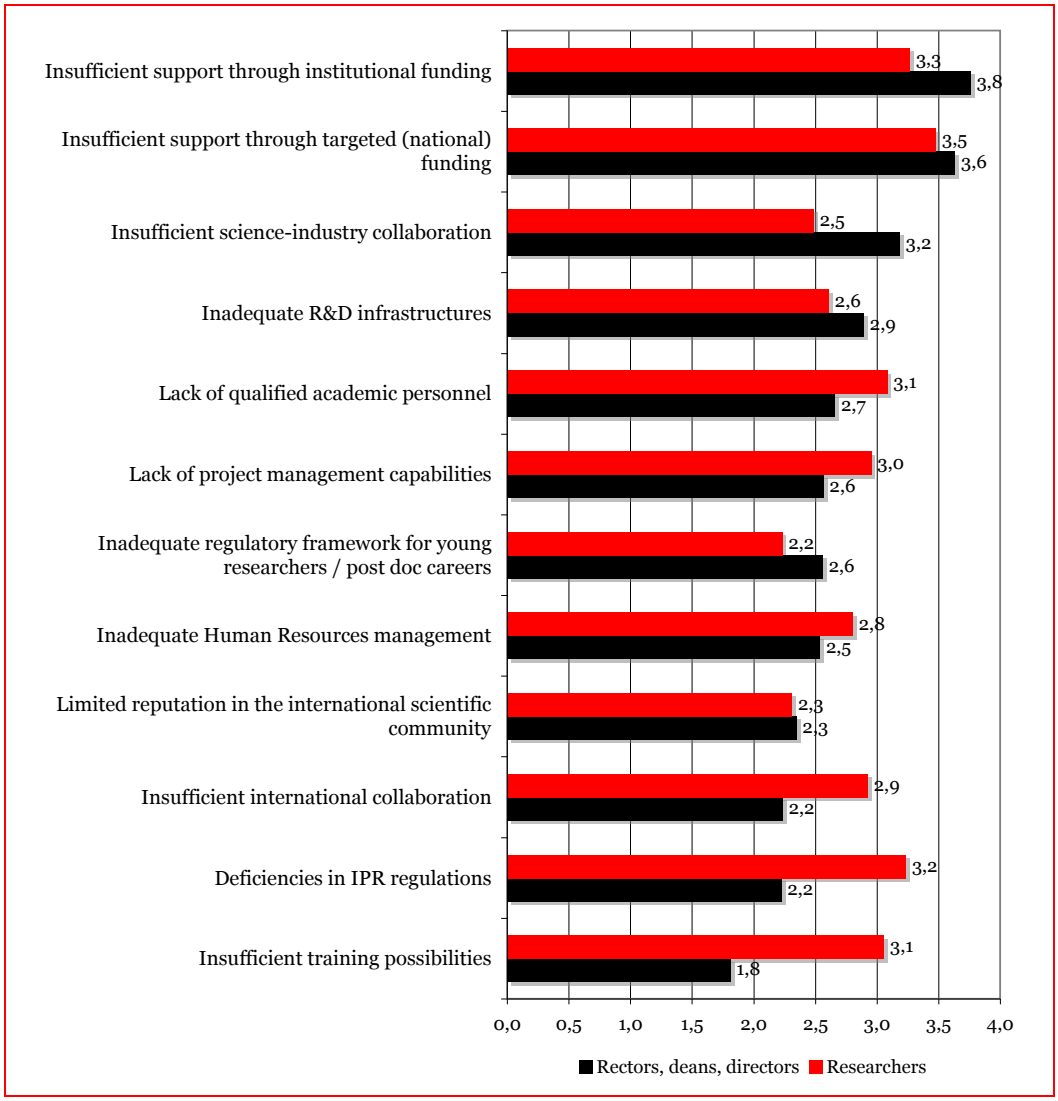
Source: survey of rectors, deans, directors, n=74; survey of researchers, n=74

However, although there is plenty of evidence from the interviews and the international peer reviews showing that research management is an important bottleneck in the Czech RDI system, it is not necessarily perceived as such by research organisation leaders. We asked rectors, deans, and directors what they considered the most important barriers for RDI performance in the Czech Republic. By far the most important barriers concerned funding – insufficient institutional funding and insufficient targeted funding, while the more management related barriers such as inadequate human resource management or inadequate regulatory framework for young researchers/postdoc careers ranked in the middle and insufficient training possibilities at the very bottom (Figure 3). However, the international peer reviews showed that the management of research and the awareness of its importance could often be improved – including human resource management, careers of young scientists and training possibilities.

Having said this, we are not downplaying funding problems; we have indeed met with some very unfavourable funding situations in the international peer reviews that threaten institutes' high-quality work. But unfavourable funding situations call for even better management to mitigate the adverse effects of too little funding.

What is most interesting about Figure 3, however, is that research organisation leaders and researchers do not agree on the barriers to RDI performance existing in the Czech Republic. While rectors, deans, and directors perceive insufficient levels of funding as the largest barriers, researchers – though not ignoring funding problems – underline human resource orientated issues, such as insufficient training possibilities, lack of qualified academic personnel or lack of project management capabilities. The latter is also in line with the findings of international review of research institutes.

Figure 3 Barriers to Czech RDI performance, as seen by research organisation leaders and researchers*



*Arithmetic mean on a scale from 1=not at all to 5=to a very large extent

Source: survey of rectors, deans, directors, n=73; survey of researchers, n=616

To summarise, the survey results show that career and recruitment policies are not quite as widespread as they should be. As the main input to research is human capital, every research organisation *must* have career and recruitment policies. For recruitment and career policies not to be a 'paper tiger', all they need to do is a give a convincing answer to the question: "Why should a talented researcher want to work for our research organisation?".

Nonetheless, research organisation leaders perceive insufficient funding levels as the most important barrier while researchers though acknowledging insufficient levels of funding also emphasis human resource orientated barriers. While we do not want to

downplay funding problems – we have indeed encountered some very critical funding trajectories in our international peer reviews – we also think that management issues should not be ignored either.

2.5 Conclusions on the quality of research in the Czech Republic

The notion 'quality of research' represents a variety of meanings: The quality of the topics dealt with and their coherence (priority setting, strategic management), the standing and reputation of the researchers involved (track record), the outcomes of research activities (publications), training and scientific careers (PhD), not least the context, in which decisions are made and research is performed (infrastructure, funding, management, governance).

In this conclusion we will present our main observations on the quality of research in the Czech Republic. We firstly will deal with the question of outcome, thus of actual achievement. Secondly, we will direct our attention to the quality of the context of performing research and thus the factors contributing to the quality of outcomes.

- Bibliometric analysis shows that there is **an increasing international visibility** of the output of researchers from the Czech Republic and a remarkable improvement in the quality of the scientific publication output of the Czech Republic over time – especially over the last decade. Moreover, there is an **increasing trend in international and especially national cooperation** – the latter indicating an **improvement of the internal cohesion** in the Czech research system.
- When looking at the institutional level, the **university sector and the Academy of Sciences are the main players** in the publication activities of the country. There are **two particular actors** on the international scene in the Czech Republic, the Academy of Sciences, participating in more than 45% of all the scientific publications of the country, followed by Charles University in Prague. Besides them, other major institutes and universities that publish a lot are the Masaryk University, the Institute of Chemical Technology in Prague, the Czech Technical University in Prague, or the Palacky University in Olomouc (all of them with more than 3000 publications during the period 1993-2009) among others.
- A comparison with the research system of Denmark, Finland, the Netherlands, and Sweden, which all have a strong scientific performance, makes clear that **the Czech research system can still improve**.
- Despite the fact that research performance and quality of research overall is lagging behind compared to the strongest research-performing nations in Europe, there are groups of high international visibility in terms of publication, reputation, or strategic partners the Czech Republic. There is no doubt that Czech research teams are well recognised members in the international research community.

The only practical way to influence improvement in the amount and quality of research outputs is through the improvement of the environment for doing research. This encompasses a number of factors contributing to research quality. If these factors are not taken care of, they will seriously harm the quality of research. Because the research environment can be influenced by research management and policy making, the factors are particularly interesting from policy makers' point of view.

- At policy level, we have observed a wide range of systems and instruments in place. **The Czech Republic has developed a wide range of policy institutions, instruments, and processes over time**. However, some of them need change, e.g. the Evaluation Methodology; others need improvement, a few have to be implemented from scratch. Therefore, in general improvement and learning is on the agenda.
- Likewise, at the level of research performing institutions, many of them have formally implemented specialised management systems and tools. This observation is consistent with the self-perception by leaders of institutions

(rectors, deans, directors). At the same time, when going into details, there is plenty of evidence that **the way research management is practiced needs to be improved**. The international peer reviews in particular indicated a serious under-perception of the role of and impact of proper management and management systems: organisational structure, size of units, composition of staff, age structure, leadership, recruitment, career development, and mobility.

- We consider the **limited awareness for organisational and managerial issues one of the major bottlenecks and shortcomings** in the Czech research performing institutions. Leaders of institutions (rectors, deans, directors) perceive funding as the major problem in the Czech research system. The focus on funding seems to prevent a more in-depth examination and discussion of institutional, organisational, and managerial issues. This under-estimation can be perceived both at the level of policy makers as well as at the level of representatives of research performing institutions.
- **To increase the quality of research, the organisational / managerial performance of the research institutions has to be improved.** For this to happen and to make plans over longer periods, research organisations need a stable funding environment.
- The international peer review has highlighted a number of issues, which are to a large extent prevailing throughout the different types of institutions and fields. Therefore, they are robust and thus worth to be considered as structural issues in the Czech research system. In a nutshell, these are the most relevant ones:
  - There is **a tendency towards small units, which leads to limited opportunities for innovation**: to explore new fields, to enter into collaboration with strategic partners or to enter into interdisciplinary fields. As these units are embedded in larger autonomous institutions (universities, faculties, institutes), the first level of responsibility is within these institutions' internal structure, leadership, and management capacity.
  - There is **little awareness and room for manoeuvre in human resource issues.** Quite a high number of research units suffer from a small size (cf. above), from an unhealthy age structure, problems in recruitment with a strong dominance of 'self-breeding', and mobility. Most of these problems can be seen as missed opportunities rather than institutional barriers. Again, it is the small size of many units which typically causes a systematic problem: they have difficulties managing a 'portfolio' of projects in which both young and more experienced staff is involved, they have difficulties in actively managing mobility.
  - **Limited collaboration between universities and Academy of Sciences**. Universities do have a monopoly in awarding academic degrees. Some (mainly Academy of Sciences institutes) perceive this division of labour as a massive barrier, while others, in the same type of institutes, consider it a cheap research workforce. However, it is hardly ever resolved in a creative way, e.g. by implementing joint programmes. The **implementation of 'research schools'** could be such an opportunity and a giant step in the training of young researchers in the Czech Republic. The problems are less that of insufficient or poor policy making, rather than a lack of imagination and entrepreneurship.
  - **Funding policy as a major cause for short-termism.** Many research institutes suffer from the 'funding regime', i.e. the Evaluation Methodolgy. However, it is not primarily the limited availability of funding in general – there should always be a certain level of complaints about limited resources as it is an indicator of a productive research environment. Nonetheless, there are problems and these arise at two levels: at the policy level as well as at the level of the funded institutions. The major consequences of the current funding regime: lack of continuity, small size and short-term thinking. It is important

to note that the absence of leadership at the top level of institutions and poor management at lower levels do play their roles exacerbates the problem of short-term, small-scale, discontinuous funding. While in the majority of cases the global budget is more or less linearly passed on to the sub-units; there have been some who allocated institutional funds in a more creative way reflecting strategic choices.

- **The most promising advice, however,** is to acknowledge that there is always at least one unit that has found a proper solution to a problem with which others are still struggling. Thus look for the **'positive deviants'** and learn from them.

3. The Evaluation Methodology

With the introduction of the Evaluation Methodology in 2004 the Czech Republic developed an instrument to assess the quality of research. We have analysed the 2004 Methodology as well as the most recent versions of 2010 and 2011, as the latter were given the additional role as a tool for decision-making on the allocation of institutional funding. The aim was to see if the Evaluation Methodology is an appropriate tool for assessing the quality of research, and based on this, allocating institutional funding and give recommendations for improvement.

The analysis of Czech policy documents and international literature will show some of the major problems of the Methodology (sections 3 and 4). This first analysis will be followed by a description of Czech researchers' and research managers' view of the Evaluation Methodology and an analysis of its effects on the Czech research community (section 5). Based on four international case studies, we assess the Evaluation Methodology against international practice (section 6). We also describe alternative approaches to the Evaluation Methodology found in the Czech Republic (section 7).

All this information results in a SWOT analysis of the Evaluation Methodology (section 8) and recommendations for its reform (section 9), or rather for the two main issues it addresses: evaluating and developing the quality of research and allocating institutional funding on the basis of performance.

3.1 The 2004 Evaluation Methodology (EM 2004)

The methodology to be applied for the evaluation of R&D is outlined in the 'Resolution of the Government of the Czech Republic' of 23 June 2004, no. 644 on the evaluation of research and development and its results'.[15]

The objective of this Evaluation Methodology document was to tackle the perceived failure in evaluation quality in the R&D system. The document states, "The current R&D support system does not allow one to distinguish between the quality of attained results, professional standards and performance of individual organizations, departments and individuals and to take advantage of these distinctions to facilitate changing the system (the amount of the provided support, organisational changes, personnel classification, salary and promotion orders, etc.)." The alarming decline in Czech R&D performance in international comparison and the persistence of failures in the RDI system at national level convinced policy-makers that changes to the system were needed.

The main problems that were identified in the way evaluation was practised in that time period were

[15] For the Czech version of this document see http://www.vyzkum.cz/storage/att/4095103B3DF675FBB4E74B73874615F5/Metodika%20hodnoceni%20vav.pdf. Our analysis is based on an English translation of this document.

- The complexity of evaluation
- Extensive administration
- The request of unnecessary data that often was not used for evaluation
- Most importantly, the fragmentation of the evaluation system and lack of standardisation

The 2004 Evaluation Methodology document first reproduces the evaluation guidelines that were included in the National R&D Policy for the years 2004-2008 (section III.1) and subsequently complements these guidelines with more detailed information (section III.2).

The 2004 document is very ambitious in aiming to improve evaluation on all the levels of the research system at the same time. It diagnoses a strong bias towards ex ante evaluation and tries to balance ex ante and ex post evaluation in response. Similarly, it stresses that the development of the Czech Evaluation system should be based on international experience with research evaluation and foreign experts included.

The 2004 Evaluation Methodology introduced to the Czech Republic the concept of quantitative results evaluation, stressing the importance of respecting the differences between disciplines when evaluating research results. However, it is not entirely clear what the balance should be between quantitative and qualitative evaluation. Sometimes 'mechanical' evaluation based on the counting of and point attribution to R&D outputs is indicated as only part of a broader evaluation system; sometimes it stresses that research results need to be quantified. Nevertheless, quantification concerns numbers on both input (funding) and output (R&D results).

And although it is said that the evaluation of results should be taken into account when distributing state budget for R&D it is certainly not clearly stated that it should be the only basis for institutional funding.

In the sub-sections below we describe the key characteristics of the methodology proposed in 2004, first listing the established key principles for evaluation in the Czech Republic and subsequently listing the evaluation best practise it recommends to be implemented at the various levels (state, policy, programme, project, etc.)

3.1.1 Key principles for evaluation

The resolution stated that evaluation at all levels needs to reflect the following base lines, specifying that "Without these basic principles being observed, it is not possible to carry out evaluation of R&D and its results."

- Evaluation is to be regular and repeated after some time (i.e. not just the entry proposal appraisals but also continuous assessment and final assessment, including ex-post analyses).
- Specific objectives – and measurable goals - need to be set out at all levels (programme, project, research plan, institution, policy, etc.) that can then guide the evaluation.
- Evaluation criteria should be:
  - Known and binding,
  - Clear-cut (not contradicting)
  - Quantifiable, measurable, evaluable,
  - Related to a given objective

3.1.2 The use of the R&D Information System

In line with the 2002 R&D Support Act, the R&D Information System is indicated as the basic tool for the collection of information for evaluation. Information in the register is considered to be of key importance:

- For the overall evaluation of R&D in the Czech Republic (or EU)
- As a source of information on the efficiency of R&D, research activity, institutions and departments
- As information for the evaluation committees

It was envisaged in the EM 2004 that based on the results of the evaluations, changes in the state budget on R&D could occur, in particular in the medium term. The results in the data register were also to be used for the analysis of the R&D situation in the Czech Republic and for the preparation of National R&D policies.

The 2004 Guideline suggested including in the R&D Information System only the highest quality results into the R&D information system, thus reducing the burden of input of information that was not usable for evaluation purposes. Depending on the nature of the evaluated research, **high quality results** were defined as follows:

- Articles in impact and selected peer-reviewed scientific journals; scientific books and articles in professional books and peer-reviewed collections in foreign languages (and in social sciences and in the Czech language),
- Results protected by industrial legal protection and other similarly protected results (patent applications, granted patents, industrial samples, recognised varieties of agricultural crops, recognised breeds of livestock, etc.), licenses.
- Applied outputs
- Other outputs provided that they are predetermined in the draft programme sanctioned by the Government (allowing one to report the results in specific R&D activities – e.g. a research report such as the results of procurement of research for the needs of state administration).

Based on the comments of the Academy of Sciences and the Grant Agency of the Czech Republic the outputs in the field of social sciences were expanded by professional books and peer-reviewed collections published in the Czech language. An option to use other specific outputs was included, on condition that these outputs would be clearly defined beforehand in the programme proposal approved by the Government.

In addition, it was felt that the evaluation scale should be standardised in order to be able to compare evaluation results. **Evaluation criteria** were specified per area:

- Technology and engineering (inanimate nature): applied results, patents, licenses, new technological products and technologies, articles in impacted and selected peer-reviewed scientific journals, scientific books and article collections and books in foreign languages.
- Living nature (biology, medicine, agriculture, environment, etc.): new treatments, recognized breeds of livestock, recognized plant types, etc., articles in impacted and selected peer-reviewed scientific journals; professional books or articles in scientific books or peer-reviewed collections in foreign language, applied results, patents, licences and new technological products and technologies
- Social sciences (humanities, social, economic, legal and historical): articles in impacted and selected peer-reviewed scientific journals; professional books or articles in scientific books or peer-reviewed collections in Czech language, applied results
- Mathematical sciences (mathematics, cybernetics, information technologies, physics and its applications, chemistry): articles in impacted and selected peer-reviewed scientific journals; professional books or articles in scientific books or peer-reviewed collections in a foreign language, applied results, patents, licences and new technological products and technologies.

Finally, the policy document states, “Quantified results of R&D activity must be given weight in the allocation of funds for R&D in the next solution period and used as one of the main criteria for proving the quality of the workplace during the evaluation of

project proposals, research plans and other activities and in announcing programmes for providers."

In 2004 all types of results were weighted by factor 1, meaning that there was no differentiation between results. This was changed in the EM 2005 when weighting war introduced: Articles in the highly cited journals (with a high impact factor), patents and selected applied results received a weight of 2. Articles in journals with an average impact factor, books or chapters and other applied results still received 1 point. Articles in journals with low impact factor and in journals not included in Web of Science received 0.5 points.

3.1.3 Guidelines for evaluation at the different levels

Evaluation of R&D programmes

While the rights and obligations in the evaluation of the programme proposals were covered in detail in the 2002 R&D Support Act[16], the evaluation of programme results was only very generally regulated. The 2004 Evaluation Methodology acts upon this gap and proposes to use for programme evaluations similar procedures as are used for the approval of programme proposals. Programmes should be assessed after their completion, in order to evaluate the extent of achievement of their objectives, results and cost-effectiveness. As the Government approves programme proposals, the evaluation results should be submitted to the Government for feedback or approval, especially if a similar programme is to follow the evaluated one.

In this context, the National R&D Policy 2004-2008 stressed the evaluation of results and effects of the programmes, including the whole range of social functions and effects of research on the economy, society, education and knowledge. It considered that programme evaluations (National Research Programme, departmental programmes) should summarise the expected, ongoing and attained results - economic, social and other - of individual projects.

The National R&D Policy also hints at the influence of the evaluation results on future funding. It states, "The results and findings of ex post evaluations will serve as feedback for future decision making on the selection of projects and their solvers. […] The evaluation results (with an emphasis on the results of long term re-evaluations) will be a crucial criterion for the allocation of available funds. Those who achieve better results long term will be preferred."

Final evaluation of projects, research intentions, and other R&D activities

The 2004 Evaluation Guidelines also proposed that in the case of projects, research intentions, and other R&D activities, the principles and mechanisms valid for the proposal appraisals would also apply to the evaluation of the results. It stipulated that the character of the research activity performed would be taken into account during the final assessment. In the case of basic research, publications in renowned journals were required, while in the case of applied research, patents, licenses, technological processes, medical treatments, etc. would be counted as outputs.

The 2004 guidelines stressed the importance of assessing the R&D results by referring back to the explicit goals of the research and to focus the final evaluation of the results on quality assessment and the applicability of the R&D results (usefulness).

In 1998, the 'Research Intentions' system had been introduced in order to connect institutional support for research to specific objectives. Subsequently, the Ministry of Education, Youth and Sport started funding the Research Intentions of the institutions in 2001. They funded a little less than 60 Research Intentions. In 2004 a new set of Research Intentions was funded across all ministries and the Academy of Sciences.

[16] In the section "Public tender in research and development" of Act no. 130/2002 on the Funding of Research and Development

These new intentions were implemented in 2005. New legal arrangements allowed funding not only of the traditional research performers (universities and institutes of the Academy of Sciences) but also the funding of non-profit private research institutes and companies. It is estimated that about 18 research intentions from private companies were funded. Since 2005, a further tranche of research intentions was funded (98% at the universities), and in 2007 a further 30.

Evaluating the effectiveness of institutions in research and development

The lack of a standardised evaluation methodology across the various funding institutions was considered to be a major barrier to the assessment of the actual quality of the R&D results and therefore for an evaluation of the effectiveness of R&D funding as a whole.

In order to overcome this, the EM 2004 proposed the evaluation of the aggregated inputs (specified as public funding for R&D regardless of whether they were supplied by the project, the research plan, procurement, etc.) and outputs (specified as all quality results in the database, also irrespective of whether they were reported as a result of the project, the research plan, procurement, etc.).

It was suggested that the result of this evaluation would gradually translate into proposals for the allocation of expenditure on R&D.

The National R&D Policy 2004 – 2008, however, took a broader stand and stated that in the evaluation of research institutions, "appropriate attention is to be paid not only to the efficient use of received public funds, but also to the ability of those institutions - on the basis of their results – to obtain and valorise funds from the private sector."

Evaluation of National Policy in R&D

In the field of national policy on R&D, the Evaluation Methodology document stipulates paying more attention to the responsibility for implementing policy (also paying attention to reasons for non-compliance and countermeasures taken), the analysis of outstanding issues, and the efficiency assessment of individual policy instruments and their implementation. Independent experts are expected to be more involved and most importantly, the importance of the evaluation as a 'learning tool' is stressed, i.e. the results of the evaluation of the National R&D Policy will be taken into account during the preparation of other national policies.

3.2 The Evaluation Methodology in 2009 & 2010

Since the EM 2004, an amended new version has appeared every year.[17] Although some continuity can be observed over the years, the following transformations in the evaluation system are particularly noteworthy:

- In the 2009 version of the Methodology for Evaluation, the broad outline of a comprehensive Evaluation Guideline covering all the different layers of the research system in 2004 is replaced by a very specific evaluation methodology that focuses almost **solely on the quantification of research outputs** for the evaluation of research organisations and research programmes.
- The 2009 version marks the introduction of the adoption of the metrics-based evaluation of R&D results as a **Performance-Based Research Funding System**, albeit only at the level of funding bodies.
- The 2010 Methodology enforces and expands the use of the metrics-based evaluation of R&D results as a funding system at the level of research institutions. It also introduces 'dampening factors' to limit large shifts in funding among disciplinary fields and results categories.

[17] All versions of the EM can be found at: http://www.vyzkum.cz/FrontClanek.aspx?idsekce=18748. We based our analysis on English translations of the documents for 2004, 2009 and 2010

3.2.1 The 2009 Evaluation Methodology (EM 2009)

The 2009 version of the Evaluation Methodology was developed by the Commission for the Evaluation of R&D Results, which is an advisory body of the Research and Development Council. The evaluation of R&D is based on the Government resolution No. 287, also known as the 2008 R&D reform.

The document outlines changes and commonalities compared to earlier versions; describes the various steps in the process of results evaluation; and details the modalities for the evaluation of the results of research organisations and the evaluation of R&D programmes.

It describes in detail the procedure for the transformation of research results into points and the discussion concentrates on the weighing of the different research results and attempts to unify the evaluation. Accounting for disciplinary differences has become a major point of discussion, as this is not easily reconciled with quantitative evaluation.

The document furthermore states, "The chief purpose of the evaluation of results of research organisations is to propose a division of funds for the institutional support of R&D in accordance with the R&D Reform."

Changes and commonalities compared to earlier versions

The introduction of the document outlines the changes and commonalities between the 2009 version and earlier versions:

- A fundamental change in the 2009 version is that the point evaluation of individual results will now be done for the year that the result has been realised, whereas it used to take into account the year in which it was put into the R&D Information System (RIV)
- The 2009 version for the first time splits scientific disciplines into two large groups: For results of disciplines registered in the National Excellence Reference Framework NERR (including Philosophy and Religion, History, Archaeology, Anthropology, Ethnology, Political Science, Administration, Legal Science, Linguistics, Literature, Mass media, and Audiovisuals, Art, Architecture, Cultural Heritage, Educational science), evaluation is carried out on the national level in particular. In these fields, articles published in peer-reviewed Czech scientific periodicals (J_{neimp}) will get a higher point evaluation than in all other disciplines. The same approach will be applied to results of scholarly books (category B).
- However, the point system has remained essentially the same as it was in 2008 (see Figure 4, below), as both have the following general principles:
  - The efficiency of the recipient and provider has not been evaluated
  - Only research organisations that can be recipients of institutional support of R&D and the results are included in the evaluation of results of research organisations
  - The evaluation of results of research organisations according to this Methodology is one of the criteria for the allocation of institutional support
  - All results of research organisations claimed within the past five years will be included in the evaluation regardless of the source of funding.

Figure 4 The 2009 point table

Result types			I- NERR specializations [0]	II – other specializations
J_{imp}	article in impacted magazine [1]		**10 to 305** [2]	
	article in prestigious impacted magazine (Nature, Science, Proc. Natl. Acad. Sci. USA) [3]		**500**	
J_{neimp}	article in non-impacted magazine	world-renowned database [4]	**12**	**8**
		list of critiqued periodicals [4]	**10**	**4**
B	scholarly book	world language [5]	**40**	**40**
		other languages		**20**
D	article in proceedings [6]		**8**	
P	patent	European or international patent (EPO, WIPO), patent of USA and Japan	**500**	
		Czech or national patent excepting patent of USA and Japan, used on the basis of a valid license contract	**200**	
		other patents [7]	**40**	
Z	pilot plant, confirmed technology, species, breed		**100**	
F	usable sample		**40**	
	industrial sample		**40**	
G	prototype, functional sample		**40**	
H	Applied results		**40**	
N, L	certified methodologies and procedures, specialized maps with scholarly content		**40**	
R	software		**40**	
V	research report containing classified information		**50**	

0. NERR includes specializations (according to R&D IS codebook: AA – Philosophy and Religion, AB – History, AC – Archaeology, Anthropology, and Ethnology, AD – Politology and Political Science, AE – Administration – AG, Legal Science, AI – Linguistic Science, AJ – Literature, Mass media, and Audiovisuals, AL – Art, Architecture, and Cultural Heritage, AM – Education and Schools.)
1. publications indicated in the following database Web of Science of the company Thomson Reuters: Science Citation Index Expanded (SCI-EXPANDED) – 1945 – present; Social Science Citation Index (SSCI) – 1980 – present; Arts & Humanities Citation Index (A&HCI) – 1980 – present; Index Chemicus (IC) – 1993 – present; Current Chemical Reactions (CCR-EXPANDED) – 1986 – present
2. evaluation $J_{imp} = 10 + 295 \times$ Factor, where:
   Factor = (1 - N) / (1 + (N / 0,057)), where N is the normalized order of magazine, N = (P - 1) / (Pmax - 1)
   P = order of magazine in the given specialization according to Journal Citation Report in succession ordered descending according to IF
   Pmax = order of magazine in the given specialization according to Journal Citation Report
   In the case that the magazine will be registered to more specializations, the normalized order of magazine N will be calculated as an arithmetic average of normalized orders of the magazine in all specializations where it appears
3. this means multidisciplinary (i.e. open specialization) magazines Nature (ISSN 0028-0836), Science (ISSN 0036-8075) and Proceedings of the National Academy of Science of the USA (ISSN 0027-8424)
4. the distinction between "world renowned database" and "list of critiqued periodicals" is indicated in part B.3.1.2. of the Methodology
5. world language means English, Chinese, French, German, Russian, and Spanish
6. the proceedings must be registered in the database Conference Proceedings Citation Index – Science or Social Science & Humanities (previously ISI Proceedings) of the company Thomson Reuters.
7. Czech or other national patent awarded (hitherto unused) or used by the patent owner.

Source: Evaluation Methodology 2009, Annex 1

Evaluation at different levels

The 2009 Evaluation Methodology covered both the evaluation of the results of research organisations and the evaluation of R&D programmes.

In relation to the **evaluation of the results of research organisations**, the following points should be mentioned:

- The evaluation of the results of research organisations is understood as the translation of the results of all research organisations to a common numerical scale (i.e. quantification of results).
- The evaluation of results is carried out exclusively on the basis of valid data provided to the R&D Information System.
- In the case of universities, the evaluation of results is first done at the level of the institution. It can then be extended to the second level of the university units (departments, faculties, etc). In the case of governmental bodies, the organisational units are evaluated individually.

In the introduction of the document it is first stated that the evaluation of the results of research organisations is merely a way of sorting according to a unified criteria and does not contain any recommendations. According to its goals the evaluation of results constructs a comprehensive set of basic data to inform on the results of research organisations. However, a few lines lower, the document says that the chief purpose of the evaluation of results of research organisations is to propose a division of institutional support of R&D, in accordance with the 2008 R&D reform.

The second part of the 2009 Evaluation Methodology is devoted to the **evaluation of completed research programmes**.

The programme funding body should include in its evaluation basic data on the programme, basic data on the realisation of the programme, indication of results, information on the used methods to obtain the results, a comparison of results with the aim of the programme, and a comparison of the results of the programme with the state abroad. The results of all projects funded by the given programme and archived in the central Information System are assigned to the programme. Furthermore, evaluation should be performed based upon the objectives/aims of the programme.

For each completed programme, the funding authority should develop a summary report on the evaluation and submit it to the R&D&I Council. For this evaluation, the Index SR indicator is used as a characterisation of the effectiveness of the utilisation of public support. This indicator is based on the total score (sum of points) of achieved results per CZK1 million of funding.

Each year, the R&D&I Council publishes a Summarised Assessment of all completed programmes. This report contains an overview of the evaluated programmes, information on the timetable and progress of the evaluation, the results of the evaluation, a comparison among the completed programmes based upon the Index SR, and a summary and proposal of measures.

3.2.2 The 2010 Evaluation Methodology (EM 2010)

After some iterations of the draft version, the Government approved the 2010 Evaluation Methodology on August 4, 2010.

Our analysis indicates that it is *similar to the recent versions* of the Evaluation Methodology; the 2010 Methodology remains mechanistic and based on counting the various types of results according to the point system. The two main differences compared to the former version are an extra chapter dedicated to the allocation of funding to research organisations, as well as a division of research in field groups.

In the EM 2010, the rating of various types of results ranges from 4 to 500 points and the results are sorted into two large groups according to the research fields of their origin. Philosophy and Religion, History, Archaeology, Anthropology and Ethnology, Political Science, Management and Administration, Legal Sciences, Languages, Pedagogy and Arts are included in the National Reference Framework of Excellence (NERR). The results of these fields are evaluated according to the preferential rules.

Results of all other research fields are evaluated according to the same procedure as previous years (see Figure 5, below).

The first **main difference** between the Evaluation Methodology 2010 and previous methodologies is a change in categories of results: a new type of journal article has been introduced named J_{rec}. As a result there are now 4 recognized types of journals:

- Journals monitored (included) in the database WoS Thomson Reuters (J_{imp})
- Journals included in other recognized databases as Scopus or ERIH (J_{neimp})
- Journals in other peer review journals (J_{rec}).
- A fourth class are super journals - Nature and Science.

Figure 5 The 2010 point table

Type of result					I – NERR specializations[0]	II - other specializations
J_{imp}	article in an impacted journal [1]				**10 to 305** [2]	
	article in a prestigious impacted journal (Nature, Science) [3]				**500**	
J_{neimp}	article in a reviewed journal	world-renowned databases [4]	SCOPUS *		**12**	
			ERIH	A	**30**	**12**
				B	**20**	**11**
				C	**10**	**10**
J_{rec}	article in a Czech reviewed journal	list of reviewed journals [5]			**10**	**4**
B	monograph	world language	English, Chinese, French, German, Russian, Spanish		**40**	**40**
		other languages				**20**
D	article in a proceedings [6]				**8**	
P	patent	"European" patent (EPO)**, patent in the US (USPTO) or Japan			**500**	
		Czech or national patent (except US and Japan) in use based on a valid license agreement			**200**	
		other patents [7]			**40**	
Z	pilot plant, authentic technology, variety, stock or breed				**100**	
F	utility model [8]				**40**	
	industrial design				**40**	
G	prototype, working sample				**40**	
H	grantor-realized results				**40**	
N	certified methodologies or processes, specialized maps with expert content				**40**	
R	software				**40**	
V	research report which is a result containing classified information [9]				**50**	

0. NERR includes (as per RDIS catalogue: AA - Philosophy and Religion, AB - History, AC - Archaeology, Anthropology and Ethnology, AD - Politology and Political Science, AE - Management and Administration, AG - Legal science, AI - Linguistics, AJ - Literature, Mass-media and Audiovisuals, AL - Art, Architecture and Cultural Heritage, AM - Pedagogy and Education).
1. Publications listed in the following Web of Science databases of Thomson Reuters: Science Citation Index Expanded (SCI-EXPANDED) – 1945-present; Social Science Citation Index (SSCI) – 1980-present; Arts & Humanities Citation Index (A&HCI) – 1980-present; Index Chemicus (IC) – 1993-present; Current Chemical Reactions (CCR-EXPANDED) – 1986-present.
2. Evaluation of $J_{imp} = 10 + 295 \times$ factor, where:
Factor = (1 - N) / (1 + (N / 0,057)), where N is the normalised rank of the journal, N = (P - 1) / (P_{max} - 1)
P = rank of the journal in the given specialization according to Journal Citation Report in a sequence sorted

Type of result	I – NERR specializations)	II - other specializations

descending by IF. The IF value used is the IF value valid in the year of result realization, not IF-5.
P_{max} = total number of journals in the given specialization according to Journal Citation Report
If a journal is listed under multiple specializations, the normative rank N will be calculated as the arithmetic mean of the normative ranks of the journal for all the specializations where it occurs.
* If the IF is introduced in the SCOPUS database, the results for the following period will be attributed point scores similarly to J_{imp}.

3. These are the multi-disciplinarian journals (journals of open specialization) Nature (ISSN 0028-0836), and Science (ISSN 0036-8075).
4. The world-renowned databases are ERIH category A, ERIH category B, ERIH category C, or SCOPUS.
5. The list of reviewed non-impacted journals published in the Czech Republic including information on origin and validity is available at www.vyzkum.cz.
6. The proceedings must be registered in the Conference Proceedings Citation Index – Science or Social Science & Humanities (formerly ISI Proceedings) of Thomson Reuters (New York, USA).

** EPO - European Patent Office is an IGO founded in 1977 based on the European Patent Convention. The EPO guarantees non-discrimination when inventors or companies submit their application for patent protection in up to 39 European countries.

7. Czech or other national awarded patent, so far not in use or in use by the owner of the patent.
8. Result type S submitted into the RIV before 2008 has a value of 40 points.
9. This only concerns results containing classified information as per special legislation (e.g. Act No. 148/1998, in the wording of later amendments, Act No. 412/2005, in the wording of later amendments).

Source: Evaluation Methodology 2010, Annex 1

The second and most important difference with previous versions is the addition of a chapter in the 2010 Evaluation Methodology that sets out the application of the results evaluation as a Performance-based Research Funding System (further: PRFS) at the level of research organisations.

It introduces a 'damping factor' for groups of disciplines in order to limit large shifts in funding among fields and results categories:

- Between basic and applied research (limit 1.5% change)
- Among 10 large research fields (limit 15%)
- Among various categories of results, with exception of J_{imp} and J_{neimp} (the limit is a 150% change)

This 'damping factor', however, can only slightly mitigate the potential negative effects of adopting the results evaluation as a performance-based research funding system at the level of single organisations. In fact,

- There are still considerable inter-field differences. These differences relate to different publishing patterns among disciplines even within one disciplinary category (e.g. economics and history) and to different relevance of outputs for different types of research organisations (e.g. university and applied research centre, see also survey results (5.1). These differences are not taken into account.
- J_{imp} journals are excluded from the dampening factor. Not only are these the outputs that can least easily be manipulated, they are also the research results that weigh most for the achievement of points in the system. Figure 6, below, illustrates that in 2009, articles published in journals covered by the WoS accounted overall for 65% of the points achieved.

Figure 6 Weight of the R&D results for the achievement of points – results of 2009

Type of result		Percent of total in 2009
J_{imp}	Article in journals covered by WoS	65.1%
J_{neimp}	Article in journals covered by SCOPUS or ERIH (non WoS)	5.0%
	Article in (Czech) peer-reviewed journals listed in the List of Periodicals (nonWoS)	3.5%

B (+C)	Book or Chapter in Book	8.1%
D	Article in Proceedings (included in the ISI Proceedings)	1.6%
P	Patents	1.9%
V	Research report containing secret information	0.02%
T*	Trials, Verified technologies, prototypes etc.	2.5%
Z	Trial operation, Verified technology, Variety, Breed	2.7%
S**	Prototype, certified (applied) method, functional sample, authorized software, utility model, industrial design	5.9%
G	Prototype, Functional Model	1.6%
N	Certified Methodology	0.5%
F	Utility Model	0.3%
L	Specialised Maps	0.2%
R	Authorised Software	1.0%
	Total	**100.0%**

* Category valid until 2006; substituted by categories S and Z

** Category valid in 2007 and 2008; substituted by new categories in 2009

Source: Technology Centre

3.3 Funding streams at public universities

The Evaluation Methodology does not allocate 100% of funding. In this section, we present an overview of the funding streams existing in Higher Education Institutions (HEI), thus putting the Evaluation Methodology in a broader context.

3.3.1 Main HEI budget items

Public higher education expenditure predominantly supports public universities, which form more than 90% of all HEIs in the Czech Republic. They comprise more than 90% of students and the vast majority of their revenues come from public sources (over 90% of all income at 17 of the 24 Czech public HEIs). Other income sources of the public universities include property revenues, services to students, extra teaching activities, R & D activities and study related fees.[18]

Essentially the funding of HEIs comes from two different budgets: the budget for R&D and the budget for higher education/teaching. In 2010 and 2011, institutional funding for R&D made up around 20% of total HEI budget (see Figure 7). In contrast, institutional funding at the ASCR was 52% of total ASCR budget in 2010[19].

[18] OECD report: Pabian, P.; Melichar, M. and Šebkova, H. (2006) Funding Systems and Their Effects on Higher Education Systems, COUNTRY STUDY – CZECH REPUBLIC, available online at www.oecd.org/dataoecd/53/9/36765218.pdf

[19] Annual report of the ASCR 2010. Total budget including ASCR's own resources from activities related to licensing etc.

Figure 7 Higher education funding for years 2010 and 2011

Funding type	2010	2011	2010	2011
Teaching-related funding	21,615,242	20,686,641	81.1%	79.6%
R&D related funding / non-competitive	5,033,617	5,285,548	18.8%	20.4%
Research Intentions (Institutional funding)	2,624,073	1,705,021	9.8%	6.6%
Development of Research Organisations (Evaluation Methodology) (Institutional funding)	1,471,941	2,675,771	5.5%	10.3%
Specific university research (within Targeted Funding)	937,603	904,756	3.5%	3.5%
Total	26,648,859	25,972,189	100%	100%

Source: Breakdown of the budget of HEIs in 2011, MEYS

As shown in Figure 7 above, non-competitive funding of R&D in the institutions is based on the following budget items:

- Research Intentions (Výzkumné záměry) included in the Institutional funding budget
- Development of Research Organisations (Rozvoj výzkumných organizací) in the Institutional funding budget. This is the funding based on the Evaluation Methodology ("coffee mill"). Funding allocated based on the Evaluation Methodology accounted for 10.3% of HEI funding in 2011, as compared to 5.5% in 2011 (+ 81.8%), reflecting the transition from allocation based on Research Intentions to allocation based on the Evaluation Methodology.
- Specific university research funding (Specifický výzkum) currently included in the Targeted Funding budget (while previously in the institutional budget). Specific university research is defined in the Act on the support of research and development (Act number 130/2002 Coll.) as "research performed by students as part of accredited doctoral or masters study programmes and which is directly related to their studies". It is therefore research that focuses on concrete tasks related to deepening knowledge that involves students.[20] Specific university research funding is allocated based on a formula, which uses EM results as one indicator.[21]

3.3.2 Higher Education teaching-related budget

In 2010 and 2011 respectively, the teaching-related funding formed 81.1% and 79.6% of the overall university funding (Figure 8). The main teaching-related budget item is the **normative budget part** that forms the largest share of teaching related-funding. Provided to the university as a lump sum, the normative part of the HEI budget is based on a formula, taking account of the scope and content of the educational activity of the HEI, as judged by the number of study programmes (80%)[22], graduates (10%)[23],

[20] Specific university research however does not include those student research activities that are funded through research intentions, projects of the Czech Science Foundation, development fund of higher education institutions and others. MEYS provides funding for specific university research to universities who ran an internal tender for such projects and apply for such funding

[21] Other indicators are number of Master and Doctoral students and Master graduates, number of professors and associate professors. The indicator 'external funding' was replaced by points obtained in the Evaluation Methodology in 2010, see e.g. OECD, Performance-based funding for Public Research in Tertiary Education Institutions: Country experiences, Paris, 2010

[22] **A+B1**:"Study programmes"= Normative number of financed students x normative base. Number of financed students = "special students"(overstudying) x 0 + newly admitted students x 1.0 + half year students (those who will start to over study in half year) x 0.5 + other students x 1.0. Approved number of financed students used for the calculation is then limited and agreed between the ministry and the university. Normative number of financed students = Sum of Approved numbers of financed students in each category (based on how financially demanding it is) multiplied by the financial coefficient of that

and qualitative outputs (10%). The qualitative outputs encompass scientific performance (50%), based in equal parts on points obtained in the Evaluation Methodology and targeted funding for research. Other criteria are qualification of teaching staff and doctoral studies (20%)and internationalisation and international cooperation (30%).

The second teaching-related budget item is the **social student affairs part.** This part of the budget consists of doctoral students' scholarships[24], grant for students' meals[25], students' social grants[26] and grant for students' accommodation[27]. All of these budget items are calculated based on formulas.

The third teaching-related budget item is the **development of higher education institutions**, which includes educational development projects of the University Development Fund and the Development Programmes.

- Grants for development projects are awarded to universities that won a tender defined and announced by the University Development Fund Committee. The projects aim to creatively develop educational activities at universities. Eligible applicants for these projects are researchers and students at universities. In 2011 there were two categories of projects. The investment projects supported mostly modernisations of laboratories and other systems, while the non-investment projects mainly funded the development and improvement of teaching modules, methodologies and other education-related activities.[28]
- The objective of the development programmes is to contribute to the achievement of priorities set in the Long-Term Plan of the Ministry. There are two types of programmes within this funding mechanism - an institutional development programme based on a institutional development plan submitted by the university to the ministry and centralised development programmes usually undertaken by consortiums of universities. Thematic priorities for the year 2012 are for example support for sharing capacities and creation of university networks in the Czech Republic.[29]

The fourth and last teaching-related budget item is the **International cooperation and other.** This part of the budget forms a relatively small part of the budget and consists of a combination of formula-based and non-formula based funding mechanisms. Specific items within this part of the budget include:

- International cooperation: grants provided for fulfilment of the obligations resulting from international agreements that do not include indicators already

---

particular programme. Normative base = Cost of study programme per one student in accredited study programme with financial coefficient of 1.0 specified by the ministry.

[23] **B2:** bonus for graduates = [nr. of Bachelor graduates x 1; number of Master graduates x 1.5; number of post master graduates x 0.5 (bonus base x financial coefficient of study programme)] + [nr. of PhD graduates x (bonus base x financial coefficient of study programme) x 2 if finalised within 4 years, every half year after that reduces the coefficient by 0,25]+[bonus for graduates that do not continue in the study is received for every student that does not study 2 years after finishing bachelor degree, funding for such students is multiplied by coefficient of 2]+[employability of students finishing in the last 3 years].

[24] Number of on site PhD students not studying more than one year above standard study time (no restrictions on enrolment) multiplied by the set funding per student. HEIs set their own criteria for the distribution of the grant to the PhD students

[25] The total sum is allocated to HEIs according to the number of meals they distributed over last year

[26] Means-tested grants to students whose family income meets criteria in paragraph 91 of the Act on higher education. Introduced in 2006

[27] The financial base multiplied by the number of students residing outside the district where they study; redistributed to students according to criteria set by individual HEIs

[28] A share of this funding is provided to the Agency of the Council of HEIs at Charles University

[29] Similarly to the University Development Fund a share of this funding is provided to the Agency of the Council of HEIs at the Charles University

included in the normative part. The level of support is defined based on materials drafted by the relevant national agency offices.

- Fund for educational policy: funding for securing educational activity of new universities and support for projects and activities not financed already from normative funding. This stream of funding can also be used for activities that support developmental intentions of a university or the ministry according to the long-term intention of a university or the long-term intention of the ministry. Generally, this budget item will be used for one-year educational projects. The decision on provision of this support is based on an application by the HEI to the deputy minister of education.
- Education policy fund/extraordinary activities: funding for extraordinary activities such as projects to support inter-institutional cooperation or projects of special interest to particular institutions and the ministry of education. Unpredictable expenditures of HEIs (for example resulting from changes in law).

Most of the formulas used to allocate budget are based on input indicators (volume indicators) but output oriented indicators – most importantly points obtained in the Evaluation Methodology – have increased with the new items in the normative part of the budget (see Figure 8).

Last but not least, given the variety of funding streams and the bodies deciding upon them (Figure 8), the question arises whether governance of universities is not very difficult.

3.3.3 Share of funding allocated based on the Evaluation Methodology

In 2010, the total share of HEI budget that was determined by the Evaluation Methodology (Development of Research Organisations within Institutional funding) equalled 1.47 billion CZK, which is 5.5%. In addition, 50% of the B3 part of normative funding was based on the Evaluation Methodology, corresponding to 815 million CZK or 3% of total HEI budget. Therefore, the total share determined by the Evaluation Methodology was 8.5% of total HEI budget.

In 2011, the budget item (Development of Research Organisations within Institutional funding) rose to 2.68 billion CZK (10.3%, as compared to 5.5% in 2010), while 25% of the B3 part of normative university funding was based on the Evaluation Methodology. This represented 413.8 million CZK (1.6%) of the overall HEI budget. Therefore, in 2011, the total university budget determined by the Evaluation Methodology was nearly 12% of the overall HEI budget[30] and will continue to increase with funding based on Research Intentions running out.

Likewise, in 2010 at the ASCR the share of funding determined by the Evaluation Methodology was 7.9% of the total ASCR budget[31] (8.5% at universities). Again, this share is bound to increase, all the more so as institutional funding makes up around 50% of the Academy's total budget.

The percentage of funding allocated through the Evaluation Methodology may seem small. However, the Evaluation Methodology is increasingly applied to other funding streams, and it is the part of funding with the highest relevance for institutional development. Hence, its allocation should be based on sensible principles.

[30] This calculation excludes funding in the specific university research stream, which is also partly based on Evaluation Methodology points. Hence, the total university budget determined by the Evaluation Methodology is somewhat higher than 12%

[31] Sources: Annual report of the ASCR 2010 (total budget) and Approved R&D budget for 2010 (level of funding determined by Evaluation Methodology). Because the information in the annual reports do not precisely correspond to the information in the R&D budget, the percentage is subject to error

Figure 8 Composition of public expenditures on public HEIs – Funding and shares group per 'OKRUH'

	Budget item label	Formula/Non-formula	Indicators used	Decision maker	HEI Budget 2010 ('000s CZK)	Share of 2010 HEI Budget	HEI Budget 2011 ('000s CZK)	Share of 2011 HEI Budget	Annual change
1. Normative part of the budget	Teaching Activity	formula	input / output	MEYS	17,391,341	65.3%	16,598,609	63.9%	-4.6%
2. Social student affairs	Doctoral Students Scholarships	formula	input	MEYS determines the base	1,013,279	3.8%	1,054,455	4.1%	4.1%
	Grant for Students' Meals	formula	input	MEYS determines the total sum	217,770	0.8%	198,340	0.8%	-8.9%
	Students' Social Grants	formula	input	Cabinet	90,081	0.3%	60,503	0.2%	-32.8%
	Grant for Students' Accommodation	formula	input	MEYS	1,010,507	3.8%	970,301	3.7%	-4.0%
3. Development of HEIs	Higher Education Development Fund	non-formula	-	Jointly managed by MEYS and Council of HEI	334,000	1.3%	319,638	1.2%	-4.3%
	Development Programmes	non-formula	-	MEYS upon recommendation of Development Programmes Steering Committee	1,131,885	4.2%	1,083,213	4.2%	-4.3%
4. International Cooperation and other	International Cooperation	mixed	input/output	MEYS/cabinet	311,379	1.2%	310,082	1.2%	-0.4%
	Fund of Educational policy	non-formula	-	MEYS	100,000	0.4%	81,500	0.3%	-18.5%
	Education Policy Fund / Extraordinary Activities	non-formula	-	MEYS	15,000	0.1%	10,000	0.0%	-33.3%
R&D-related funding	Research Intentions within Institutional funding budget	non-formula	-	MEYS upon recommendation of expert committees	2,624,073	9.8%	1,705,021	6.6%	-35.0%
	Development of Research Organisations within Institutional funding budget (Evaluation Methodology)	formula	output	R&D&I Council	1,471,941	5.5%	2,675,771	10.3%	81.8%
	Specific research currently within Targeted Funding budget	formula	input/output	Cabinet upon joint recommendation of MEYS and RD&I Council	937,603	3.5%	904,756	3.5%	-3.5%
Total					26,648,859	100%	25,972,189	100%	-2.5%

Source: Breakdown of the budget of HEIs in 2011, MEYS

3.4 Summary and first conclusions

Our summary and first conclusions are based on an in-depth analysis of the Evaluation Methodologies 2004, 2009 and 2010 Methodology. In the last 5 years, the policy approach to the evaluation of R&D saw the following fundamental changes:

- The 2004 Methodology introduced 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. The 2009 version marks the adoption of the metrics-based evaluation of R&D results as a **Performance-Based Research Funding System -** albeit only at the level of funding bodies. The 2010 Methodology explicitly recognizes the possibility to use the metrics-based evaluation of R&D results as a mechanism for allocating institutional funding to individual research organisations. Currently, the Evaluation Methodology has therefore a two-fold role. Of course, there is a close link between these two roles: at its 'practical level', the Evaluation Methodology defines what results are eligible, how the data are collected and how they are converted into point values, which then form the basis to allocate institutional funding for R&D,
- A progressive restriction of the **scope of the evaluation guidelines** can be noted: while the 2004 Methodology was characterised by a broad outlining of a comprehensive Evaluation Guideline covering all the different layers of the research system, the 2009 and 2010 Methodologies focus almost exclusively on the quantification of research outputs for the evaluation of research organisations and research programmes
- There is a substantial restriction also in the **scope of the envisaged evaluations** themselves: the 2004 Methodology stresses the importance of evaluating research programmes in terms of their results and effects in the socio-economic spheres (and taking into account their specific socio-economic objectives); the 2009 and 2010 Methodologies focus *exclusively* on the counting of R&D outputs. The implementation of ex-post impact analyses is never mentioned - neither in the Methodology nor in other policy documents
- Curiously, the Czech Evaluation Methodology seems to have been developed in almost complete **isolation** from developments going on in other countries. Other countries have been struggling with the same questions as the Czech Republic, and most have come up with different answers from the Czech Republic. Hence, important differences can be noted between the current Evaluation Methodology in the Czech Republic and international practice. We will discuss these in the sections to come.

4. The Evaluation Methodology as a Performance-based Research Funding System (PRFS)

In this section we focus on the use of the metrics-based results evaluation system as a Performance-based Research Funding System.

In a first stage we sketch the policy background to this additional role of the evaluation methodology. Secondly, we set the PRFS in the Czech Republic in the international context, describing key characteristics of PRFSs and their typical effects. We then report on our preliminary analysis of the effects on fields of research in the Czech Republic and relate on the outcomes of an expert panel study on practices for assessing university-based research, commissioned by the European Commission. Finally, we draw some conclusions.

4.1 Policy Background to the PRFS in the Czech Republic

4.1.1 The Policy Context

As mentioned in the preceding section 3.1, the concept that funding allocation should be guided by performance and results achieved had already been voiced in the 2004 Evaluation Methodology and the National R&D Policy 2004 -2008. In these policy documents, however, the R&D results were considered to be only one of the relevant factors – albeit at times a critical one.

The shift occurred in the 2008 Reform of the R&D&I System where it was stated, "the institutional support to departments will be allocated **at the level of budgetary chapters** according to the results of research organisations within their competence achieved over the past five years." In other words, the budgets made available to the institutions responsible for institutional funding (the Ministries and the Academy of Sciences) would be determined by the aggregated R&D results of the research organisations in their field of competence.

As a rationale for this change in the budget allocation, the 2008 Reform stated "*The allocation of institutional funds among the budgetary chapters cannot stem from only the amount of funds that were allocated in the previous year, but must be based on **more objective criteria**, which are the results that have been achieved in the past five years. Unlike the targeted (project) support directed at the attainment of set targets, which is therefore directed at the future, the institutional support is intended to produce a long-term conceptual development of research organisations and must bring along corresponding results, too (regardless of the source, from which they were financed).*"

Apparently, the intention of the 2008 Reform was not that the metrics-based results evaluation would determine institutional funding at the level of the individual institutions. In fact, the policy document stated *"Within the respective budgetary chapters, it will be possible to modify the allocation of funds on the basis of a more detailed evaluation of research organisations using internationally recognised methodologies, the results of which will be published."*

However, all Ministries declined this offer and opted for the application of the metrics-based results evaluation to also be used for the distribution of institutional funding at **the level of individual institutions**. Only the Academy of Sciences is currently organising an alternative internal evaluation.

It can be envisaged – and interviewees have confirmed – that the results evaluation will also be applied at lower levels within the institutions concerned, e.g. within universities at the level of faculties and departments, and even down to the individual researcher. In this context it is telling that the Charles University felt the need to mitigate the effects and foresees – internally to the university - the attribution of extra points and additional funds to books as a research result.

One should therefore note that when indicating the modalities for the allocation of institutional funding at institutional level, the 2010 Methodology actually reacts upon an accomplished fact.

4.1.2 From 'Research Intentions' to 'Long-term Conceptual Development'

The change in policy-making in relation to institutional funding did not only regard the more pronounced focus on the achievement of results as objective criteria for funding allocation; policy-makers also stepped away from the concept of 'Research Intentions'.

Essentially, policy-makers in 2008 substituted a radically prospective funding system (the Research Intentions) with a radically retrospective funding system (the achievement of results in the past).

The Research Intentions scheme, launched in 1998, is generally perceived to have had some very positive attributes. Most important, it provided a measure of stability and security for research organisations. In addition, organisations had considerable freedom to design the Research Intentions according to their needs and wishes. In other words, institutes with good plans could do a lot. However, the Research Intention scheme also gave rise to criticism. As Research Intentions cover long periods (5-7 years), this can restrict progress or development within institutions because they are locked into the programmes that are defined in the intentions. Moreover, the Research Intentions divided university research staff into those 'on board' (being members of the Research Intention team) and those 'left behind' (not having the Research Intention). Even if someone's research results were excellent, it was impossible to become a member of the Research Intention plan if the university did not have a Research Plan in the particular field. This was perceived as very unfair by many university researchers. In addition, just like the Evaluation Methodology (see 5.2.1), the Research Intentions gave rise to 'gaming' of some team leaders. Finally, the evaluation of the Research Intentions was also criticised, which was perceived as not very objective: almost all Research Intentions were evaluated as 'excellent' or 'above average'.

The evaluation of the Research Intentions scheme voiced in the 2008 Reform document was that *"The introduction of a research plan-based institutional financing did not bring the necessary dynamics into the organisational structure of public research. While institutes abroad in this area are born and die, their focus (research programmes or plans) considerably changes, their management radically changes, etc., the changes in the Czech Republic are basically negligible."*

The 2008 Reform document considered that the decision no longer to tie institutional funding to research intentions would result especially in *"the transfer of decision-making in funds for long-term development of organisations from the level of ministries (where decisions were taken in funds and often in the research plans orientation) to the organisations themselves." In essence – or at least in theory - the research organisations are given full autonomous decision power. The 2008 Reform states: "Whether the organisation uses these funds to attract (or retain) skilled workers, investments into new apparatuses and equipment, cooperation with other organisations, etc. will depend on the organisation itself with only one single condition – money must be used on research and development and bring results."*

4.2 Performance-Based Research Funding Systems in the International Context

4.2.1 Key characteristics of PRSFs

There is growing international interest in performance-based research funding systems. Performance-based models have been implemented in the UK, Spain, Slovak Republic, Hong Kong, Australia, Poland, Italy, New Zealand, Flanders, Norway, Denmark, Finland and Austria. They operate **at different levels**, for example, the Spanish system measures and rewards individual performance. The UK and Hong Kong RAEs have 'units of assessment' that correspond more or less to departments or research groups. Most PRFSs operate at the level of institutions. Unlike in the unique case of the Czech Evaluation Methodology, they are generally applied to one type of institution only, most often to public universities, i.e. they are used for allocating funding to a relatively homogeneous group of organisations.

Intervals between successive allocation exercises vary among countries. In the Czech system, the decision about budget allocation is updated annually, taking into account the past five years of research results. This means that funding in 2011 will be based on the evaluation of results in 2009 covering results published or produced from 2004 to 2008; funding in 2012 will be determined by the evaluation of results in 2010, covering results of the time period 2005-09. For comparison, the Austrian universities receive their block grants every three years.

Research evaluation processes tend to focus on four output measures: **volume; quality; impact; and utility**[32]. The earlier systems tend to be heavily based upon peer review but – as in the Czech Republic – there is growing interest in indicator-based systems. However, while the use of indicators in the Czech Republic appears to be driven by a desire to de-politicise and de-personalise the evaluation and funding process, elsewhere the interest in indicators reflects a desire to simplify and reduce the cost of assessment. Even among indicators-based systems, **cost** is a major consideration. While Norway chose to establish a national system of grading journals and to require researchers to input their publications into a central database, Sweden has opted to focus on ISI journals and ISI-derived indicators in order to put a system in place more quickly and economically than was the case in Norway[33].

In performance-based research funding, scientific publications and other research outputs typically make up a **small part of the funding formula** only. In Norway, publication is only one of four indicators that drive institutional research funding, the others being: PhD production; EU research funding; research funding from the Research Council of Norway.[34] In Finland, scientific publications make up only 5% of formula-based core-funding for research and researcher education. The other indicators used are research person-years, total number of doctoral degrees determined in the performance contract between the Ministry and the university, number of doctoral degrees completed at the university, Academy of Finland research funding, Tekes research funding, international competitive research funding, and researcher mobility (see case study on Finland in section 6.5). Where research outputs are used, far fewer types of outputs are used in comparison to the Czech System, most probably because the funding system is used for one type or organisations with rather homogeneous missions and targets. The Austrian funding formula even does without any of the research outputs used in the Czech Republic.

Finally, the **percentage of funding** that depends on the research evaluation is a key feature of performance-based research funding systems:

In the Czech Republic, the metrics-based results evaluation as a Performance-based Research Funding System (in principle at 'budget chapter' level) was planned to be implemented in various phases, predominantly depending on the conclusion of the Research Intentions. As is illustrated in Figure 9, in 2010 the results evaluation ('RVO-Hodnocení') will determine the full institutional funding for those research organisations that concluded their Research Intentions in 2009 ('Ukon□ení VZ') and 1/3 of the institutional funding for those who concluded their Research Intentions in 2010 or will conclude them in 2011 provided they were launched in 2005 or 2006 ('Zahájení VZ').

Without going into further detail, the table below illustrates that according to this original timetable, in 2011, for the majority of the research organisations the institutional funding is determined by the new PRFS – if not entirely at least for 2/3.

[32] Aldo Geuna and Ben Martin, "University research evaluation and funding: An international comparison," *Minerva*, 41, 2003, pp27-304

[33] Håkan Carlsson, "Allocation of research funds using bibliometric indicators `- asset and challenge to Swedish Higher Education Sector," *InfoTrend*, 64 (4), 2009, 82-88

[34] Gunnar Sivertsen, "A performance indicator based on complete data for the scientific output at research institutions," *ISSI Newsletter*, 6 (1), March 2010; A Rodrigues-Navarro, "Sound research, unimportant discoveries: Research, universities and formal evaluation of research in Spain," *Journal of the American Society for Information Science and Technology*, 60 (9), 1845-858

Figure 9 Introduction of the Performance-based Funding System, 2010 - 2012

Zahájení VZ	Ukončení VZ	2010		2011		2012	
		RVO (Hodnocení)	VZ („CEZ“)	RVO (Hodnocení)	VZ („CEZ“)	RVO (Hodnocení)	VZ („CEZ“)
2005	2009	1	0	1	0	1	0
2006	2009	1	0	1	0	1	0
2004	2010	1/3	2/3	1	0	1	0
2005	2010	1/3	2/3	1	0	1	0
2005	2011	1/3	2/3	2/3	1/3	1	0
2006	2011	1/3	2/3	2/3	1/3	1	0

Zahájení VZ	Ukončení VZ	2010		2011		2012	
		RVO (Hodnocení)	VZ („CEZ“)	RVO (Hodnocení)	VZ („CEZ“)	RVO (Hodnocení)	VZ („CEZ“)
2007	2011	0	1	0	1	1	0
2007	2012	0	1	0	1	1/3	2/3
2007	2013	0	1	0	1	1/3	2/3
2009	2013	0	1	0	1	1/3	2/3
2009	2013	0	1	0	1	1/3	2/3

Source: Approval of the proposal for state budget expenditure on R&D for the year 2010 with view on the years 2011 and 2012

A precise calculation of the amount of overall funding that would be influenced by the PRFS in 2010, 2011 and even 2012 looks like a titanic exercise. In the same year, a rough calculation of the total 'basic' national public funding for universities (teaching funding + institutional funding) and for the Academy of Sciences (investment funds and institutional funds) leads to the result that overall, institutional funding accounts for **approximately 30% of the total 'basic' national public funding** (including approximately 80% of the public funding of the Academy).

A consistent finding by analysts is that PRFS typically move **small** amounts of money around each time they are performed [35] - in strong contrast to the situation in the Czech Republic. In terms of shares of the total 'basic' national public funding, for example, in 2008 the PRFS in Australia and New Zealand governed 10% of university funding, in Italy and Norway 2%, and in Sweden 12.5%. Alternatively, as the case studies show, there typically are 'correcting factors' to maintain stability in the system. The UK Higher Education Funding Council operates a 'moderation fund' to help institutions cope with changes, while both Austria and Finland have maximum upper limits, to limit budget changes (especially reductions) for universities.

4.2.2 The broader context of performance-based funding systems

There are some common themes running through the rhetoric around the introduction of performance-based research funding systems. The Czech Evaluation Methodology is no exception. Of course, all systems are seen as a means for selectively distributing research funds. But most also seek to use it to drive particular behaviours, most commonly an improvement in the quality of research undertaken; or to increase accountability on the expenditure of taxpayers' money.[36] These two prominent aims can also be seen in a broader context.

Scholars and practitioners mainly concerned with understanding research and innovation tend to speak of globalisation, competitiveness, and the knowledge

[35] Sivertsen, *Op Cit*

[36] Linda Butler, Impacts of Performance-Based Research Funding Systems: A review of the concerns and evidence, OECD, 2010

economy as rationales when explaining the introduction of national research evaluation systems.[37] An appropriate statement can be found on the UK government website introducing the new Research Excellence Framework (REF):

> *Through the REF, the UK funding bodies aim to develop and sustain a dynamic and internationally competitive research sector that makes a major contribution to economic prosperity, national wellbeing and the expansion and dissemination of knowledge.*[38]

Similarly, in Finland the explicit aim of the University Act 2009, which grants universities more administrative and financial autonomy, has been to create level playing fields in the competition with the best universities worldwide. To this end, universities have been given the economic and administrative means for enhancing the quality and impact of research and teaching and for participating in international cooperation. Moreover, research and researcher training have increased in importance in the Finnish funding formula, in an attempt to develop research universities and to increase research-based competitiveness. Ultimately, the PRFS are about increasing the quality of research undertaken, in order to improve the competitiveness of universities and the economy as a whole. This is also one of the aims of the Evaluation Methodology in the Czech Republic. The (large) weight the EM gives to applied research outputs testifies to the political goal to improve the 'valorisation' of research.

In contrast, scholars and practitioners who primarily study higher education and who are concerned with education, accreditation and the like tend to trace the introduction of performance-based research funding systems to the spread of **new public management (NPM)**[39,40]. A core idea of NPM is to enhanced accountability on the expenditure of taxpayers' money. This has certainly been the case in the UK with the research assessment.

> *The move to quality assurance and assessment was to assume an increasingly important role in (the then) government's determination to impose some of the disciplines of the market on higher education, including competition, an increase in the power of consumer demand and the concept of universities as well-managed corporate enterprises. The broader policy context was that government also expected that resources should be managed with maximum efficiency and accountability, if necessary creating structures for effective decision-making, transparent measurement of performance and ultimately for resource allocation with reference to that performance.*[41]

Enhanced accountability means focusing on outputs and outcomes rather than inputs, processes and structures. Measuring research output and distributing funding based on the results is clearly meant to enhance accountability and to increase productivity. This is also called **management by results**. The whole Finnish public administration has been reformed along these lines in the last 25 years, starting with the higher education system. In the Czech Republic, enhanced accountability seems to have been a particular concern, along with improved transparency and a desire to de-

---

[37] Diana Hicks, Overview of Models of Performance-Based Research Funding Systems, OECD, 2010

[38] http://www.hefce.ac.uk/research/ref/ (11/1/2011)

[39] New public management is a management philosophy used by many governments since the 1980s to modernise the public sector. The main hypothesis in the NPM-reform wave is that more market orientation in the public sector will lead to greater cost-efficiency for governments, without having negative side effects on other objectives and considerations. One of the first countries to reform the public service along the ideas of new public management was New Zealand.

[40] Diana Hicks, *Overview of Models of Performance-Based Research Funding Systems,* OECD, 2010

[41] Valerie Bence and Charles Oppenheim, 'The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions', in Journal of Educational Administration and History, 2005, 37 (2), 137-155

politicise and de-personalise evaluation and funding. The Evaluation Methodology seemed to be the solution, offering (seemingly) objective criteria for the allocation of institutional funding and embodying an increased focus on concrete R&D results.

However, the shift to performance based funding is part of a broader movement to make universities more autonomous and introduce more strategic university management. This shift can be seen in all the international case studies and, of course, also in the Czech Republic, where the introduction of the Evaluation Methodology represents a shift towards greater autonomy of institutions by replacing research intentions with a performance-based research system.

4.2.3 Effects of Performance-Based Systems

In principle, performance-based research funding systems have important benefits

- The institutions have stronger incentives to facilitate research for their researchers
- Research is now perceived as a common and institutional responsibility not only as an individual task
- New publications receive attention not only from external peers but also internally from the institution
- Research management improves with the aid of bibliometric information about the research activities[42]

Figure 10 Advantages and drawbacks of performance-based university research funding

Advantages	Drawbacks
• Performance-based – 'meritocratic' in that it links resources to performance, rewarding good research • Strong incentive to improve individual as well as institutional performance • Competition may lead to increased efficiency – ineffective research identified and cut • Encourages research to be properly completed and written up for wider dissemination • Provides public accountability for government funds invested in research • Encourages more explicit/coherent research strategy on part of department or institution • Provides mechanism for linking university research to government policy (e.g. to shift priorities) • Concentration of resources may enable best departments to compete with world leaders (e.g. in US)	• High cost and labour intensity (whether peer review or indicator-based) for universities and evaluating agencies • May cause 'homogenization' of research and universities – i.e. decrease in diversity and experimentation • May discourage more innovative and risky research • Encourages 'publication inflation' (e.g. 'salami publishing') and other 'game playing' (e.g. with indicators) – i.e. 'looking good' rather than necessarily doing better • May encourage traditional 'academic' research at expense of research linked to society's needs • Tends to separate research from teaching, implying lower priority for teaching • Rewards past performance not current or future potential • Reinforces research elite/status quo – may cause overconcentration • May lead to excessive government influence/'interference' in university

Source: Aldo Geuna and Ben Martin, 'University research evaluation and funding: An international comparison,' Minerva, 41, 2003, pp27-304

PRFS tend to attract academic opposition. No one likes to be measured, especially if the measurement has consequences. There are nonetheless a number of serious objections to the type of systems that have been put in place in recent years. Figure 10 lists the main advantages and drawbacks of such systems.

[42] Sivertsen, *Op Cit*

The novelty of such systems, however, means that there is so far a limited amount of **evidence** about their effects. The UK system is the oldest system and hence the best researched system while, for instance, the Austrian system which has been implemented in 2007 for the first time, is just being evaluated at the time of this audit. The UK RAE has clearly increased the quality of UK university research. It has also encouraged the UK universities to take a rigorous approach to developing and implementing their own research strategies. It has also enabled the government to maximise the research return for limited funding[43]. The reason for the RAE's success in these respects is that there was a gap of several years between successive exercises, allowing time for the system to adjust to the expectations of the RAE and for individual researchers to change their behaviour[44].

Performance-based funding can have surprisingly **large effects on collective behaviour**. This is not only a response to potential funding changes but also reflects the role of rankings and grades from performance systems as indicators of esteem. Australia introduced a simple and mechanical system based on publication numbers in 1995. A study of aggregate publication data, complemented by case studies at two universities, indicates that this resulted in an increased number of publications – indeed, Australia's contribution to the Science Citation Index increase by 25% through the 1990s. However, researchers systematically shifted their output towards lower impact factor journals, in order to achieve greater publication numbers, leading to a decline in Australia's relative citation impact in the same period[45].

Norwegian university researchers have significantly increased the volume and quality of their output since an indicator-based system was introduced, despite the fact that it only affects 2% of total university funding[46]. The Norwegian experience is also that self-reporting can lead to significant error: in a spot check of the 2006 submissions, it was found that **half** the publication points credited to one university were wrong, mainly due to false reporting of textbooks as books but also to other inaccuracies[47].

Field-specific factors

The design of performance-based funding systems is sensitive and contentious. Geuna and Martin note that the process of consultation on the UK RAE led to successive refinements – to a point where the resulting complexity led to pressure to move to an indicators-based approach[48]. Such an approach has now been piloted and subsequently rejected, as a result of which the latest RAE has been postponed by a year in order to allow time for better methods to be developed.

Systems that rely solely or heavily on output indicators are especially vulnerable to inducing perverse effects. Some fields (especially in the humanities) emphasise publish in monographs or books; others (notably the basic 'hard' sciences) in journals. Applied scientists and engineers often communicate more via conference proceedings than through learned journals or, especially engineers, in journals not listed in the Web of Science. Mathematicians write few but extensive articles; chemists produce many, short articles; and so on.[49] Performance-based research funding systems which

---

[43] T Clark, *OECD Thematic Review of Tertiary Education, Country Report United Kingdom*, Paris: OECD, 2006

[44] J Taylor and R Taylor, "Performance indicators in academia: An X-efficiency approach," Australian Journal of Public Administration, 62(2), 71-82; cited from Nicoline Frølich, *The Politics of Steering by Numbers: Debating Performance-Based Funding in Europe*, Report 3/2008, Oslo: NIFUSTEP: 2008

[45] Linda Butler, "Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts," *Research Policy*, 32 (2003), 143-155

[46] Sivertsen, *Op Cit*

[47] S Hernes, "Publiseringspoeng og referensekvalitet", *Forskerforum*, 6/2007

[48] Geuna and Martin, *Op Cit*

[49] Similarly, citation practices differ widely. Consequently, citation indices have limited informative value as indicators of research quality.

use publication as an indicator need to take account of the **major differences in 'propensity to publish' among fields.**

The UK RAE and a number of others achieve this by not putting different fields in competition with each other. Instead, similar departments compete within about 60 (the number varies from RAE to RAE) 'units of assessment' so that the RAE rewards quality **within** disciplines but do not cause competition **among** disciplines. The Swedish system uses 34 different disciplinary categories. In Austria, indicators that are equally relevant within a category of organisations, i.e. the public universities, and thus 'discipline neutral' are chosen for the formula budget.

Like the Czech one, the Norwegian system sets the disciplines against each other and tries to compensate for differences in publication behaviour by providing different weights for difference performances. There are nationally defined lists of 'Level 1' and 'Level 2' (higher) publications, with books and monographs being weighted at several ties the value of an article in each case. The committee that designed the funding system argued that it was not possible to take better account of differences in publication behaviour, even using bibliometric methods[50].

Figure 11 Effects of field adjustment on Swedish university production, 2005-7

University	Papers published (A)	Field-adjusted papers published (B)	Ratio (A)/(B)
Lund	4,193.5	3,571.9	1.17
Uppsala	3,743.0	3,193.2	1.17
Gothenburg	2,485.5	2,291.2	1.08
Karolinska	4,153.5	3,164.3	1.31
Stockholm	2,091.0	1,934.7	1.08
Umeå	1,622.0	1,411.3	1.15
SLU	1,571.0	1,316.9	1.19
KTH	2,482.5	1,990.0	1.25
Linköping	1,588.5	1,423.0	1.12
Chalmers	1,767.5	1,388.5	1.27
Luleå	416.5	370.0	1.13
Mid Sweden	218.0	214.5	1.02
Örebro	213.0	220.2	0.97
Karlstad	199.5	188.3	1.06
Växjö	135.5	140.7	0.96

Source: Sandström and Sandström, *Op Cit*; own calculations

While the bibliometrics profession tends to be fastidious about making comparisons **within** fields, precisely because of inter-field differences in publication behaviour, propensity to publish is rarely quantified. One reason for this is that a proper comparison would need to take account of the numbers of researchers who do not publish at all – and they, of course, are not visible in the bibliometric databases. Sandström and Sandström[51] have developed a method for estimating the size of the non-publishing population and used this as a basis for the field-normalisation technique used in the new Swedish system. They do not publish the field factors, but it

[50] Kirsti Koch Christensen (chair), *Vekt på forskning: Nytt system for dokumentasjon av vitenskapelig publisering*, Committee report to the Norwegian Association of Higher Education Institutions, 2003

[51] Ulf Sandström and Erik Sandström, "The field factor: towards a metric for academic institutions," *Research Evaluation*, 18 (3), 2009

appears that the differences among them are very large: a publication in the humanities and social sciences can be weighted up to 15 times as much as one in chemistry[52]. Figure 11 shows the effects of the field adjustment at the university level. The effects of the adjustment are especially large for Karolinsaka, the medical university, and the two leading technical universities: KTH and Chalmers. The other big universities have a wider range of disciplines so field adjustments tend to cancel out to a greater degree. The small universities tend to be more orientated towards social sciences, humanities and professions than the large ones. The Swedish system is based on ISI papers (equivalent to the Jimp papers in the Czech system, which in recent exercises drove 65% of the funding allocation). Clearly, variability in publication behaviour is a major influence on the number of journal publications produced, which therefore needs to be handled in a performance-based research funding system that uses publications as an indicator.

Analysis of changes in Flanders also underscores the importance of field effects. The formula used to allocate the Flemish institutional funding for universities was changed in 2003. Previously, this was allocated using a formula based on PhD production (50%), other graduate production (35%) and the volume of public research funds attracted (15%). The old formula was retained for 70% of the funding. The rest was allocated based on a combination of SCI publication and citation indicators, with the new formula being phased in over three years. A notable result was a significant increase in funding to a university specialised in biomedical sciences[53].

4.2.4 Effects of the PRFS on fields of research in the Czech Republic

Figure 12, below, illustrates a simulation of the effect that the use of the metrics-based research results evaluation as a funding system for the distribution of institutional funding. The simulation compares percentage allocations between several tens of scientific fields by the Research Intentions system ('CEZ') and by the Performance-based Research Funding System (PRFS – 'Hodnocení'), based on calculations referring to 2008.

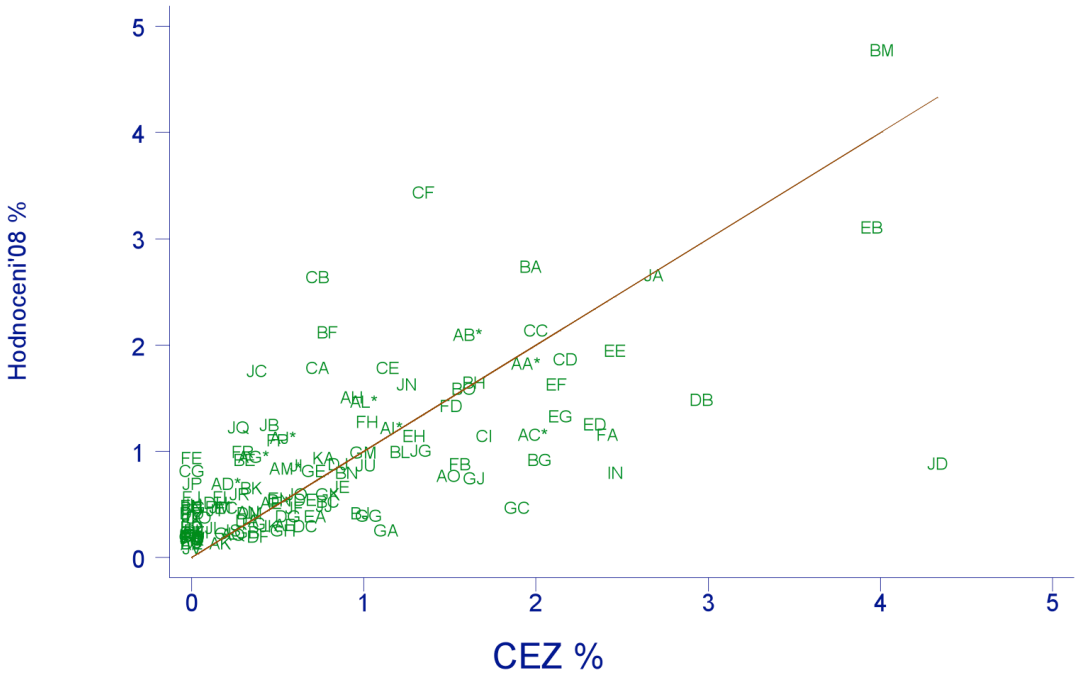
The data used for this simulation were listed in the report of the "Project for the preparation of the Methodology to evaluate the results of research institutions and of programmes finished in 2010", a report published by the Secretariat of the Board of the R&D&I Council. As such, the simulation below does not take into account the changes introduced in the Methodology 2010. However, the effects of the 'dampening factors' introduced in the Methodology 2010 are bound to be limited – especially due to the exclusion of the J_{imp} factor (see the preceding Section 3.2.2).

The graph below shows that for many fields the difference between the two funding systems is only a few percentage points, but several scientific fields will experience considerable changes – in the positive or negative sense. One can notice, for example, that the share of field 'DB' (Geology & Mineralogy) on funding based on the Research Intentions obtains 3% of total funding under the old system and drops to 1.4% if funded applying the PRFS, i.e. loose 53% of institutional research funding. By contrast, 'CF' (Physical & Theoretical Chemistry) institutional funding would increase to 143%. In any case, such radical changes of institutional funding undermine the very idea behind institutional funding, i.e. to provide a basis for institutional development

[52] Carlsson, *Op Cit*

[53] Koenraad Debackere and Wolfgang Glänzel, 'Using a bibliometric approach to support research policymaking," *Scientometrics*, 59 (2), 2004, 253-276

Figure 12 Field specific shares (in %) of institutional funds allocation by Research Intentions (x-axis) and by the Evaluation Methodology (y-axis)



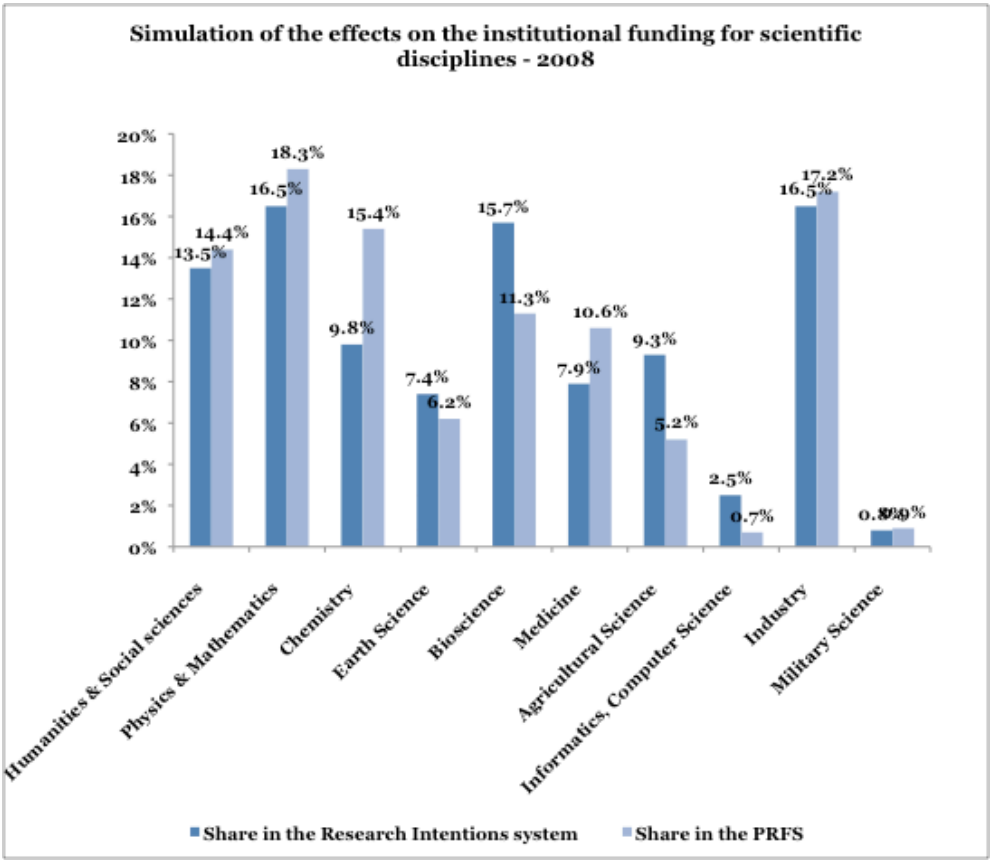
Source: Elaboration of data listed in the report of the "Project for the preparation of the Methodology to evaluate the results of research institutions and of programmes finished in 2010", Secretariat of the Board of the R&D&I Council, 2010

These envisaged changes in funding allocations between the scientific fields imply a **shift in focus** for the institutional funding of research, illustrated in Figure 13, below.

Many of these changes are influenced (also) by differences in propensity to publish among fields, described in the preceding section:

- The application of the PRFS is expected to be to the advantage of especially the scientific disciplines grouped under *Chemistry* and *Medicines*, followed by *Physics & Mathematics*.
- It is expected to be to the disadvantage of research in *Bioscience*, *Agricultural Sciences* and *Informatics & Computer Science*.

Figure 13 Simulation of the effects on the institutional funding for scientific disciplines



Source: Elaboration of data in the report of the "Project for the preparation of the Methodology to evaluate the results of research institutions and of programmes finished in 2010", Secretariat of the Board of the R&D&I Council, 2010

Figure 14 below, illustrates the actual change in institutional funding in terms of its distribution over the various Scientific and Technological fields. The figure shows extreme changes in the amount of funding similar to the simulations above.

- Under the Research Intentions system, the most financed scientific fields were *Computer Applications & Robotics (4.3% of total institutional funding for R&D), Solid Matter Physics & Magnetism, Genetics & Molecular Biology (both 4%)* (highlighted in dark orange), followed by Geology & Mineralogy (3%), Microbiology & Virology (2.5%), Informatics & Computer Science (2.5%). Cardiovascular diseases (2.4%), and Physiology (2.3%). In terms of whole scientific fields, *Physics and Mathematics* and *Industrial research* were the most financed (both 16.5%), followed by Biosciences (15.7%).
- When applying the PRFS, *Solid Matter Physics & Magnetism (5.3%)* becomes the most financed area, followed by Genetics & Molecular Biology (3.7%) , General Mathematics (3.5%) , and Physics & Theoretical Chemistry (3.5%). In terms of whole scientific fields, *Physics and Mathematics* (21.3%) has become by far the most financed area.

Figure 14 Effect of the PRFS on the funding distribution over the specific scientific and technological fields

Code	Science field	Share of institutional funding under Research Intentions 2008	Share of institutional funding under PRFS 2010	Change
AC	Archaeology, Anthropology, Ethnology	2.0%	1,0%	-48%
AD	Political Sciences	0.2%	0,6%	222%
AG	Legal sciences	0.4%	1,0%	156%
AJ	Letters, Massmedia, Audiovision	0.5%	1,1%	114%
AK	Sport & Free-time Activities	0.2%	0,1%	-58%
AO	Sociology, Demography	1.5%	0,6%	-60%
A	**Humanities & Social Sciences**	**13.5%**	**14,2%**	**5%**
BA	General Mathematics	2.0%	3,5%	74%
BE	Theoretical Physics	0.3%	1,0%	225%
BF	Elementary Particles and High Energy Physics	0.8%	1,8%	130%
BG	Nuclear, Atomic and Molecular Physics, Colliders	2.0%	1,0%	-48%
BJ	Thermodynamics	1.0%	0,4%	-63%
BM	Solid Matter Physics & Magnetism	4.0%	5,3%	32%
B	**Physics & Mathematics**	**16.5%**	**21,3%**	**29%**
CA	Inorganic Chemistry	0.7%	1,6%	122%
CB	Analytical Chemistry, Separation	0.7%	2,2%	208%
CC	Organic Chemistry	2.0%	1,8%	-9%
CD	Macromolecular Chemistry	2.2%	1,5%	-31%
CE	Biochemistry	1.1%	1,8%	68%
CF	Physics & Theoretical Chemistry	1.4%	3,5%	148%
C	**Chemistry**	**9.8%**	**14,3%**	**46%**
DB	Geology & Mineralogy	3.0%	1,5%	-52%
DC	Seismology, Volcanology, Earth Structure	0.7%	0,4%	-43%
DF	Soil Science	0.4%	0,1%	-70%
DG	Atmosphere Sciences, Meteorology	0.6%	0,3%	-54%
DH	Mining, incl. Coal Mining	0.1%	0,1%	28%
D	**Earth Science**	**7.4%**	**6,0%**	**-20%**
EA	Cell Biology	0.7%	0,3%	-62%
EB	Genetics & Molecular Biology	4.0%	3,7%	-7%
EC	Immunology	0.2%	0,5%	131%
ED	Physiology	2.3%	1,1%	-54%
EE	Microbiology, Virology	2.5%	1,9%	-24%
EF	Botanics	2.1%	1,8%	-13%
EG	Zoology	2.1%	1,4%	-34%
E	**Bioscience**	**15.7%**	**12,0%**	**-23%**
FA	Cardiovascular Diseases incl. Cardiothoracic Surgery	2.4%	1,5%	-36%
FB	Endocrinology, Diabetology, Metabolism, Nutrition	1.6%	0,9%	-43%
FI	Traumatology, Orthopedics	0.1%	0,2%	129%
FL	Psychiatry, Sexuology	0.2%	0,5%	161%
FP	Other Medical Disciplines	0.5%	1,1%	116%
FR	Pharmacology & Medical Chemistry	0.3%	0,8%	156%
F	**Medicine**	**7.9%**	**12,2%**	**54%**
GA	Agricultural Economics	1.1%	0,2%	-79%
GC	Agronomy	1.9%	0,3%	-82%
GF	Plant Pathology, Vermin, Weed, Plant Protection	0.1%	0,5%	408%
GG	Livestock Rearing	1.0%	0,3%	-72%
GH	Livestock Nutrition	0.5%	0,2%	-62%
GJ	Animal Vermins & Diseases, Veterinary Medicine	1.6%	0,7%	-56%
G	**Agricultural Science**	**9.3%**	**4,9%**	**-48%**
IN	Informatics, Computer Science	2.5%	2,0%	-21%
I	**Informatics, Computer Science**	**2.5%**	**2,0%**	**-21%**
JB	Sensors. Measurement, Regulation	0.5%	0,9%	82%
JC	Computer Hardware & Software	0.4%	1,4%	239%
JD	Computer Applications, Robotics	4.3%	0,6%	-85%
JJ	Other Materials	0.8%	0,3%	-58%
JK	Corrosion & Surface Treatment of Materials	0.5%	0,2%	-60%
JL	Materials Fatigue, Friction Mechanics	0.1%	0,2%	116%
JM	Building Engineering	0.2%	0,5%	144%
JQ	Machines & Tools	0.3%	0,6%	109%
JT	Propulsion, Motors & Fuels	0.1%	0,2%	110%
J	**Industry**	**16.5%**	**12,8%**	**-23%**

Source: Elaboration of data in the report of the "Project for the preparation of the Methodology to evaluate the results of research institutions and of programmes finished in 2010", Secretariat of the Board of the R&D&I Council, 2010; Technology Centre.

The figure above also lists those scientific fields that have experienced major shifts in their institutional funding.

- Scientific fields with percentages highlighted in red are those that experience **major drops** in their institutional funding of 50% or more. They are spread over all scientific disciplines – with the exception of Chemistry and Medicine, but show a significant effect especially for the research areas in Earth Science and Agricultural Science. Areas included are Sociology, Demography; Thermodynamics; Soil Science; Cell Biology; Cardiovascular diseases; Agricultural Economics; Agronomy; Livestock Rearing & Nutrition; and last but not least, Computer Applications & Robotics
- Scientific fields with percentages highlighted in green are those that experience a **significant increase** in their institutional funding of double or more. We see that these research areas are concentrated especially in the field of Medicine, Industry, Physics & Mathematics and Chemistry; interestingly, it includes also some research areas in Humanities & Social Sciences such as Political Sciences.

R&D&I policy in the Czech Republic is strongly influenced by its strategic objectives – a sign of good practice in R&D governance.

The next step in our analysis was therefore to consider to what extent these changes in funding distribution over the S&T fields, provoked by the adoption of the PRFS, are in line with the strategic objectives for research set out in the *National Priorities for Applied R&D&I* for the years 2009 – 2011.

For the time period 2009-2011, eight national research priorities were defined:

- Biological and ecological aspects of a sustainable development
- Molecular biology and biotechnology
- Sources of energy
- Material research
- Competitive engineering
- Information society
- Security and defence
- Priorities for the development of the Czech society.

Our findings show that the Czech PRFS can be expected to have significant negative impacts on the funding of research in Scientific and Technological fields that were explicitly indicated in the policy documents as key technologies for the achievement of the National Priorities for Applied R&D&I 2009 -2011:

- The drastic drop in institutional funding for research in **Computer Applications & Robotics** (one fifth of the funding allocated through the Research Intentions) is in strident contrast with the importance attributed to this research area – and especially robotics – in the description of the R&D objectives for the research priorities Competitive Engineering and Information Society
- The same consideration is valid for the reduction of institutional funding in **Other Materials and Corrosion & Surface Treatment of Materials,** key technological fields for the R&D objectives in the material research priority
- We notice the considerable drop in funding for several of the research fields in **agricultural sciences** which seems in contrast with the importance attributed to these specific fields of research in the research priority Biological and ecological aspects of a sustainable development

- The reduction in funding for **Biosciences** (-23%), in particular **Cell Biology and Physiology,** is certainly not in line with the high importance attributed to the research in this field in the context of the research priorities Biological and ecological aspects of a sustainable development and Molecular biology and biotechnology.
- There is also a noticeable drop in institutional funding for **Informatics & Computer Sciences**, the key technological field for the R&D objectives in Information society priority.

4.3 Observations and first conclusions

Different governments use performance-based research funding in different ways, depending on the needs and the national context. Depending on their design, performance-based research funding systems can act as a competitive source of discretionary income, as a reward for quality and/or volume of output, as an instrument of policy or – as in the case of the UK RAE – all three[54].

In the Czech Republic, the adoption of the Performance-based Research Funding System needs to be set against the context of a search for objective criteria to guide the allocation of institutional funding and an increasing – close to exclusive - focus among policy-makers on concrete R&D results. We can notice a move towards granting research institutions 'autonomy' on how to manage their funding, thereby stepping a top-down strategic steering of research financed through institutional funding.

Internationally, the use of performance-based research funding systems has mixed consequences but does certainly bring benefits in terms of apparent increases in quality and incentives to improve research strategy and management. A great deal of care has to go into the design of the system in order to avoid unintended – and sometimes perverse – consequences.

Because these exercises are intended to **affect** and not just reward or punish performance, there needs to be a sufficient time between them to allow the research community to adapt its behaviour.

Inter-field differences in publication behaviour are large and need explicitly to be tackled in any system. Another approach is to perform some tie of field normalisation in the process of calculating the indicators or in making the transition from performance indicators to money allocation.

The effects of the PRFS on the funding of research in S&T fields in the Czech republic clearly point to changes in funding influenced by the differences in propensity to publish among fields, leading to a significant increase in funding for scientific disciplines grouped under *Chemistry*, *Medicine*, and *Physics & Mathematics*, and a decrease in funding for research in *Agricultural Sciences, Bioscience, Industrial Research* and *Informatics & Computer Science.*

Our analysis furthermore shows that the PRFS leads to significant negative impacts on the funding of research in Scientific and Technological fields that were explicitly indicated in the policy documents as key technologies for the achievement of the National Priorities for Applied RDI 2009-2011. This demonstrates clearly that the mechanistic approach of the Czech PRFS cannot accommodate other (more) important policy considerations in connection with research.

Our international comparison of PRFS for institutional funding shows that each specific system is designed and applied exclusively **for one particular type of organisation**, e.g. only for universities or only for universities of applied sciences

[54] P. Bourke, Evaluating University Research: *The British Research Assessment Exercise and Australian Practice*, Commissioned Report No. 56, Canberra: National Board of Employment, Education and Training, 1997

(see the case studies in section 6). Thus, each PRFS can be modelled in line with the specific mission and activities of these institutions.

In contrast to the current intentions in the Czech Republic where by 2012 the PRFS will govern approximately 30% of the total 'basic' national public funding (including approximately 80% of the public funding of the Academy), international practice indicates that **modest amounts** of money need to be moved in any one exercise, precisely in order to minimise the damage that such consequences can do to the research system.

Based on international experience of **indicator**-based funding systems, it is easy to agree with van Raan's observation[55] that "Ranking of research institutions by bibliometric methods is an improper tool for research performance evaluation, even at the level of large institutions." He explains that the methods are not good enough and laments policymakers' tendency to try to buy 'cheap and dirty' solutions that are way behind the state of the art and produce misleading results. There is a clear need for indicator-based allocation systems to be 'damped' **through the use of indicators other than research output**.

5. The Evaluation Methodology in the eyes of the Czech research community

5.1 Survey of researchers and research organisation leaders

5.1.1 Assessment of the Evaluation Methodology by those affected by it

In order to gain a good overview of how the Czech research community views the Evaluation Methodology, we asked researchers as well as research organisation leaders (rectors, deans, directors) in a survey to assess a number of statements describing the Evaluation Methodology and its effects.

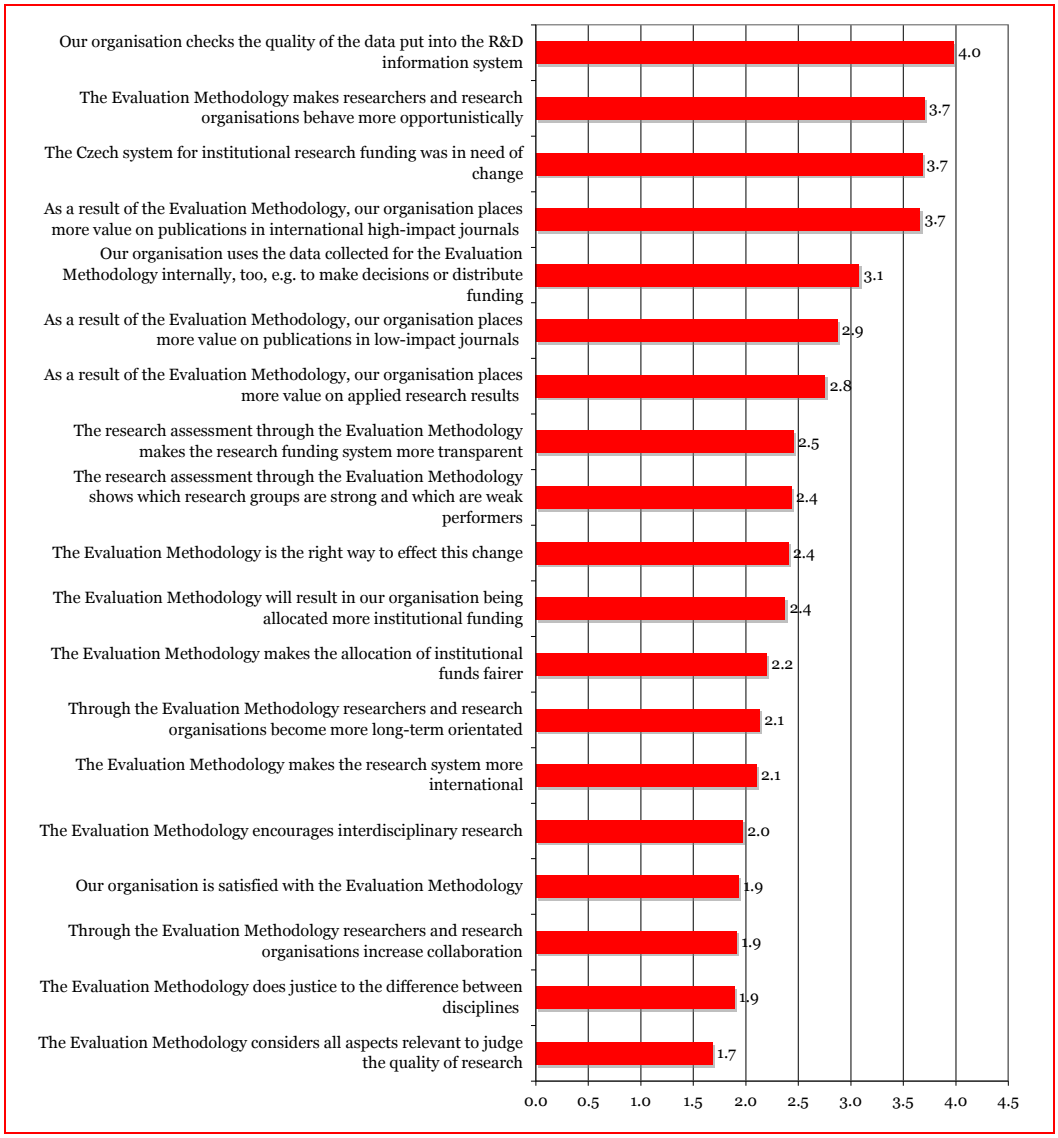
As can be seen from Figure 15, researchers (excluding rectors, deans, directors) view the Evaluation Methodology in a rather unfavourable light. In particular, they feel that the Evaluation Methodology does 'not at all' or only 'to a limited extent' consider all aspects relevant to judge the quality of research, do justice to the differences between disciplines, and encourage collaboration.

On the positive side, most respondents report that their organisation checks the quality of the data put into the R&D information system (76% agree 'to a large extent' or 'very large extent'). On the negative side, a large majority thinks that the Evaluation Methodology makes researchers and research organisations behave more opportunistically (63% agree to 'a large extent' or 'a very large extent', 21% 'to a moderate extent').

58% of respondents agreed to 'a large or very large extent' that the Czech system for institutional funding was in need of change but only 20% agreed to 'a large or very large extent' that the Evaluation Methodology was the right way to effect this change. Indeed, organisations are not very satisfied with the Evaluation Methodology (mean 1.9). Researchers from the Czech Academy of Science are significantly less satisfied with the Evaluation Methodology than researchers from universities and other research organisations (63% not at all satisfied compared to 24% at universities and 29% at other research organisations).

[55] Anthony FJ van Raan, "Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods," *Scientometrics*, 62 (1), 2005, 133-143

Figure 15 Researchers' assessment of the Evaluation Methodology*



*Arithmetic mean on a scale from 1='not at all' to 5='to a very large extent'

Source: researcher survey, n=543

Organisation type[56]

Respondents from the Academy of Science generally view the Evaluation Methodology less favourably than their counterparts in other research organisations and especially universities. This is true for 13 items out of the 19 statements (68%) in Figure 15. Tying in with this pattern is the result that significantly more universities (64%) agree 'to a large or very large extent' that the Czech system for institutional research funding was in need of change compared to research organisations (59%) and the Academy of Science (52%). Other research organisations are often situated in the middle between the Academy institutes and universities in the extent to which they agree or disagree with the statements.

[56] We left out the category 'industry' in our analysis because we had only a handful of cases in this category. Leaving them in would have affected the robustness of results.

Respondents from the Academy of Science viewing the Evaluation Methodology less favourably than their counterparts may be due to the perception that Academy institutes fare worst with the Evaluation Methodology: 21% of respondents from universities agreed to a 'large or very large extent' that they would receive more institutional funding as a result of the Evaluation Methodology, while only 13% from research organisations and 7% from Academy institutes said so. In contrast, 67% of respondents from Academy institutes, 52% from research organisations and only 42% from universities did not agree at all or only to a limited extent that the Evaluation Methodology would result in their organisation being allocated more institutional funding.

Another interesting result from the analysis is that Academy institutes seem to 'accommodate' less to the Evaluation Methodology, compared to universities and research organisations. While researchers from universities report that their organisations place more value on publications in international high-impact journals as a result of the Evaluation Methodology, researchers from other research organisations state that their organisations place more value on applied research results. In contrast, researchers from Academy do much less so: While 77% of respondents from universities agreed 'to a large or very large extent' that as a result of the Evaluation Methodology their organisations place more value on publications in international high-impact journals, the corresponding figures for other research organisations is 64% and for Academy institutes is 51%. Similarly, while 54% of respondents from other research organisations agreed 'to a large or very large extent' that as a results of the Evaluation Methodology their organisations placed more value on applied research results, only 30% from universities and 13% from Academy institutes said so.[57]

Disciplines

There are also differences in perception among the different disciplines: Most importantly, more than 60% of respondents from the social sciences and humanities (SSH) feel that the Evaluation Methodology does not at all do justice to the differences between the disciplines, compared to 37% from the natural sciences and life sciences and 39% from the engineering/technical sciences.

This may be linked to the observation that researchers from different disciplines rate the importance of different types of outputs differently (see below in section 5.1.2), i.e. the value the Evaluation Methodology attaches to a certain type of output (through the point score) is not necessarily the same as the value this output has for the research community in a certain field.

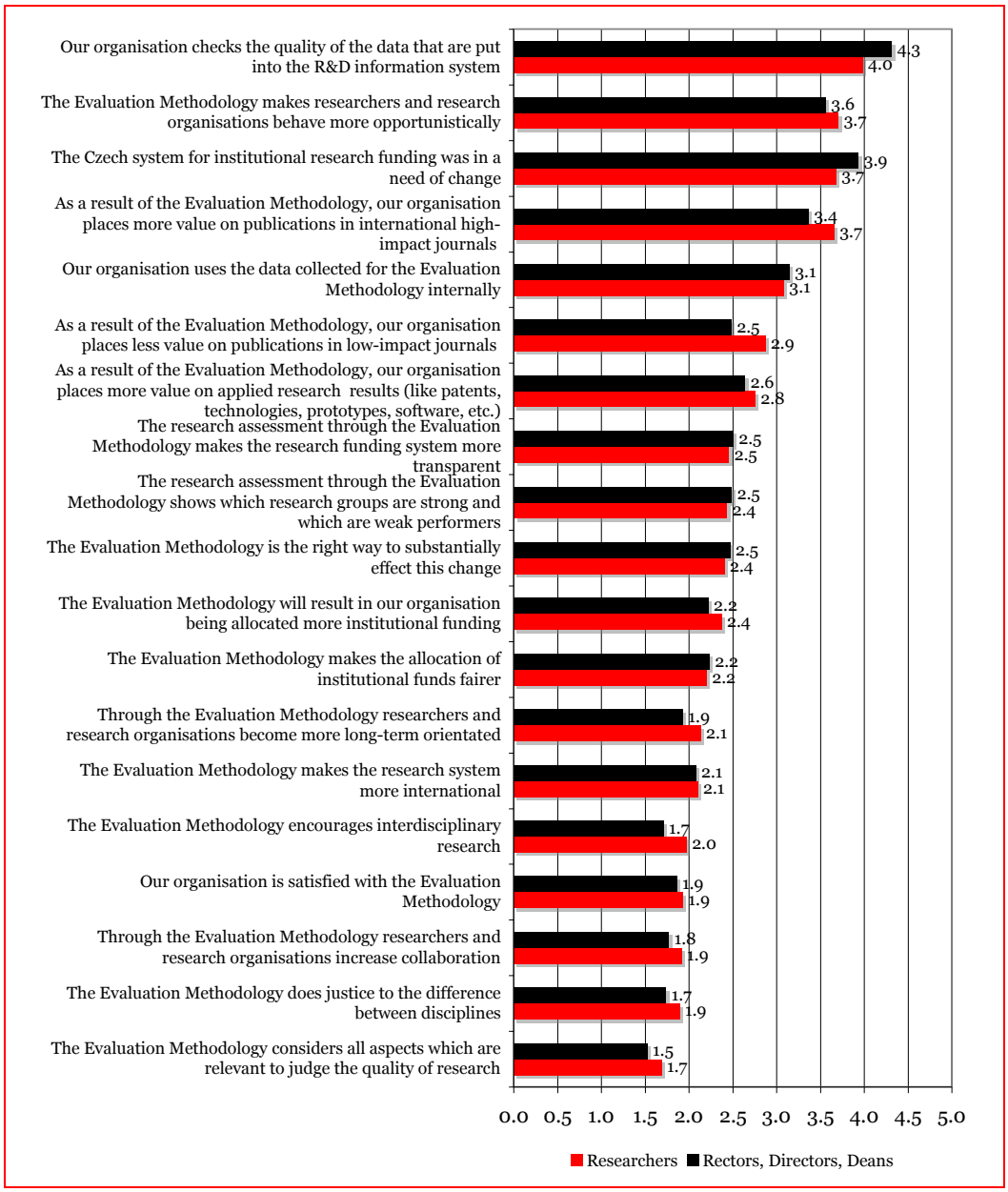
As a consequence, many researchers from the SSH felt that they are losing out most as a result of the Evaluation Methodology: 38% of respondents from the SSH reported that the Evaluation Methodology would 'not at all' result in their organisation being allocated more institutional funding. Percentages for the natural sciences and life sciences are 24%, and for the engineering/technical sciences 21%. Conversely, while 15% of respondents from the natural sciences, 14% from engineering/technical sciences and 11% from life sciences agree 'to a large extent' that the Evaluation Methodology will result in our organisation being more institutional funding , only 2% from the SSH do so.

[57] In recent years, the publication growth rate at universities has been larger than at Academy institutes. Some of our interviewees have put forth the hypothesis that the different evaluation systems used in the two sectors – the Evaluation Methodology in the university sector, peer-review based evaluation in the ASCR sector – influenced the publication rate of the two sectors. Indeed, the fact that Academy staff 'accommodate' less to the Evaluation Methodology seems to support this hypothesis. However, others argue that the fact the universities' output grew more quickly than the Academy's is often at the expenses of AVCR, with research paid and performed by the ASCR, equally counted by the universities because they "own" the PhD students involved? Also, we can surmise that there might be a 'catching up' effect at work, as research became part of universities' missions only after the Velvet Revolution. We cannot verify these hypotheses but it shows that there are different factors influencing publication growth rates

Similarly, while most disciplines, in particular the engineering/technical sciences place more value on applied results, the SSH do so to a much smaller extent, presumably because it is not possible in their disciplines.

We also asked rectors of universities, deans of faculties, and directors of research institutes to assess the Evaluation Methodology. The sample of was small, so figures should be viewed as tendencies and comparisons between researchers on the one hand and rectors, deans, and directors on the other hand should not be overestimated. In fact, a comparison between researchers on the one hand and rectors, directors and deans on the other hand shows that both groups view the Evaluation Methodology similarly (Figure 16).

Figure 16 The Evaluation Methodology as assessed by researchers and research organisation management (rectors, directors, deans)



*Arithmetic mean on a scale from 1='not at all' to 5='to a very large extent'

Source: researcher survey, n=543; survey of rectors, directors, and deans, n=71

5.1.2 Importance of research outputs

We also asked respondents about the importance their organisations placed on a number of research outputs, *irrespective* of the Evaluation Methodology. The research outputs we asked them to assess were mainly research outputs for which the Evaluation Methodology gives points and research outputs typically considered in knowledge transfer.

Researchers report that their organisations give by far the highest importance to articles published in journals included in the Web of Science and to articles published in Nature, Science, and PNAS (Figure 17). In other words, irrespective of the Evaluation Methodology, organisations place the highest importance on research results that give the highest points in the Evaluation Methodology, implying that the incentive structure of organisations and the Evaluation Methodology overlap in the case of articles published in a journal included in the Web of Science or published in Nature, Science, and PNAS.

The least important research results are the organisation of exhibitions, audiovisual works/electronic documents, and research reports containing classified information. The latter is a research result for which the Evaluation Methodology confers points.

However, if we break the answers down by disciplines, we find that out of the 24 research outputs listed, 20 (83%) are assessed differently by respondents from different disciplines.[58] In most cases, the SSH display different patterns compared to the natural sciences and the life sciences, with engineering/technical sciences situated in-between. For example, to the SSH and to a lesser extent also engineering/technical sciences, articles in Nature, Science, PNAS are less important, while articles in (Czech) reviewed journals listed in the 'List of Periodicals' are significantly more important.

Similarly, if we break the answers down by organisation type, we find that 19 out of 24 research results (80%) are assessed differently by respondents from different types of organisations[59]. In this instance it's mostly the research organisations that differ from the universities and Academy institutes. The typical pattern is that the more applied research outputs are more important to other research institutes and to some extent also the universities but less so to the Academy institutes.

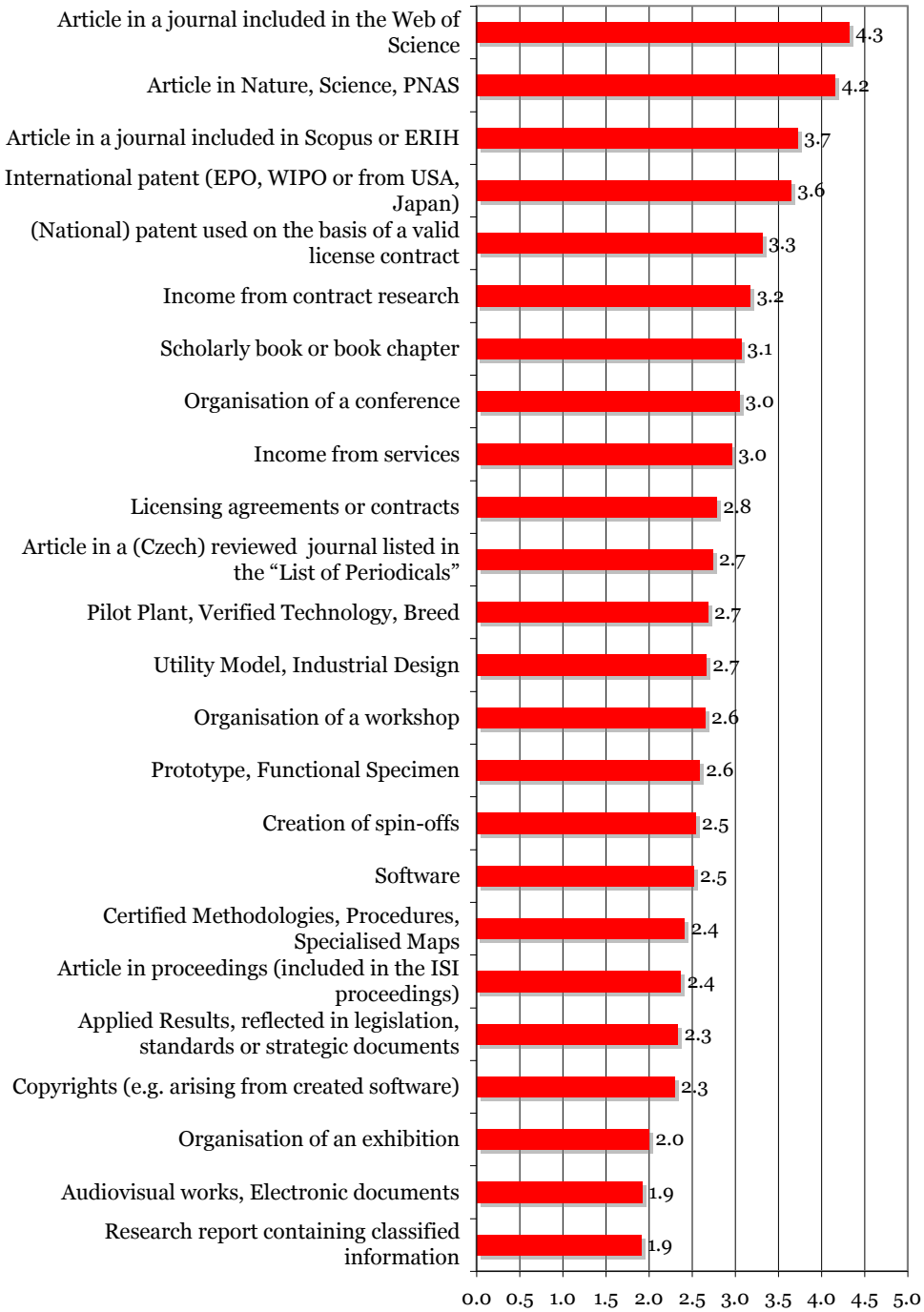
What do these results tell us? Of course, survey results always have to be taken with a pinch of salt. Nonetheless, given the differences in assessment by different disciplines and different types of research organisations, it seems fair to say that it is problematic to lump together different disciplines and different types of organisations, as is done in the Evaluation Methodology. The newer versions of the Evaluation Methodology, namely 2009 and 2010, are trying to come to grips with differences among disciplines by applying a differentiation according to disciplines. But surprisingly, the fact that different types of organisations are assessed in the same way by the Evaluation Methodology appears not to be an issue at all.

---

[58] The four research results having the *same* importance for the different disciplines are: article in a journal included in Scopus or ERIH; applied results, reflected in legislation, standards or strategic documents; audiovisual works, electronic documented; organisation of a workshop

[59] The five outputs that have the same importance to different types of organisations are: article in a journal included in Scopus or ERIH; scholarly book or book chapter; organisation of exhibition; licensing agreements or contracts; creation of spin-offs

Figure 17 Importance of research results irrespective of Evaluation Methodology

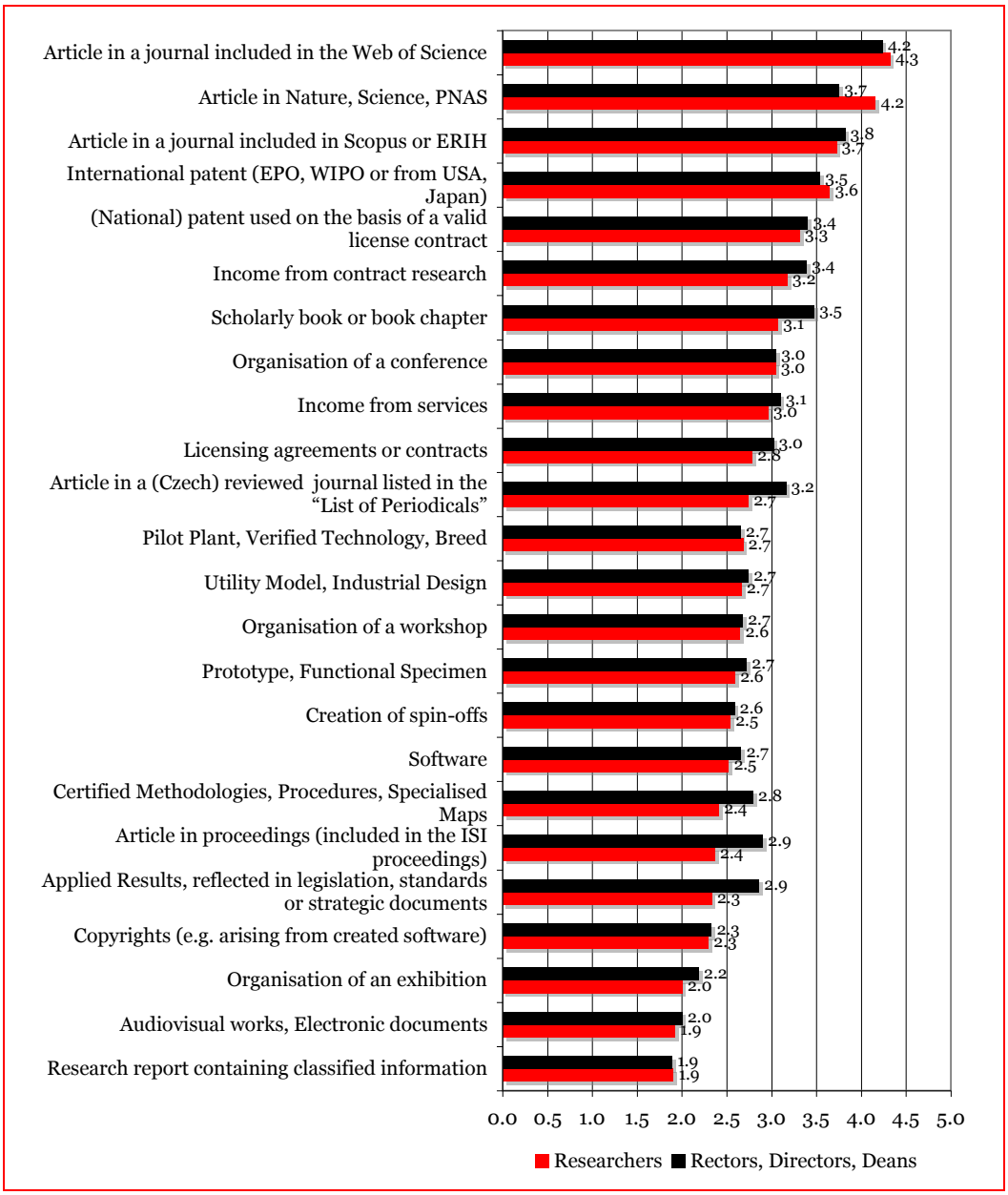


*Arithmetic mean on a scale from 1='no importance' to 5='very high importance'

Source: researcher survey, n=484

We also asked rectors of universities, deans of faculties, and directors of research institutes how their organisations rated various research outputs. The most important result is that researchers and research organisation management give more or less the same importance to research outputs (Figure 18). Given the small sample size for rectors, deans, and directors differences should not be overrated.

Figure 18 Research outputs as assessed by researchers and research organisation management (rectors, deans, directors)



*Arithmetic mean on a scale from 1='no importance' to 5='very high importance'

Source: researcher survey, n=484; survey of rectors, directors, and deans, n=65

5.1.3 Summary and conclusions from the survey results

What becomes obvious from the survey we conducted among the Czech research community is that both researchers and research organisation management (rectors, deans, directors) view the Evaluation Methodology in a rather unfavourable light. In general, respondents from the Academy of Science view the Evaluation Methodology less favourably than their counterparts in other research organisations and especially in universities. This may be due to respondents' perception that Academy institutes would fare worst with the Evaluation Methodology. Another interesting result fitting the picture is that Academy institutes seem to 'accommodate' less to the Evaluation Methodology than universities and research organisations.

The disciplines also differ in perception of the Evaluation Methodology. Most importantly, mainly respondents from the SSH feel that the EM does not do justice to the differences between the disciplines. This may be due the fact that researchers from different disciplines rate the importance of different types of outputs. This may also be due to a feeling that the SSH would be losing out the most as a result of the EM.

We also asked respondents about the importance of a number of research outputs irrespective of the Evaluation Methodology. Again, both researchers and research organisation management (rectors, deans, directors) assessed research outputs quite similarly. Interestingly, but not surprisingly, most research outputs are assessed differently by respondents from different disciplines. In most cases, the SSH differ dramatically in their assessment from the natural sciences and life sciences, with engineering/technical sciences situated somewhere between the two. Similarly, different types of organisations assess research outputs differently. Generally, the more applied research results are more important to other research organisations and to some extent also to the universities but less so to the Academy.

These results are in line with our own expectations as our analysis of the Evaluation Methodology came to very similar conclusions. In particular, we think it is problematic to lump together different disciplines and different types of organisations, as is done in the Evaluation Methodology. While the problem the different disciplines have with the EM has been recognised, the fact that different types of organisations are assessed in the same way by the Evaluation Methodology does not appear to be an issue at all.

5.2 Consequences on researchers' behaviour and the Evaluation Methodology's blind spot

In this section we are discussing two main results from the survey in more detail, integrating them with results from interviews.

5.2.1 Opportunistic behaviour

From the survey results we know that a large majority of researchers (84%) think that the Evaluation Methodology makes researchers and research organisations behave more opportunistically (63% agree to 'a large extent' or 'a very large extent', 21% 'to a moderate extent'). Our literature analysis has shown that opportunistic behaviour has also been triggered by other simplistic PRFS, most notably in Australia (see section 4.2.3). In our interviews we were given various examples of such opportunistic – also often called illegitimate – behaviour, which can be synthesised into a number of patterns.

- **Adapting outputs to make them 'countable':** A first typical behaviour is to adapt outputs that are produced during the normal activities of a research institute but do not give points, making them into 'countable' outputs. An example would be textbooks that are published as scientific monographs. Other examples would be re-publishing older works, establishing working paper series and promote them as if they were refereed journals or strategic citation behaviour to boost the impact factor of a journal. Some interviewees tell us that such 'fake' outputs make up a substantial part of the outputs submitted in their disciplines. Indeed, while all types of publications increased in numbers between 2009 and 2010[60], those publications that are most easily manipulated – Czech reviewed journals, books and book chapters, as well as proceedings –grew at the highest rate (Figure 19). (Similar increases can be seen with the applied research results, see text below.) This problem is particularly pronounced in the SSH because the SSH produce much greater shares of outputs that are easily manipulated while

[60] Unfortunately, longer – and hence more valid – times series were not possible due to new categories being taken up and categories being redefined

giving fixed numbers of points. This benefits low quality research within the SSH, discriminating against excellent SSH research.

Figure 19 Distribution of points 2006-2010

	Evaluation year	2010	Growth rate (%)	2009	2008	2007	2006
	Years counted	**2005-9**		**2004-8**	**2003-7**	**2002-6**	**2001-5**
Jimp	**Article in WoS journal**	35617	8	33056		29773	25478
	Article in SCOPUS or ERIH journal	14113	14	12352			
	Article in Czech journal-reviewed	19263	**30**	14824			
Jneimp	**Article in non-WoS journal- Total**	33376	23	27176		47445	46581
J	**Article in journal-Total**	68992	15	60232	40124	77218	72059
B,C	**Book, chapter**	21096	**61**	13094	13111	17756	18740
B	**Book**					7164	6468
C	**Chapter**					10592	12272
D	**Proceedings**	7481	**66**	4501	2730	104340	83713
P	**Patent**	229	-38	371	276	562	363
F	**Utility model, industrial design**	566	**210**	183			
G	**Prototype, functional model**	2225	**143**	915			
H	**Results implemented into legislation or standards**	183	**215**	58			
N	**Certified method**	1325	**393**	269			
R	**Software**	1692	**192**	580			
V	**Secret report**	8	-98	400	2		
S	**Prototype, applied method**	3065	-7	3284	3133	1077	
Z	**Trial operation, variety, breed**	902	52	593			
T	**Prototype, trial operation**	352	-36	551			
Z*	**Trial operation, verified technology, variety, breed**	1253	10	1144	887	1676	1471
L	**Specialized maps**			105			
	Total number of items	*108116*	28	*84744*	*60263*	*202630*	*176350*

Numbers in black are taken from the webpages of RVVI, numbers in red are Technology Centre calculations, based on the above data

*This category was named Technologies (T) in 2006 and may include also some other types of results.

Source: Technology Centre

- **Going for 'cheap points':** Researchers not only adapt their research outputs to make them countable, but also employ strategies to significantly increase their amounts of points, often without actually improving the quality of their research. Virtually all researchers we interviewed stress that in the current Evaluation Methodology, a large number of mediocre results can weigh much more than (single) outstanding contributions that could eventually change a whole field. One reason for this is that it is easier to pick a journal at the top of a Evaluation Methodology category that has a low impact factor but gives a lot of points rather than publishing in a journal with a higher impact factor that gives fewer points. E.g. in the medical field, the journal 'Ultrasound in Obstetrics and Gynaecology' has an impact factor of 3.154 but because it tops the list of 28 journals in its category, it is awarded 305 points according to the Evaluation Methodology. In contrast, the journal 'Cancer Cell' with an impact factor of 25.288 ranks third of 166 journals in its category and hence get awarded 250.32 points. Researchers underline that the correct strategy for increasing research quality would be to

produce fewer high-quality results rather than many mediocre or sub-standard ones. However, just the opposite is encouraged by the Evaluation Methodology.

There is also the issue of local versus international citations. Local citations are used to boost the impact factor of a journal that actually publishes poor-quality articles. These journals are typically located in Eastern European countries, including the Czech Republic, and are often in the SSH.

A specific problem that encourages the publication in low quality journals is that in some fields the Evaluation Methodology does not sufficiently discriminate between mid-level journals and low-level journals. It is insensitive to the fact that many relatively respectable journals are ranked approximately as median journals, which earn relatively few points. This is a disincentive to increase quality, in particular for institutions who are increasingly orientating themselves internationally on their way to international excellence. In other words, *the ambitious get punished* in the current system.

Another example of 'going for cheap points' is for a Czech researcher to have his / her name added as a co-author on a paper with a large number of international co-authors (who do not qualify for the Evaluation Methodology because they are foreign), without having contributed much.

Finally, *applied research outputs* seem to pose a particularly important problem with regard to 'cheap points'. Various applied outputs seem to be quite easy to produce, e.g. utility models or the national 'small patent', which only requires an administrative act to file. A prototype receives the same number of points, regardless of whether it has been created by one person in one year or by a whole team over several years. A device consisting of ten subsystems can be registered as one item or as ten devices (one for each subsystem). What makes it worse, is that the applied research results are not independently assessed by anonymous (and international) peers. Originality and functionality are not required, so that existing and non-functioning solutions are registered. In short, some applied research outputs can be produced in large amounts as no rigorous review process with internationally recognised scientists are involved. Figure 19 shows that most applied research outputs have indeed grown at an amazing rate since their introduction.

- **Producing outputs for outputs' sake:** Next to presenting research results in a form that that delivers the most points, researchers also engage in the production of extra outputs that give points but are not directly related to performed research or the mission of the research organisation. For instance, a Czech researcher has his / her name added as a co-author on a paper with a large number of international co-authors (who do not qualify for the Evaluation Methodology because they are foreign), without having contributed much. Another example are patents which are produced regardless of whether or not they are needed, providing the best way to bring a result to application in the sector, or whether they are being applied at all.
- **Using by-products:** A fourth typical behaviour is to get points for by-products that arise in the normal course of a research institute's activities, i.e. products that are a means to an end and not a research result in its own right. A typical by-product would be software[61] which is often produced as a tool to support research, where the ultimate outcome would be a research result gained, among other things, with the help of the software. By-products bind resources (time, attention)

[61] The new rule with regard to software requires closure of an appropriate agreement on use/realization of the software between the author of the software and the user of the software, so that a financial benefit of the software of at least 1 million CZK can be proved. This condition is in effect starting Jan 1st 2011. Researchers fear that this condition will exclude the possibility of registering software. Also it will act as an incentive to stop producing open source software.

that would be better used for producing end products that are central to the research institute's mission, e.g. academic papers.

Adapting such opportunistic behaviours does not mean that research actors in the Czech Republic are devious or 'evil'. On the contrary, the Evaluation Methodology creates certain opportunities for behaviour which research actors, quite understandably, take advantage of, as they want achieve the highest possible institutional research income. To put it in a nutshell, they 'play the system' to achieve points but they cannot really be blamed because the system sets the incentives to do so. If research actors refuse to 'play the system', they are actually acting rather stupidly because they risk being out-competed by other research actors who do. In other words, everyone in the Czech RDI system has an incentive to 'play the system'.[62]

What is more, the unpredictability of the Evaluation Methodology increases the 'hunt for points': as research organisations do not know how much money they will earn for their points in the following year, they only race for more of them with uncertain results. Of course, this results in inflating the total number of points, with the result that a point is worth less and less.

Ultimately, the result of the various types of opportunistic behaviour is that it prevents the Evaluation Methodology from distinguishing between weak and strong research units and from increasing research quality.

Last but not least, not enough attention is paid to the process of *verifying* research outputs that enter the evaluation, points assignments and financial allocation at the very end. No budget is set aside for verification. Hence, there is no proper check whether research outputs fulfil the definition of a research output as described by the Evaluation Methodology or whether they exist at all. By way of example, books should be physically verified for existence and meet some basic requirements. There was a time-consuming verification process taking place in summer 2009 but we were told the process was flawed and that there was definite room for improvement.

5.2.2 Different organisation types

The survey results also showed that different research outputs have different relevance for different types of organisations. Similarly, our own conclusion was that the different organisations cannot be lumped together, as is the case in the Evaluation Methodology but not in the institutional PRFS in other countries (see section 4.2.1).

However, in the political discussion in the Czech Republic, applying the same Evaluation Methodology to different types of research organisations does not to appear to be an issue at all. The primordial concern seems to have been to keep the system simple and 'automatic'. At the political level there has never been a *constructive* discussion about what institutions should play what role in the Czech RDI system. Of course, there is a discussion about the Academy but it seems to have been more at a superficial level about whether or not to abolish it. From our involvement in the international peer reviews of Czech research institutes (WPdii, see section 2.3 and the OP Research and Development for Innovation (OP R&DI, see section 7.2) contract negotiations with the institutes selected for funding we also know that few research institutes themselves have really thought about their mission and position in the Czech RDI systems, their linkages with external parties, etc. Last but not least, there has not really been a dialogue between policy-makers and institutions about different institutes' roles in the Czech RDI system in the sense of asking 'What is good for the country?'

[62] Of course, not all behaviour engendered by the Evaluation Methodology is necessarily devious or involves cheating. We have heard of teams who have invested in their research capacities, trusting in their abilities to produce excellent results, which will, simply speaking, help them to produce more points and thus make money they can use for further development

In such a situation comes in the Evaluation Methodology, which also ignores theses issues and thus exacerbates them. The Evaluation Methodology fails to do justice to the different types of organisations (and disciplines for that matter) in two instances.

- **What outputs are defined as outputs?** At first glance, the list of research outputs seems quite comprehensive, but as with any exhaustive list, there will always be cases that are not covered by the list. For example, the applied results are mostly products of industrial development; this discriminates against applied research that serves society as a whole (e.g. parts of environmental research or agricultural research or Czech language studies) and that uses other ways to communicate its results (e.g. local and / or sector-specific media) and produces different types of outputs (e.g. policy advice, input to regulation, guidelines, grammar rules). Of course, the Evaluation Methodology contains the 'H type results', i.e. results transposed into legislation, regulations and standards, but their definition is too narrow to cover the thematic range of research outputs (e.g. Hknoc is only counting strategy inputs into *research and development* policies). What's more, factors external to research determine whether or not a certain output is counted in the evaluation: results must be used unchanged and the respective law or regulation must be published within a year after the research result has been produced. These conditions are not determined by the researchers but by a political or regulatory process. In fact, H type results are impacts of research rather than outputs like in the other categories. It does make sense to evaluate the impact of research at some point, but basing budget allocation on impacts for some actors and on mere outputs for others, is unfair.

  There are some indications that quite a number of Czech research institutes are unproductive, and the Evaluation Methodology was seen as a way to make them more productive or to get rid of eventually (by withdrawing institutional research funding from them). This is of course a very legitimate concern. However, the question is whether these institutes are really unproductive or whether they produce outputs that are not covered by the list. This is a point that will need to be identified in an appropriate evaluation.

- **How many points are given to an output?** The number of points given to different outputs and the differences in points between two outputs are quite arbitrary, sometimes concealing a political decision about what outputs are desirable and what outputs are less desirable from a political point of view. However, a certain type of output can be of high importance and relevance for one organisation and not relevant at all for another – and rightly so, determined by their very mission, activities and target groups. Patents, for example, are certainly not a relevant research output at an institute specialised in basic research (e.g. in nuclear physics), where the main target users of results are other scientists (e.g. in nuclear physics but e.g. also in reactor physics). Similarly, an applied research institute that develops inputs to policy makers as a main output would not reach its target audience with patents. Even in an applied institute in engineering which works for industrial users and where patents are an important output, simply increasing the absolute number of patents is no sensible strategy to deal with intellectual property. The Evaluation Methodology inevitably fails to accommodate this, as a result of which the (high number of) points assigned to patents are attainable for some research actors only.

5.2.3 Comparing Evaluation 2009 by Metodika Hodnocení with the informed peer review exercise done during the Czech Audit

Results from the Evaluation Methodology 2009 and the international peer review performed during our audit can only be compared to a very limited extent. It is important to keep in mind that although both exercises are called 'evaluation', they serve different purposes: The EM is used to make funding decisions in an automated way. Our peer review was formative, i.e. done in order to learn about the situation of research in the CR and, at the level of the participating institutions, to provide

feedback as a basis for organisational learning, development and improvement. Consequently, the scope and content of coverage, they type of results and the type of feedback given are different. This limits comparability.

Where the findings can be compared roughly, they are not in line, i.e. the more favourable ranking of the two applied institutes by the EM is not confirmed by the results of the peer reviews. Yet, we have taken the analysis of actual EM scores a little bit further in order to illustrate some or the main issues we have presented in the SWOT analysis in a very condensed form. For details please see Appendix B.

6. Research Evaluation and Institutional Funding Abroad: international case studies

6.1 Introduction

In agreement with the terms of reference, we have carried out four international case studies. These case studies have different foci, as can be seen in Figure 20. Although some case studies have more of a focus on either the evaluation of research or on the allocation of funding, they also discuss the other topic and how the two are linked.

One of our main selection criteria was that there should be enough experience with an evaluation or a funding system. Many countries experiment with performance based research funding systems but many systems are quite recent (e.g. Sweden, Denmark, the new Australian system)

In addition, we include two case studies from the Czech Republic. They demonstrate that there is already experience in the Czech Republic with approaches to research evaluation and the allocation of funding other than the one used in the Metodika Hodnodení. The two case studies are:

- The Academy of Sciences system for the evaluation of research;
- The implementation of performance contracts for Centres of Excellence and Regional Centres funded under the Operational Programme 'R&D for Innovation'.

We would like to point out that the case studies are not evaluations. We use them to illustrate different practices of institutional funding allocation and research evaluation, looking for generic key issues. We did ***not*** conduct the case studies in order to find the 'perfect model' (which does not exist) and ***not*** in order to present the case studies as examples of 'best practice' (which also does not exist) but to learn from experiences inside and outside the Czech Republic.

Figure 20 The international case studies

Country	Main focus	Selection criteria
Netherlands	Standard Evaluation Protocol: evaluation of research in universities and research organisations	• Internationally renowned evaluation system • Inspired the Academy of Science in their internal evaluation of research
UK	UK Research Assessment Exercise: evaluation of research quality and allocation of research funding	• The oldest performance based research funding system. Hence, a great deal of experience exists
Finland	Allocation of funding based on performance contracts and funding formula and their evolution over time	• Finland introduced performance contracts in the early 1990s and performance based research funding in the late 1990s. This was part of a larger modernisation of public services. Hence, a great deal of experience exists
Austria	Allocation of funding based on performance contracts and funding formula	• Cultural closeness • Was in a similar situation as the Czech Republic 15 years ago, catching up in terms of RDI

Country	Main focus	Selection criteria
		• Czech actors showed an interest in the Austrian model

Source: Technopolis

6.2 Netherlands

6.2.1 Evaluation of research in the Netherlands: Standard Evaluation Protocol

6.2.1.1 Background

The steering philosophy of the Dutch Quality Assurance for public research organisations is founded on the Higher Education and Research Act from 1993 (Dutch abbreviation WHW). The Act codified enhanced institutional autonomy and introduced the principle of self-regulation. Since then, the policy framework for the Dutch universities and research organisations has revolved mostly around quality assurance and funding. The Act introduced quality assurance as a policy instrument in the steering philosophy. In exchange for more autonomy, public research organisations were expected to play an active role in the establishment of a new quality assurance system for research (and teaching).

In order to establish a new quality assurance system for publicly funded research, the Standard Evaluation Protocol (SEP) was introduced in 1994. The Standard Evaluation Protocol has been formulated by the three main Dutch organisations responsible for publicly funded research in the Netherlands: the Royal Netherlands Academy of Arts and Sciences (KNAW), the Netherlands Organisation for Scientific Research (NWO) and the Association of Dutch Universities (VSNU). Over time, some changes were made in SEP. In 2009 the fourth SEP was accepted, following the protocols of 1994, 1998 and 2003.

The SEP aims to assess the quality of research in universities, and to facilitate both (public) accountability and improvement of research and research management. The SEP provides common guidelines for the evaluation and improvement of research and research policy, based on expert assessments. Quality assurance is based on self-evaluation reports prepared by the institutions and site visits carried out by experts (peers) for each disciplinary area in a six year cycle. More than the previous systems, the SEP and the self-assessment focus on bringing about a dialogue involving the assessor and the institution – as opposed to a detached assessment – and on improving the quality of research. The relevant literature distinguishes between a 'jury' model of assessment and a 'coach' model.

Up until 2003, the VSNU had been responsible for organising nation-wide evaluations at discipline level. Since 2003 the universities themselves have organised the research assessments. During its six-year cycle, an independent expert committee (the Meta Evaluation Committee) reviews the Standard Evaluation Protocol and the research assessment processes based on the protocol. KNAW, NWO and VSNU organise this review. The last one was in 2008, the next one will be organised in 2013. The results of the review are made public.

6.2.1.2 The Standard Evaluation Protocol for Public Research Organisations

The SEP 2009-2015 aims at two objectives with regard to the evaluation of research and research management:

- Improvement of research quality based on an external peer review, including scientific and societal relevance of research, research policy and research management
- Accountability to the board of the research organisation, and towards funding agencies, government and society at large

The objective of improvement is aimed at both the research and its management. External evaluations can be of great value to the institute and its researchers, since

international experts in the field formulate recommendations regarding the research, including the strategy and policies which direct and provide the conditions for conducting research.

With the external evaluation, the institute and its research groups account for their research activities to the board of the university, KNAW or NWO. In a broader sense, the external evaluations inform funding agencies, government and society at large of the quality and relevance of research activities, thus accounting for the public investments made in scientific research.

This protocol is primarily directed toward the evaluation of scientific research. Traditionally, such evaluation focuses on the quality of work according to the standards of scientific disciplines and the ways in which results are communicated to a scientific audience. However, the scope of the term 'research' is not limited to the research results. Research management, research policy, research facilities, PhD training and the societal relevance of research are considered integral parts of the quality of work in an institute and its programmes. With these elements, the evaluation of research has a broad scope.

The evaluation process

Assessment following the SEP means that every six year research activities will be assessed through an external evaluation, based on a self-evaluation and a site visit. In addition, universities must carry out an internal mid-term review: a self-evaluation of their research activities. This means that self-evaluation takes place every three years. In the SEP, guidelines regarding assessment criteria, minimum information requirements and the procedure of the external review are formulated. It is important to note that the assessment is both retrospective and prospective, attending to past performance as well as future plans.

The Standard Evaluation Protocol entails three main characteristics:

- Two levels of assessment: The assessment takes place at two levels of research organisation, i.e. the level of the institute (or faculty or research school) and the level of research groups or programmes.
- Three vital tasks: The assessment relates the three vital tasks of research organisations, i.e. producing results for the academic community; producing results that are relevant for society; and educating and training the next generation of researchers.
- The assessment is based on four criteria:
    1. Quality (including international academic reputation and PhD training)
    2. Productivity (the relationship between input and output)
    3. Societal relevance (including valorisation)
    4. Vitality and feasibility (the ability to react adequately to important changes in the environment).

The external assessment covers not only the content of the research programme but also the management, strategy and mission of the research centre where it is carried out. The evaluation protocol leaves scope for assessment of one or more research centres (institutes) within the same university or for comparison with similar centres at home or abroad.

The conclusion of the assessment should be summarized in a single term according to a five-point scale, 'excellent' meaning world class research, and 'unsatisfactory' meaning below acceptable standards. In reality there is a tendency to only give a 4 or 5, with a 3 as the minimum.

After the site visit, the evaluation committee will report its findings to the board of the research organisation. The board will publish the report after internal discussion with the assessed research unit and will make its position regarding the evaluation

outcomes public. The evaluation report and the position of the board together constitute the results of the evaluation.

The management of the evaluation

The governing boards of KNAW, NWO and the universities determine the units that will be evaluated and provide an overall schedule for all evaluations. The executive board of the university is further responsible for the planning of each individual evaluation and its follow-up.

At the start of any external evaluation process, the board will provide a planning document. This includes all major steps to be taken from the start of the self-evaluation to the eventual mid-term review. It consists of:

- Arrangements for the self-evaluation report and other documentation, such as the outcome of the previous mid-term review and other relevant evaluation results
- Selection and configuration of the external evaluation committee
- Planning of the site visit
- Publication of the evaluation results
- Arrangements for the follow-up of the evaluation

The unit to be evaluated provides a self-evaluation document, including a SWOT analysis, which is to be endorsed by the board. After approval by the board, the self-evaluation is sent to the external evaluation committee.

The board is responsible for the selection of the chair and further configuration of the external evaluation committee. The unit to be evaluated is invited to suggest committee members. The board may also consult third parties to reflect on the impartiality and independence of the committee chair and members.

Accounting for different disciplines and interdisciplinarity

- Taking different disciplines into account

"The SEP is the basis for research evaluation in all research areas, fields or disciplines. It is of great importance that research activities are assessed according to the standards of the specific discipline. The specific character of each field may require emphasis on some elements of the SEP, while other elements may be less relevant to a certain discipline. The fields of languages & culture, humanities & social sciences, natural & life sciences, medicine & health sciences, design & engineering and agriculture & food sciences may each require different approaches to the evaluation. Within these fields, approaches may also vary among sub-disciplines. While the description of the evaluation criteria and information requirements in the SEP are based on the common ground of these disciplines, the institute is invited to present the specific characteristics of the discipline that inform its research and identity in its self-evaluation and facts & figures".[63]

- Taking interdisciplinarity into account

"Research in the Netherlands and abroad is of an increasingly multi-, inter-, or trans-disciplinary nature. Institutes and research programmes with multi-, inter-, or transdisciplinary research require special attention in the evaluation. It is, for instance, often more difficult for these groups to show their results through traditional indicators based on publications in high impact journals, and therefore review committees should include evaluators who have a solid experience in assessing such research. The board responsible for evaluating multi-, inter- or transdisciplinary research should therefore see to adaptations in the standard procedures necessary to assess these particular aspects of an institute's mission, for

[63] SEP 2009-2013, p. 6

example with regard to the composition of the evaluation committee or to specific, more qualitative, criteria and indicators".[64]

6.2.1.3 Evaluating Research in Context

For years, politicians and those involved in scientific policy have been interested in the societal impact of scientific research. In recent years, however, concern for this topic has grown. The Lisbon Agenda with its focus on the knowledge economy and the growing attention to the valorisation of scientific research – the 'third mission' of universities and public research institutes – are the driving force behind this. Within the framework of quality assurance an important question is: how exactly can this societal impact be measured? And how can it play a role in research and science policy? Assessing the social impact of research is a relatively new phenomenon and covers a broader description than is currently requested. In order to promote the measurement of societal impact, a new platform, Evaluating Research in Context (ERiC), was set up in 2006.

ERiC emerged from a project of the Consultative Committee of Sector Councils for Research and Development concerned with how to measure the social impact of research. This project yielded the successful measuring method sci_Quest. Following on from this, a broader platform representing the full spectrum of higher education was set up, the EriC platform. Since then the Royal Netherlands Academy of Arts and Sciences (KNAW), the Netherlands Organisation for Scientific Research (NWO), the Netherlands Association of Universities of Applied Sciences, and the Quality Assurance Netherlands Universities have participated in the project, while the Hogeschool Utrecht, the Ministry of Education, Culture and Science and the Rathenau Institute have been involved as observers.

ERiC's main objectives are stimulating the exchange of knowledge and developing a methodology at both a national and international level. In order to do so, ERiC undertakes various activities:

- ERiC develops and disseminates information about how to measure the social impact of research
- ERiC raises awareness of the possibilities for assessing the social impact of research through workshops, seminars and the website
- ERiC develops methods for measuring the social impact of research, by carrying out projects with universities and universities for applied sciences
- ERiC carries out several pilots at various universities, in order to 'test' the methods that are being developed
- ERiC responds to issues faced by the research community

One of the successes of ERiC is that in the latest SEP (used from 2009 onwards) there has been a stronger emphasis on measuring the societal impact of research. In brief, the SEP no longer just assesses the scientific quality of research but also uses indicators to measure its societal impact. Examples are cooperation with the private sector and memberships of social organisations and policy bodies.

6.2.1.4 Review of SEP

Research assessment and quality assurance are generally well developed in the Netherlands, and this country is often looked upon as a positive example. SEP is broadly accepted and utilised by the publicly funded research organisations. The new quality assurance system has also been effective. The quality of research has improved in three respects: in general, the Netherlands scores well internationally, low-quality

[64] SEP 2009-2013, p. 6

research has been weeded out, and the assessment system has become more transparent.[65]

Although the system is believed to function well, it recently has become the object of criticism (after a critical report of the Meta Evaluation Committee in 2008 about the evaluation of SEP over the years 2003-2009). One of the reasons for this criticism is the fact that it is hard to discover what institutions actually do with the results of assessments. University administrators seem to consider it inexpedient for various reasons to be transparent about this matter. When questioned, they admitted that little was in fact done with the results of assessment in the current context, either because financial or employment law offered no scope for this, or because of a lack of decisiveness. Also the government does not translate the outcomes of the quality assessments into its budget allocations. In essence, the higher education institutions themselves and their professionals are playing the leading role in quality assurance and the government has no means to interfere. It made the acceptance of the system easy, but the downside is that it is not a steering instrument for the government.

A second reason for concern is that subsidising bodies regularly circumvent the SEP and make use of protocols of their own that differ from those of the SEP. Over time new arrangements have arisen, such as the national research schools or (public-private) innovation programmes which both have their own evaluation systems. Overall there is an accumulation of procedures that differ somewhat from one another, thus putting additional pressure on researchers, who can consequently devote less time to the primary process of research. The critics underline that the administrative burden for researchers should be diminished.

A third critical remark is that institutions are free to choose the units to be assessed and they are also free to include societal relevance and specific management aspects. This has led to a whole range of units open to assessment, varying from whole disciplines to faculties (and parts of faculties), research institutes, and research schools. This multiplicity of evaluation research units makes it almost impossible to benchmark them.

The fourth point is that external evaluation shows that the definition of 'relevance' is not clear. SEP does require attention to social and economic relevance, but the criterion 'relevance' stands for both scientific and societal relevance. The Meta Evaluation Committee points out that more emphasis should be placed on societal relevance.

Last but not least an erosion of scores (by the external committee) is recognized. Both the overall scores and the scores at the programme level are often very high (four or five) – an observation also made in the UK RAE, where the phenomenon is called 'grade inflation'. This could be the indication of the good quality of the Dutch research groups, but also the erosion of scores. A low score indicates that a group is underperforming while a high score does not necessarily stand for an excellent research group (from an international perspective).

6.2.2 Research funding in the Netherlands

6.2.2.1 Funding of scientific research

The amount of funding provided for scientific research in the Netherlands was estimated in 2006 at some EUR 9 billion; approximately half of this is provided by businesses. Government is the second-largest provider of funds, accounting for more than a third. Over the years, the proportion of funding derived from foreign sources has increased substantially, now accounting for more than 10%, with most of that coming from foreign companies.

[65] KNAW, *Evalueren. Het beoordelen van wetenschappelijk onderzoek in de praktijk* (2009)

Government funding for scientific research carried out in the Netherlands is provided in a number of different ways:

- Provision of a fixed contribution to institutions ('institutional funding' or 'basic funding'), for which there may or may not be management responsibility;
- Funding of research via intermediary organisations (such as NWO, KNAW, and SenterNovem);
- Funding of research via the ministry's own knowledge institutes (research organisations), for example at the Ministry of Justice and the Ministry of Health, Welfare and Sport;
- Direct funding of policy-oriented research;
- Funding of international organisations (some EUR 200 million, goes to international organisations like CERN or ESA or to foreign researchers, specifically from the Ministry of Foreign Affairs (Department for Development Cooperation).

6.2.2.2 University funding

In the Netherlands there are two main types of regular higher education, namely research universities and universities of applied sciences (*UAS)*. The research universities focus on providing scientific instruction and conducting scientific research, while the UAS specialise in technical and vocational training.

Research universities

In the funding of research universities, three flows of funding are distinguished, defined by their origin. The first flow concerns direct government funding, the second consists of research project funding (competitive funding), and the third is money coming from third parties, both public and private (such as companies, the EU etc.). The latter can be contract research but also public money for research programmes and projects (mostly on a competitive basis). The research part of the first flow constitutes the largest part of research funding – about 60% of all research funding (€1647.8 million in 2007) – and its only structural basis. Money coming in via the second (10%) and third flows (30%) – together accounting for 40% of university funding for research, the second flow being about a third of the size of the third flow – consists of temporary funds.

First flow

The first flow – institutional funding – is a payment by the Dutch Ministry of Education, Culture and Science (in the case of Wageningen University and Research Centre by the Dutch Ministry of Agriculture, Nature and Food Quality).

The lump sum (block grant) allocation is based on different measures, known as the *BAMA model*, named after the BA and MA degrees that were introduced from the year 2002 onwards. The BAMA model is largely formula-based; it distributes a given sum of money (set by Parliament) across the 13 research universities. The formula takes into account the relative performance of each university (as compared to the other universities). Unlike in Austria and Finland, the formula is not combined with performance contracts.

The BAMA allocation consists of a *teaching* component and a *research* component, but this distinction is for calculation purposes only. In fact, the Executive Board of the university is free to use its own model in distributing the first stream funding (and the tuition fees received directly from its students) across teaching and research activities. The teaching component is 42% of the lump sum (excluding the Academic Hospital allocation), and the research component makes up the remaining 58%. The *teaching component* consists of

- a new entrants allocation (about 15% of the teaching grant)
- a diploma (BA/MA) based allocation (about 60%)

- a basic allocation (about 25%)

For individual universities, these shares may differ, due to their relative performance. The *basic allocation* consists of fixed amounts per university. Actual amounts differ across universities; they have a historical basis. The emphasis on performance increased in 2000, as degrees received a higher weight in the formula.

The *research component* of the BAMA funding model consists of six parts:

- a strategic component (same amount of money every year based on history)
- basic funding for each university depending on the number MA diplomas, allocation for PhD degrees and designer certificates,
- allocation for (top) research schools
- a dynamic fund (sum taken away from basic funding and reallocated based on the success in the 2nd and 3rd funding stream)

Figure 21 shows the first flow of funds for research comprising six components for the year 2006.

Figure 21 Research component of BAMA funding model in 2006

Compartment	Bases (indicators)	Size (in million €)	%
Strategic component	Historical basis	853.9	60.3
Basic funding	Number of MA diplomas	245.9	17.4
Dissertations	Number of finished PhDs	175.0	12.4
Research Schools	Participation in research schools	45.3	3.2
Top Research Schools	Participation in top research schools	45.4	3.2
Dynamic fund	Success in 2nd and 3rd funding stream	49.9	3.5

Source: VSNU (Association of Universities in the Netherlands)

Second flow

The second flow of funds consists of research council funding. The Netherlands Organization for Scientific Research (NOW, the national research council) is responsible for allocating the second flow. NWO receives funding from the Education ministry and the Ministry of Economic Affairs (the latter supports the natural/technical sciences). NWO then awards project funds after reviewing the research proposals submitted by researchers. Competition for this type of prestigious funding is high. Only universities can win competitive research council grants. In recent years the first flow funding was cut back (M€100 less annually) by the new cabinet. This money was added to the NWO budget to support excellent individual researchers through competition-based funding (second flow). However, still the second flow is relatively small.

NWO has several forms of funding:

- Free competition: bottom-up competition for the advancement of innovative and high-quality scientific research.
- The Innovational Research Incentives Scheme: this scheme is directed at individual researchers at various stages of their careers. It includes three forms of grant: Veni (for researchers who have recently completed their doctorates), Vidi (for experienced researchers) and Vici (for researchers of professorial quality).
- Research facilities: two funding schemes are dedicated to research facilities (large and medium size facilities)
- Thematic programmes: the thematic programmes are intended to create focus and mass and they cover multidisciplinary research programmes in certain (limited) subjects.

Third flow

The third flow of funds consists of a heterogeneous mix of revenues from activities such as contract research (approximately half of the third stream), contract teaching, consultancies, research commercialization, endowments and renting out university facilities. Clients are private businesses, government, non-profit organizations and the European Union, as well as individual students and staff.

6.2.2.3 Funding of the universities of Applied Science

The funding model for the UAS so far has remained largely intact from the early 1990s onwards. Until now UAS receive recurrent funds on the basis of a formula. The formula takes into account the number of registered students, the number of graduates (BA only), the number of dropouts, and the duration of enrolment for students. The funding formula stresses performance, especially in terms of graduation rates.

The UAS budget covers recurrent costs (mostly staff) and the costs related to infrastructure. UAS so far do not receive recurrent funding for research, although they can apply for competitive funds set aside for applied research that is closely connected to the professional training they do. Six 6 years ago funds were introduced for a new staff position in UAS: the *lector*. This is a kind of associate professor who is expected to engage in practice-oriented research collaborating with the regional business sector. Special funding is awarded to UAS through an intermediate body (a foundation).

The UAS receive a amount of money form the ministry for each student. The amount of money depends on the period of registration and the success rate (diploma or premature departure) (see Figure 22). There is a strong monetary incentive for UAS to make sure that students finish their degrees.

Figure 22 Funding per student at UAS

	If a student leaves the UAS prematurely (without diploma)			If the student succeeds and gets a diploma		
Period of registration	**Funding from government (in €)**	**Tuition fees (in €)**	**Total (in €)**	**Funding from government (in €)**	**Tuition fees (in €)**	**Total (in €)**
1st year	5.900	1.400	7.300	5.900	1.400	7.300
2nd year	4.500	2.800	7.300	7.400	2.800	10.200
3rd year	3.100	4.200	7.300	20.300	4.200	24.400
4th year	1.800	5.600	7.400	18.900	5.600	24.400
5th year	400	7.000	7.400	17.500	7.000	24.400
6th year	-1.000	8.400	7.400	16.100	8.400	24.500
…						

Source: Ministry of Science, Education and Culture

6.2.3 Outlook

A new harmonized funding model for both universities and UAS will be introduced in 2011. This funding model will be based on number of students registered (60%), the number of diplomas (20%) and a block grant that is not formula-based (20%). In the new system, the universities will not receive funding for students who already have a degree and want to do a second degree. In the old system there is a barrier of 30 (i.e. no funding for student aged 30+). Also the government might introduce a bonus for the best performing universities on top of the block grant.

The rationale for the reform is that the current system is very complicated, so that hardly anybody in the Netherlands really understands it. This reduces the incentivising function of the formula. Other aims are to attract more students (30+) and a better positioning of Dutch universities in the international competition. Finally,

the current funding system contains some oddities, like the strategic component for research funding which is not strategic at all but the same amount of funding allocated every year based on historical grounds. Such oddities will be eliminated in the reform.

The whole thinking is to make the funding systems less complex, more transparent and reward the best institutes (for example with a bonus). During the last years, there has been a debate about a new system, and several committees have been established. In the end it has been decided to postpone the new funding system (i.e. a new law for higher education) as there is a new government in power.

6.2.4 Lessons learnt from the Dutch case study

- A broadly accepted quality assurance system is very important: it can be very effective in increasing the quality of research.
- The four criteria of SEP – quality, productivity, relevance and viability – seem to be a very useful framework for evaluating institutions and research groups.
- There should be attention of the red tape involved with the evaluation system. It should be 'lean and mean'.
- Continuity is important, but the evaluation system should be evaluated on a regular basis in order to adopt new developments (like the emphasis on valorisation and societal relevance) and to improve the system as a whole.
- The Dutch evaluation system is not linked to the funding system but the conclusions of the evaluations are very well adopted by the research communities itself. A close link can encourage strategic behaviour rather than stimulating excellence (see RAE UK).
- Complexity of the formula has increased over time so that it has lost its incentivising function. Simplicity in the formula is important!

6.3 United Kingdom[66]

6.3.1 Introduction

In the UK research quality is evaluated in the Research Assessment Exercise. The system was introduced in 1986; subsequent research assessments took place in 1989, 1992, 1996, 2001, and 2008. The UK was one of the first countries both to

[66] This case study is based on the following literature: Evaluation Associates Ltd, Interdisciplinary Research and the Research Assessment Exercise, a report for the UK Higher Education Funding Bodies, March 1999, http://www.rae.ac.uk/2001/Pubs/1_99/ (3/1/2011); Ian McNay, The Research Assessment Exercise (RAE) and after: 'You never know how it will all turn out', In *Perspectives: Policy and Practice in Higher Education*, 2(1), 19, 1998; Review of research assessment, Report by Sir Gareth Roberts to the UK funding bodies, May 2003; Valerie Bence and Charles Oppenheim, 'The Evaluation of the UK's Research Assessment Exercise: Publications, Performance and Perceptions', in Journal of Educational Administration and History, 2005, 37 (2), 137-155; Higher Education Funding Council for England, Scottish Higher Education Funding Council, Higher Education Funding Council for Wales, Department for Employment and Learning Northern Ireland, Research Assessment Exercise 2005, Guidance on submissions, Ref RAE 03/2005, June 2005; RAE, Generic statement on criteria and working methods, 2006; Jochen Gläser, 'The social orders of research evaluation systems', in Richard Whitley, Jochen Gläser (ed), *The Changing Governance of the Sciences*, 2008, pp. 245-266; Ben R. Martin, Research Assessment in the Management of Research and Innovation Systems, presentation held at the Round Table on 'International Practices in the Governance and Management of Research and Innovation Systems, Czech Ministry of Education, Prague, 28 April 2010; B.R. Martin and R. Whitley, 'The UK Research Assessment Exercise: a Case of Regulatory Capture?', in R. Whitley, J. Gläser and L. Engwall (eds), *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation,* Oxford University Press, Oxford, 2010, pp.51-8); Linda Butler, Impacts of Performance-Based Research Funding Systems: A review of the concerns and evidence, OECD, 2010; Hanne Foss Hansen, Performance indicators used in performance-based funding systems, OECD, 2010. In addition, the following websites were consulted: http://www.rae.ac.uk/ (30/12/2010), http://www.hefce.ac.uk/ (7/1/2011)

institutionalise university research assessment over 20 years ago *and* to link it to the allocation of institutional funding.

The RAE looks at performance based on the quality and volume of research. It is largely peer-review based and has been so since the beginning in 1986. The RAE was perhaps the first occasion where peer review was extended to assess entire university departments on a systematic nation-wide basis.

The UK system is the oldest performance-based research funding system and hence the best researched system. There is a wealth of material available on the UK RAE, making this case study a bit longer than the other international case studies.

6.3.2 Rationale and aim

A first driving force for introducing research assessment was the growing costs of research at a time of increasing pressures on public expenditure in the 1980s. Rapid rises in costs and overheads meant that it was increasingly important for universities to maximise their research incomes. Hence, the aim was to have a mechanism for providing greater selectivity in the allocation of institutional funding.

A second major driving force consisted of political demands for greater accountability and better 'value for money'. This demand was associated with a shift to a more managerialist approach to the public sector and to the development of what became known as 'New Public Management' (NPM). Nowadays governments require accountability mechanisms for almost all areas of public spending. In the UK, the political pressures for accountability and for selectivity and concentration have arguably been greater than elsewhere, an apparent legacy of the 1980s Thatcher government.

6.3.3 History of the UK RAE

The first Research Selectivity Exercise in 1986, as it was initially termed, was based on a relatively simple methodology. Each university department or 'unit of assessment' was asked to complete a questionnaire on research income and expenditure, research planning priorities and output. It also requested units to identify their five best publications from the previous five years. The responses were considered by subject subcommittees, along with a number of assessors. From the information submitted, the subject committees classified units on the basis of a simple four-category ranking (outstanding, above average, average, below average). At the time, the methodology was heavily criticised, the main reason being that the five best publications biased the RSE in favour of big departments.

A second RSE was carried out in 1989, this time with a more sophisticated methodology. Units were asked to supply publication data including bibliometric details of up to two publications for each full-time member of faculty. Units were also asked to provide data on research studentships, research grants, and contracts, and total numbers of publications in relation to the number of full-time staff. Approximately seventy discipline-based panels were established to evaluate each subject (or 'unit of assessment'). An attempt was made to standardise ratings across subjects by asking panels to judge groups on the basis of 'attainable levels of excellence' but in the event there were large variations between the average ratings awarded by different panels. The early RAEs demonstrated that methodologies for assessing research output were the main areas of difficulty for both academics and assessors alike. Problems centred on what to collect and what measurements to use.

By 1992, the RAE approach had begun to 'settle down'. Although 'tweaked' in subsequent RAEs, the approach remained broadly the same from here onwards. In 1992, units were given the opportunity to include only 'research active' staff, as those units with a long tail of weaker researchers had reportedly been penalised with a lower rating in earlier exercises. Units of assessment were also asked for information on the research environment and future plans, and for quantitative data on *all* publications classified under specific headings. However, according to the Higher Education

Funding Council for England, the latter proved not very useful because only limited use was made of the lists of publications and put the focus on quantity rather than quality.

By 1996, the RAE had moved away from quantity measures, i.e. collecting full lists of publications. In contrast, units were requested to submit up to four publications per member of 'research active' staff. They could also include 'indications of peer esteem' (such as journal editorships and invited conference presentations). Submissions were assessed by some sixty subject panels. The chairs of these were appointed by the funding councils, with other panellists being selected from nominees put forward by 1,300 learned societies, professional bodies, and subject associations.

In 2001, panels were expected to treat each publication on its merits, i.e. irrespective of the medium of publication such as the journal and its status, imposing a huge reading burden on panellists. As a consequence, some panels decided to take a short-cut and assume that work already subjected to rigorous peer review (e.g. that published in 'top' journals) should be given more weight. After 2001, there was a growing sense that the RAE had become much too cumbersome and costly.

The main change in the 2008 RAE was a switch from a single rating on a seven-point scale to a 'quality profile' for the research of each department, based in large part on what proportion of its publications were judged to be of national or international quality, but also taking account of other data included in the submissions (such as esteem indicators). The next section explains the 2008 RAE in more detail.

6.3.4 The 2008 RAE

6.3.4.1 Introduction

The RAE 2008 was conducted jointly by the Higher Education Funding Council for England (HEFCE), the Scottish Higher Education Funding Council (SHEFC), the Higher Education Funding Council for Wales (HEFCW) and the Department for Employment and Learning, Northern Ireland (DEL). The four higher education funding bodies use the quality profiles produced in the RAE to determine their grant for research to the institutions which they fund with effect from 2009-10.

6.3.4.2 RAE 2008 Panels

As with previous RAEs, the assessment process was based on expert review: panels used their professional judgement to form a view about the quality profile of the research described in each submission, taking into account all the evidence presented.

For the purpose of the 2008 RAE each academic discipline was assigned to one of 67 units of assessment. Work submitted to the exercise was assessed by experts, drawn from universities and the wider research community. To ensure greater consistency, a formal two-tiered panel structure was introduced in the 2008 RAE: 67 sub-panels of experts, one for each unit, worked under the guidance of 15 main panels. Each main panel included broadly comparable disciplines whose subjects had similar approaches to research. This system provided a strategic overview of the work of the sub-panels and a more consistent approach both to setting criteria and to the assessment of work in related fields.

Main panels were made up of a chair, the chairs of each of the sub-panels within the main panel area, and a number of international and additional members. The international membership of the main panels ensured that international standards were maintained consistently across the exercise. Each sub-panel had a chair and on average about 15 other members, who had expertise that covers the full range of research in that subject area. The main panels were responsible for:

- reviewing and endorsing the criteria and working methods to be used by the sub-panels
- deciding on the quality profile to be awarded to each submission

- communication and joint working with the other main panels.

The sub-panels were responsible for:

- preparing draft statements of relevant criteria and working methods
- undertaking the detailed assessment of submissions from universities
- making recommendations to main panels on the quality profiles to be awarded for each submission.

Panel members were nominated by subject associations and other stakeholder organisations, including users of research. They were appointed by the UK higher education funding bodies, which received just under 5,000 nominations from almost 1,400 bodies. In the end, there were over 1,000 panel members. They were chosen for their standing in the academic and wider research community, their extensive research experience, and their understanding of the needs of research users and commissioners of research from both the public and commercial sectors.

In January 2006, panel criteria and working methods were published for each panel. In November 2007, the UK funding councils published a complete list of the panel membership. Panel membership was also regularly updated via the RAE web page to take into account any changes in membership since the publication of the hard copy list. Panels assessed submissions between January and November 2008.

6.3.4.3 Submissions

For the 2008 RAE, 2,344 submissions were made by 159 universities. For each submission, universities provided data about research activity. Each submission contained the following data:

- Overall staff summary: summary information on research active staff selected (FTE and headcount) and related academic support staff (FTE) in the unit of assessment
- Research active individuals: detailed information on individuals selected by the institution for inclusion as research active
- Research output: up to four items of research output (e.g. books, chapters in books, journal articles, conference contributions) produced during the publication period (1 January 2001 to 31 December 2007) by each individual named as research active and in post on the census date (31 October 2007)
- Research students: numbers of full-time and part-time postgraduate research students and degrees awarded
- Research studentships: numbers of postgraduate research studentships and source of funding
- External research income: amounts and sources of external funding
- Textual description: including indicators of esteem and information about the research environment. The latter included:
  - information about the university's strategic investment in the unit of assessment
  - their strategies for promoting and developing research staff, particularly those new to research
  - their strategies in relation to collaborative research with academic and non-academic bodies and with overseas universities
  - their strategies in relation to interdisciplinary research
  - the significance of their research on a range of academic and other audiences, including, where appropriate, practitioners, businesses and other users.

- Individual staff circumstances. Information describing any individual staff circumstances that significantly adversely affected their contribution to the submission.[67] Panels used the information supplied confidentially in assessing submissions against their published criteria. This information was not published.
- Category C staff[68] circumstances. Information demonstrating that staff research is clearly and demonstrably focussed in the department.

All information provided by universities in submissions to the 2008 RAE had to be capable of verification. A proportion of submissions from each institution was checked as a matter of course. Panel members were also asked to draw attention to any data that they wanted to have verified. Universities therefore had to be able to provide detailed justification for all information submitted.

Where possible, submission data was compared with other available datasets including Higher Education Statistical Agency (HESA) returns, other surveys conducted by the four funding bodies, and information held by the Research Councils.

Sub-panels assessed the research submitted against agreed quality standards, which were established within a common framework that recognised appropriate variations between subjects in terms of both the research submitted and the assessment criteria. Sub-panels then made a recommendation to the main panel for endorsement.

An underpinning principle was that sub-panels should assess each submission in the round: The were not meant to make collective judgements about the contributions of individual researchers but about a range of indicators relating to the unit, research group or department put forward for assessment.

The sub-panel's judgement indicates the proportion of the research that met each of four quality levels or is unclassified. Figure 23 illustrates the different quality levels and their definitions.

Figure 23 Definitions of quality levels

Quality level	Definition
Four star	Quality that is world-leading in terms of originality, significance and rigour.
Three star	Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence.
Two star	Quality that is recognised internationally in terms of originality, significance and rigour.
One star	Quality that is recognised nationally in terms of originality, significance and rigour.
Unclassified	Quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purposes of this assessment.

Source: HEFCE et al., RAE 2008, Guidance on submissions, June 2005

'World-leading' quality denotes an absolute standard of quality in each unit of assessment. Each submission was assessed against absolute standards and was not ranked against other submissions. Work that fell below national quality or was not recognised as research was unclassified. The 'international' criterion equates to a level of excellence that it was reasonable to expect for the unit, even though there may be no current examples of such a level in the UK or elsewhere. It should be noted that 'world-leading', 'international' and 'national' refer to standards, not to the nature or geographical scope of particular subjects.

[67] Such circumstances might include, but are not restricted to: matters covered by legislation including maternity leave, part-time working, engagement on long-term projects, early career researcher status, prolonged absence due to secondment, career break, ill health or injury

[68] Independent investigators active in research who do not meet the definition for Category A staff (i.e. academic staff in post and on the payroll of the submitting institution on the census date) but whose research on the census date is clearly and demonstrably focussed in the department

The RAE quality profiles present in blocks of 5% the proportion of each submission judged by the panels to have met each of the quality levels defined (Figure 23). In each case, the panel took account of three overarching components of the submission – research outputs, research environment and indicators of esteem. The results are published as a graded profile for each unit for each submission. An example of quality profiles for two hypothetical submissions can be found in Figure 24.

Figure 24 Sample quality profile*

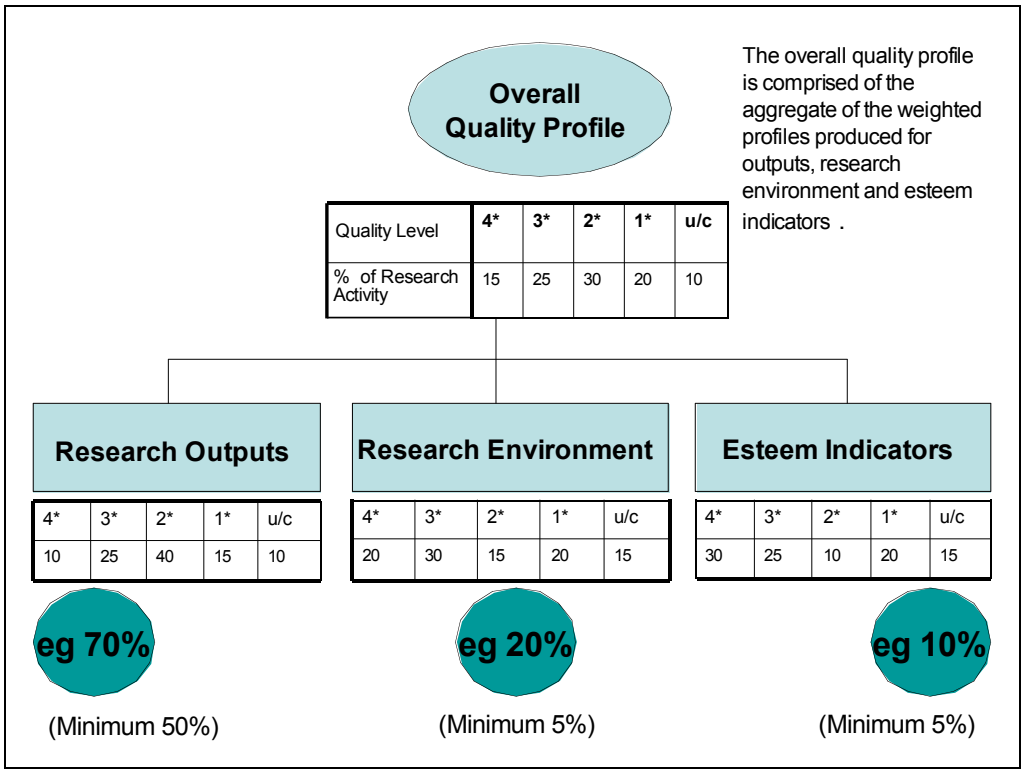
Unit of assessment A	FTE Category A staff submitted for assessment	Percentage of research activity in the submission judged to meet the standard for				
		Four star	Three star	Two star	One star	Unclassified
University X	50	15	25	40	15	5
University Y	20	0	5	40	45	10

* The figures are for fictional universities. They do not indicate expected proportions.

Source: HEFCE et al., RAE 2008, Guidance on submissions, June 2005

Figure 25 shows how the three overarching components of the submission – research outputs, research environment and indicators of esteem – are built into a quality profile for a unit of assessment. For each quality profile, the weighting of each of the three overarching components was up to the panels. However, minimum weightings for each of the components had been defined, with research outputs making up at least 50%, and research environment and esteem indicators at least 5% of a quality profile.

Figure 25 Building a quality profile



Please note that the percentage weightings to the three elements are illustrative. Panels should allocate these. The minimum weighting for the research outputs profile is 50%. In this example the overall quality profile shows 15% of research activity is at 4* level. This is made up of 70% x 10 (research outputs), 20% x 20 (research environment) and 10% x 30 (esteem indicators) or more precisely: (0.7*10)+(0.2*20)+(0.1*30)=14, rounded to blocks of 5%.

Source: HEFCE et al., RAE 2008, Guidance on submissions, June 2005

The results of the 2008 RAE were published on 18 December 2008. They are available on a website[69], either by subject unit of assessment or by institution. Alongside the quality profile, the funding bodies also published the full-time equivalent (FTE) number of staff included in each submission (submitted staff). The results of the RAE are not subject to appeal.

6.3.5 Allocation of institutional funding over time

Every succeeding RAE has seen extensive consultations and considerable changes, with effects on funding becoming progressively more marked. While the effects of the 1986 RSE on institutional research funding were initially rather limited, the Universities Funding Council decided that in the 1989 RAE there would be a larger, more explicit link between rankings and research funding. Of the 33% of institutional research funding, nearly half was allocated on the basis of the 1989 ratings, with the remainder being determined mainly by student numbers.

By 1992, over 90% of research funds (QR) were distributed by the funding councils[70] on the basis of RAE ratings, with no funds being awarded to units given the lowest grade (research was graded on a five-point scale). The amount given depended on the quality rating, the number of staff declared 'research active', the amount of research income and an element on future research activity.

In 1996, the eventual funding awarded was based on the overall quality grade multiplied by the number of research active staff in the department. However, a shortage of funding council funds meant that the founding councils had to introduce much larger differentials in funding, with no funds at all for the lowest two grades 1 and 2 (1, 2, 3B, 3A, 4, 5, 5*).

In 2001, because grades had gone up and the HEFCE budget remained the same, funding was maintained for 5* departments only, for all lower ratings it was cut. This meant that there were steep differentials between 3A, 4, 5 and 5* departments and much higher than in Scotland and Wales. There was no funding for 1,2 and 3B rated departments. As a consequence, low-ranking departments were less and less likely to bother submitting to the RAEs.

Over time, RAE results have been used to justify the progressive withdrawal of funds from lower-rated departments (with ratings 1, 2, and 3B). At the same time, the differential unit of resource between 3a, 4, 5, and 5* departments has sharply increased (particularly in England). The net effect has been to concentrate resources on top departments.

This development was reversed in 2008, when the results showed that the great majority of the units submitted ended up with an average 'score ' of between 2 and 3 (1*, 2*, 3*, 4*, unclassified), resulting in a 'flatter' distribution than in the previous RAEs. As a consequence, research funding has been spread slightly more widely, with some middle-raking universities experiencing the largest increased in 2009/10. This shift was thought to be a result of the use of quality profiles rather than single 'star' ratings for a department.

As in all RAEs, the exact relationship of QR funding to RAE results was only announced by HEFCE after the ratings were announced. Universities were thus forced to take part in each RAE without knowing what the financial consequences would be. Given this intrinsic uncertainty, universities often concentrated exclusively on achieving the highest possible grade for each unit (see follow-on effects of the RAE).

---

[69] www.rae.ac.uk

[70] University Funding Council by then had been replace by separate councils for England (HEFCE), Wales, Scotland, and Northern Ireland

Despite all the attention paid to the RAE, it is important to emphasize that in larger research-intensive universities, the scale of RAE funding in generally relatively small compared with other research income (e.g. less than 10% of research income). More generally, the RAE drives an allocation of funding for each institution that currently accounts for around one third of non-capital public spend in science and research allocated through the UK's dual support system[71].

6.3.6 Description of allocation of funding

6.3.6.1 Quality-related funding

The Research Assessment Exercise informs the distribution of funds by the UK higher education funding bodies[72]. Any UK university that is eligible to receive research funding from one of these bodies can participate in the exercise. Once funding levels for institutions have been set, these are used for the annual allocation of funding until the next round of RAE.

The results of the RAE are combined with other indicators to derive an allocation of funds for each institution. The funds allocated based on the RAE and other indicators are called quality related funding (QR funding), the largest of several types of institutional funding. The so-called 'mainstream QR' distributes a grant based on the quality, volume and relative cost of research in different areas.

The peer review provided by the RAE accounts for approximately 70% of the allocation. The other indicators in the funding formula are: volume of postgraduate research supervision (funding for research degree programme supervision); charity-related funding; volume of business research (business-related funding); and funding for national research libraries (Figure 26).

Figure 26 The components of quality related research funding 2010-2011

	QR funding (in £)	Share of total QR funding (in %)
Mainstream QR allocated to reflect the quality, volume and cost of research at institutions in different subjects	1,130 million	70.2
Research degree programme supervision fund allocated to support postgraduate research	205 million	12.8
Charity support element allocated on the basis of eligible charity research income	198 million	12.4
Business research element	64 million	4.4
National research libraries	6 million	0.4

Source: HEFCE

- **Funding for research degree programme supervision.** This is allocated to reflect postgraduate research numbers and the relative costs of the subjects they are studying.
  - Funding for RDP supervision is provided on the basis of postgraduate research student FTEs in all departments that receive mainstream QR funding for research. The allocation is based on the numbers of eligible students in each department, weighted by the subject cost weights and the London

[71] I.e. block grant ('lump sum') plus project funding

[72] The Higher Education Funding Council for England (HEFCE), the Scottish Funding Council (SFC) for Scotland, the Higher Education Funding Council for Wales (HEFCW) in Wales and the Department of Education and Learning Northern Ireland (DELNI) for Northern Ireland. Total allocated funding for the four UK HEFCs in 2007/08 was £2,170 million

weighting percentages (see below). There is no minimum quality threshold for eligibility for this funding.

- **Charity-related funding.** Many charities (e.g. the Wellcome Trust) support research in higher education, particularly in medical disciplines, but they are not always able to meet the full economic costs of research. HEFCE therefore provides additional funding to institutions in proportion to the income they receive from charities for research.
  - Funding is provided on the basis of research income from charities in departments that receive mainstream QR funding for research for 2010-11. The allocation is made in proportion to the amount of eligible income from charities reported in the 2008 and 2009 Research Activity Surveys, weighted for London weighting (see below). There is no minimum quality threshold for eligibility for this funding.
- **Business-related funding.** HEFCE also provides funding to support institutions undertaking research with business and industry. This is allocated in proportion to the income they receive from business for research.
  - The allocation is based on the amount of research income institutions receive from UK industry, commerce and public corporations, using data for the two most recent years available reported by institutions on the HESA Finance Statistical Record.
- **Funding for national research libraries.** This is allocated to five research libraries on the basis of a review carried out during 2007.

Institutions receive most of their funding as a 'block grant' (lump sum). They are free to spend this according to their own priorities. HEFCE does not expect universities to model their internal allocations on its own calculations because universities are autonomous bodies that set their own strategic priorities.

The block grant does come with strings attached though. HEFCE has a Financial Memorandum which sets out the general requirements made of institutions, and each year it issues a funding agreement that details further particular conditions. These include e.g. targets relating to student numbers. Accountability is in terms of what is delivered, by way of the numbers of students taught and audit outcomes and the quantity and quality of research.

To help maintain stability, HEFCE phases in changes by moderating increases or decreases in teaching and research funding to institutions that would otherwise be affected by large fluctuations. So-called **moderation funding** is provided to smooth year-on-year changes in teaching and research grant and to ensure any reductions are manageable for institutions. HEFCE does not provide moderation funding if it amounts to less than £100,000 for an institution. The moderation fund was £20 million for 2010-11.

6.3.6.2 Components used in mainstream QR funding

There are a number of different components used in HEFCE's mainstream QR funding method.

- **A volume measure:** The volume measure in HEFCE's research funding method is the number of research-active staff employed by the institution (counted in FTE terms) multiplied by the proportion of research that meets a 2*, 3* or 4* quality threshold in the RAE (Figure 24). These staff numbers are fixed between RAEs at the so-called census date, to ensure they are consistent with the quality ratings. This means that there is a change in volume at each RAE.
- **A quality profile:** The quality profile is based on the 2008 RAE. For each submission made, the panels determined a quality profile, identifying what proportion of the research met certain quality thresholds. These were on a five-point scale (Figure 23), ranging from four-star ('world-leading') to unclassified.

- **Subject cost weights:** There are three cost weights.
  - High-cost laboratory and clinical subjects 1.6
  - Intermediate-cost subjects 1.3
  - Others 1.0
  - The subject-related cost weights have been determined from analysis of data on expenditure in different subjects reported by institutions to HESA. For research, HEFCE identifies the relative expenditure per staff FTE. This approach is consistent with the volume measure used in the funding method.
- **London weighting:** HEFCE also provides a London weighting as a percentage of the funding calculated for mainstream QR, to reflect higher prices in London. This is 12% for institutions in inner London, and 8% for institutions in outer London.

6.3.6.3 Calculation of mainstream QR funding

First, HEFCE determines how much funding to provide for research in different subjects, and then divides the total for each subject between institutions. These decisions take account of the volume of research (using research-active staff numbers), the relative costs (reflecting, for example that laboratory based research is more expensive than library-based research), any government policy priorities for particular subjects, and the quality of research as measured in the RAE.

More precisely, there are three stages to the allocation of mainstream QR funds.

- Stage 1: determining the amount provided for the 15 main RAE panel disciplines

HEFCE's first step in distributing mainstream QR is to decide how much to allocate to different subjects. The total amount has been divided between the subject fields of the 15 RAE main panels in proportion to the volume of research in each field that has been assessed as meeting, or exceeding, the 2* quality level in the 2008 RAE, weighted to reflect the relative costs of research in different subjects.

In line with government policy priorities[73], HEFCE then adjusted these totals for 2009-10, to ensure that the proportion of mainstream QR allocated to each main panel group in science, technology, engineering and mathematics (STEM) was not less than in 2008-09. HEFCE has continued this practice for 2010-11 and also extended it, in a more limited way, to geography and psychology, because around half of the research activity in these two disciplines could reasonably be regarded as more analogous to work in STEM disciplines than in the other social sciences.

- Stages 2 and 3: distributing the main panel totals between units of assessments and then institutions

The next steps are to disaggregate the totals for each main panel subject group between its constituent units of assessments, and then to disaggregate the totals for each unit of assessment between institutions. For both calculations, this is in proportion to the volume of activity assessed to reach each of the 2*, 3* and 4* quality levels in the 2008 RAE respectively, multiplied by quality weights (see Figure 27), and also taking cost weights into account where these vary within a main panel group. For example, in Figure 24, 15% of research activity of a submission made by the University of X is judged to meet the standard for 4*, 25% for 3* and 40 for 2*. HEFCE applies

[73] The previous Secretary of State wrote in his grant letter to HEFCE in January 2009 that he expected HEFCE to continue to recognise and respond to the high cost and national importance of STEM subjects. HEFCE therefore ensured that there was no reduction in the share of funding for each of the main panel areas associated with these subjects. The funding for the main panel areas in STEM subjects was increased as a result. This meant that non-STEM subjects did not get as large an increase in funding as they otherwise would have received, but they still had significant increases over the period 2002-03 to 2010-11

weightings to research volume attributable to each RAE quality rating, ranging from zero to nine, to ensure that funding of research is highly selective.

Figure 27 Research quality weights

Quality rates	Funding weights
4* (world-leading)	9
3* (internationally excellent)	3
2* (recognised internationally)	1
1* (recognised nationally)	0
Unclassified (below the standard of nationally recognised work)	0

Source: HEFCE

6.3.7 Effects of the RAE

One of the main – intended – directs effects of the RAE has been to make institutional funding more selective, concentrating it on those institutes that perform-high quality research (see section 6.3.5). However, there have been a number of other effects, both more positive and more negative.

Focused attention on publishing research results in good journals

Vastly different systems can have quite similar impacts. There are only limited ways in which actors can respond to the stimuli they are faced with. Increased publication output appears to be a common impact of performance-based research funding systems, irrespective of the model used.

Even though the focus of the 1996 RAE was on quality, nearly two-thirds of respondents to a survey agreed that it had encouraged them to publish more, and that the focus was predominantly on peer-reviewed journals. However, many researchers believe that the RAE has lead to premature publications. Work is being fast-tracked for publication and may not be as thoroughly researched as in the past.

Quality of research

There is a widely held belief that the RAE has been a driver of improved research quality in the UK and has made UK research more competitive and meritocratic. Bibliometric analyses confirm the improvement in UK science in terms of relative citation impact. The data shows a downward trajectory in citation impact until the introduction of quality-based assessments for the 1992 RAE. The improvement in UK research has continued with each subsequent round of the RAE. Although most bibliometricians caution against asserting that a causal relationship has been proved, the evidence is mounting and there are few doubts that this is what the data is revealing.

Similarly, in 1992, 23% of staff worked in 5 rated departments. This percentage rose to almost 31% in 1196 (in 5 and 5*) and rose again in 2001 to almost 55%. These figures seem to show a steady increase in the quality of UK research activity but they could also mean that institutions have learnt who, how and what to submit to the best effect. It could also simply be due to 'grade inflation', with panels becoming more generous in their assessments. There is also evidence that 5* tended to go to institutions represented on the panel.

More strategic approach to research and HR policy

In general, universities and universities departments have developed a more strategic approach to research. This has prompted some institutions to consider career paths for researchers, to provide bridging funds and to introduce more conscious management and appraisal of research staff. Since 1986, recruitment and promotion have become much more explicit components of institutional strategy, with recruitment and promotion of individuals increasingly taking account of (actual or potential) RAE contributions.

Repositories of knowledge

A particularly beneficial, unintended consequence of the introduction of the RAE is that it has led universities to expend considerable effort on improving or establishing institutional repositories for housing publications likely to be included in the assessment.[74]

Managing RAE submissions

By 2001, there was mounting evidence of universities actively managing RAE submissions, with many excluding staff with few publications as 'research-inactive'. In the extreme, this could lead to enforced 'early retirements'. Although such exclusions reduced a unit's 'volume factor', they could significantly increase the probability of a higher RAE rating.

A study by the association of University Teachers found that the selection of research-active staff was biased against women – men were 1.6 times more likely to be entered in the RAE when compared to their female counterparts.

Another study found significant impact on morale of staff and collegiality of institutions after the 1996 RAE when institutional submissions became more selective and only 'research active' staff were submitted. It was characterised as the most 'traumatic' effect of the RAE.

Transfer market

Another sign of active management of RAE submissions was the 'transfer market' that developed in the period before the RAE census date, where one institution would 'poach' staff from another institution close to the RAE census. One survey showed that only 2% of staff had moved institutions in the 18 months prior to the 1996 RAE. However, senior administrators commented that the 2% were all top-level staff. Similarly, a number of 2008 RAE panels believed that some universities excluded research-active staff in order to gain a higher ranking, while others drafted in 'research stars' on 'unusual' contracts who were not fully integrated members of the of the research team. But this can also be seen as a just reward for high achievers – something that was missing prior to the introduction of the RAE.

Disciplines

Generally, the main concerns in relation to the discipline mix of performance based research funding systems centre on the real or perceived advantages or disadvantages seen to flow to particular discipline groups. Any quantitative assessment is generally seen to favour science, technology and medicine subjects, at the expense of those in the humanities, social sciences and arts.

Some disciplines seem to have been quite well suited to the RAE (e.g. laboratory-based sciences or economics). In contrast, for many humanities and social sciences, the cultural and structural consequences were dramatic. In such fields, the RAE brought heightened emphasis on research 'output', on journal publications, and on bringing in research income to produce more outputs. Likewise, for professionally related disciplines such as engineering, law accountancy, and nursing, reconciling their work with the demands of the RAE often proved difficult.

There remain some concerns about the comparability of results across disciplines, particularly between the sciences and social sciences. These same concerns about comparability remain even *within* these two subject groups, and do not appear to have been alleviated by changes to the panel structures for the 2008 RAE (i.e. the introduction of the two-tier panel structure), which was designed to alleviate these concerns.

---

[74] The same effect can be seen in Australia

Interdisciplinarity

A major concern of performance-based research funding systems, based either on quantitative indicators or peer assessments made through panels convened along discipline lines, is the impact on interdisciplinary work. Trying to gain a clear handle on whether such systems are indeed biased against such research is extremely difficult.

There has been evidence that discipline-based panels have found it hard to assess interdisciplinary research. Yet, two studies undertaken focussing on this concern found no evidence to support the assertion that the RAE adversely affected interdisciplinary work. One study found that while the number of staff who reported that they had moved away from interdisciplinary work was relatively small, almost half of those surveyed felt that the RAE hindered interdisciplinary work. Another study found that while nearly a quarter of researchers believed that the RAE was a strong inhibitor of interdisciplinary research, in reality the ratings achieved by departments with a high proportion of interdisciplinary researchers showed that there was no such discrimination.

There was no doubt that many academics perceived the UK system to favour disciplines-based research, but clearly alternative policy drivers outside the RAE were ensuring that interdisciplinary work continued to thrive.

Collaboration

In a survey to the 1996 RAE, researchers felt that the RE did not encourage them to work more collaboratively, particularly with researchers or users outside the university sector. However, many countries have specific schemes aimed at stimulating collaborative activities. Most EU funding programmes also require extensive collaborative networks. These work to soften any adverse impact of the RAE on collaboration.

Effects on teaching

Because RAE incentives are perceived to be much stronger than those for teaching, many departments have witnessed a widening split between teaching and research, with negative effects on the former.

Research undertaken by the Association of University Teachers appears to substantiate claims that, in the lead-up to the 2008 exercise, less research-active staff were being moved across to teaching-only contracts. The belief was that this was an adverse reaction to the need to select 'research active' staff for submission to the RAE, and carried the implication that teaching-only staff had lower status in the university system than those with research activities.

Game-playing

The heightened sense of competition engendered by the RAE has encouraged widespread 'game-playing', i.e. efforts designed to make the department 'look good' while not necessarily improving research quality.

Shifts in power

The introduction of the RAE has lead to various shifts in power. In particular, the RAE has increased the authority of government, the funding councils, disciplinary elites, and university management.

- The continuing government emphasis on ever greater concentration of research resources has increased the degree of stratification in the higher education sector, leaving universities to compete ever more fiercely for government funding in a 'game' in which the state and its agencies, in particular the funding councils, set the rules and determine the financial outcome.
- The RAE has also strengthened the authority of the funding councils. In particular, they determine how RAE results are to be translated into funds. Over time, the

funding councils, and in particular HEFCE, have chosen to increase greatly the differential between 'excellent' research and the rest.

- The RAE has increased the power of disciplinary elites, who have been able to exert influence on research in their discipline, in particular on the criteria used for judging 'excellence'. This, together with the fact that most 'top journals' tend to focus on research in the disciplinary mainstream, means that the RAE has reinforced the emphasis on conventional mainstream research, discouraging new strands of research.
- The RAE has reinforced the shift from collegial governance towards 'managerialism' in most universities. The RAE has given university administrators a means of comparing the 'quality' of departments, and of legitimating the development of research strategies with differential funding between fields.

Departmental restructuring

One consequence of research assessment that has been directly linked to the RAE has been the internal restructuring of universities after the results of the latest exercise have been released. In some cases, universities responded to poor RAE results by closing departments.

Researcher autonomy

For individual researchers, the growing significance of the RAE means that they face more constraints on the type of research they choose to pursue. In a survey, nearly one-third of staff felt that the direction of their research was increasingly conditioned by the collective priorities of their group or department rather than their individual preferences. Changed and distorted behaviour of researchers, e.g. avoidance of and discouragement from non-mainstream or 'risky' research, have also been reported. Moreover, individual academics have been subject to growing publication pressures.

Follow-on effects

Performance-based research funding systems can have significant follow-on effects outside the immediate funding allocations. This is because such systems do not operate in isolation. Most countries operate dual support funding systems, with significant amounts of money distributed through Research Councils. The RAE results have considerable 'symbolic power, signalling to funding and other external bodies as well as the wider academic community where the 'best' departments are, and hence the relative standing of the institution in the status hierarchy.

6.3.8 Other public sector research

It is important to note that *only* universities are subject to the RAE while (sectoral) research institutes/government labs are not. The research organisation sector as a major performer of research comprises the **Public Sector Research Establishments,** i.e. government laboratories that have not been privatised and the institutes of some of the Research Councils[75]. These receive funding from the government in order to undertake research relevant to the respective policy needs of their sponsoring department.

The size of this sector has been considerably reduced in recent years through the privatisation or semi-privatisation of government laboratories. In addition, and partially as a consequence of this reduction, civil spending on R&D by Government departments has declined over recent years but remains substantial although it is now disbursed primarily on a competitive basis. Nonetheless, the former government laboratories remain the major performers of this research.

[75] Notably the Natural Environment Research Council, Biotechnology and Biological Sciences Research Council and Medical Research Council

6.3.9 Outlook

The mechanism by which the assessment is made has changed over time and has been the subject of protracted debate and a number of extensive reviews, for example the review led by Sir Gareth Roberts following the 2001 RAE. One of the major criticisms of the process is the enormous amount of staff time and resources that universities have to devote to the process of preparing RAE submissions.

Hence, in the 2006 pre-Budget Report the Government, "in recognition of the burden imposed on universities by the Research Assessment Exercise"[76], announced the development of a revised scheme for assessment of quality and allocation of funding: the **Research Excellence Framework (REF)**[77], which will replace the Research Assessment Exercise after 2008. The first proposal put forward by HEFCE suggested a largely metrics-based system – utilising bibliometric approaches and indicators of external research income generated and number of research students for the sciences, engineering, technology and medicine, and light touch peer review informed by metrics for the arts, humanities, social sciences, mathematics and statistics. After extensive consultation[78], taking into account the scientific community's reservations about a purely metrics-based approach, the proposal for the REF was modified.

Hence, the REF will also be a process of expert review, informed by indicators where appropriate. Expert sub-panels for each of 36 units of assessment will carry out the assessment, working under the guidance of four broad main panels. Units of assessments will be assessed in terms of:

- **The quality of research outputs:** This will continue to be the primary factor in the assessment. The quality of research outputs will be assessed by the expert panels against international standards of excellence. HEFCE expects that some of the panels will make use of citation information to inform their review of outputs and will make citation information available to some REF panels, as follows:
  - Each sub-panel will be invited to decide whether it wishes to use citation information to inform its review of outputs. Panels will set out in their criteria statements whether or not they will use such data, and if so how. This will be within a framework of central guidance, to ensure appropriate use of the data and to avoid any potential disadvantage to outputs for which citation data are unavailable.
  - HEFCE will provide citation data to those panels using them in a standardised and simplified format. The aim is to procure the data and make them available to panels in a way that is transparent and available in an appropriate form to institutions at minimal additional cost to institutions.
  - HEFCE will reconsider whether the benefits of incorporating citation information into the REF outweigh the costs if only a small minority of panels request citation information, the costs are high, or if the equalities implications cannot be effectively mitigated.
- **The wider impact of research.** Significant additional recognition will be given where researchers build on excellent research to deliver demonstrable benefits to the economy, society, public policy, culture and quality of life. Impacts will be assessed through a case-study approach. HEFCE has completed a pilot exercise to develop and test the method for assessing this.

---

[76] http://webarchive.nationalarchives.gov.uk/+/http://www.hm-treasury.gov.uk/pre_budget_report/prebud_pbr06/report/prebud_pbr06_repindex.cfm (Chapter 3: Meeting the productivity challenge, p. 57)

[77] The Research Excellence Framework is work in progress. This description is based on the system as it was planned in April 2010

[78] For the consultation see http://www.hefce.ac.uk/pubs/consult/outcomes/ref.asp (7/1/2011)

- **The vitality of the research environment:** The REF will take account of the vitality of the research environment in supporting a continuing flow of excellent research and its effective dissemination and application. This will include, for example, the research strategy, staff development and training of postgraduate researchers, and engagement with research users and the public.

HEFCE expects the outcomes of the overall assessment to be fine-grained enough to identify excellence 'wherever this may be found'. Panels will produce a sub-profile for each element (outputs, impact and environment), to be combined into an overall excellence profile. The profiles will show the proportion of submitted work at each point on a five-point scale, as used in the 2008 RAE (1* to 4* plus unclassified). Institutions will be invited to make submissions during 2013 and the assessment will take place during 2014.

Compared to the RAE, the new REF appears to be much broader, taking into account a number of aspects unacknowledged in the RAE. However, we doubt that the new REF will be 'lighter' than the RAE. On the contrary, given the inherent difficulty of identifying and attributing impacts, the burden and costs could even be even higher.

6.3.10 Lessons learnt from the UK case study

- The RAE places a huge administrative burden on universities. It is no surprise then that the periods between RAEs have become longer and longer. However, any system capable of providing the 'necessary fairness' will probably be relatively time and labour intensive.
- The RAE was criticised for being too narrowly academic. The new REF tries to take this into account by recognising all aspects of excellence in research, such as pure intellectual quality, value added to professional practice, applicability, and impact within and beyond the research community.
- Importance of peer review for evaluating the quality of research – rather than just using metrics.
- Importance of consultation: For the new REF there have been several consultation rounds. These rounds are important for the acceptance of the evaluation system (see also the Dutch case study).
- Importance of clear rules and transparent procedures
  - Panels:
    - Need for transparency in panel selection
    - Consistence of practice across panels vs. flexibility for assessors to develop methods appropriate to their subject
    - Ensure that panels adhere to their own criteria
  - Submission rules:
    - it is important to define what is assessable and how it is assessed.
    - It is also important to be clear which staff should be submitted in the assessment system ('transfer market', equal opportunities)
- The RAE has had profound effects – positive and negative – on the research system as a whole, on the way research is conducted, on research culture, on power structures in the research system. Such effects can be observed in Finland too.
- Probably the most direct effect is the concentration of funding in a relatively small number of excellent universities.
- With its moderation funds, HEFCE has built stability into the system. Moderation funds help universities manage changes.

- Political prioritising on science, technology, engineering, and maths (STEM) subjects. There is room for prioritising in Finland as well: 'science policy considerations'. Any system for allocating institutional funding needs to have room for political priorities.
- Importance of data: A large amount of data necessary to implement such a system.

6.4 Austria

6.4.1 Introduction: The Higher Education Sector in Austria

This case study describes the current legal regulation and practices of public funding and of research evaluation at Austrian public universities. The present status is based on a fundamental reform of the University Act in 2002 which granted far-reaching autonomy to universities and completely changed governance and financing as of 2004. The case study has to be mainly **descriptive** in nature because at the time of this study, the new funding system and its effects had not been evaluated yet.[79]

The higher education sector (HES) in Austria comprises public universities, the Austrian Academy of Science, Universities of Applied Science ('Fachhochschulen'), teacher training colleges and private universities. It is the second largest research performing sector in Austria, receiving 23.8%[80] of all R&D expenditures (70.6% are spent for R&D in the business sector). Within the HES, the 22 public universities play by far the largest role as research performers: they consume nearly 91% of the sector's R&D budget. The Austrian Academy of Sciences is the largest non-university academic research institution and receives another 5%, and 3% go to the Universities of Applied Sciences which are young and growing players with a focus on applied R&D[81]. It is fair to say that the 22 public universities are the backbone of basic research in Austria.

These public universities are a heterogeneous group of organisations in many respects, in terms of history, size, and subject specialisation:

- Five 'classical' universities (but without medicine and the arts)
- Five technical universities with an exclusive or predominant specialisation in engineering sciences
- One university specialised in economics
- Four medical universities, one of them specialised in veterinary medicine
- Six universities of the arts
- One university for continuing education which offers exclusively post graduate education and plays a small role as a research performer. This university is governed by a law of its own and will not be dealt with any further in this case study.

---

[79] This case study is mainly based on the following sources: Official Austrian university reports published by the Federal Ministry in charge of universities: *Universitätsbericht 2008*, *Universitätsbericht 2005* and *Universitätsbericht 2002*; University Act 2002: *Bundesgesetz über die Organisation der Universitäten und ihre Studien*, Formula-budget Ordinance 2006: *Verordnung der Bundesministerin für Bildung, Wissenschaft und Kultur über das formelgebundene Budget der Universitäten*; Intellectual Capital Report Ordinance: *Verordnung der Bundesministerin für Bildung, Wissenschaft und Kultur über die Wissensbilanz*, Heinz Kasparovsky, Ingrid Wadsack: *Das österreichische Hochschulsystem*, BMWF 2007; BMWF: *uni:data Statistisches Jahrbuch,* Status on December 31, 2009

[80] This and the other figures in this chapter were calculated on the basis of Statistik Austria, *Erhebung über Forschung und experimentelle Entwicklung in Österreich*, 2007. The basic data are available in German and English on http://www.statistik.at in the sections about R&D and innovation

[81] For more information please see the sections 'Research Funding System' and 'Research Performers' in the Austrian country profile on the ERAWATCH website

The lion's share of universities' income, 76%, is funded by the state, mainly through institutional funding (block grants) and, to a far smaller extent, through competitive public funding. No more than 5.8%% of all university research are funded through contracts from industry, and 4.8% from international sources including the EU. Obviously, public institutional funding plays by far the biggest role for the universities' budgets.

6.4.2 Governance and Public Funding of Universities

The Austrian higher education system has changed radically during the last two decades. As a result, there is an unprecedented institutional diversity, ranging from public universities, universities of applied sciences, universities of teacher education, and private universities.

The governance of Austrian universities has been reformed in several steps during the last decades, most notably in 1993 when first steps towards more autonomous universities where taken by increasing universities' leeway for decision making, and ultimately in 2002. With the University Act 2002, universities were granted full autonomy as a legal person under public law as of 2004. They were given a new organisational structure as well as full decision-making power and the way of allocating institutional funding has been changed radically. The new funding instruments are now the only way for the state to govern its universities and to steer them towards the desired goals. It is therefore crucial that these instruments set the right incentives and provide the necessary means to the universities in order to fulfil their mission.

The Federal Ministry of Science and Research (Bundesministerium für Wissenschaft und Forschung, BMWF) is in charge of public universities and provides their institutional funding through block grants ('Globalbudget', lump sum) for three-year periods. While in many countries (also in the Czech Republic), funding for teaching is separated from institutional research funding, this is not the case in Austria. It is each university's responsibility to distribute the funding internally across the different organisational units and in principle they are free to design their budgeting in accordance with their needs. In practice, a large share of the block grant is spent on more or less fixed costs, mainly staff and infrastructure.

The BMWF allocates the total available budget for institutional funding ('General University Fund') to the universities through three different instruments:

- Firstly, the lion's share of public institutional funding in Austria is allocated through three-years performance contracts. This is the basis for financing day-to-day operation. Moreover, targets and projects are agreed between the ministry and each university in the performance contract, which always contains agreements for the future. Performance contracts can be considered a cooperative steering element, with the ministry as the 'principal' and the university as the 'agent' agreeing on what is attractive and desirable for both parties.
- Secondly, universities receive institutional funding through the formula budget, which is about one quarter the size of the performance contract budget. In contrast to the performance contract, the formula budget rewards past achievements retrospectively. Moreover, it aims at improving teaching, research and societal services delivered by universities by counting on competition between universities. The major objectives are to reduce the average duration of studies (which is far longer than the required minimum duration), to increase success rates in exams, to increase the number of graduates at all levels as well as the share of female professors, student mobility and the competitive income for research and the development of the arts. Together, the performance contract money and the formula budget provide the 'Globalbudget', i.e. the block grant each university receives for periods of three years.
- Thirdly, a small share, up to 2% of the „General University Fund' can be kept aside and spent flexibly by the Minister of Science and Research, e.g. to complement

specific projects related to the implementation of performance contracts. So far, these monies have been allocated to universities in a competitive procedure: universities could submit projects for developing their profile alone or in cooperation with other universities. The proposals were assessed and selected for funding by a jury of independent experts.

In addition, a special budget is reserved for funding wage increases of those university staff who still have the status of civil servant, additional financial outlays at university clinics as well as expenses for the provision of additional premises. This budget is not included in the block grant and it is exclusively used for covering the costs related to the listed items.

These different mechanisms of allocating funding are driven by (implicit or explicit) assumptions on how the Austrian public universities should develop and what objectives they should (ideally) achieve. Some objectives are 'translated' into the formula budget, others in the performance contracts and several objectives can be found in both.

We will describe the new funding tools in more detail below.

Before the reform of the public universities introduced three-years funding periods, institutional funding was granted annually, mainly based on past budgets with incremental changes (and on negotiation skills). At times, additional funding was distributed e.g. in order to renew universities research or teaching infrastructures. This was done, for instance, in the course of the so-called 'Offensivprogramm für F&E' (R&D Initiative), when in the early 2000s additional money was made available for R&D in the wake of the Lisbon strategy.

The new funding system was fully implemented for the first time in 2007 for the years 2007 – 2009. The first three years of the new regime, 2004 – 2006, were a transition period during which the institutional funding was still allocated based on historical trajectories. This period was necessary to prepare the basis for the new instruments: design the formula budget in detail, make development plans, collect data, negotiate contracts etc. The first budget allocated on the basis of performance contracts and the formula budget was for the years 2007 – 2009.

Stability and planning security have been built into the new system and secured by law in the University Act 2002 by limiting the maximum amount of budget cuts: a university's block grant for a given three-years funding period must not be less than 96% of the block grant in the preceding period. Moreover, maximum budget cuts for the annual appropriations are restricted, too. A cut of 4% over three years might seem small at first sight but it can account for a large share of the 'disposable' budget a rector can use, given the large share used to cover more or less fixed 'running costs'. Hence, even this seemingly small cut certainly sends out a strong signal.

In return for autonomy, the universities have to report to the Ministry of Science and Research in a new way and to deliver two different annual reports[82]: In their 'Leistungsbericht' (performance report) universities give an account of the progress made with respect to the different projects and targets defined in the performance contract, and in their so-called 'Wissensbilanz' (intellectual capital report) they present a large set of standardised monitoring data covering the indicators used for the formula budget and for quantitative targets of the performance contract as well as extensive sets of other data, e.g. about research outputs. These reports are the ministry's main tools for monitoring each university's performance against the contract and the achievement of their goals. In addition, representatives of the ministry and each university meet on a regular basis for so-called 'Begleitgespräche' (interim talks) in order to discuss achievements and plans.

[82] From the reporting year 2010 onwards the two reports have been merged and partly restructured.

The second performance contracts for the years 2010 to 2012 were signed in 2009 and the formula-budget has been calculated for the second time, too. Therefore, it is still too early to fully assess the effects of this new governance and funding system. However, a first evaluation of the smaller part of the institutional funding, the formula-based budget, is undertaken at the time of this study.

In contrast to the Evaluation Methodology in the Czech Republic, the *evaluation of research* is not linked to decision-making on allocating institutional funding. Instead, the University Act 2004 has assigned the evaluation of research as a compulsory task to each university. Evaluation is embedded in the wider context of quality development: Each university is obliged to establish an appropriate system of quality development and control covering all tasks and activities of a university. The related projects are agreed upon in the performance contracts on an individual basis. Whether and how a university uses evaluation results for decision making internally differs from case to case, as it is at the sole disposition of each university to handle evaluation and to make use of related results.

Other public funding sources

The main source of competitive public funding for Austrian universities is the Austrian Science Fund FWF which provides competitive grants. FWF's programmes are open to all fields of science and committed to scientific standards only – they are a key element of the Austrian tradition of thematically open, bottom-up research funding system. As all proposals are evaluated by international peers, these grants can be considered an important contribution to ensuring the quality of research at universities. In addition, the programmes managed by the Austrian Research Promotion Agency FFG have become increasingly important for university researchers, in particular for the natural sciences and engineering subjects, who benefit from numerous opportunities for engaging in collaborative research projects with industry. In addition, universities are also supported by provincial governments or municipalities, typically for enhancing their built infrastructure and their research equipment, but also through direct funding for new professorships.

Other public R&D institutions

The 21 public universities are not the only research organisations receiving institutional funding from the state. On the contrary, there are numerous other research performers that receive a smaller or larger share of their budget as a block grant from one of the ministries involved in R&D policies as well as by sectoral ministries in charge of research institutes. The allocation of these institutional monies is very often based on history and negotiations on an annual basis and is often neither transparent nor related to performance. However, awareness for the importance of state-of-the-art governance of research institutes is growing, and in the meantime, performance contracts have been set up or are under preparation also for (still few) non-university research institutes.

6.4.3 A Closer Look at the Performance Contracts

The performance contracts define the services to be provided by the university: teaching, research, mobility of researchers and students, co-operation, strategy, specialisation etc. All universities have had to work out strategies for their long-term development ('Entwicklungsplan') which are also used as a basis for the negotiations with the BMWF.

The content of the performance contracts as specified in the University Act 2002 is:

1. the services to be performed by the university according to its objectives, guiding principles and tasks in the following areas:
   i) strategic objectives, development of an institutional profile and human resource development
   ii) research and arts-based research

   iii) study programme and postgraduate education
   iv) societal objectives (e.g. increasing the share of women in leading positions, offers for students with special needs, knowledge and technology transfer, research in fields of high societal relevance)
   v) internationalisation and mobility
   vi) interuniversity cooperation
2. the state's contractual obligations: allocation of the block grant
3. content and scope of targets and timing of target achievement
4. schedule of budget appropriation
5. measures for the event of non-performance
6. reporting

The focus of the first performance contract period has been to continue the existing range of activities. In addition, parts of the budget were dedicated to activities that would take universities towards a higher degree of specialisation and towards a more distinct institutional profile ('Profilbildung'). Consequently the performance contracts also comprise activities for increasing the number of professors in the fields of specialisation and for improving conditions for students e.g. by investing in infrastructure, reducing the number of students per lecturer, or establishing PhD programmes and graduate schools.

According to the University Report of 2008, most intentions, 85%, agreed upon in the first performance contract period have been implemented as scheduled, and 13% were either delayed or modified in content. Those intentions which the universities considered not realisable were discussed in the interim talks with the Ministry and modified accordingly. Out of the targets agreed upon 80% were reached in 2007. All in all, the Ministry of Science and Research concluded in this report, the performance contracts are considered a successful tool for governing autonomous universities.

6.4.4 The Formula-based Budget: Concept and Practice

Formula based allocation of funding to universities has become a common practice in many countries, which is also illustrated by the other case studies in this report. An increasing number of countries uses some sort of formula budget, generally in combination with other mechanisms, mainly negotiations and performance contracts, while incremental budgeting (i.e. allocation based on history) is losing ground[83].

Formula-budgets differ widely in their financial scope as well as in the indicators used and their relative importance. The Austrian formula-based budget in particular contains incentives towards objectives in the three key fields of universities' activities: teaching, research, and societal responsibility. The formula budget tackles several problems of the Austrian university system, above all the above-average duration of studies which is far longer than scheduled in curricula, the high drop-out rate (i.e. the high share of students that do not complete their studies), the comparatively low average share of competitive research funding, or the low share of women among professors. Consequently, the formula budget provides incentives for universities to reduce the real duration of studies to reach the scheduled duration, to increase students' success rates (i.e. reduce the drop-out rates), and to raise research income from competitive sources, and to promote women in research and the mobility of students. These objectives have been 'translated' into 11 numeric indicators which are

[83] CHEPS, IoE London, Technopolis: *Progress in Higher Education Reform in Europe: Governance and Funding Reform*, Studie im Auftrag der Generaldirektion Bildung und Kultur der Europäischen Kommission, 2008

weighted according to their relative importance (see Figure 28). There are four indicators for teaching related objectives, three for research and four indicators for societal objectives. The groups of teaching and research indicators weigh 45% each and 10% of the formula budget are based on the societal indicators. Universities evidently are places of teaching and research, both equally important. Moreover, the universities have to contribute to societal objectives in a way compatible with their key mission.

The relative importance and purpose of the formula budget is defined in the University Act 2002, whereas the indicators actually used, their definition and measurement, and the translation of these indicators into budgets are settled in a separate ordinance, the so-called Formula Budget Ordinance (Formelbudget-Verordnung). This ordinance is valid for a limited period of time and has to be redefined on the basis of an evaluation which is ongoing at the time of this study.

Figure 28 The indicators used in the Austrian formula budget

Area	Indicator	Description of Indicator	Weight[84]
Teaching			***45%***
	Indicator 1	Number of active[85] students who study in the timeline of the curriculum (plus one 'tolerance semester'). All bachelor and master studies the „old' diploma studies are counted and weighted by subject group factors[86]	15%
	Indicator 2	Number of bachelor, master and diploma degrees, weighted by subject group factors	10%
	Indicator 3	Share of bachelor, master and diploma degrees that were completed according to the programme duration scheduled in the curriculum plus one 'tolerance semester' out of all degrees in the same programme	10%
	Indicator 4	Success rate of students in bachelor, master and diploma programmes	10%
Research and arts-based research			***45%***
	Indicator 5	Number of doctoral degrees, weighted by subject group factors	15%
	Indicator 6	External income[87] from research projects and from arts-based research[88] projects funded by the Austrian Science Funds FWF (research council) or by the European Union, in Euros	15%
	Indicator 7	External income[89] from research projects and from arts-based research projects funded by other sources than those defined in Indicator 6, in Euros	15%
Societal objectives: promotion of women and students' mobility			***10%***
	Indicator 8	Share of women among university professors	6%
	Indicator 9	Number of doctoral degrees earned by women, weighted by subject group factors	1%
	Indicator 10	Number of regular students participating in outgoing programmes for international mobility	2,5%
	Indicator 11	Number of accredited master and doctoral students without an Austrian bachelor, master or diploma degree	0,5%

[84] According to the Formelbudget-Verordnung, Annex 2, paragraph 2.4

[85] 'Active' means that the student must have passed a defined quantum of exams

[86] The factors are 5 for arts and medicine, 3 for engineering and natural sciences, and 1 for social sciences, humanities, law, economics, and theology.

[87] 'Income' is defined in the University Act 2002, §26, subsection 1, and §27, subsection 1, numbers 1 and 3

[88] The former art colleges have become universities through the University Act 2002. The term 'arts-based research' refers to the specific characteristics of research by artists

[89] 'Income' is defined in the University Act 2002, §26, subsection 1, and §27, subsection 1, numbers 1 and 3

Source: Formelbudget-Verordnung (Formula Budget Ordinance)

How would the Austrian 'ideal university' look like according to the formula budget? It is a university of unspecified size and subject specialisation, populated by student studying actively (i.e. passing exams). Almost all of them (90%) graduate, i.e. no more than 10% of them drop out, and no less than 80% of the students graduate within the duration scheduled in the curriculum (plus 'tolerance semester'). At the 'ideal university', many students also complete a doctoral degree and the university strives for a high share of women among them. The 'ideal university' earns large sums for research and arts-based research from competitive funding sources such as the prestigious Austrian Science Funds (FWF), EU programmes, and other sources. 50% of professors are women, and a high share of students spends some time at a university abroad – small wonder as the 'ideal university' attracts large numbers of students that have earned their first academic degree abroad and want to continue their studies here.

However, although these indicators reflect legitimate objectives of higher education policy, they could still yield undesired side effects (perverse behaviour): a university governed strictly and exclusively by the formula budget would probably neglect persons that cannot complete their studies according to schedule, e.g. due to family obligations or because they have to earn an income next to studying – as this would reduce the indicators 1 and 3. Such a university could also be prone to prefer research topics that attract large volumes of contract research incomes and neglect less profitable fields, regardless of their societal importance, or the university could count on routine R&D assignments at the expenses of long-term research endeavours- both strategies would increase the indicators 6 and 7. The performance contracts contain targets and activities in order to avoid such undesired effects of the incentives given through the formula budget, e.g. activities in support of students with special needs.

Complementarity of formula budget and performance contracts

This illustrates that the formula budget covers only a certain subset of those targets that Austria aims at for its universities. However, this is not a problem because the formula budget does not stand alone as a governance instrument. In particular, the formula budget and the performance contract complement and support each other.

Once the state has defined the objectives for universities, practical considerations become important when deciding which objectives should be governed through which steering instrument, i.e. which objectives should be included in the formula budget:

- Relevance: Is the target relevant (and attractive) for all 21 universities (although maybe to different extents), regardless of their starting position, their subject specialisation and their size?
- Attainability: Can each university approach this objective through its own efforts?
- Feasibility: Can the objective be expressed in a quantifiable indicator which can be standardised and measured easily?

Calculation of formula

Setting targets and defining indicators for monitoring them are the first steps towards a formula budget. The next key setting has been that the Austrian formula budget is no 'price model'[90] but a distribution mechanism that allocates a certain available budget to all universities in a competition among each other. Furthermore, it has been decided that for no indicator the absolute value achieved will be decisive, but the change of the indicator in the given three-year period compared to the previous period. This mechanism is intended to support a development towards a goal, no matter where the starting point of each university is. In calculating these changes,

[90] 'Price models' are used in several countries especially for allocating teaching grants to universities

three year averages are used in order to level short term fluctuations of the indicator. The monitoring period for each indicator is three years – identical to the duration of the performance contract period.

In calculating the formula budget, some indicators are weighted by subject group factors, and some by type of study programme (bachelor, master, PhD). Subject factors distinguish three different groups: (i) arts and medicine, (ii) natural sciences and engineering, and (iii) all other subjects. The different weights attached to these groups reflect the state's intention to cater for different resource needs of different studies. The situation is similar for types of programmes. In practice, to put is simply, this would mean that 100 active students of fine arts would result in an indicator score of 500, 100 equally active students of chemistry score 300, and 100 students of Slavonic studies score 100 points. The weighting system has to be seen against the background that Austrian universities receive a block grant for all their activities, including teaching, whereas in many European countries, parts of the universities' institutional funding is based on a certain allocation per student. However, no assessments of the actual costs and the actual differences for these different categories have been performed.

For each of the indicators, the data for the current value (i.e. the average of the three last years) and the reference value (i.e. for the previous period of three years) are collected.

The next step is to convert these data into a funding decision, and it is a complicated step. To our knowledge, these calculations are quite unique among formula budgets. First, the absolute values of each indicator are transformed into function values which are calculated by means of a sigmoid function. The particular function applied makes it possible to use the same indicators for all universities because by adapting the parameters of the function the different starting points of universities for any given indicator can be compensated. Moreover, the parameters for some indicators are defined in such a way that they award the development towards a certain target value but not beyond. For example, the share of female professors differs widely between 5.1% and 51.5%[91]. The target value has been set at 50% for all universities and for this particular indicator any step towards the target value is rewarded in the same way, regardless of the starting point, but increases beyond 50% will not yield any further incentive. The parameters are defined separately for each of the indicators in the Formula Budget Ordinance.

Once the function values for both the current value and the reference value of each indicator have been calculated, the function value for the current value is divided by the function value of the reference value. The result of this operation is a dimensionless number for each indicator which, in the next step, is multiplied with the weight of the indicator according to Figure 28. These points are added up across all indicators for each university. In the following step, the actual size of the university is taken into account. Size has been defined as the block grant allocated in the preceding performance contract period, and this number is multiplied with the summary indicator score, resulting in a size-scaled score for each university. These scores are added up to a summary score, and ultimately the formula budget available is divided among the universities according to the share of each university's size-scaled score within the summary score. It is this final step of the calculation when all universities and their achievements are assessed against each other.

Preliminary assessment of the Austrian formula budget

This (simplified) description of the Austrian formula budget and the mechanisms used for translating indicators into budget allocations clearly shows that this model is far from a simple model. Nonetheless, we can safely assume that decision makers at

[91] BMWF: *uni:data Statistisches Jahrbuch,* Status on December 31, 2009; headcounts.

universities are well capable of having an eye on each individual indicator and of developing and (perhaps less easily) implementing measures towards the related objectives. However, most likely it is very difficult to anticipate the overall result of the formula budget and use it for planning due to the high complexity of the model and due to the fact that, ultimately, the financial result depends not only on each university's own achievement individually. This high complexity might well be an advantage, as perhaps a complicated system is less easily undermined through 'serving the indicators'. On the other hand, the formula budget is also intended to send out clear signals and provide unambiguous incentives. The question is can it so in its current complex form.

At the moment, it is not possible to assess the real effects of the formula budget and to what extent it has contributed to the development of Austrian universities towards the desired objectives. It is not known yet whether the system is really doing justice to all universities, given their different sizes, subject specialisations and starting positions for the different indicators. It is equally unclear if the formula budget has had unintended (negative or positive) side effects. Moreover, many of the typical activities and targets included in the performance contracts demonstrate that there is a link between the performance contracts and the formula budget: provided these activities agreed upon in the performance contract are successful, they can have a positive effect on the related quantitative indicators in the formula budget in the subsequent funding period(s). E.g. many performance contracts include activities for increasing competitive income which is also taken into account in the formula budget. In parts the two instruments strive for the same objectives of Austrian university policy, one in the form of a contract about a target to be reached in the future, the other as a decision on past achievements. At present, it is not yet possible to assess the balance between these two approaches due to a lack of evidence.

Therefore, it may be worth while for Czech policy-makers to study the results of the ongoing evaluation once they are available and to enter into an exchange of experience with the Austrian ministry in charge as well as with universities.

6.4.5 The Evaluation of Research at Austrian Universities

As already mentioned, the University Act 2002 has granted far reaching autonomy to Austrian universities and this includes the responsibility for the quality of their performance. According to §14 of the Act, the universities are responsible for setting up their own tailored quality management system comprising all their activities. Moreover, they have to evaluate all their performance in teaching, research and arts-based research according to subject specific international standards at least once in five years. Apart from these specifications, the University Act 2002 gives the universities plenty of rope in implementing these clauses according to their specific needs and preferences. In July 2011 a new law was enacted which stipulates that in the future all universities need to have their quality management systems audited regularly by an independent accreditation agency (e.g. an agency listed in the European Quality Assurance Register for Higher Education (EQAR).

An analysis[92] of the status quo of research evaluation at Austrian universities in 2009 has come to the following conclusions:

Austrian universities have evaluated research already before the university reform of 2002. The debate about evaluation started in the late 1980s. During the 1990s, a number of large scale evaluation exercises were performed, covering entire subjects across universities (and in some cases also other research institutes), e.g. in physics, mathematics, electrical engineering, to name just the first ones. Before these evaluations started and while they were going on, an intensive debate took place, with

[92] B. Tiefenthaler, F. Ohler: *Dokumentation der Evaluierungen von Forschung an Universitäten*, Study commissioned by the Federal Ministry of Science and Research, Technopolis, 2009

studies, concepts, workshops, and conferences. This debate involved not only the scientific communities but also policy makers and the ministries, and it has been crucial in the development of a now well established Austrian culture of evaluation in research and in research policy making.

The early evaluations were endogenous, i.e. they were initiated by the scientific communities themselves, often managed through the respective scientific society. The responsible ministry provided funding and acted mainly as a (supportive) observer.

Peer review was the dominant method in these early subject evaluations. They were often performed in German and consequently a large share of peers came from Germany. The evaluations were based on questionnaires to be completed by the participating institutes and site visits. The peers' judgements were 'eminence based' rather than evidence based, as the data collected were generally not analysed in great detail (nor with simple statistics).

After the reform, the universities have embraced the assignment given in the law. They have chosen different ways of handling the task and they develop and implement their solutions at different paces. When the 'slowest' university had just begun systematic evaluation, the fastest university had already implemented a sophisticated evaluation system and reached the second evaluation circle. The universities have established the necessary organisational capacities and defined the procedures. All these systems, regardless of their present status, are flexible and will be developed further on the basis of the experience made in the course of their implementation.

All universities have defined evaluation as a part of their quality culture. Quality management and evaluation are generally integrated into the wider governance system of the universities. Evaluation is used as one of several sources of information for decision making within the university, especially for internal target agreements or performance contracts. Universities understand evaluation as an instrument of self-reflection and a support for learning. Its main purpose is to support the improvement and development of the unit evaluated. Despite the abundance of indicators (to be collected for the 'Wissensbilanz', the intellectual capital reports), mechanistic use of indicators or rankings rarely play a role in research evaluation.

Informed peer review is the state-of-the-art evaluation methodology applied by most universities. This typically involves a self assessment of the unit evaluated with standardised data as inputs (i.e. not just a description), a site visit of the international peers and a written evaluation report. In most cases, the unit evaluated has the chance to comment on the reports. Generally, these reports are for internal use only but some universities have decided to publish them fully or in summaries (e.g. the University of Vienna[93] and the University of Graz[94]). Those universities that have included a thorough self assessment of the unit evaluated have often experienced considerable additional learning effects through this exercise, often leading to improvements before the peers even visit the site.

In general, evaluators make recommendations but it would be naive to take such recommendations literally and to turn them into decisions mechanically. Even the best evaluation system has its limits and cannot answer all questions needed for decision making. Therefore, the more advanced universities have carefully designed the process that translates evaluation findings and recommendations into decisions and they put much emphasis on gaining the acceptance of those evaluated.

Most universities have already set up (or are in the course of doing so) extensive information systems, with a high awareness for the importance of data quality (completeness, correctness, up-to-dateness). In the evaluation of research, they use

[93] http://www.qs.univie.ac.at/evaluation-von-fakultaeten-zentren-und-dle/evaluationsberichte/ (Jan. 21, 2011)

[94] http://www.uni-graz.at/ffowww/ffowww_forschung/ffowww_evaluierung.htm (Jan. 21, 2011)

these standardised data as inputs for the self-assessment reports. These data comprise different types of output data which are also required for the intellectual capital reports where they are reported at a highly aggregate level, which has been heavily criticised for not doing justice to different practices in different subjects. Within a university however, and in research evaluation, the same output data are normally not used to compare across different disciplines. Rather, they are used in order to assess research at the level of organisational units or disciplines, where their specificities can be taken into account and the outputs are assessed in the context of a particular organisational setting, budget situation, infrastructure, management and staff.

The big early subject evaluation exercises took long, two years in average, some even longer. This has changed dramatically in the new system, where a research evaluation of a particular unit rarely lasts longer than half a year.

The process owners of the early subject evaluations in Austria were either scientific societies or groups of universities and both lacked the authority to act on the evaluation results. Clearly, this has changed and this major drawback has been overcome in the new legal setting: universities are not only responsible for research evaluation, they have also decided to place responsibility high up in the hierarchy, usually with the rector or the vice-rector in charge of research. 'Ownership' of evaluation is now closely (but not mechanically) linked with decision making within the university. This is an important feature for legitimating evaluation and for making the best possible use of a valuable (and resource intensive) instrument. While the external costs are normally not too large, the internal costs can be substantial: self-assessment, participation in hearings, co-ordination, development and negotiation of consequences. Ideally, benefits are larger: knowing and understanding one's own 'business' better, helpful feedback from outside, well founded decisions. This means that evaluations should not take place at short intervals nor apply inappropriate procedures, nor remain without consequences, if their benefits are to outweigh their (financial and social) costs.

While it is clearly useful for universities to have responsibility for quality development and evaluation at the autonomous universities, there are issues that cannot be solved by individual universities themselves. Even if universities enjoy far reaching autonomy, the ministry needs to be a responsible principal for each of them and for the entire system of tertiary education in Austria, and it needs to tackle overarching issues, e.g. concerning coordination, duplication of activities or gaps, or the evaluation of entire disciplines.

6.4.6 Lessons learnt from the Austrian case study

- The reform of 2002 has actually re-invented universities as strong institutions with a powerful manager-rector on top, a new management structure and many new responsibilities and tasks, i.e. caused a great deal of transformation and change. However, the stability of institutional funding is rated highly and it is embedded even embedded in the law.
- The two main mechanisms for allocating institutional funding aim at balancing competition between universities (i.e. the formula budget) and the individual development of each university, (i.e. in the performance contracts). This system is based on past achievements and on (contractually agreed!) plans for the future.
- Both the performance contracts and the formula budget in their present form are extensive and complex and it remains to be seen whether or not they can actually fulfil their task as steering instruments well. In other countries such as the Netherlands and Finland complexity in the formula has become such problem that the funding formula is going to be revised.
- Completely autonomous universities imply that the ministry has to direct more attention at overarching issues that cannot be dealt with by individual universities.

- The universities have embraced the tasks of quality management and evaluation of research (and their other activities) but at different paces and with different attitudes and approaches, depending very much on their history and experience with evaluation and on their subject specialisation. Evaluation is primarily being used as a tool for getting feedback and for learning. Where it is linked to decision making, the connection is not linear: a critical feedback can also imply more resources (instead of less) have to be allocated if a given field is crucial for the university's development.
- Even now, six years after the new University Act has entered into force, no full account of its impacts can be given, and evaluation of parts of the reform has only just begun – and for a reason: it takes several funding cycles before the effects of such a comprehensive reform can be duly assessed.

6.5 Finland[95]

6.5.1 Introduction

The Finnish institutional funding system for universities is based on performance contracts in combination with a funding formula. Targets set for institutional activities and the resources needed for their implementation are determined in negotiations between the Ministry of Education and each university. Performance contracts date back to the early 1990s, while the allocation of (part of) university funding on the basis of a formula began in 1998, with the current system dating from 2010.

In Finland, the transition to result-based higher education management policy began in 1986 when the Finnish government decided to grant universities a 15% increase in their research and teaching budget. However, a precondition for the increase was, among others, that conditions for result-orientated management be improved. This turning point of government steering was quite dramatic as Finland had had one of the most centralised higher education systems in Europe.

The development of government steering and institutional management in the Finnish higher education system has followed international trends, especially those in the OECD countries. Finland has adopted a steering-by-results model, which has to be seen in a broader context of New Public Management. Interestingly, the basic elements of the new government steering strategy for the university sector were introduced before a more general steering reform for the public sector was implemented.

A striking aspect in Finnish public management development is the high degree of continuity since the beginning of 1990 despite changes in government. Another typical feature is the use of gradually progressing projects and pilot projects to test and introduce changes.

6.5.2 Context

In 2009, Finland passed a new Universities Act, which reforms the legal status of all Finnish universities starting from 1 January 2010. The reform increased the administrative and financial autonomy of the universities and has made them independent legal persons, with universities becoming either corporations under

[95] This case study is based on the following literature: Seppo Hölttä, Eila Rekilä, 'Ministerial Steering and Institutional Responses: Recent Developments of the Finnish Higher Education System', in: OECD, *Higher Education Management and Policy*, Volume 15, No. 1, 2003; Ulla Mäkeläinen, *Efficiency and effectiveness of public expenditure on tertiary education in the EU, Country Fiche Finland,* Joint Report by the Economic Policy Committee (Quality of Public Finances) and the Directorate-General for Economic and Financial Affairs, European Economy Occasional Papers No 70, 2010; OECD, *Performance-based Funding for Public Research in Tertiary Education Institutions: Country Experiences. Summary of country questionnaire responses,* Paris, June 2010. In addition, an interview with Seppo Hölttä was conducted

public law or foundations under private law[96]. The Act contains provisions on the functions and administration of the universities, the financing and steering of operations, and university research and teaching, students and staff. The operations of both the public and foundation universities will continue to be primarily funded by the state, and the universities will continue to perform the public duty assigned to them in legislation.

The motive for the reform was to provide a level playing field for Finnish universities to compete with the best international universities. To this end, universities have been given the economic and administrative means for enhancing the quality and impact of research and teaching and for participating in international cooperation.

6.5.3 Steering of the university system

The main element in relations between the Ministry of Education and the universities is steering based on financing, legislation and information. The key means to this end include performance contracts concluded by the Ministry of Education and the universities, and a feedback and monitoring system, especially the KOTA database on universities (see below).

Through performance steering, the Ministry of Education advances key policies set by Parliament and the Government. The performance contract procedure ensures that the agreed targets support the strategic development of the entire university system and that the government is adequately informed by university strategies and profiles.

A unique feature of the steering and management reform in Finnish higher education was the establishment of a national university database. For getting better information on university inputs and outputs the open database KOTA was built up in the mid 1980s. The database contains statistical data on universities. Data collected every year are the number of applicants, the number of entrants, student mobility, degrees, graduate placement, median graduation times, teachers, other staff, annual accounts, expenditure by performance areas, university premises, teacher and researcher visits, scientific publications etc. The database is maintained by the Ministry of Education but the data are collected by the institutions themselves and updated at the beginning of each calendar year. Without the database it would have been impossible to create contracting and funding models and a planning system, as they all heavily depend on reliable institutional information.

6.5.3.1 Performance contracting system

The main idea of management-by-results is that the goals set for institutional activities and the resources needed for their implementation are determined in the negotiations between the Ministry of Education and each university. The funding is allocated to the universities in a lump sum to implement the contract. The budgeting system has been developed to support management-by-results so that most of the goals and appropriations are inter-linked.

Höltä and Rekilä[97] describe the performance contracting process, as instituted before the 2009 Universities Act, as follows:

---

[96] The reform reduced the number of universities from 20 to 16. Of these 16 universities 14 are public universities (legal persons under public law) and two foundation universities under private law – Tampere University of Technology and Aalto University which is a merger of the Helsinki School of Economics and Business Administration, the University of Art and Design and Helsinki University of Technology. In the public sector, the Universities of Joensuu and Kuopio have merged to form the University of Eastern Finland and the University of Turku and Turku School of Economics have merged to become the new University of Turku. There is no difference between public universities and foundation universities in terms of content or disciplines. The foundation universities also perform a statutory public mission

[97] Seppo Hölttä, Eila Rekilä, 'Ministerial Steering and Institutional Responses: Recent Developments of the Finnish Higher Education System', in: OECD, *Higher Education Management and Policy*, Volume 15, No. 1, 2003

- The contracts between the Ministry and the university are concluded for three-year periods. This means that the target outcomes and the basic grounds for allocating resources are also determined for this period. The final sum of annual budget is dependent on the ministerial budget frame, and financing is revised annually in a supplementary protocol. Also some smaller elements, like new development projects, are agreed upon annually.
- The universities send their proposal for the contract to the Ministry in February. An activity and finance plan for the next three-year period is submitted to the Ministry as well as background material for contracting. The Ministry comments on the proposals and the universities have a possibility to provide the Ministry with more information if needed.
- Negotiations are conducted in April. By then, most issues will already have been agreed on in the dialogue taking place before the formal negotiation. Every three years, there is a detailed negotiation round for the coming planning and contracting period.
- Unofficial seminars and meetings between the Ministry and university leaders as well as deans and planners are conducted before the negotiations. The idea is to introduce both institutional and disciplinary views as well as transparency to the negotiations.
- The agreements are signed after approval of the state budget by Parliament.
- Reporting is an important element in the steering process. The universities submit an annual report describing the achievement of goals. The core of this process is the annual updating of the KOTA database.
- From the year 2002 onwards the Ministry has provided universities with formal written feedback on their performance.

The goals determined in the contract include 1) the general goals for the whole higher education system and 2) specific goals for the individual institution in question, lines of education and central objectives. The *general goals* set for the university system concern mainly the quality and impact of education and research, and the total number of degrees to be awarded. The *specific goals* for each university primarily include the targeted number of Master's and doctoral degrees in each study field, which are central elements in the finance allocation formula. In addition, the negotiations also cover universities' strategic issues relating to the development of operations. The performance contract takes into account the special characteristics of each university, with the exception of the common goals for the system of higher education.

Under the new University Act, the regular negotiations held between the Ministry of Education and the universities continue to constitute a key element in the steering process. The performance contracting process is for the most part the same after the new Universities Act.

However, with the new Universities Act, the steering process has been streamlined, with special emphasis on the long-term nature of the objectives in the sector and autonomous decision-making of universities. The performance contracts have become more strategic. This has to be seen against the background of negotiations being very comprehensive, compared to those in other European countries. They cover all aspects of an institution.

Furthermore, the performance contract periods were extended to four years to coincide with the legislative period. Thus, the contracting procedures and contract

periods are better linked to national goals and to the more comprehensive steering reform project in the administrative sector of the Ministry of Education.[98]

The agreement negotiations are held in the year preceding the agreement period or during it if deemed necessary by the Ministry of Education. Hence, the goals and appropriations for the universities operating as legal persons were agreed upon for the first time in 2009 with respect to the budget year 2010.

The structure of the contracts to be drawn up in the negotiations between the Ministry of Education and the university is the same for all universities. Negotiations are not pro forma, they are real negotiations. For example, rectors have to be realistic about degree targets because if their university does not achieve them, it will be 'punished' in the next round and e.g. get less funding. After all, the performance contracting system is a zero-sum game. If one university gets more, another loses out.[99]

6.5.3.2 Public funding of universities

Although the new Universities Act converts universities into public or private legal persons, the government still remains the primary source of financing of universities. State funding ensures that all universities are able to carry out their statutory duties. In funding, the same principles are applied to all universities.

As before the Universities Act, the Ministry of Education and the universities agree on state funding during the performance negotiations. A sizeable part of institutional funding (75%) is allocated based on a formula (see Figure 29).[100]

The funding model aims to enable universities to undertake long-term development and encourage them to develop their profiles and continuously enhance quality in education and research, while aiming at the best productivity and cost-effectiveness possible.

During the transition period beginning in 2010, the aim is to keep universities' proportional funding unchanged, to maintain stability in the university system. In the longer term, however, the aim is to reward individual universities for effective activities. Future increases in the resources of individual universities will partly depend on how well they succeed in their basic mission, as measured by the indicators contained in the formula. The allocation of public formula-based funding is expected to benefit universities with above-average quality, output and impact.

The budget appropriation allocated to the universities, with the exception of non-recurring items, will be raised in accordance with the annual rise in the cost level based on a *university index*. The university index comprises the general index of wages and salaries, the consumer price index and the wholesale price index.

6.5.3.3 The funding formula

The current funding formula has to be seen in the context of the new University Act 2010, which grants universities increased administrative and financial autonomy. According to the Ministry of Education, strengthening the financial autonomy of the of the universities requires steering and the funding instruments to be more strategic.

---

[98] The four-year agreement periods will be introduced from 2013. For the transition period, three-year contracts were negotiated (2010–2012)

[99] Budgets for higher education have increased steadily over the last twenty years, with the exception of the early 1990s when Finland was hit by a severe recession. These increases have mitigated the competition over resources

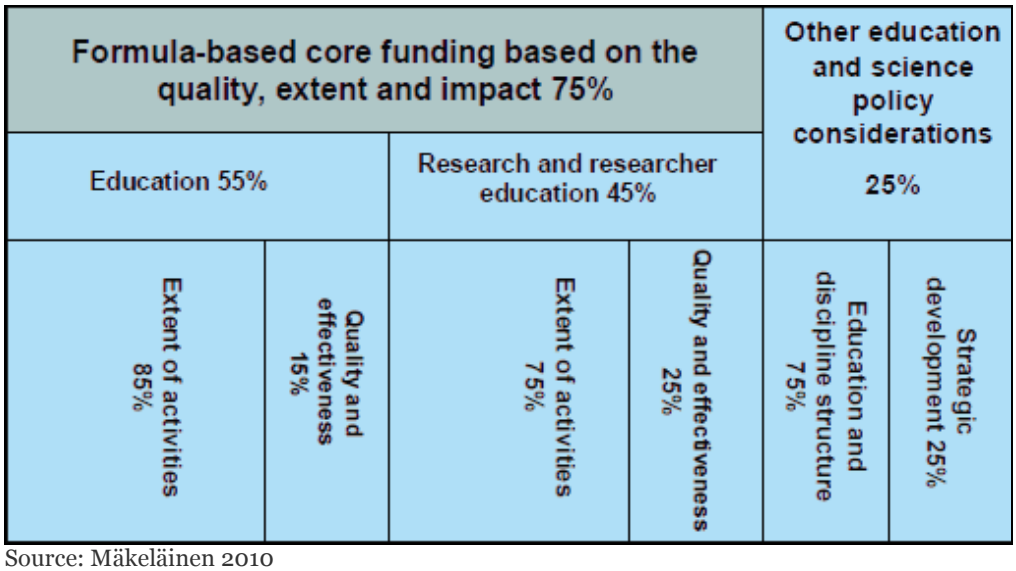
[100] In the early 2000s, an even larger share of funding was allocated on the basis of a formula. Core funding allocated based on annual institutional targets for Master's and doctoral degrees covered about 90% of universities' operational expenditure. So-called performance-based funds made up 3-5% of the operational expenditure. The rest were funds for special purposes allocated to programmes of great national relevance and institutional development projects, such as opening a new study programme and improving institutional facilities

Having said this, though developed over time, the formula has remained relatively stable since its introduction in 1998. Over the years, the formula has become more and more complex, also because limits have been built in to make sure that changes in funding are not too large. If a university loses out 'too much', the Ministry of Education applies a correcting factor. Furthermore, over time new policy targets have been integrated into the formula. Finally, the formula has been set up in consultation with universities, thus representing a compromise.

The formula-based core funds are allocated internally at the discretion of the university according to its strategic choices, although with reference to target outcomes agreed in the performance contract with the Ministry of Education. However, universities typically imitate the funding formula in their internal allocation of funding.

Finland's universities receive 75% of their core funding calculated on the basis of the extent and quality of activities in education and research and researcher education, and 25% based on other education and science policy considerations (Figure 29). The same criteria are used in formula-based funding with regard to all the universities.

Figure 29 Finnish university core funding formula



Source: Mäkeläinen 2010

In the current formula, more attention has been given to output-based measures compared to input-based measures. In fact, the amount of funding allocated on the basis of the quality and effectiveness of operations (i.e. on the basis of output measures) has increased from 3-5%[101] to about 20% of formula-based core funding or 15% of total core funding.

Education and research, researcher training and artistic activities are the most important statutory duties of universities, on which a significant part of the allocation of appropriations is based. Interaction with and impact on society are mainly realised through these.

The educational function continues to be a major element but for policy reasons research and researcher training have become more important considerations than before (55% compared to 45%, see Figure 29). The shift has to be seen against the background of attempts to increase research-based competitiveness and develop research universities. Finnish policy is in line with European policy in this instance.

[101] So-called performance-based funding, as used in the formula in 2003

In education, the extent of activities is a more important factor than in research and researcher training (85% compared to 75%, see Figure 29). This is meant to create favourable conditions for education, as it is more dependent on direct state funding than research. In contrast, more than half of universities' research funding is competitive funding (targeted funding).

Education: extent of activities

Degree targets and their attainment continue to play a key part in the model (Figure 29) because they are the key outputs of universities. However, the focus on degree-based funding has shifted from targets to outputs, in order to find a balance between plans and reality. The idea is to have incentives in place. Making the number of degrees awarded a criterion in funding encourages universities to organise their activities in such a way as to make it possible for students to complete their degree studies within the normal time.

Previously[102], the institutional funding component of the budget was directly based on annual institutional *targets* for Master's and doctoral degrees, as agreed with the Ministry of Education for each main field of study offered by the university. Target figures were simply multiplied by a field-specific cost factor, which was also agreed for the three-year contract period.

With a view to balancing the annual variations in the number of degrees awarded by the smaller universities, the average number of degrees over several years will be considered. The differences in the cost structure of different fields of education (including the specific nature of the arts; required equipment) and in teacher training colleges will be taken into account in the funding model as part of the educational and disciplinary structure funding element, which forms part of 'other education and science policy objectives' (see right side of Figure 29).

Education: quality and effectiveness

Figure 30 Quality and effectiveness indicators for education

	Weighting	
1. The quality of education and functioning of study processes	**80%**	**Of which**
a) On the basis of the Centre of Excellence proposed by the Finnish Higher Education Evaluation Council, 300,000 euro annually/centre.		
b) The number of students studying for first- and second-cycle higher education degrees completing at least 45 ECTS credits in one academic year		50%
c) The number of student graduates who started studying for their first degree in x after 7 years have passed		50%
2. Internationalisation of education	**20%**	
a) Number of outgoing and incoming exchange students in Finland (duration of exchange over 3 months)		50%
b) Number of ECTS credits completed in education in a foreign language (and the number of ECTS credits completed abroad included in the degree)		13%
c) The number of ECTS credits acquired abroad and included in the degree is included in the calculation when the data collection of the statistical material is complete.		12%
d) Number of international degree students		25%

Source: Mäkeläinen 2010

The quality and effectiveness of education is assessed on the basis of monitoring indicators in line with education policy objectives that are of strategic importance to all universities. The basic principle is that all universities will have the opportunity to

[102] This refers mostly to the early 2000s

obtain funding based on quality and effectiveness[103]. The quality and effectiveness of university education is determined according to the following indicators (Figure 30).

In the past, quality of education was evaluated on the basis of Centres of Excellence in teaching, Centres of excellence in adult education, length of studies, graduate placement and international activities. In other words, indicators for assessing the quality and effectiveness of education have evolved over time but have not changed radically.

Research and researcher education: extent of activities

The following indictors are used to determine the extent of research and researcher education at universities:

- teaching and research person-years (50%)
- total number of doctoral degrees determined in the performance contract between the Ministry and the university (25%)
- number of doctoral degrees completed at the university (25%)

Research and researcher education: quality and effectiveness of research and researcher education

As with the quality and effectiveness of education, the quality and effectiveness of research and researcher education is taken into account on the basis of monitoring indicators in line with science policy objectives of strategic importance to all universities. All universities will receive a share of funding determined on the basis of various criteria.

With a view to encouraging universities to enhance the quality and effectiveness of research, the allocation of resources will be mainly based on the amount of competitive research funding acquired by the university (60%) and the extent of scientific publication (20%) (see Figure 31). Hence, scientific publications make up only 2.2% of total formula-based core funding or 5% of formula-based core-funding for research and researcher education.

Figure 31 Quality and effectiveness indicators for research and researcher education

	Weighting	
1. National competitive research funding	**60%**	
a) Academy of Finland (Finnish research Council) funding for the university		75%
b) Tekes (Finnish innovation agency) funding for the university		25%
2. Scientific publications	**20%**	
a) Number of refereed international publications		60%
b) Number of other scientific publications		40%
3. Internationalisation of research	**20%**	
a) Amount of international competitive research funding*		60%
b) Overall amount of teacher and researcher mobility**		40%
*excluding enterprise funding (contract research) or EU Structural funds **The overall amount of mobility to and from Finland (minimum two weeks over the period 2010-2011,		

[103] With the exception of funding for Centres of Excellence. The 'Centres of Excellence in University Education' is a programme administrated by the Ministry of Education. Its aim is improve the quality and relevance of university education and to encourage universities to carry out long-term development. The nomination of Centres of Excellence in education is thus one means to promote the enhancement of university-level education and to highlight the importance of the quality of education also through economic incentives. Their counterpart in research – the 'Centres of Excellence in research' – are awarded by the Academy of Finland

	Weighting	
minimum one week from 2012)		

Source: Mäkeläinen 2010

Interestingly, university leaders would like research output to play a greater part in the formula. This is a remarkable development, as it shows the change in values having occurred in the university sector in the last ten years: When the formula was introduced in the 1990s, universities felt that it violated their academic freedom.

However, universities' high hopes and expectations for research outputs have been disappointed so far because it is enormously difficult to implement a sensible system. Generally, there is the hope to increase the importance of publications in the formula. Disciplinary working groups have been established to deal with differences in publication patterns etc. in the different disciplines. They are currently drawing up lists of journals in order of their significance/prestige.

In the past, quality of research was assessed on the basis of Centres of Excellence in research, competitive funds granted by the Academy of Finland and other external (competitive) funding. Again, indicators have changed over time but have not become radically different.

Calculation

The basic calculation is updated annually on the basis of statistical data from the KOTA database. The values are determined on the basis of information available during the year in question. Three-year averages are used as far as possible in the calculation in order to balance out annual changes.[104]

In the calculation, the allocation between the universities takes place directly in proportion to the weighting of the criterion. Thus, there is no need for a separate coefficient for the size of the university. In other words, size is already accounted for through the use of absolute numbers in the indicators.

Documents are not very clear about how funding is allocated but as far as we can see the general idea is that the total core budget gets apportioned according to the formula, i.e. 75% are allocated to the component based on the quality, extent and effectiveness of activities and 25% to other education and science policy objectives, and so on. Within the 'portions', universities get funding based on their share of points of total points.

6.5.4 Effects

6.5.4.1 Effects of steering-by-results

In a study conducted in 2003[105], rectors and administrative directors of universities reported a wide range of autonomy in their daily operations. No one would be willing to return to the old steering system. They also reported that processes related to institutional planning and funding had simplified remarkably. Moreover, steering-by-results had quite effectively supported institutional profile building and sharpened universities' missions.

Steering by results was also seen to have increased competition between universities but in general university leaders felt that competition had been healthy and improved the functioning of the higher education system.

[104] So far, quantitative data from university databases have been used by the ministry. As there is a wish to have more qualitative data, the ministry is currently building up a new database with more qualitative information

[105] Seppo Hölttä, Eila Rekilä, 'Ministerial Steering and Institutional Responses: Recent Developments of the Finnish Higher Education System', in: OECD, *Higher Education Management and Policy*, Volume 15, No. 1, 2003

While seeing the benefits of the new steering system, university leaders worried about the additional reporting to the Ministry they had to do. Formal reporting was necessary in the existing steering model to guarantee accountability and involved not only the KOTA database but also reporting about national programmes and institutional projects.

The transfer to block grant (lump sum) budgeting was a major reform advancing the financial autonomy of universities in the early 1990s. However, university leaders felt that real financial autonomy had decreased, the more the implementation of the formula-funding system progressed, as the implementation of degree goals and national development programmes[106] used up an increasing amount of resources.

Maximisation of degree goals in the negotiations with the Ministry had been the rational strategy for each institution. As a consequence, funding per degree had decreased, as did non-earmarked funding. On the other hand, unit costs and productivity of the higher education system had improved dramatically.

6.5.4.2 Effects of formula budgeting

In general, the formula is seen as a success and a positive experience. Efficiency of universities has increased dramatically. For example, from 1985 to 2001 the increase in student numbers was 79.4%, in student intake 61.4%, in numbers of teachers 5.4%, in Master's degrees 43.6% and doctoral degrees 312%. It has also been effective in implementing priorities.[107]

However, some negative consequences have also been observed. The formula has forced universities to adapt their strategies. There is some evidence that as a result the innovative potential of universities has been not been fully deployed. Universities like to play it safe rather than do something risky because risky behaviour may be punished by the formula. In a similar vein, the formula has increased the homogeneity of universities.

There is also a feeling that the 'softer' subjects such as the social sciences and the humanities are losing out under the formula budget as the formula is better suited to the natural and technological sciences and medicine.

As mentioned before, though developed over time, the formula has been relatively stable since its introduction. However, it has lost the transparency it had at the beginning. Also, the incentives have come to be difficult to understand for universities. For this reason, a bigger reform will take place in 2013. The outcome is completely open but it is unlikely that there will be no formula-budget any more.

6.5.4.3 Expected effects of the 2009 University Act

The 2009 Universities Act has reformed and strengthened university governance and leadership to enable the universities to give a better and more flexible response to the challenges and opportunities of their new financial situation. The status of the rector and the university's own academic decision-making are expected be become stronger. The new, stronger financial and administrative autonomy will open up new

---

[106] In the previous system before the 2009 Universities Act, there were also national development programmes (also called targeted funding programmes). They had been in place since the late 1990s, e.g. in areas such as ICT, information society, training for teacher education. The targeted funding programmes were part of the performance contract negotiations. With the new Act targeted funding programmes were stopped. The funding is now given to universities as a lump sum, so that universities can sharpen their profile. It will be interesting to see how this strategic funding will be allocated and used internally

[107] The Finnish formula does not contain any ‚efficiency indictors', i.e. indicators that relate input to output. As far as we know (see Hanne Fosse, Performance indicators used in performance-based research funding systems, Paris, OECD 2010), ‚efficiency indicators' are not used anywhere in the world. However, there need not necessarily be efficiency indicators in a formula for a system to become more efficient. In Finland, the volume indicators linked to a certain sum of money were enough to make the system more efficient

opportunities but also give the universities new kinds of responsibilities. For example, finances and leadership will require new kinds of competences, as the responsibility for solvency and financial standing rests with the university.

6.5.5 Evaluation

The Finnish national quality assurance (QA) system in higher education has three components: government steering through performance contracts and formula budgeting (Ministry of Education), institutional quality assurance (universities), and national quality audits (through the Finnish Higher Education Evaluation Council FINHEEC[108]). The national quality assurance system is only linked to the allocation of institutional funding by the first component – government steering through performance contracts and formula budgeting.

According to legislation, universities are responsible for quality assurance. The quality of the activities is taken into account in the target output of the universities and in the monitoring of their attainment. Hence, the database KOTA is an important element in quality assurance. However, as there is no unambiguous and comprehensive way to describe quality in statistical monitoring, evaluations have their place too.

Every university has its own QA system covering education, research, societal interaction and support services. Each institution independently determines its own objectives, the organisation and methods of its own quality assurance (QA) system and defines the quality it aims to attain in its strategies.

The QA tools used by universities include quality handbooks and other documentation; quality criteria and indicators for education and research; descriptions of core and support processes; definition of quality responsibilities; student, teacher and stakeholder feedback; periodic internal and external evaluations; and systematic use of feedback given by management or teachers and improvement measures.

Most universities have procedures for periodic monitoring and review of programmes and awards. For instance, many universities regularly organize international evaluations of education and research.[109] Some of them carry out internal audits of study programmes and units.

The FINHEEC audits the QA systems[110]. The QA system are evaluated against institutional strategic objectives. As the audits look for evidence of comprehensiveness, transparency and effectiveness of each QA system, the institutions are invited to present concrete examples of improvement. An important evaluation criterion is that the QA system should cover all operations, not only education but also research and support services. The FINHEEC audit criteria emphasise quality improvement (formative evaluation).

The audits are performed on a registration basis. The university either passes the audit or, if it has major shortcomings in its QA, it has to take part in a re-audit within two years. One fourth of the audited universities have been re-audited so far. The first round of audits will end in 2011, after which it will start anew. A QA audit is valid for six years. Audits do not have any direct links to financing.

---

[108] FINHEEC, operating under the auspices of the Ministry of Education and the Academy of Finland, is an independent expert body assisting the higher education institutions and the Ministry of Education in matters relating to evaluation. The Council consists of 12 members representing universities, polytechnics, students and employers (see http://www.finheec.fi/)

[109] For example, education and research at the University of Helsinki undergoes international evaluation every six years

[110] FINHEEC also conducts thematic and programme evaluations in selected fields. It also organises evaluations relating to Centres of Excellence in education

University research is evaluated by the Academy of Finland. It issues a review of the state and quality of science and research in Finland once during the three-year term of its research councils[111]. In addition, the Academy conducts more detailed evaluations of fields of research (one every one or two years), mainly using international panels.

The results of the completed evaluations must be published. The reports published by FINHEEC and universities are all available at their websites and in the form of publications.

6.5.6 Lessons learnt from the Finnish case study

- It takes time to introduce results-based management, in Finland it started in the mid 1980s.
- Importance of a proper database: Finland started its data base in the mid 1980s. Performance contracting and formula budgeting are not possible without appropriate data to monitor the achievement of objectives in performance contracts and to calculate the formula budget.
- Continuity is important if you want to change the steering and the funding of the higher education system. Although governments changed, all political forces were willing to go through with the modernisation of higher education steering and funding.
- Complexity of the formula has increased over time so that it has lost its incentivising function. This seems to be a problem in the Netherlands too. Simplicity in the formula is important!
- Research outputs – especially scientific publications – play a very small role in the formula. Only 5% of institutional funding for R&D is allocated based on scientific publications, not least because it is so difficult to implement a sensible system. Working groups have been established, trying to get to grips with difficulties such as the different publishing behaviours in different disciplines. So far, the philosopher's stone has not been found yet. However, external funding is often used as an indicator to assess quality of research.
- Evaluation system not linked to funding system. Quality is partly monitored through funding system (performance contracts & formula budget) but a simple monitoring system is not enough, as evaluation comprises more than just monitoring.

6.6 Assessment of the Evaluation Methodology in the light of international practice

In this section we take the two functions of the Evaluation Methodology, evaluation of research and allocation of institutional funding, and assess each separately in the light of international practices. We also mention issues cutting through both functions of the Evaluation Methodology.

6.6.1 Evaluation of research quality

An EU expert group recently reviewed practices for assessing university-based research across thirteen different countries, mostly in Europe[112]. Among its more striking conclusions is that "There is no single set of indicators capable of capturing the complexity of research and research assessment." It stresses a number of aspects of assessment systems, notably

[111] Research Council for Biosciences and Environment, Research Council for Culture and Society, Research Council for Natural Sciences and Engineering, Research Council for Health

[112] Expert Group on Assessment of University-Based Research, Assessing Europe's University-Based Research, DTD.C4, EUR 24187 EN, Brussels: European Commission, 2008

- Consultation of higher education researchers in the development of assessment systems to ensure procedural fairness, transparency and a high level of acceptance
- The use of peer review panels, to ensure a broader understanding of the research being assessed, as well as of its contribution to knowledge, and to facilitate the assessment of research in emerging new disciplines and of interdisciplinary research
- The combination of peer assessment and bibliometric indicators
- The use of information about process and impact, including impact on teaching, to balance the focus on research output
- Self-evaluation as a key component in the assessment process
- Experiments designed to facilitate the measuring of societal impact
- Focus on units of assessment positioned somewhere between the individual researcher and the entire institution
- Unintended consequences of assessment exercises, e.g. that stakeholders make decisions contrary to the original objective(s) pursued, or that research quality is made the focus of attention to the detriment of other university functions.

The Expert Group proposed the following 'good practices'

- **Combine indicator-based quantitative data with qualitative information**, for example information based on expert peer assessment. This enables the quantitative information to be tested and validated within the context and purpose of the assessment, with appropriate reference to the discipline and disciplinary practice
- **Recognise important differences across research disciplines**. Peer–reviewed journal articles are the primary publication channel for practically all disciplines, but the complexity of knowledge has led to a diverse range of output formats and outlets
- **Include assessment of impact and benefits**. Because research does not exist in isolation, assessment should include indicators which are capable of capturing and recognising this. This differs for different disciplines. Stakeholder esteem indicators can show how research is viewed by the wider community
- **Integrate self-evaluation** as a useful way to include the research community proactively in assessing their own contribution, but also as a means of placing the research process – which includes the organization, management, and developments over time – into context and related to institutional mission.

The international case studies we have presented show that in particular the Dutch and the UK systems integrate these 'good practices'. In Austria, many universities have already implemented quality management and evaluation systems in line with these recommendations. This is much less the case in the Czech Evaluation Methodology: While qualitative information and self-evaluation are absent, the Evaluation Methodology is recognising the differences between the disciplines, trying to get to grips with them but in our opinion still not to a sufficient extent. With regard to the assessment of impact and benefits, the Evaluation Methodology does include research outputs but research outputs are direct research results and impact or benefits are not taken into account.

The Czech Evaluation system evaluates research **purely based on research outputs** produced. As such, it is unique. Other systems for evaluating research may

use a number of indicators to inform peer reviews[113], as does the UK RAE. The overview of evaluation practices in the Netherlands, the UK, Finland and Austria show other considerable differences from international practice:

- There is a worldwide trend towards more formative evaluation of research (not allocation of funding!). An evaluation system should provide useful feedback to participants to help them improve and develop their research. Both the Dutch and Finnish research assessment exercises adopt a **formative approach**. The Dutch system and most Austrian universities also allow research units evaluated to comment on the assessment they receive. Unlike the Austrian, Dutch and Finnish systems, the Czech Evaluation Methodology has no formative element, being retrospective and purely metrics-based. Thus, it does not create any opportunities for researchers, research managers and research institutions to learn and to improve quality. All it does is create incentives to produce more points.
- All research assessments we have looked at have a **larger scope** than the Czech one, taking into account dimensions that go beyond research outputs. They examine research quality in a broader sense, the Dutch system also taking into account research management, research policy, research facilities, PhD training, and societal relevance of research, while the UK RAE looks at research strategy, staff development, training of postgraduate researchers, and engagement with research users and the public. Austrian universities pursue similarly comprehensive approaches; some even evaluate organisational units in their entirety, looking all their tasks and activities (teaching, research etc.) in context.
- The research assessments give research institutes **sufficient time** to the react to the signals of the evaluation and take measures to improve quality in whatever field it was judged to be insufficient. For example, the Dutch system evaluates research quality in six year cycles, and at Austrian universities evaluation takes place once in five years. In contrast, the Czech Evaluation Methodology is applied every year.
- Since 2009, the Dutch Standard Evaluation protocol has taken into account **societal impact**, using the ERiC systems. An increased focus on societal impact is also planned in the UK Research Evaluation Framework. Indicators of societal impact used are cooperation with private sector and memberships of social organisations and policy bodies. In contrast, the Czech system only takes into account applications (mainly patents) with a strong conceptual bias towards industrial applications.
- While the Czech evaluation methodology is purely retrospective, the Dutch system is **both retrospective and prospective**, taking into account past performance as well as future plans.
- The Dutch Standard Evaluation Protocol has to be seen in a context of autonomy. In exchange for autonomy, universities are subject to **quality assurance**. In Austria, universities are fully autonomous in setting up quality management systems and in the evaluation of research but they are legally obliged to carry out evaluations. The same is true for Finnish universities.
- Last but not least, the examples show us that there does not necessarily need to be a (direct) **link** between the assessment of research and the allocation of funding. In the Dutch system there are no links with funding at all, while in the UK there is. In Norway there is also a direct link but only a very small part of overall resources (2%) is affected. In Finland and Austria, there is an indirect link between the two roles: Austrian universities have to report to the Science Ministry about the setting up of evaluations systems (but not about the results of evaluation), while in

[113] This is called 'informed peer-review' by the OECD, cf. Hanne Foss Hansen, Performance indicators used in performance-based funding systems, OECD, 2010

Finland universities' quality assurance systems are regularly audited by an independent body. Another interesting case is the French system, where there is also an indirect link with funding: the research assessment is used for negotiating performance contracts between the university and the ministry. The French system has the advantage of allowing capacity building.

Where evaluation is linked to funding decisions the link is generally not mechanistic. Unlike in the Czech Evaluation Methodology, the evaluation results are not automatically translated into budget allocations. Rather, they are used as one of several inputs into decision making.

6.6.2 Allocation of institutional funding

Our case studies and literature analysis have shown that the Czech Evaluation Methodology is quite a unique model of a performance based research funding system (RRFS). No other PRFS known to us and still in use is equally radical in its exclusive focus on research outputs produced in the past and in ignoring differences between different types of institutions and disciplines. Many other systems rely on quantitative indicators as well but they are all less radical than the Czech Evaluation Methodology because a) they use indicators other than research outputs if they use research outputs at all and b) they all are combined with a prospective funding approach, often a performance contract.

It is difficult to talk of best practice in the context of performance based research funding systems because countries are still experimenting with their systems, learning and improving them. Moreover, even if there were a best practice system it would still need to be adapted to another country's circumstances. Nonetheless, it emerges that a system combining a performance-based research funding system (funding formula) and a performance contract is quite common.

We have identified a number of critical issues:

- Both as the UK and the Finnish cases show, it takes time to introduce results-based management. Both started out in 1986. **Continuity** is important too. A striking aspect in Finnish public management modernisation is the high degree of continuity since the beginning of 1990 despite changes in government. In the UK too, the 1997 change in government did not result in abandoning the RAE.
- The performance based research systems we looked at in our case studies are applied in the **university sector** only. There are also performance based research funding systems for universities of applied sciences, but they typically employ different indictors as the universities of applied sciences have a different role in the research and innovation systems. The Dutch case study is a good case in point and also the Austrian universities of applied sciences have their own public funding system.
- Research funding is normally allocated based on **types of research organisations**, not disciplines, with each group of organisations – universities, universities of applied sciences, applied research institutes etc. – having their own funding regime.
- All the systems we have looked at either only move very small shares of money for institutional funding or they build in correcting factor to make sure that changes in institutional funding are not too great. The UK provides institutions with so-called moderation funding as soon as the change is larger than £100,000 and in Austria, annual cuts of the block grant for public universities are limited by law. All these measures are to make sure that institutions can cope with changes, and to **stabilise the system**.
- Having **appropriate monitoring data** is a necessity. Without a proper database it is impossible to create contracting and funding models and a planning system, as they all heavily depend on reliable institutional information.

- Research outputs – in particular scientific publications – play a small role because it is so difficult to establish with a sensible system that does justice to the different disciplines. A more typical – and simpler – indicator used for assessing the quality of research is **external funding** (distinguishing between grants and contract research).
- All the countries we looked at take into account that research (and teaching) in **some disciplines is more expensive** than in others. The reasons are obvious: some disciplines require expensive equipment while others need just 'paper and pencil'. The UK and Finland collect data on the costs of different disciplines and group them according to cost intensity. These data influence budget allocation. In the Czech Republic the points earned are directly converted into money, without any discipline-specific consideration.
- As the UK RAE shows, performance based research funding systems have **wide-reaching impacts** on the research and higher education system as a whole, on the way research is conducted, on research culture, and on power structures in the research system. They are both positive and negative. In Finland, the formula budget has clearly increased the efficiency of the higher education system. Steering-by-results has brought universities a large amount of autonomy in their daily operations. In any case, it takes years before the impacts of a new funding system can be understood and assessed fully. Therefore, patience is needed in order to overcome the temptation to meddle with the system prematurely. For instance, the Austrian formula budget is valid for a duration of two full funding periods and can only be adapted for the third period, based on an evaluation.
- Any system for allocating institutional funding needs to have a **prospective element** in order to allow institutions to build up capacity or to venture into new research directions. As the case studies show, this is typically done through performance contracts. Prospective elements can also be included in the funding formula. Again, this is not currently the case with the Czech Evaluation Methodology.
- Similarly, any system for allocating institutional funding needs to have room for **political prioritising**. In the UK, the government prioritises science, technology, engineering, and maths (STEM) subjects. While there is also room for prioritising in Finland, that is currently not the case with the Czech Republic with the Evaluation Methodology. In fact, the 2010 Evaluation Methodology produces results that contradict research priorities in the Czech Republic (see 4.2.4).
- Performance-based research funding system (formula) have a tendency to become **complex** over time. There are several reasons for this, not least that more and more policy concerns and correcting factors are built into the formula. In the Dutch case, the funding formula is more or less opaque to anyone not an expert in university funding systems. In the Finnish case, the formula has lost its incentivising function because due to its complexity universities do no longer understand it. Our observation is that the Czech funding formula has become more complex over time too. Complexity is a danger that should not be underestimated. Any new funding formula needs to be simple in the sense of having a clear message!

6.6.3 Cross-cutting critical issues

Based on the different analyses we have conducted, we have also identified a number of **critical issues** which are valid for both roles of the Evaluation Methodology and remain stable despite the efforts made to differentiate and detail the system from 2004 to present:

- Different types of institutions, their different missions, tasks, activities, and funding situations are assessed according to identical criteria despite institutions being very heterogeneous.

- Differences between disciplines are not taken into account accordingly:
  - Different types of results are not equally important for different disciplines and for different modes of performing research
  - Chances of actually succeeding in the scheme differ, as not all types of results are equally achievable for all scientific disciplines, especially the highly rated patents
  - Scientific disciplines differ in their publication cultures and patterns both in terms of publication type and frequency
- The exclusive focus on countable research results in widespread opportunistic behaviour, encouraging research actors to produce results that are 'fake' or irrelevant or of inferior quality.
- The Evaluation Methodology does not consider input factors, i.e. the different starting positions of the institutions are not taken into account.

7. Examples of international based practice in the Czech Republic

The two Czech case studies below describe and analyse evaluation and funding systems in the Czech Republic *other* than the Evaluation Methodology. They show that the Czech Republic already has some experience with procedures and methods of research evaluation and allocating funding that are more in line with international practice. They can be used as a first step towards an alternative system of research evaluation and for allocating research funding.

7.1 Internal evaluation at the Czech Academy of Sciences

7.1.1 Introduction

The 2004 Resolution of the Czech Government on the evaluation of research and its results that launched the evaluation of research in the Czech Republic was mainly written because there was a need for ex-post evaluation of research in the Czech Republic. This need concerned research institutes as well as research programmes and projects. While the evaluation of programmes and projects – entailing both qualitative and quantitative evaluation– seems to have become part of the policies of funding agencies, the evaluation of research institutes has acquired a purely quantitative character with the Evaluation Methodology versions of 2009 and 2010.[114] Moreover, the Evaluation Methodology has been primarily connected to the allocation of institutional funding, while evaluation of research in other countries is explicitly used to improve the quality, strategy and management of research groups and institutes.

The Evaluation Methodology leaves institutes free to develop an internal evaluation system. We have tried to find examples of research evaluation in different types of research institutes in the Czech Republic, but only found an evaluation tradition in place within the Academy of Sciences, which in fact goes back to the early 1990s. Universities and applied research centres are not used to evaluations of research other than the quantitative Evaluation Methodology. However, the Evaluation Methodology is not an evaluation tool in the proper sense of the word because all it does is count research outputs, and evaluation of research is more than just counting research outputs. Consequently, research evaluation is not an integral part of the research culture in the Czech Republic.

[114] See for the Czech versions of the EM: http://www.vyzkum.cz/FrontClanek.aspx?idsekce=18748

In line with policies of other European countries, it would be recommendable to make evaluation of research part of the scientific culture in the Czech Republic in order to improve the quality of research. However, international experience teaches that an evaluation culture and evaluation practices need time to develop. As the experience with evaluation in the Academy of Sciences can possibly be used as a first step in the building of evaluation capacity, this section describes and analyses the way in which evaluation in the Academy of Sciences has evolved. Starting with the history of evaluation in the Academy of Sciences, this case study presents an overview of current practice and the ways in which the evaluation in the Academy of Sciences can function as a first step towards the set-up and implementation of research evaluation in the Czech Republic.

7.1.2 The history of research evaluation in the Academy of Sciences

From 1993 onwards the Academy of Sciences has performed evaluations every four-five years. While being inspired by international practice, the ways in which the evaluations have been performed has changed over the years

7.1.2.1 The 1993 evaluation

Evaluations in the Academy of Sciences of the Czech Republic (ASCR) started in 1993 when the Czech House of Deputies significantly cut the budget for the ASCR.[115] As this meant that not all the ASCR institutes could be supported anymore, the ASCR set up an evaluation of its institutes. The objective of this evaluation was to obtain a verdict on the fate of the institute: should institutes be retained in the ASCR system or should they be closed down. The evaluations were performed by three ad hoc committees of experts from the Academy (internal) and Czech universities (external) who looked at the different divisions within the Academy:

- Mathematics, physics and earth sciences
- Life and chemical sciences
- Humanities and social sciences

At the time there were no international experts involved and there were also no databases with research results consulted. The final statement was based on personal knowledge about scientific activities and the assessment of scientific quality and productivity of the institute. As a result of these evaluations 26 institutes were closed down and the number of ASCR employees decreased by approximately one half to about 6000 people. However, it is stated that given this substantial cut, publication activity in the ASCR increased in the period after the cuts.

7.1.2.2 The 1995 evaluation

In 1995 there was another round of evaluations in the ASCR. This time there was no external reason to pursue these evaluations, but the ASCR wanted to increase the scientific quality and productivity of its institutes. Inspired by international evaluation practices, the Academy set up its own Evaluation Body which consisted of a mix of national and international experts (one-third versus two-third). The evaluations were based on a detailed and structured self-assessment report written by the institute and bibliometric information. In addition, site visits and presentations of teams took place and institutes were able to comment on the reports of the expert panels. While the result of the evaluation was again a choice between retaining or closing down an institute, all the evaluated institutes were retained in the ASCR system. The advice of foreign experts was seen as very valuable and a control system was set-up to monitor how identified weaknesses were remedied in the institutes.

---

[115] Information on the evaluation practices in the ASCR is derived from documents on this evaluation given by the ASCR and communications with the person responsible for evaluations in the ASCR at the moment, Prof. Petr Ráb

7.1.2.3 The 1999 evaluation

In 1999 there was another round of evaluations. The evaluation was performed by three committees for every ASCR division composed of internal experts from the ASCR (1/3) and external experts (2/3). The external reviewers were usually four persons, of which two or three were international experts. Institutes wrote a detailed and structured self-assessment report, summarizing research results of the period 1995 – 1999 and giving a research strategy over the period 1999-2003. In addition, site visits took place. The following evaluation criteria were used:

- publication activity and indicator analyses
- visibility and level of research (first application of impact factor calculations)
- research strategy and outlook on quality improvements
- success rates when applying for grants
- human resource policy (focus on diploma and PhD students)
- participation in scientific communities
- popularisation of research results

As a result of the evaluations the committees divided the institutes into four categories, with the last category equalling the earlier decision to close down an institute. The final verdict of the panel, in combination with a statement of the institutes was used to make the budgeting decisions of the ASCR. It resulted in differential budget support for institutional financing, thereby using the so-called coefficient adding 11%, 7% or 4% to the final annual budget of the evaluated institutes, in accordance with the results of the evaluation.

7.1.2.4 The 2004 evaluation

In 2004 the Evaluation Methodology Act provided the legal basis for another evaluation in the ASCR. The procedure and implications where exactly the same as in 1999, but the institutes' self-assessment reports covered the period 1999 – 2003 for research results and 2004 - 2011 for research strategies. In addition to the internal evaluation of the ASCR, another evaluation took place by seven Committees from the Ministry of Education. The two evaluations led to a division of ASCR institutes into four categories with implications for institutional financing. Again, differential budget support was granted using coefficients. In addition, seven ASCR institutes were substantially reorganized or closed down.

7.1.2.5 The 2008 evaluation

In 2008 the evaluation round was based on the governmental Evaluation Act and the internal decision of the Academy Council to check the quality of research in ASCR institutes. However, the way in which this evaluation was conducted was different from previous rounds. No external reviewers were involved, institutes did not write self-assessment reports, and there were also no site presentations of institutes. Instead there was a strong quantitative element to the evaluation with detailed productivity and quality indicators derived from the WoS and also from Scopus, plus a detailed overview of staff figures and success rates achieved applying for project funding. The statistics were prepared by the ASCR itself. In addition, a SWOT analysis was made by the institute itself, combined with the selection of the ten best research results of the institute and a detailed overview of student activities. The material was reviewed by a panel that gave a final verdict on the quality of research activities at the institute. Institutes were then invited to comment on the review panel's verdict. The evaluation had no effect on budget decisions. The evaluation of whole institutes instead of smaller research units resulted in minimal quality differences between the institutes.

The 2008 evaluation procedure was inspired by the Dutch Evaluation Protocol, which for instance also entails a SWOT analysis. However, it also has to be noted that there are still some important differences to the Dutch system. For example, the Dutch

evaluations do include international experts, on site visits and the careful examination of research strategies.

7.1.3 The current evaluation of research in the ASCR

At the time of this study, again an internal evaluation is taking place in the ASCR. This 2010/2011 evaluation is based on the internal decision of the Academy Council to increase the quality of research in ASCR institutes and to identify the weakest segments of ASCR research. This should also help with the internal division of institutional funding given to the ASCR by the government, based on the EM 2010.

The evaluation covers the period 2005-2009 and the process resembles the last round of evaluations in 2008 to a large extent. Nine committees have been set up, with external evaluators involved, and guidelines have been established that contain information on the background materials that institutes need to provide and on the evaluation criteria.

7.1.3.1 Evaluations committees

Nine committees for all the ASCR divisions were composed, three for every division (Figure 32).

Figure 32 Evaluation committees in the current evaluation in the ASCR

Division Mathematics, Physics and Earth Sciences
Section of Mathematics, Physics and Computer Science
Astronomical Institute of the ASCR, v. v. i. Institute of Computer Science of the ASCR, v. v. i. Institute of Information Theory and Automation of the ASCR, v. v. i. Institute of Mathematics of the ASCR, v. v. i. Institute of Physics of the ASCR, v. v. i. Nuclear Physics Institute of the ASCR, v. v. i.
Section of Applied Physics
Institute of Hydrodynamics of the ASCR, v. v. i. Institute of Photonics and Electronics of the ASCR, v. v. i. Institute of Physics of Materials of the ASCR, v. v. i. Institute of Plasma Physics of the ASCR, v. v. i. Institute of Scientific Instruments of the ASCR, v. v. i. Institute of Theoretical and Applied Mechanics of the ASCR, v. v. i. Institute of Thermomechanics of the ASCR, v. v. i
Section of Earth Sciences
Institute of Atmospheric Physics of the ASCR, v. v. i. Institute of Geology of the ASCR, v. v. i. Institute of Geonics of the ASCR, v. v. i. Institute of Geophysics of the ASCR, v. v. i. Institute of Rock Structure and Mechanics of the ASCR, v. v. i.
Division Life and Chemical Sciences
Section of Chemical Sciences
Institute of Analytical Chemistry of the ASCR, v. v. i. Institute of Chemical Process Fundamentals of the ASCR, v. v. i. Institute of Inorganic Chemistry of the ASCR, v. v. i. Institute of Macromolecular Chemistry of the ASCR, v. v. i. Institute of Organic Chemistry and Biochemistry of the ASCR, v. v. i. J. Heyrovsky Institute of Physical Chemistry of the ASCR, v. v. i.
Section of Biological and Medical Sciences
Institute of Animal Physiology and Genetics of the ASCR, v. v. i. Institute of Biophysics of the ASCR, v. v. i. Institute of Biotechnology of the ASCR, v. v. i. Institute of Experimental Botany of the ASCR, v. v. i. Institute of Experimental Medicine of the ASCR, v. v. i. Institute of Microbiology of the ASCR, v. v. i.

Institute of Molecular Genetics of the ASCR, v. v. i. Institute of Physiology of the ASCR, v. v. i.
Section of Bio-Ecological Sciences
Biology Centre of the ASCR, v. v. i. Institute of Botany of the ASCR, v. v. i. Institute of Systems Biology and Ecology of the ASCR, v. v. i. Institute of Vertebrate Biology of the ASCR, v. v. i.
Division Humanities and Social Sciences
Section of Social and Economic Sciences
Economics Institute of the ASCR, v. v. i. Institute of Psychology of the ASCR, v. v. i. Institute of Sociology of the ASCR, v. v. i. Institute of State and Law of the ASCR, v. v. i. Library of the ASCR, v. v. i.
Section of Historical Sciences
Institute of Archaeology of the ASCR, Brno, v. v. i. Institute of Archaeology of the ASCR, Prague, v. v. i. Institute of Art History of the ASCR, v. v. i. Institute of Contemporary History of the ASCR, v. v. i. Institute of History of the ASCR, v. v. i. Masaryk Institute and Archives of the ASCR, v. v. i.
Section of Humanities and Philology
Institute of Czech Literature of the ASCR, v. v. i. Institute of Ethnology of the ASCR, v. v. i. Institute of Philosophy of the ASCR, v. v. i. Institute of Slavonic Studies of the ASCR, v. v. i Institute of the Czech Language of the ASCR, v. v. i. Oriental Institute of the ASCR, v. v. i.

The committees did not only evaluate on the institute level, but also on the level of research teams. However, the units of evaluation were not always clear. Negotiations between the evaluation committees and the ASCR institutes resulted in the identification of 406 research units that were subject of evaluation.

Altogether the nine committees consist of 62 panel members that participate in the review. Six of the panel members are internal members from institutes of the ASCR responsible for the main organizational work of the evaluation. The majority of panel members are external peers from the Czech Republic, and altogether nine international experts are involved.

The selection of the external committee members was based on a combination of three sources:

- Proposals from the ASCR Institutes
- Proposals from the Academy Council
- Inviting of reviewers from previous rounds that proved to be suited for the job and who had a good understanding of the research and innovation system in the Czech Republic.

The external committee members were asked to evaluate the research units close to their disciplines. However, this did not result in full coverage, and some units of evaluation were only reviewed by one external reviewer.

7.1.3.2 Background materials for the current evaluation

The institutes and smaller units of evaluation were asked to prepare background materials for the Evaluation committees. The guidelines of the evaluation describe the content of these materials that focus on five areas of evaluation:

A) The quality and results of the scientific work and the recency of the topic with regard to world scientific trends and their innovation potential. The evaluation will take into account the scientific community's response to the results and the standing of the evaluated unit (institute/department) in an international context.

B) The importance and the specific contribution of the institute/department to society, taking into account social, cultural, or economic needs of the Czech Republic, or with regard to maintaining data files collected on a long-term basis, public service delivery, etc.

C) The participation of the institute/department in international collaboration, including the participation in international projects, information networks, organisation of important conferences, etc. The results of international cooperation and their significance for the work of the institute.

D) Human resources, material, and organisational issues in the activities of the institute and the prospects for development, including:

- division of the institute into departments, roles of the departments in the philosophy of the institute, evaluation of the importance and quality of departments by the management of the institute;
- internal promotion and evaluation procedures, and mechanisms for the determination of research priorities;
- age structure of the employees, especially with regard to the quality and results of the scientific work of employees likely to remain at the institute for the long-term;
- SWOT analysis of the institute and its departments carried out by the management of the institute.

E) Other activities and additional information, including:

- participation in grant and programme projects of the Czech Republic, listing particularly significant projects with regard to the scientific philosophy of the institute;
- educational activities of the institute, PhD training, presentation courses, preparation of textbooks, etc.;
- participation in the scientific community (membership in commissions related to scientific and research activities, editorial boards, grant agency bodies, science councils, or other participation in science administration and popularisation).

7.1.3.3 Evaluation criteria

According to the guidelines, the evaluation commissions will evaluate the institutes and their departments according to the following three criteria:

- The quality and number of research results achieved at the institute/department, innovation potential of the results, response from the scientific community (areas A, B, C).
- The standing of the institute/department in an international or national context of the field (area C).
- The prospect of the institute/departments, importance of the departments for the scientific strategy of the whole institute, feasibility of the research programme, the quality and reputation of the leading personnel in the departments, potential for the participation in new scientific topics (area D).

7.1.3.4 The timeline of the evaluation

At the beginning of 2010 the institutes submitted the required documents. Afterwards the experts were selected, and the institutes wrote their self-assessment reports. At the

beginning of 2011 the site-visits are taking place. This process will result in a final verdict of the panel and a statement of the institute, and a rating and clustering of institutes and their teams (research units) into five categories ranging from excellent to unsatisfactory. The evaluation results are expected to be available in March/April 2011.

At the moment there are discussions going on about the ways in which the research evaluations will affect the budgets of institutes. While the final results of this discussion are not known yet, the aim is to use the evaluation to come to a profile of all the institutes concerning their quality and future perspectives. The budget from 2012 onwards is now envisioned to have three components:

- A bonus for above-standard and standard groups and a cut for below-standard groups.
- A basic contribution to the running costs of institutes
- A programme orientated budget for broader research areas, like for instance information and communication science and technologies; energy resources and energy efficiency; nanosciences, new materials and technologies; biomedicine and quality of life; biodiversity, bioresources and ecology; climate changes, risks and nature; astronomy, space research and technologies; cultural, historical heritage and national identity; man and society.

However, as discussions are still going on this budget proposal is subject to change.

7.1.3.5 Comparison with our review of star research groups and institutes

Comparing the grades given to research groups and research institutes in the course of the latest ASCR evaluation with the results of our own reviews of star groups, we can safely say that *ASCR grades are in line with the assessments of the* reviews conducted in the framework of this international audit. The only exception is the Institute of Czech Literature, which received rather *better* feedback in our review, because the panel underlined the value of the Institute of Czech Literature as a locus of cultural memory.

Having said this, a direct comparison between the grades given in the ASCR evaluation and our qualitative review reports is not easy because, seeing the need for formative evaluations in the CR, we explicitly refrained from using grades in our own reviews.

7.1.4 Lessons learnt from the research evaluations in the ASCR

From the history of research evaluations and the current state of affairs it becomes clear that the ASCR has already quite a long tradition in the evaluation of research. Moreover, the evaluation of research in the ASCR is often – but not always – in line with international practice, combining quantitative elements (e.g. bibliometric information) and qualitative elements (e.g. peer review), in order to come to a conclusion on the quality of research and ways in which institutes can improve. With regard to the learning process the ASCR went through, it certainly is an inspiring example.

During the peer reviews of research institutes that was part of this international audit, it also became apparent that the research groups and institutes of the ASCR were familiar with research evaluation: the writing of a self-assessment report and site-visits in which the group/institute was presented to international peers, followed by a discussion on the strengths, weaknesses and opportunities of the group/institute.

Nevertheless, we want to emphasize that the evaluations at the ASCR are certainly not perfect and that there is room for improvement, for instance with regard to the independent selection of peers, the number of international peers involved, the selection of proper units to evaluate, etc. Therefore, we recommend learning from the experience of the ASCR with the evaluation of research, together with international experience to construct an evaluation system for the entire Czech research system,

which combines quantitative and qualitative information with international peer-review and site-visits.

Positive lessons learned from the evaluation practice of ASCR that we recommend:

- Broad definition of research quality: The evaluations do not only look at research outputs like the Evaluation Methodology. Besides examining research quality in the strict sense of the word, the AS evaluations also take into account research management, research strategy, research facilities, PhD training, and societal relevance of research, etc.
- Expertise of reviewers: In the evaluations of the ASCR it has been a challenge to find a way to match the very broad range of research activities in the institutes with the expertise of the participating reviewers. Balancing the expertise of the reviewed groups and the expertise of the evaluators is indeed complicated. On the one hand, it is important that the research groups or institutes are evaluated by reviewers that have the expertise to evaluate the quality of research of this unit. On the other hand, if there are in total 406 research units subject to evaluation, as is the case in the recent round of evaluations in the ASCR, it is clearly not possible to have one or more experts on all of these research units. Therefore, the ASCR has composed committees covering different sections of institutes. Although this is a good first step to come up with adequate evaluation committees, it is important that these committees are composed of reviewers with the required expertise. For that reason, we suggest a second step that makes sure that for every institute there is one reviewer with expertise in the subject that the institute covers who will be able to make a good judgment of the quality of research of the institute. In addition, it is important to note that a good evaluation of research *need not* exclusively be done by reviewers with expertise that exactly mirrors the expertise of the evaluated unit. As the evaluation not only concerns research quality but also the functioning of research groups, research strategy, human resource policy, etc. reviewers with expertise in related fields are also perfectly able to contribute to a good evaluation. And in case the specific review committee feels that they lack the expertise to judge the quality of the research of a certain sub-unit of an institute, they can and should use their own connections to find a person who can inform them on the quality of the specific research.
- Participation of foreign experts: Their inclusion in the evaluations is seen as a positive development, both by Czech scholars and in relation to international standards. However, we highly recommend to increase the number of international peers and take care that they are not only coming from countries close to the Czech Republic, but are representing a wider international scientific community. In addition, the selection of international peers should be done by an external and independent party.
- SWOT analyses: Based on Dutch evaluations practice the ASCR evaluation includes an analysis of strong and weak points of the research institute, as well as opportunities and threats. This analysis gives the reviewers a good insight into the position of the institute, but it also helps the institutes to reflect thoroughly on the state-of-affairs of the institute.
- Research strategy: the inclusion of a strategy of research was not always part of the evaluations of the ASCR. However, including a strategy is recommendable as it obliges the institute to think about the future of their research and get valuable feedback on their plans from international experts in their field.
- Human resource management: The evaluation documents of the ASCR also pays attention to the quality of the staff and the human resource policy practices in the institute. For example, there is a focus on MA and PhD students, and also on key scientific personnel. As human capital is the key resource of research, the focus on staff profiles and human resource management is very important. However, all the

different types of personnel should be included, also technicians and administrative personnel.

- Evaluation cycle: It is a very good practice that the 1995 evaluations in the ASCR were followed-up by another round of evaluations and that evaluation has become a regular practice. However, AS institutes get evaluated quite often compared with international standards. An evaluation every 5-6 years should be enough. This would give institutes enough time to react to the evaluation.
- Improve, reorganise or close down weak institutes: After the evaluations of 1995 the AS set-up a control system to monitor how identified weaknesses were remedied in the institutes. In line with this good practice, we suggest using evaluation as a basis for identifying weaker institutes in the Czech Republic, improving and/or reorganising them or in some cases closing them down. However, it is important to recognise that some institutes are vital for the Czech research system – because their task is required (e.g. meteorology) or they are unique (e.g. Czech literature or Czech language) – and they should be part of a healthy national scientific culture.

7.2 Performance contracts in the Operational Programme Research and Development for Innovation (OP R&DI)

We chose the Operational Programme Research and Development for Innovation (OP R&DI) because it demonstrates a way of allocating institutional funding that is new to the Czech Republic. The OP R&DI is a competitive programme with a limited duration and the particular aim of establishing new structures. As such, it differs from 'normal' institutional funding. On the other hand, the thorough evaluation of all proposals and the contract negotiations with those selected for funding has revealed a number of issues that are highly relevant for governing and funding **all** publicly funded research organisations in the Czech Republic.

7.2.1 The OP R&DI in numbers

- Duration: 2007-2013, operations until 2015
- Total budget: €2,436,095,160
- Priority axis 1: €806,347,498, Priority axis 2: €806,347,498
- Sources of funding: 85% European Commission, 15% national public sector
- Objects of funding: Centres of Excellence (Priority axis 1) and Regional Research Centres (Priority axis 2)
- Number of applications: 111 (15 PA 1 + 96 PA 2)
- Number of funded projects: 43 (8 PA 1 + 35 PA 2)

7.2.2 The background of the OP R&DI

By joining the EU, the Czech Republic was included among the member states that receive financial support from the EU Structural Funds. For the period 2007–13, all regions of the Czech Republic (except the capital city Prague) are eligible for funding under the 'Convergence' objective. In compliance with the objectives of European Regional Policy, the priorities of the Czech Republic are to strengthen the competitiveness of the state and to foster the orientation towards a knowledge economy.

The Operational Programme Research and Development for Innovation[116] is one of the important operational programmes in the Czech Republic. Together with the Operational Programme Enterprise and Innovation and the Operational programme Education for Competitiveness, the OP R&DI represents a mutually interconnected system of interventions, which aims at ensuring the long-term sustainable competitiveness of the Czech Economy and the targeted regions within the 'Convergence' objective.

The specification of the OP R&DI is based on a rigorous SWOT analysis of the Czech research and innovation system and embedded into numerous national strategies, particularly the National Development Plan of the Czech Republic 2007–13 (NDP), the National Strategic Reference Framework (NSRF), the National Programme of Reforms of the Czech Republic (NPR), the Economic Growth Strategy of the Czech Republic (EGS), the National Innovation Policy of the Czech Republic for years 2005–10 (NIP), and the National Research and Development Policy of the Czech Republic for years 2004–08 (NR&DP).

7.2.3 The objectives of the OP R&DI

The global objective of the OP R&DI is to strengthen the research, development and innovation potential of the Czech Republic which shall contribute to economic growth, competitiveness and the creation of highly qualified workplaces. The OP R&DI has chosen a two-tier strategy for supporting R&D:

- **Centres of Excellence**: A relatively limited number of high quality interdisciplinary research centres. They are to use and strengthen the potential of existing high-quality research teams to become internationally visible research partners. The Centres of Excellence are expected to contribute in an essential way to the creation of new knowledge, to the training of (young) researchers and to the advancement of cutting edge science and technology in their respective field. The focus is clearly international and the centres of excellence have to be orientated towards the international research community.
- **Regional R&D Centres**: A larger number of sector-specific application-oriented, demand-driven research institutions that already have, or have the potential to develop, strong partnerships with the application sector. Their contribution is expected primarily in networking and close collaboration with the users of their results through contract research and the provision of services meeting the demand of the application sector, and in the training of people for highly qualified positions in R&D.

A key feature of both Centres of Excellence and Regional R&D Centres is an emphasis on performance. The Managing Authority negotiates a performance contract on an individual basis with each of the centres which forms the basis of the strategic orientation and operational activities of each centre. The performance contracts contain binding targets for the centres to achieve. The contracts are set individually, taking into account the different research profiles and related user groups. In its essence the performance contract is based on promises for future performance which have to be based (i) on credible past achievements of key staff, and (ii) on a proper governance and management system. The performance orientation is built into the research programme as well as into the governance of each individual centre. This new feature, the performance contract, is of crucial importance for the overall strategic orientation and steering of the research activities supported through the OP R&DI and it represents a completely new, progressive mode of operation in the Czech research policy.

[116] The Czech Republic, Ministry of Education, Youth and Sports, Operational Programme Research and Development for Innovation, Prague, July 24, 2008

7.2.4 The decision making process for funding projects in the OP R&DI

In traditional research funding two criteria prevail: (i) quality of the research programme and (ii) qualification of key staff. As the objects of funding in the OP R&DI are highly autonomous research organisations the funding decisions have to reflect a wider range of criteria: not only the quality of the research programme and the qualification of key staff, but also the application potential and attractiveness for users, the management and governance model, human resource policy, budget and funding, and, not least, criteria reflecting environmental, spatial, and cohesion policy considerations.

Accordingly, a decision maker would have to be skilled in a set of rather heterogeneous fields – which is obviously difficult to find in one person. To overcome this bottleneck the decision making process has been transformed into a sound process of subsequent checks, balances, negotiations, and decisions involving different actors with different expertise at different steps.

Novelty 1: A combined evaluation and negotiation process. The evaluation of proposals for funding is mainly based on the knowledge and experience of peers. At the same time the availability of expertise is limited by the availability of experts which are rather reluctant to spent more than a couple of days abroad. Taking into account these restrictions a sequence of several distinct and in some parts overlapping steps has been designed:

1. Submission of proposal
2. Formal evaluation and evaluation of eligibility
3. Expert evaluation
   - Evaluation of construction and technical aspects
   - Evaluation of application potential, including synergies and financial sustainability
   - Panel for applications
   - Evaluation by international experts
   - International panel
4. Selection for negotiation
5. Specification of performance parameters ('negotiation')
6. Funding decision

The main emphasis in this process is on evaluation and negotiation. The evaluation part itself is composed of a series of steps involving different types of expertise and thus of different types of experts. After the formal eligibility check, evaluation is based on two pillars:

- Evaluation of the application potential performed by Czech experts mainly from industry and other application spheres, and by experts on technology transfer, financing of R&D, and management of R&D. Their task is to assess the application potential within the Czech industry and other sectors of application.
- Evaluation of the quality of the proposal, in particular the quality of the research programmes and of the key staff, the specification and understanding of users, human resource development, management, and budget & financial sustainability. These experts from abroad are experts in the respective research field and experienced in the management of research programmes or institutes.

The novel element in the funding process is the negotiation of performance contracts. Due to the volume and complexity of the projects hardly any proposals fulfil all relevant criteria sufficiently. There is always some room for improvement, particularly in those areas in which academic researchers typically have their blind spots:

orientation of their research activities towards carefully selected user groups on the one hand, and governance, management and human resource development on the other hand. At the same time, it has proven important to direct the attention of key personnel, i.e. of the centres' directors and the managers of research programmes to manageable outputs rather than to uncontrollable impacts. This disburdens them from feeling responsible for the well-being of the greater system and establishes a high level of commitment to and responsibility for outcomes. Accordingly, the negotiation about types and volumes of outputs such as publications, patents, income from research contracts and grants etc. immediately results in increased (final) performance. At the same time, the case-by-case approach taken in the negotiations makes it possible to tailor these output targets to the justified specificities of every research field.

Novelty 2: A limited number of evaluation criteria. Common sense tells us that the evaluation of complex systems require a complex set of criteria. This is certainly true. Unfortunately, in practical terms, this often leads to lengthy lists of evaluation criteria which causes a number of serious problems: the more criteria, the more difficult it is for evaluators to maintain an overview; the more criteria, the higher the propensity of evaluators to reward mediocre proposals as they typically fulfil all criteria 'somewhat' ('the average drives out the best'); the more criteria, the higher the propensity to aggregate individual scores by statistical methods rather than by consensus processes. To overcome these methodological problems, a rather simple but at the same time comprehensive set of evaluation criteria has been used:

1. Quality of the proposed research programme
2. Qualification of the key staff / team
3. Understanding and specification of users
4. Management and governance
5. Human resource policy
6. Budget & financial sustainability

These criteria guided the entire process: the applicants had to substantiate these questions in their proposals; evaluators had to assess the proposals along these criteria and to comment on each of them, and the negotiations with the centres selected for funding were structured accordingly. To maintain a certain consistency, up to ten 'guiding questions' per chapter supported all parties. Both the applicants and the evaluators had a certain liberty to respond to those guiding questions which were most suitable for 'telling their story'. This approach encouraged the applicants to write a convincing story, composed of six chapters and evaluators had to judge whether the story was convincing. Of course, both had to make extensive use of numbers, facts, milestones, quantifications etc. to substantiate their stories and their assessments.

7.2.5 Negotiation of performance contracts and specification of project performance parameters

One of the most outstanding experiences in making funding decisions is that hardly any proposals will be accepted without further adaptations. Taking into account that projects to be funded within the OP R&DI are not just (large) research projects but fully fledged research centres with a size ranging from 50 to 1000 employees, it is clear that further improvements here and strict conditions there have been the rule rather than the exception. Accordingly, proposals which were generally accepted for funding by national and international panels still had to undergo substantial adaptations in particular aspects.

While evaluators are perfectly able to address critical issues and indicate the direction of change, they are not in a position to ad hoc determine the required changes in detail, mainly for two reasons. One is mainly cognitive: the complexity of the proposals and the importance of contextual factors vis-à-vis the limited resources (time, information) does not allow an in-depth analysis and determination of adjustments in

panel meetings. The other reason is more psychological: it has turned out that applicants' acceptance of the final performance contract is higher if the applicants themselves are invited to propose adjustments and negotiate these proposals rather than to react to propositions by the negotiation committee. It is not necessary for the negotiation committee to determine the type and rate of change in advance, while it is certainly helpful to perform an effective negotiation.

7.2.6 Experiences and observations in the evaluation and negotiation process as a mirror of the 'state of affairs' in major parts in the Czech research system

In the course of the evaluation and the contract negotiations we have made a number of recurring observations in the proposals, across applications from all organisation types and subject specialisations. In the panel review of research institutes which was performed during this study, we found many of the same dominant characteristics (and shortcomings). Hence, the following summary of these observations mirrors some important issues of the Czech public research system.

- **Lack of focus of individual research programmes and lack of coherence between them**. Many of the proposals have put together (i) too many, (ii) poorly connected, and (iii) ill focussed research projects and programmes. Accordingly, adaptations have been demanded to reach a more compact research profile. As the composition of the research programmes are to a certain extent defined by contextual factors, neither the evaluators nor the negotiation committees were in the position to make specific suggestions for the re-direction of the research agenda. The only way out was to request adjustments from the applicants and to discuss and negotiate the up-dated version. Research teams / organisations with a more coherent research agenda have been rewarded as they have established preferred conditions for exchange and collaboration and thus for learning and higher productivity in the future.
- **Dominance of supply-side orientation vis-à-vis poor understanding of the target group**. Target groups are often either poorly defined or, even more, the proposals are missing the specific mode of access to and interaction with the chosen target group which can range from the international research community to the local farmers association, depending on the research topic. A lack of coherence in the research programme is a strong indicator for the dominance of supply-side thinking in planning research activities. Research teams – ceteris paribus – which have a clear orientation to their user groups are undoubtedly in a better position to achieve high levels of performance. Again, applicants themselves had to propose adaptations which were then discussed with the negotiation committee.
- **Overly complicated management and governance models**. Two observations are prevailing: (i) An underestimation of the role and power of management and governance and (ii) overly complex and inefficient management systems. Underestimating the power of well functioning management systems leads to a risk of missing opportunities. The most frequent kind of bad management models is an unclear separation between executive, supervisory, and advisory functions at the cost of a restricted range of action. In these particular cases proposals for improvement have been made by the negotiation committee, mainly aiming at a strict separation of the mentioned functions and a leaning of the bodies, particularly by empowering the top executives. We noticed that many applicants perceive management as yet another administrative burden and underestimate the critical role of management and governance as a supportive framework for research: "Management is the difference which makes the difference!" Thus negotiation can help to change people's minds and increase the long-term performance of research institutions.
- **Underdeveloped human resource policies**. Many, particularly academic, researchers exhibit a rather defensive attitude towards human resource policy. Rather, they consider it a tiresome administrative burden. Again, the benefits

from a more explicit recruitment policy, career development policy, mobility policy, not to mention of a well-thought gender policy are often not perceived as a source of performance gains and job satisfaction. Awareness for a systematic development of young researchers is particularly lacking. A poor or missing HR policy is often linked to and an indication of a lack in long-term thinking.

- **Poorly focused or poorly linked research programmes**. Those applicants which exhibited a lack of focus and / or coherence in their research programmes had to face substantial cuts, particularly in their investments in advanced equipment. Accordingly, those investments which could not convincingly related to an accepted research agenda, experienced severe cuts.
- **Research contracts and / or grants as indicators of performance**. Both research contracts and research grants are not only sources of income but also indicators of quality. Many applicants are hardly aware of the indicator role research contracts and grants can play. Therefore the re-balancing between the different sources and levels of research funding – institutional funding, research contracts, research grants – has been a permanent issue dealt with in the negotiations. As a consequence, the generation of income from reputable contracts and from certain grants shall be considered as an indicator of performance. This means that input indicators sometimes can be considered pretty good indicators of performance.
- **Low research productivity.** Increasing the production of (tangible) outputs (publications, income from grants and contracts, patents, training, etc.) has been an important issue in the negotiations. The OP R&DI aims at strengthening the research, development and innovation potential of the Czech Republic which shall contribute to its economic growth, competitiveness and to the creation of highly qualified workplaces, but the individual centres and their intentions are not measured against these superordinate goals. Rather, they are oriented at goals which can be directly influenced by management actions, i.e. putting the right people at the right place, allocating money in a suitable way, implementing policies for the recruitment and development people, setting specific goals for different subunits, etc. The most important decision was thus to set tangible goals which can be directly influenced by managers. Accordingly applicants are determined to achieve outputs rather than impacts. Specific targets for publications of different type and quality, income from grants and contracts, patents, training, particularly of PhD students, third party certification of laboratories, etc. have been agreed upon individually for each centre. Depending on their fields of research and on their target groups, the centres had the chance to customise their output profile.

  To provide an example: A regional R&D centre which performs research contributing to the safety of roads, or to fodder crops better adjusted to soil and climate in Eastern Moravia, or to more efficient production machinery for Czech manufacturing companies, has to publish in the Czech language in any case in order to reach its main target users – Czech policy makers in the transport sector and in agriculture, farmers and their associations, or small companies in Northern Bohemia. Furthermore, the researchers have to present their findings and solutions in national and regional workshops and seminars to their target users as otherwise they will have no serious opportunity to create any impacts. This research centre would indeed miss its mission if it failed to be present in local meetings and mainly published in international journals with high impact factors because their main target groups do not read international journals with a high impact factor.

  The negotiation process had in fact two effects: it raised awareness for a proper balance between the mission of the centre, its target groups and the profile of its outcome and performance. At the same time, the respective levels of outcomes have been significantly increased – on the average by more than 30%, which can

be considered as a considerable leverage, which even increases if one takes justified budget cuts into consideration.

7.2.7 Lessons leant

The experience made in the evaluation of applications for research centres and the negotiation of performance contracts have clearly shown that it is right to focus on tangible outcomes in order to raise the productivity of the Czech research system to an internationally competitive level. Increasing output numbers was no end in itself: understanding the target users of the outputs and tailoring the related plans towards them have been equally important. However, experience has also shown that weak attention to governance, management, and human resource policy ends up in numerous missed opportunities and under-performance of the centres. Moreover, funding decisions without negotiations would not only result in an underdeveloped organisational performance (management, governance, HR policy, IPR policy), but also in missed additional output performance – 30% on the average.

8. SWOT analysis of the Evaluation Methodology

8.1 Strengths

The idea of a transparent system for allocating institutional funding

The Evaluation Methodology with its reliance on quantitative indicators is driven by a desire to de-politicise and de-personalise the funding process. By relying only on (proven) past achievements, it aims to act against nepotism, corruption and lobbying.

Clear signal

The Evaluation Methodology sends out a clear signal. To put it simply, its message is: no outputs, no institutional funding. As such, it tries to address the issue of low productivity in some research organisations in the Czech Republic. At the same, it is intended to improve the quality of research outputs. However, while the Evaluation Methodology may be the right tool to increase the *number* of research outputs it is not necessarily the right tool to increase the *quality* of research outputs.

The idea of linking evaluation to decision making and policy making

Evaluation needs to have consequences if it is not to be an empty exercise, all the more so as evaluation is resource-intensive. However, for various reasons set out in this report, the way in which the Evaluation Methodology evaluates research and the way it is linked to allocating institutional funding is far from perfect.

8.2 Weaknesses

Reductionism

Both as an evaluation tool and as a mechanism for allocating institutional funding, the Evaluation Methodology is reductionist because it only considers outputs. It counts them and awards points with the only target of maximising the number of outputs. As an evaluation tool, it ignores key factors, such as research management, research strategy, and human resource policy – these are factors that any evaluation worth its name needs to take into account. By focusing on countable outputs only, it is not in line with international good practice.

As a mechanism for allocating institutional funding, it only considers past performance, not leaving room for institutions to build up capacity and venture into new research directions. This is exacerbated by the "automatic" translation of evaluation results (i.e. the point scores) into the amount of funding allocated. Moreover, at a political level, the Evaluation Methodology fails to take into account other relevant (thematic) policy priorities in the Czech Republic or in effect even contradicts them. Again, this is not in line with international good practice. Hence, our

conclusion is that the existing Evaluation Methodology is inappropriate for both the evaluation of research quality and the allocation of institutional funding.

Not addressing differences between disciplines sufficiently

The EM has tried to take into account differences between disciplines by including different research outputs, by introducing separate point ratings for National Reference Framework of Excellence (NERR) specialisations and in 2010 by introducing 'dampening factors' among ten disciplinary groups. However, disciplines are still treated unequally: unequal treatment occurs when defining what counts as research result and how many points this research result gets. The present system ignores important (legitimate) differences between disciplines: they differ significantly in their publication patterns both in terms of publication type and frequency, and not all types of results defined in the EM are equally achievable for all disciplines (e.g. patents). Moreover, there are still considerable differences between disciplines within the ten disciplinary groups. Also, as a basis for funding decisions, the EM does not take into account that some disciplines are more expensive than others. In short, the differences between disciplines cannot be captured in a formula, no matter how complex it becomes.

Not differentiating among different types of organisations

The Evaluation Methodology treats all research organisations in the same way, ignoring their different missions, goals, roles and tasks. In the current EM each type of research output is awarded a certain number of points. This number is the same for all institutions without regard to the importance this particular output has (or should have) for a given institution. As a result, unequal treatment occurs when defining what counts as research result and how many points this research result gets. In other countries, evaluation systems as well as performance-based research funding systems take care to make a distinction between different types of research organisations to account for their different missions (e.g. emphasis on research, teaching or application). In fact, performance based *research* funding systems have almost exclusively been tried for universities. There are also PBRFs for universities of applied sciences, but *in line with their mission*, they mostly focus on teaching, and PBRFs for applied research institutes typically emphasise contract research. The Czech EM lumps together different types of research organisations despite their different missions and roles in the Czech RDI system – missions and roles that have not even been subject to debate in the development of the Evaluation Methodology.

Eligible outputs and their value partly arbitrary

The outputs that count as well as the number of points given to those outputs are arbitrary. There is no explanation why certain outputs are chosen and others not. What is more, there is an implicit funding decision embedded in the points given to outputs, in particular regarding the value of applied (industry-oriented) research versus basic research. This decision has not been made explicit.

8.3 Opportunities

The Evaluation Methodology as a break with the past

Researchers and policy-makers both felt that a new system was necessary to evaluate research and to allocate institutional funding. The introduction of the Evaluation Methodology has been an intervention that brought the change needed, but it has not brought the proper change. Nevertheless, it has opened the floor for the introduction of a more appropriate system that is closer to good international practice.

Opening up a debate

Dissatisfaction with the Evaluation Methodology is great, as we have seen in the survey and interviews. However, considering the Evaluation Methodology's visibility, this could be a great opportunity to start a debate about institutional funding and its purpose. Such a discussion seems to be necessary as we found that it is not always

clear for research-performing actors what institutional funding is for and what project funding (targeted funding) is for. Furthermore, given that the Evaluation Methodology is not really a tool for *evaluating* research but much more a mechanism for allocating institutional funding (and as such a rather bad one), this could also open a window of opportunity for debating research evaluation and how to build up an evaluation culture. Last but not least, this debate should take into account international experience and learning from others.

8.4 Threats

Widespread opportunistic behaviour

The Evaluation Methodology provides incentives for opportunistic behaviour, that is unintended behaviour that perverts the intention of the Evaluation Methodology (ie to increase research quality) and encourages cheating. As a result of widespread opportunistic behaviour, the Evaluation Methodology is unable to distinguish between good 'gamblers' and good researchers, or, put more prosaically, between weak and strong research groups.

Unpredictability of institutional funding

The Evaluation Methodology causes large – and unpredictable – changes in institutional funding, making institutional funding unreliable and planning for an institution difficult. Furthermore, as research organisations do not know how much money they will earn for their points in the following year, they have an incentive to hunt for as many points as possible. Of course, this results in inflating the total number of points, with the result that a point is worth less and less.

Instability of research system

As mentioned before, the Evaluation Methodology can cause large changes in institutional funding. Large changes in institutional funding are a problem, as radical increases are as difficult to accommodate as radical reductions. The changes at an institutional level have a big impact on the stability of the research and innovation system as a whole. Therefore, all other systems of allocating institutional funding that we looked at have built-in mechanisms to make sure that the RDI system maintains stability.

Short-term thinking encouraged, long-term planning discouraged

Because of the instability and the unpredictability of institutional funding, institutes start to focus on short-term solutions to solve immediate problems. However, long-term planning is important to create high-quality research results, especially in a research environment where time horizons are long. For example, it takes years to set up a PhD programme that delivers highly qualified researchers and the publications of a new research project will only appear years after the project started.

No room for improvement and building up of capacity

Because it is purely retrospective, looking at past performance only, the Evaluation Methodology does not leave room for improvement and building up of capacity – one of the purposes of institutional funding. As a purely retrospective mechanism, it is also very sensitive to initial conditions, i.e. an organisation's conditions at the point of introducing the Evaluation Methodology. If an organisation was having a good, successful period, this was frozen and possibly reinforced by the Evaluation Methodology. If it was in a weak trajectory, this was frozen and reinforced too. What is more, the Evaluation Methodology does not give feedback, e.g. on research management, research strategy or human resource policy, which we identified as *the* key issues for the development of Czech R&D. Hence, the Evaluation Methodology does not support learning and development of institutions, research groups, and researchers.

Contradiction between strategic objectives and results produced by the Evaluation Methodology

Our analysis shows that the Evaluation Methodology leads to significant negative impacts on the funding of research in scientific and technological fields that were explicitly indicated in the policy documents as key technologies for the achievement of the National Priorities for Applied RDI 2009-2011. This demonstrates clearly that the mechanistic approach of the Czech PRFS cannot accommodate other (more) important policy considerations in connection with research.

Patching up the Evaluation Methodology

The Evaluation Methodology has been changed annually. Attempting to correct the most obvious imbalances, the changes often represented detailed refinements or restrictions. We have found that the frequent changes of the Evaluation Methodology cause confusion, with researchers no longer being sure what counts and what does not. More importantly, patching up the Evaluation Methodology only makes it more complicated and cannot remedy the more fundamental problems.

Weak ownership

The Evaluation Methodology has little support among researchers and research leaders as well as policy-makers. This can lead to frustration and low morale among researchers, which is not conducive to research quality and productivity. Ultimately, it can also lead to brain drain – researchers either leaving research or leaving the country. Needless to mention that it takes 15 years to train a researcher.

Using the Evaluation Methodology within institutions

All the threats are exacerbated when the Evaluation Methodology is applied within institutions because opportunistic behaviour, unpredictability, instability and short-termism will then increase within the walls of institutions, causing a research environment that is not conducive to improving the quality research. A lack of good leadership and professional management in many Czech research organisations exacerbates this threat.

Unknown effects

Not all the effects of the Evaluation Methodology on the scientific community are well understood. This is because effects take time before they become apparent, and the Evaluation Methodology has not been in operation for very long yet. What is more, there typically are different forces at play that make the prediction of effects difficult. For example, the survey shows that the Evaluation Methodology is not conducive to collaboration. This makes sense as individual publications deliver more points than co-publications, so researchers might choose not to collaborate. However, it might be argued that in time collaboration will increase as collaboration normally leads to more publications. As features such as collaboration - but also others like interdisciplinary research - are important aspects of scientific practice, the intended and unintended effects of new measures should be carefully considered before implementing a new system.

9. Recommendations for allocating institutional funding and evaluating research

9.1 Introducing the recommendations

As we have stressed throughout the report, the Evaluation Methodology has two functions. On the one hand, it allocates institutional funding; on the other, it evaluates the quality of research. Of course, there is a direct and automatic link between these two functions, whereby institutional funding is allocated based on the points research

organisations collect in the research evaluation; the research evaluation itself consists of counting the organisation's research outputs.

As our analysis in this reports shows, the Evaluation Methodology does not fulfil either of these functions well. With regard to the allocation of institutional funding, the Evaluation Methodology introduces significant threats and is likely to cause discontinuities in the Czech RDI system. With regard to evaluating research quality, the Evaluation Methodology does not reach international good practice and, in particular, fails to deliver the kind of information needed for the advancement of Czech research organisations.

For this reason we strongly recommend discontinuing the Evaluation Methodology. Improving the Evaluation Methodology bit by bit, as is currently being done, is not a solution, as these incremental improvements will not remedy the Evaluation Methodology's most fundamental flaws, such as the Evaluation Methodology not differentiating among types of research organisations despite their different missions and roles, or the Evaluation Methodology treating the different disciplines the same although they produce different research outputs and have different publishing patterns, or the Evaluation Methodology not being forward-looking, thus leaving no room for improvement and capacity building (see also the SWOT analysis in the preceding section 8).

For the allocation of institutional funding we recommend replacing the Evaluation Methodology with a combination of a new *funding formula* and *performance contracts*.

- A *funding formula* is a set of indicators based on which (a part of) institutional funding is allocated. How much institutional funding a research organisation receives, depends on the scores it achieves. The funding formula takes up the idea of the Evaluation Methodology by basing the allocation of institutional funding on past performance.
- *Performance contracts* are a contract between a principal (e.g. a ministry) and an agent (e.g. the research organisation). They define objectives that the research organisation needs to attain and are embedded in proper procedural steps to prevent nepotism. Performance contracts are not purely prospective: past performance is the basis for credible and achievable, yet ambitious objectives. As such, they differ from the research intentions. Indeed, performance contracts must not be confused with research intentions. As our description shows, the two instruments differ substantially (see sections 7.2 on the OP RD&I and 9.3.2).

For the evaluation of research quality we recommend replacing the Evaluation Methodology with *informed peer review* with independent selection of (international) peers. We do not recommend using the assessments of the peer reviews for allocating institutional funding. Not doing so will permit the Czech RDI system to build up a culture of evaluation and allow policy-makers and research organisations to come to see evaluation as a tool for learning and improving research.

There will *not* be a direct, automatic link between the two functions, i.e. the allocation of funding (through formula budget and performance contracts) and the evaluation of research (through peer review). There can be an indirect link, which we will discuss below (section 9.4.3).

Before spelling out our recommendations in more detail, first attending to the allocation of funding (including the implementation trajectory) and then the evaluation of research, we will start with the fundamental ideas which have lead us to the recommendations outlined above.

9.2 Fundamental ideas

Below are the fundamental ideas on which we base our recommendations for the allocation of institutional funding and for the evaluation of research.

Performance and Achievements

It appears that an important motive for introducing the present Metodika hodnocení was the observation that some organisations receiving institutional funding had not reported any outputs in the RIV database, often for several years. External peers engaged in the evaluations of proposals in the OP R&DI also found cases of low productivity.

We support the basic idea of linking institutional funding with performance – performance accomplished in the past and performance agreed on for the future. To assess performance it is necessary that the principal (e.g. the ministry) and the agent (e.g. the research organisation) have a shared understanding of what the performance is which has to be achieved with institutional funding. Hence, the fundamental question that will have to be asked is: What do we (as an organisation) want to achieve?

Building Institutional Capacity

Our ultimate aim is to base institutional funding to a large extent on performance contracts. We would welcome performance contracts immediately but think that the Czech administration is not yet professionalised enough to implement them right away. Expertise is clearly a bottleneck.

Hence, institutional capacity will need to be built up. Ministries and research organisations alike will have to go through a learning process and learn how to draw up and negotiate a performance contract. Just as Western European countries started to learn how to handle performance contracts some ten years ago, the Czech Republic will have to go through the same process of learning and building up expertise. However, for expertise to build up, there needs to be continuity in ministries (and research organisations); people need to be allowed and encouraged to build up expertise.

There is already a basis for such a learning process in the Czech Republic: The Centres of Excellence and the Regional Research Centres within the Operational Programme 'R&D for Innovation' base funding on performance contracts, the related negotiations have just been concluded. In other words, there already is some knowledge about performance contracts in the Czech RDI system and there will be more as these Centres move on. We will recommend actively sharing these – and international – experiences.

Retrospective and prospective elements

For the allocation of institutional funding, we recommend introducing two complementary systems – a formula budget and performance contracts. With regard to the formula budget, we recommend using a variety of (simple) indicators instead of just research outputs as is currently the case; this is the retrospective element in the allocation of institutional funding. As every funding system also needs to have a prospective element, to build up and develop capacity and venture into new research directions, we also recommend introducing performance contracts in the long run.

The advantage of the mixed funding system we are suggesting is that it combines a prospective (performance contracts) with a retrospective element (formula). As such, the two elements are complementary.

Transparency

The Evaluation Methodology with its reliance on quantitative indicators is driven by a desire to de-politicise and de-personalise the evaluation and funding process.

The mix of funding formula and performance contracts we are proposing to introduce takes up the idea of transparency. While it is immediately obvious why a formula budget – with its reliance on indicators – is transparent, this is also the case for performance contracts as long as objectives are transparent and monitored and negotiations are embedded in an appropriate structure which prevents nepotism and

ad hoc solutions. Hence, we recommend introducing them step by step and putting in place a number of safeguards.

As for *research evaluation*, we do not think that it can do without peer review, as the examples from other countries show. We will recommend putting safeguards in place to make sure that the peer review process is transparent, fair and as objective as possible. These concern mainly the independent selection of – preferably international – peers. However, we refrain from recommending a peer review based system for allocating institutional funding like in the UK or in New Zealand, as such a system would prevent the Czech RDI system from building up a constructive understanding of research evaluation and particularly peer review.

Different models of governance and funding for different groups of organisations

The present Evaluation Methodology is valid for more or less three different groups of institutes: universities, the Academy of Science, and research institutes (both public and private). There is also a string of other organisations that receive institutional funding such as hospitals, museums etc. These have different missions and tasks and also access to other funding sources, and research is not always their key activity.

In the current Evaluation Methodology each type of research output is awarded a certain number of points. This number is the same for all institutions regardless of the importance this particular output has (or should have) for a given institution. At first glance, the scoring system seems rather arbitrary. In fact, there is an implicit funding decision embedded in the points in terms of how much money should go to applied research (mostly applied research institutes) and to basic research (mostly universities and Academy of Science institutes).

Different types of research institutions have different missions, different targets, different roles and different tasks. Evaluation and funding systems need to be in line with organisations' missions – research evaluation basically has to answer the question of whether or not an organisation is fulfilling its role and research funding needs to be such that research organisations can fulfil their roles in the Czech RDI system.

The new funding and evaluation systems we are proposing are able to accommodate different missions and roles. However, before they can do so, the Czech RDI system will need to discuss and agree upon the different roles of research organisations.

We are wondering whether this is a blind spot in the discussion on the Evaluation Methodology so far.

Stability

In our report we argue that the transition from one funding system to the other should be organised in such a way that the budgets of the individual institutions affected do not change dramatically from one year to the other.

Indeed, we think that the Czech system is in need of stability. The allocation of institutional funding has been changed repeatedly over the last years. As a result, some organisations have experienced dramatic changes in the amount of institutional funding they have received – and radical increases are nearly as difficult to accommodate as radical reductions. So a first aim of ours will be to stabilise the system, and to make funding more predictable – this is the basis for serious planning and for institutional development.

We will recommend introducing multi-annual funding periods rather than an annual allocation of budget. Moreover, we recommend establishing the new funding system stepwise over at least one or several transitory funding periods (see section 9.3.3). In the 'steady state', a funding period should not be shorter than four years, but during the transition phase, the periods can be shorter (2-3 years). Finally, we will suggest limits to budget cuts within the regular allocation of funding.

Evaluation as a tool for learning and institutional development

Research evaluation should not be used for funding decisions between ministries and research organisations but as a tool for institutional development. It can and should be used to inform and support decision making about and within institutions, but not in a linear and mechanistic way. Setting up a quality management system, which includes evaluation, will be a task for all research organisations, and building the required 'culture' and capacities has to be considered a key task for the next years.

9.3 Allocation of institutional funding

We propose to allocate institutional funding through a combination of a formula budget and performance contracts. In this section we first describe the formula budget and the performance contract in more detail and then suggest the way in which they should be implemented over time.

9.3.1 Formula budget

A formula budget is a set of indicators whose scores determine how much institutional funding a research organisation receives.

In line with our argument above that funding systems need take into account organisations' missions and roles, we do recommend **not** using the same indicators for all the research organisations. Instead, we suggest using different indicators for each group of research organisations, depending on their roles and missions in the Czech RDI system. The setting up of a formula budget will require a discussion about different institutions' role in the Czech RDI system, as indicators have to be in line with the missions and objectives of universities, the Academy of Science, and research institutes.

For example, universities train young researchers and award PhDs and they receive funding for this. There is evidence that many doctoral students do not complete their PhD in time or at all. Hence, the number of PhD students who complete their degree in time could be included in a funding formula. The Academy, on the other hand, does not award doctorates but many of its institutes are involved in the training and supervision of PhD students. This is currently not reflected in institutional research funding. Hence, an indicator for the Academy's formula budget might be: number of co-supervised PhD students.

The research institutes will need to be subdivided, as they have different roles and missions, and for each of these sub-divisions there should be a specific formula. There could for example be a specific set of indicators for applied public research institutes[117] which conduct research whose main beneficiaries are the public sector and/or specific user groups, many of whom do not normally commission research of their own (e.g. farmers or road users). Currently the Metodika Hodnocení fails to do justice to this kind of research which is indeed applied in nature, but often not for industrial applications. This approach has been adopted in Norway, where the public research institutes are divided in five categories, e.g. industrial-applied, social science and humanities.

An OECD paper gives an overview of indicators currently used in performance-based research funding systems[118], describing their advantages and disadvantages. We strongly advise to consult this document when putting together formulas for the different types of research organisations. As a guiding principle we recommend using indicators that are equally relevant within a category of organisations, based on a

[117] Different types of external funding (including contract research) would certainly play the most important role in the formula. See also 9.3.3.2.

[118] Hanne Foss Hansen, Performance indicators used in performance-based funding systems, OECD, 2010

sound decision about their mission and role. This generally excludes numbers of publications or citations as indicators in formula.

A simple indicator that is often used is external funding. External funding is money allocated in a peer-review process, hence it is quality-checked. External funding consists of two strands: grants (competitive project funding, usually called targeted funding in the Czech Republic) which are normally awarded by public agencies, and research contracts where the research results normally belong to the organisation who commissions the research.

For the Czech Republic such an indicator could be four-fold, with the weight of each indicator depending on the missions of the research organisations:

- Competitive funding from the Grant Agency
- Competitive funding from the Technology Agency
- International competitive funding (e.g. from the European Framework Programme)
- Research contracts: The customers of such research services could be industry, ministries, public authorities, associations etc. – any kind of user or applicant of research results. This indicator reflects the interest of the 'outside world' in the research. It is, however, of limited use for such disciplines or organisations that find their main users within science, mostly performers of fundamental research.

The second last bullet point above brings us to another issue that requires attention. A number of Czech research organisations seem to need a more international orientation. In order to give an incentive to do so, an 'internationalisation indicator' could be used (in addition to international competitive funding). The Finnish formula-budget uses 'researcher and teacher mobility' as an indicator, the Austrian uses students' mobility.

There are two types of indicators used in a formula budget: volume indicators such as number of researchers (FTE) or number of PhD students, and goal indicators (also called quality and effectiveness indicators). Both will need to be used in the Czech Republic, and volume indicators should have more weight. Furthermore, like in the UK and in Austria, weighting should be applied for the more expensive disciplines, that is high cost laboratory and clinical subjects. As an option, changes in time (deltas) rather than volumes (stock figures) can be used in order to reflect development towards the defined objectives, to mitigate size effects and to account for different starting positions.

We do not recommend using indicators in the formula that count research outputs, publications (like in Norway or Finland) or citations (like in Flanders) because they cannot do justice to the different disciplines. Of course, weighting can be used but that makes a system complicated. International examples show that a satisfactory way to deal with disciplines' different publishing behaviours or to weigh has not been found yet. In this context, it is important to be aware that the use of bibliometrics depends on the context in which it is used: Because the bias inherent in bibliometric data is the same for all countries, it is standard to use bibliometric indicators at the macro level to compare publication output internationally. In contrast, it is highly problematic to use them to allocate funding. Hence, when using bibliometric indicators it is essential to be aware of the context in which they are used.

Last but not least, indicators also have to comply with standards for good indicators, such as the SMART or Cream tests. Smart indicators are: S Specific, M Measurable, A Achievable, R Relevant, and T Time-bound[119]. In contrast, the Cream test assesses

[119] Interestingly, the Evaluation Methodology's output indicators are not SMART. Most notably, many of the output indicators (e.g. patents) are not achievable and/or relevant for certain types of research organisations or disciplines

indicators as follows: Are indicators clear (precise and unambiguous), relevant (appropriate to the subject at hand), economic (available at a reasonable cost), adequate (provide a sufficient basis to assess performance), monitorable (amenable to independent validation)?

9.3.2 Performance contracts

Performance contracts are a contract between a principal (e.g. a ministry) and an agent (e.g. the research organisation). They define objectives that the research organisation needs to attain and are embedded in proper procedural steps to prevent nepotism.

It is often assumed that performance contracts are purely prospective and that research organisations can promise anything without basing their promises on past performance. This is not true.

a) Performance contracts define objectives

b) Objectives need to be manageable, realistic and ambitious, and whether this is the case will be assessed on the basis of past performance. Manageable means that objectives must be attainable with the resources (money, hierarchy, trust, information, leadership/orientation) available in-house

c) Attainment of objectives is monitored regularly, annually or less often depending on the objective

d) If objectives are not attained, there must be consequences

Objectives can be both qualitative or quantitative. Examples of qualitative objectives would be to set up a doctoral school with international faculty in a certain field, to close down a field of study, or to build up a strategic partnership with a specific research organisation. Examples of quantitative objectives can be to decrease the drop-out rate of doctoral students by 50% or to increase research grants (targeted funding) by 20% in the period covered by the performance contract. As these examples show, both qualitative and quantitative objectives can be monitored. What is more, combining qualitative and quantitative objects can help avoid perverse effects of simple quantitative objectives. For example, if a university is required to reduce the drop-out rate of PhD by 50%, it can of course do so by lowering standards for obtaining a PhD (= perverse effect). However, by setting the additional objective of requiring all doctoral students to have a local and an international supervisor the likelihood of such manoeuvrings can be mitigated.

This brings us to the next point: There need to be safeguards against the dangers of incompetence, ad-hoc solutions, nepotism and corruption. These can be:

- International external expertise should be used to negotiate and draw up performance contracts, e.g. from a person experienced in negotiating performance contracts. This person could chair the negotiations between ministry and research organisations. This would not only make sure that the process is fair and objective, it also helps build up knowledge
- Budget allocation (i.e. how much institutional funding, targeted funding, contract research etc. a research organisation received) and annual reports need to be published, to ensure transparency
- The process for setting up performance contracts needs to be standardised, i.e. it needs to be the same for all organisations, and known to all parties involved
- Performance contracts will be monitored during and at the end of the period, i.e. there will be a check whether objectives have been achieved. The monitoring will require data suitable for assessing the achievement of objectives defined in the performance contract. We also suggest two monitoring sessions (meetings between representatives from the research organisation and the ministry) per year.

The basis for the performance contract will be a self-assessment, which will be primarily used as a source of information about past performance and the starting point for the performance targets to be agreed upon. Moreover, each research organisation will prepare a development plan as a basis for the performance contract and present it to the ministry in charge. Both sources of information are necessary for defining objectives that are manageable, ambitious and realistic. Then there will be a series of negotiation talks between the research organisation and the ministry, discussing mission, objectives, tasks research institutions are legally required to carry out, expectations. Experience shows that it takes about 2-3 sessions to establish a performance contract.

The first round (i.e. the first performance contract) will be primarily for learning. The ministry will stick to its side of the contract regardless of whether or not the targets are met (as long as the research organisations commit no crimes), i.e. the ministry will pay the full sum agreed in the first period. However, if research organisations do not achieve their objectives in the first period this will weaken their negotiation position for the following period and they will receive less money or corrective measures will be implemented in the next period. The general rule is that the government must not withdraw first in this phase.

In the second period, after a first performance contract, funding can be negotiated between the research organisation and the ministry taking into account the achievement of objectives set in the performance contract. The basis would be the amount of funding received in the preceding period. There can also be sanctions if research organisations do not reach the objectives agreed upon in the previous performance contract.

9.3.3 Implementing the two new forms of allocating institutional funding in the Czech Republic

In this section we are describing the process required to set up the new funding system. Obviously, it takes time to change a whole funding system. As particularly the Finnish case study shows, continuity and a *political will* to go through with the changes – irrespective of the colour of the government – are key to success.

Setting up a new funding system consists of three phases: first, a planning phase where the exact mechanisms to be used are conceived. This phase includes a great deal of research and simulation. The second phase is the implementation phase where the mechanisms are implemented, possibly first as a pilot and then at full scale. The third phase is the 'settling down' phase where the system becomes 'natural' and is showing first effects.

It is mainly due to the performance contracts that it will take around a decade whole new funding system to be established and running in the sense that the new system will have become 'routine' for all parties involved. However, new funding formulas for the different types of organisations will not take very long to set up and implement. So the Czech Republic will not need to wait for ten years to have a new funding system. The first part of the system we are suggesting – the funding formulas – can be set up and implemented within a couple of years.

9.3.3.1 Transition period 1: 'Historical' allocation of funding & preparation of formula budget

'Historical' allocation

For the next few years we suggest allocating funding 'historically' based on the average yearly amounts organisations received over the last years (Figure 33). Changes have been dramatic over the last few years, so means of the last three or five years will have to be used. We suggest that research organisations will get whichever amount is greater. The 'historical' allocation of institutional funding will give the Czech system the predictability it needs and the opportunity to establish formula-based systems for the universities, the Academy of Science, and different types of research institutes.

Developing indicators for formula budget

The first transition period should be used to develop the formula budget by defining indicators and to 'organise' the data (Figure 33). A great deal of data will be available already but they have to be put together and quality checked. For example, there seems to be a problem with number of staff: It appears that a considerable number of researchers in the Czech Republic work full-time at several institutions, meaning that in effect they only work part-time at an institution. Problems like this need to be sorted out. If this particular problem cannot be sorted out at all, the solution may be in not using (stock) figures for staff.

Moreover, it might be necessary to adapt legislation for the new funding system, which itself takes time. It will not do to start the legislative process once discussion among stakeholders on the new allocation systems has been concluded. In other words, there will need to be an overlap between discussion among stakeholders and the legislative process.

After the first transition period (possibly two to three years depending on what data is available), a formula-based way to allocate institutional will have been established, with ceilings limiting budget cuts. It goes without saying that the formula budget needs to be simulated before implementing it.

As an option, some volunteer organisations (e.g. the Academy, some universities) could already set up performance contracts during this phase or a trial formula could be operated with a small number of (homogenous) research organisations.

Figure 33 Allocation of institutional funding for universities and the Academy of Science

Transition period 1 (Year 1-2)	Transition period 2 (Years 3-5)	New system: Funding cycle 1 (Years 6-9)
Funding mechanisms: how institutional funding is allocated		
'historical' allocation of funding (100%)	'historical' allocation of funding (pilot organisations: historical allocation *with* performance contract)	Funding based on performance contract
	Formula-based funding	Formula-based funding
Activities to set up new funding system		
Preparatory activities for funding formula	Monitoring of funding formula implementation Preparatory activities for performance contracts Monitoring of pilot cases	Monitoring of funding formula implementation and performance contract implementation

Source: Technopolis

9.3.3.2 Transition period 2: 'Historical' allocation of funding & implementation of formula budget

In the second transition period, one part of institutional funding will still be allocated 'historically', a second part based on the new formula-budget. The respective percentages need to be negotiated. As the international case studies show, percentages differ widely, i.e. there is no single best solution. However, all these systems have built in dampening factors or upper/lower limit for budget fluctuations. We strongly recommend the Czech system to implement similar systems.

The second transition period can be used to develop and phase in performance contracts. First of all, research organisations and the responsible ministries have to sit down together to define objectives. Defining indicators for a formula-budget in the first years will already have made ministries and research-performing actors think about missions and objectives. 'Countable' target indicators are candidates for the

formula-part of the funding system, other targets relating to research management, human resource policy etc. can be better covered in performance contracts. The two systems are not independent: while the number of PhD graduates could be used as a formula indicator, setting up a doctoral school in subject X could be part of a performance contract of University Y.

As an option, the volunteers from transition period 1 will already receive performance contract funding. Another possibility would be performance contracts within the Academy of Science, between the Academy Council and Academy institutes willing to participate in the pilot scheme. In both cases participants could receive a small financial top-up as an incentive.

For the research organisations that receive small amounts of institutional funding as a share of their budget (applied research institutes, universities), it may be an option to use either a formula budget or a performance contract. If only a performance contract is used, the institutional funding could be tied to a certain purpose, e.g. to the building up of capacity, because institutions that rely to a great extent on project funding often find it difficult to finance the completion of a PhD thesis or research visits abroad. Alternately, only a formula budget consisting of a very simple set of indicators, e.g. external funding, can be used to allocate institutional; this is for example the way it is handled in the German Fraunhofer institutes.

The database built up for the formula-budget can also be used to monitor the achievement of objectives defined in performance contracts. In general, indicators used to monitor performance contracts should have 'dual use', that is, they should also be used in daily operations because this will help make sure that indicators are relevant, cost-effective, and robust, and people will know how to use them.

At the end of the second transition period, a part of institutional funding for the subsequent period of (ideally) four years could be allocated based on a performance contract, the other based on a formula-budget. Again, the respective percentages – i.e. what percentage of money gets allocated based on the formula, and what percentage based on the performance contract – needs to be discussed and agreed upon.

Sharing experience about performance contracts (e.g. in the Czech Operational Programme and in other countries) and learning will have to take place during this period, involving ministries and research organisations alike.

9.3.3.3 New funding system in place: first funding cycle with formula budget & performance contracts

This is the 'steady state' of the new institutional funding system. International experience shows that performance contracts take about three periods to become established in the sense that the new system has become fully familiar to all parties involved. The first period is a trial phase. In the second period, negotiations will improve because negotiators will have gone through a learning process and know 'what it is all about'. In the third period, implementation will already be based on the experience and learning from the previous periods. Altogether, this means that no major changes should be made before at least two full funding cycles have been completed and assessed.

9.4 Evaluation of research

Although it has an Evaluation Methodology, the Czech Republic lacks an evaluation culture. In particular, it has no understanding and appreciation of evaluation as a formative instrument – something that helps institutions to learn and get better. The current Evaluation Methodology simply counts outputs and as such it is purely retrospective and summative. As a result we observed that the Czech research community can be somewhat hesitant towards evaluation, perceiving it as an instrument for judgement, based on one-dimensional or even inappropriate criteria and with unclear and potentially severe consequences – while actually evaluations should be a normal part of research life, not only making researchers accountable for

their work, but also giving them valuable feedback in order to improve both the quality of research and the research environment.

Therefore, our recommendations are geared towards the building of an evaluation culture in the Czech Republic, making evaluation of research part of academic routine. By separating evaluation of research from the allocation of institutional funding and by introducing evaluation as an instrument for learning, we aim to transform research evaluation into a positive and constructive experience that advances research. More precisely, by involving international peers, research evaluation will become a policy instrument that systematically gathers advice based on international good practice to improve the quality of research in the Czech Republic.

In short, we recommend implementing **informed peer-review** to evaluate research in the Czech Republic. This peer-review will be performed **by objectively and independently selected (international) peers.** Below we will give some detailed recommendations on basic elements that should be part of such evaluation, including the selection of peers and units of evaluation, and the use of indicators, self-assessments and site-visits. In addition, we will discuss the organisation and the process of evaluation.

It will not be enough to just install a new system of research evaluation. Rather, a research evaluation culture needs to be formed. Hence, we will also recommend establishing a debate among stakeholders and the research community to discuss research evaluation, different approaches to it, examples of good practice etc., to work on a common understanding of evaluation.

9.4.1 Establishing an evaluation culture

In order to establish an evaluation culture in the Czech Republic, we strongly advise to build on existing national and international experience. Next to a wide variety of international examples (see section 6 of this Annex), there already is some knowledge about what research evaluation is in the Czech RDI system, which can give information about how it is done and what the advantages and disadvantages are.

The Academy of Sciences has established an internal evaluation culture and this can be used as a starting point for learning about evaluation, although it should be firmly noted that evaluation at the ASCR needs to be brought more in line with international good practice (see 7.1). In addition, the reviews of institutes we did as part of this audit can be used as a basis for developing an evaluation culture (see 2.3). These reviews, being completely independent and international, were typically welcomed as useful exercises and an opportunity for learning. Moreover, an international peer evaluation was performed in the OP R&D for Innovation (see 7.2), and of course, there is a wealth of experience to draw upon internationally. We would not recommend using the UK RAE as a role model though because it is used for allocating institutional and as such is strongly summative. The Czech Republic, however, is in need of formative evaluation that gives research organisations concrete suggestions as to what they can do better.

Sharing information about these and other concrete examples and discussing them will give Czech institutions (ministries and research performers alike) the opportunity to get an understanding of what evaluations can be used for and what they can achieve – which would provide them with a basis for establishing peer-review in the Czech Republic.

One way to organise information sharing and building up evaluation expertise can be to create a platform for RTDI evaluation. Such a platform could be modelled on the Austrian Platform Research & Technology Policy Evaluation[120], which explicitly aims at contributing to the development of an evaluation culture in Austria and in fact has

[120] http://www.fteval.at/cms/en/home.html

contributed substantially since it was founded in 1996. To do so, the Platform organises talks and training workshops, presenting different approaches and methods of evaluation and discussing current evaluation practice on an international level. Ministries, research organisations, and evaluators can participate in and contribute to such a platform.

9.4.2 Informed peer-review in the Czech Republic

To evaluate research in the Czech Republic we suggest what the OECD calls 'informed peer-review'[121], that is peer review supported by indicators, performed by objectively selected (international) peers. This peer-review works in similar ways as peer-review for journal publications. However, while with publications peers are asked to review a paper, informed peer-review involves a group of peers that is asked to review entire research institutes or research groups. On the basis of information on the units of evaluation, with input from indicators, self assessment and site visit, the group of peers writes a peer-review report.

The exact details of the informed peer-review process can vary, depending on the goal and character of the peer-review, as well as national practice. It is therefore important that the Czech Republic will develop its own way of embedding informed peer-review in the Czech research system. In order to develop this national system, we give some detailed recommendations, covering the basic elements that should be included in the design of the Czech evaluation system.

9.4.2.1 Basic elements of informed peer-review

Basic features and suggestions that should be considered in the design of the Czech evaluation system:

- **International peers**. The quality of the peers directly determines the quality of peer-review. Obviously, reviewers should be independent, preferably international, and be selected by an independent body, as was the case in the reviews of research institutes we did as part of this audit. Reviewers should *not* be selected on the basis of suggestions made by the researchers to be evaluated. However, the researchers to be evaluated could be given the possibility to exclude a small number of persons as reviewers. This will be a safeguard against reciprocity, incompetence and nepotism and contribute to the acceptance of peer-review based evaluation in the Czech Republic. A consequence of international peers will be that the reviews will be conducted in English.

  The panel members together have to cover a larger research area than is the case in the ex ante peer review of publications, and for that reason, a panel consists of experts in different, subsidiary areas within the field that is to be evaluated.

- **Units to be evaluated.** Informed peer-review can be performed on various units, e.g. research organisation, research divisions and research groups, but also research programmes and research projects. While the evaluation of research programmes and projects is discussed in Work Package C, we will concentrate on the evaluation of research organisations, divisions and groups.

  When selecting units of evaluation, we suggest that it is best to stick to units that already exist. Which unit is most appropriate needs to be decided, taking into account different factors such as the goal of the evaluation, the budget available, and how specific disciplines are structured. E.g. when looking for a unit of evaluation in philosophy, the evaluation will probably review the whole institute/department of philosophy. In contrast, in medicine, the evaluation unit will be more specialised, e.g. an institute for virology.

[121] Hanne Foss Hansen, Performance indicators used in performance-based funding systems, OECD, 2010, p. 11

- **The period that is evaluated.** When deciding on the period under evaluation, it is important to know that in evaluation there is a common distinction between ex-ante and ex-post evaluations, or between evaluations that are taking place before a research activity starts and evaluations that take place after an activity has finished. In addition, there is often a mid-term review taking place halfway. Ex-ante and ex-post evaluations are very common when it concerns temporary research projects, as they need to be assessed in order to acquire funding and then again after they have finished evaluating the functioning and the results of the project. However, as research institutes, research divisions or research groups have a more permanent character, evaluations should combine an evaluation of past performance with the evaluation of future plans. The exact period that is taken into account can vary, but an evaluation could for instance decide to look back at the last five years and look forward to the plans for the next five years. Alternatively, a period of seven or ten years can be chosen.
- **Establishment of evaluation indicators**. In order to have an *informed* peer review, the peers should be provided in advance with information about the unit that is under evaluation. The nature of the required indicators should be negotiated when designing the evaluation process and should be relatively easy to produce. Examples of such indicators are: list of selected key publications, number of PhD graduates, external competitive funding acquired, international collaborations, etc. Information on these indicators can be provided by national authorities or by the institutes or research groups themselves (as part of the self-assessment report).

  This approach has the distinct advantage that peers can put the indicators into the context of the discipline, the organisation's mission and strategy and can give detailed feedback based on international practice but geared to the specific local circumstances.
- **Self-assessment report.** Information about the unit under evaluation should also be provided by the unit itself in the form of a so-called self-assessment report (see also section 0). This report should contain information about the research organisation, its research programme(s), key staff, achievements and users of research results, management and human resource policy, teaming and strategic partners, research funding, perceived barriers in the context of the organisation and should be presented to the peers in preparation of the site-visit.

  This self-assessment report has a dual function, as it not only informs the peers, but it also makes the research group or institutes under review reflect on their own performance, organisation, and research plans.
- **Site-visits**. During a site-visit, the peers visit the unit of evaluation to meet the researchers and see the environment in which the research takes place. The site-visit is an important part of the review process, as it provides the peers with the opportunity to meet the researchers, exchange information and discuss. Normally, a site visit includes a presentation by the head of the unit under evaluation, a discussion based on information provided, an interview with junior and senior staff of the evaluated unit, a tour around the research facilities and the presentation of first conclusions by the peers.
- **Review report.** Based on information acquired through indicators, the self-assessment report and the site-visit, the peers will write a review report that contains an overview of the research, the research environment and future strategies of the evaluated unit, as well as their conclusions and recommendations. As a general rule, peers have to be requested to base any judgement on evidence. It is important to note that it is common practice that the head of the evaluated unit receives the draft version of the report and has a chance to comment before the final version is produced and disseminated (see also 9.4.3).

  With regard to the publication of the review reports, the Czech Republic could opt for the Dutch solution, where only those parts of the report that do not contain

sensitive information are published. In this way, the objective of accountability can be met by producing a transparent and informative public report of the evaluation's outcomes. Sensitive information that should be kept confidential and not form part of the public report concerns the future strategy of the institute, matters of personnel policy or sensitive decisions.

- **The evaluation cycle.** As the evaluation of research organisations and research groups asks for a commitment from both the peer-reviewers and the evaluated unit, it is important that evaluations are timed well. This means that they should be performed regularly in order to observe the development of a research unit over time, but certainly not too often. Full-scale research evaluation of research groups, institutes or disciplines should in general not be conducted more often than once every five years, not only because it places a burden on the institute, but also in order to give the institutes involved time to react to the signals of the evaluation, i.e. to discuss the recommendations and to decide on what measures to take.

  However, this does not mean that all research organisations need to be evaluated in the same year and that only every five years evaluations take place. A five year cycle also makes it possible to organise evaluations continuously, but to focus on different scientific fields every year in such a way that all units of evaluation are evaluated within five years and every unit is evaluated every five years.

- **The organisation of informed peer-review.** As becomes clear from the international case studies (see section 6), countries have different ways to organise peer-reviewed ways of evaluating research. The main responsibility for the organisation of the reviews can reside on a national level (see for example the French AERES) or evaluations can take place in a more decentralised way, making the research organisations responsible for installing an evaluation system (see the example of Austrian public universities). In addition, there are systems where national organisations and research institutes are sharing the responsibility for review (see the Standard Evaluation Protocol of the Netherlands).

  For the Czech Republic we suggest establishing a central national organisation to build up evaluation expertise and guide the evaluation process. This organisation should be an independent public agency, and it should be a *permanent* body that can develop experience with informed peer-review over time, developing evaluation capacity in the Czech Republic. This organisation should gather existing experiences with peer-review, initiate discussions with relevant stakeholders, and develop and implement standards for informed peer-review in the Czech Republic in interaction with the research organisations. In turn, every research organisation should develop capacity for the evaluation of its activities as part of its quality assurance system.

  The question is to what extent the responsibility for evaluation will reside in the central body and to what extent responsibility will be devolved to the research organisations. At one extreme, the central national body is fully responsible for design, organisation and execution of the evaluation. In the other extreme, the responsibility lies fully with the research organisation, with the central body accrediting research organisations' evaluation systems. In the first case, there is a danger of the evaluation approach not differentiating enough among different types of research organisations and evaluation results not being appropriated and made use of by the research organisation. In contrast, in the second case, there is a danger of many actors making large efforts to design their own systems, resulting in a multitude of systems, which makes comparison between organisations difficult.

  At any rate, whatever role the Czech Republic is going to give this new evaluation body, four features need to be secured:

  – There need to be standards of evaluation, to which all evaluations have to adhere. In a centralised system, these standards will be defining the ways in which the evaluations take place, while in a decentralised system, these

standards are used by the central body to accredit the research organisations' evaluation systems.

- Evaluation always needs to take into account research organisations' missions and roles. Hence, evaluation standards need to be flexible enough so that they can be adapted to different groups of research organisations (e.g. universities, sectoral research institutes etc.).
- It has to be made sure that evaluation results are appropriated by research organisations. This means that evaluation always needs to be in cooperation between the central body and the research organisations, whoever has the ultimate responsibility for the process. It is also vital that research organisations can contribute questions that the evaluation will answer.
- Peers need to be selected independently, i.e. not by the research organisations themselves. Hence, the central body should have a defining role in it. In other countries, established research councils have considerable expertise and experience in selecting peers. In the Czech Republic the Grant Agency will already have some experience with peer-selection. This expertise should be used.

- **Rapporteur.** While the panel of peers is normally chaired by an international scientist or scholar who is an expert in the field under evaluation and has a great deal of experience with peer reviews, this person can be supported by a rapporteur who makes sure that the review process is managed properly. This person should have knowledge about research policy in general and the peer-review process in particular, as being in charge of the practical organisation of the review, gathering relevant information, preparing the site-visit, and assisting with the writing of the peer-review report. The rapporteur can come from the central body or the research organisation.
- **Costs of peer-review**. When discussing the introduction of informed peer-review with Czech researchers and policymakers, it became clear that the costs of a peer review based evaluation system are sometimes seen as a possible barrier to implementing it in the Czech Republic. However, we would like to stress that the costs involved should not be seen as a reason to not develop informed peer-review because the benefits of peer review typically outweigh the costs if it is implemented intelligently. Peer review is by far the best way for the Czech RTDI system to learn. Indeed, many countries see it as the best way to develop the quality of research and do not see the costs as an impediment.

9.4.3 Using the results of informed peer-review evaluation

We clearly recommend separating the evaluation of research from the allocation of institutional funding. However, not linking evaluation to the allocation to funding does not mean that evaluations should not have consequences: at the level of research institutions, state-of-the-art models of (research) evaluation always specify how evaluation results will be used in decision making.

The results of evaluations through peer-review should first and foremost be used by the units (e.g. the institute) that have received feedback through the evaluations in order to improve the research and the research environment in line with the recommendations. In addition, the review results can be used by research organisations themselves (e.g. the university) as information when decisions about research directions and research management need to be taken.

What is more, ministries can use research evaluations as a basis for policy decisions. In case some research organisations have exhibited low research quality over longer periods of time and appear not to have been very productive for years. In such cases, we suggest using evaluations to have sufficient information to decide which institutes should be closed down, which institutes need to improve the quality of their research, which institutes need to be reorganised etc. If results of the evaluation show that the

institute is not performing well, then closing down the institute is an option. Before doing that, however, one should think about whether the Czech Republic needs an institute of a certain specialisation. For example, it does not do to close down the meteorological office if its research quality is found wanting. Indeed, in the case of sectoral research institutes that measure, test and monitor as one of their main tasks, closing them down often is not an option. We would also advise against closing down institutes studying and preserving Czech cultural heritage (e.g. Czech literature). Ministries who have responsibility for several research institutes working in related fields also have to think about whether to put them together, whether to introduce some sort of formalised cooperation between them and how to coordinate them.

In addition, the evaluation results should be used to inform the negotiation process for performance contracts. This would give the ministries the strategic intelligence needed for a good performance contract that will lead to increasing quality or to the building up of capacity. As the evaluation provides a thorough assessment of past achievements, it not only points to changes needed but also helps set realistic and ambitious targets in the contract.

9.5 Policy learning

Policy learning means sharing experiences among policy-makers and learning from each other. Policy learning is indispensable for any government or political system to develop better polices and make sure that policies are evidence-based. There are whole international organisations devoted to policy learning, for example the OECD. In some ways this report is also an exercise in policy learning as it operates with experience from other countries.

The Evaluation Methodology seems to have developed in almost complete isolation from developments going on in other countries. In our perspective, this has not improved the Evaluation Methodology. Hence, we think it is crucial for the Czech Republic to engage in discussions with representatives from other countries. As a member state of the EU and the OECD this should not really pose be a problem.

For example, representatives from MEYS and the RDI Council could tap into the discussions currently going on at the OECD about performance-based funding systems. Policy learning can also occur on a bilateral basis with other Member States. Admittedly, as the international case studies have shown other countries do not have perfect systems, continuously improving and reforming the ones they have – while making sure to maintain the stability of the university system. Nonetheless, sharing experiences with others can be very rewarding for all parties involved.

Also, we are sure that other countries would be very interested to learn from the Czech experience. The Czech Republic has accumulated a wealth of knowledge with regard to those indicators that generally pose the greatest problems in a funding formula – research outputs. Czech representatives should share their knowledge with other countries.

Similarly, the Czech Republic likes to benchmark itself internationally. However, it does so only against easily available indictors (e.g. BERD as a % of GDP). It does not benchmark itself to whole governance systems. As a systematic view on RTDI with a good governance structure is key for a successful RDI system, we highly recommend not only to look at indicators but also to benchmark oneself against whole research and innovation systems and their governance. Of course, it goes without saying that foreign examples cannot be imported they way they are, they always have to be adapted to local circumstances.

Appendix A - Reviewed institutes and reviewers

Institutes	Reviewers
Social Sciences and Humanities	
• The Czech Institute for Egyptology of the Charles University in Prague (Faculty of Arts) • The Institute of International Studies of the Charles University in Prague (Faculty of Social Sciences) • The Institute of Czech literature of the Academy of Sciences	• Prof. dr. Christiane von Stutterheim, Institute for German as a foreign Language Philology, University of Heidelberg, Germany • Prof. dr. Ursula Verhoeven – van Elsbergen, Institute for Egyptology, University of Mainz, Germany • Prof. dr. Andrew Williams, School of International Relations, University of St. Andrews, United Kingdom
Natural sciences	
• The Necas Center for Mathematical Modeling (Charles University, Czech Technical University, Academy of Sciences) • The research group Bio-organic & Medicinal Chemistry of the Institute of Organic Chemistry and Biochemistry of the Academy of Sciences • The Department of Spintronics and Nanoelectronics of the Institute of Physics of the Academy of Sciences	• Prof. Ria Broer, Theoretical Chemistry, Zernike Institute for Advanced Materials, University of Groningen, The Netherlands • Prof. Barry Koren, Centrum Wiskunde & Informatica (CWI, Amsterdam) and Mathematical Institute, Leiden University, The Netherlands • Prof.em. Louis Schlapbach, Swiss Federal Laboratories for Material Science and Technology (EMPA), Dübendorf-Zurich, and ETH Dept. of Physics, Switzerland
Medicine	
• Laboratory of Signal Transduction, Institute of Molecular Genetics (IMG) of the Academy of Sciences • Laboratory of Biophysical Chemistry and Molecular Oncology, Institute of Biophysics, ASCR, Brno • Joint Laboratory of Cancer Cell Biology - Institute of Biochemistry & Experimental Oncology (Charles University) and Institute of Physiology (Academy of Sciences)	• Prof. dr. ir. Boudewijn M.T. Burgering, professor in Signal Transduction, Department of Physiological Chemistry, University Medical Center Utrecht (UMCU), the Netherlands • Dr. Johannes Stöckl, Institute of Immunology, Medical University of Vienna, Austria • Prof. Emile E. Voest, MD PhD, professor in Medical Oncology, Laboratory of Experimental Oncology, University Medical Center Utrecht (UMCU), the Netherlands • Prof. Thomas Seufferlein MD PhD, Department of Internal Medicine I, University Clinic and Polyclinic for Internal Medicine, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany • Prof. Leopold Öhler MD PhD, professor in Oncology and Haematology, Klinik fur Innere Medizin I am AKH in Vienna, Austria
Economics	
• Centre for Economic Research and Graduate Education-Economics Institute (CERGE-EI) • The Institute of Economic Studies, Charles University	• Prof. Koenraad Debackere, Professor in Technology and Innovation Management, Department Managerial Economics, Strategy and Innovation, Faculty of Business and Economics, Katholieke Universiteit Leuven, Belgium • Prof. Ulrich Hommel, Director of the Strategic Finance Institute, Department of Finance, Accounting and Real Estate, European Business School, Germany • Prof. Dennis Mueller, Professor of Economics (Emeritus), Department of Economics, University of Vienna, Austria

Technical sciences	
• Laboratory of luminescent and scintillation materials, Department of Optical Materials of the Institute of Physics, Academy of Sciences • Institute for Research and Applications of Fuzzy Modelling, University of Ostrava • Department of Control Engineering, Faculty of Electrical Engineering, Czech Technical University	• Prof. Andries Meijerink, Professor in Condensed Matter and Interfaces, Debye Institute for Nanomaterials Science, Utrecht University, The Netherlands • Prof. Per Larsson-Edefors, Professor of Computer Engineering, Chalmers University of Technology, Gothenburg, Sweden • Prof. Giovanni De Micheli, Professor and Director of the Institute of Electrical Engineering and of the Integrated Systems Centre, École Polytechnique Fédérale de Lausanne, Switzerland
Applied research institutes	
• Research Institute of Textile Machines (VUTS), Liberec • Transport Research Center, Division of Transport Infrastructure and Environment, Brno	• Dr. Lothar Behlau, Head of the Strategy and Programmes department, Fraunhofer Gesellschaft, Munich, Germany • Prof. Josef Hochgerner, Scientific Director, Centre for Social Innovation, Vienna, Austria • Prof. Berghold Bayer, former Managing Director of the Technologie- und Marketinggesellschaft (Upper Austrian Technology and Marketing Company), former Technology Representative of the State of Upper Austria, professor at the University of Applied Sciences in Graz and Johannes Kepler University, Linz, Austria

Appendix B - Comparing Evaluation 2009 by Metodika Hodnocení with the results of the informed peer review exercise done during the Czech Audit

Introductory remarks

We refer to the Metodika Hodnocení as Evaluation Methodology, abbreviated EM and to the informed peer review exercise performed in the course of the Czech Audit as (our) IPR exercise.

6 out of 18 units selected for our evaluation during the Czech Audit are also units of evaluation in the EM 2009. The other units in our Informed Peer Review are parts of larger entities according to the structure used in the EM – i.e. we do not have point scores from the EM 2009 for these units.

- In one of the six cases (Transport Research Institutes) we did not evaluate the entire institute but one division. For the purpose of this analysis, we consider the division evaluated representative of the entire institute.
- One of the organisations evaluated during our audit, CERGE-EI, is actually composed of two separate subunits in the structure used by the EM, CERGE (Charles University) and EI (AS CR); for the purpose of this comparison, we add the numbers of results and respective point scores for the two subunits
- The Occupational Safety Research Institute, also on the list of the 6 institutes, opted out of the IPR exercise, meaning that we cannot make a comparison in this case.

Therefore, we compare the findings for the following organisations (in brackets: thematic peer group):

- Institute of Czech literature of the Academy of Science of the Czech Republic (humanities)
- CERGE-EI, the joint workplace of the Centre for Economic Research and Graduate Education of Charles University and the Economics Institute of the Academy of Science of the Czech Republic (economics)
- Research Institute of Textile Machines (VUTS), Liberec (applied institutes)
- Transport Research Center, Brno (applied institutes)
- Institute for Research and Applications of Fuzzy Modelling, University of Ostrava (technical sciences)

For illustrative purposes, we will use concrete examples from the EM data for our comparison. We do not intend to make a judgement about the respective organisations nor do we aim at a comprehensive analysis.

It is important to keep in mind that although both exercises are called “evaluation”, they serve different purposes: The EM is used to make funding decisions in an automated way. Our IPR exercise was formative, i.e. done in order to learn about the situation of research in the CR and, at the level of the participating institutions, to provide feedback as a basis for organisational learning, development and improvement. This limits comparability, as we shall see.

Observations

Scope and content of coverage: The first and most obvious difference: the IPR collected and assessed a wider range of information and data about the units evaluated than the EM.

- The EM collects numbers of pre-defined research outputs in nearly 20 categories (definition and number has varied over the years)
- The IPR collected and assessed structured quantitative and qualitative information about each unit of evaluation, comprising the research programme, critical infrastructure, staff (numbers, composition, qualifications, roles), strategic partnerships, achievements and users of results, management and human resource policy, research funding, and perceived barriers to development.

Type of feedback given: The EM assigns a defined number of points to each output, the feedback given is the resulting point score. The informed peer review provides a structured written report typically several pages long, containing verbal feedback for each of the evaluation criteria.

Type of results: The results of the EM are presented in a list of all participating (subunits of) research organisations. The list contains (apart from identification data) the numbers of results by category, the number of points achieved for each category as well as the total number of results and points scored. The organisations are sorted by number of points achieved in descending order. The higher the point score, the higher the budget allocated through the EM (except for subunits of the AS CR).

There are 373 evaluation units in the EM 2009. For the purpose of this comparison we added the two units that together form CERGE-EI, reducing the number to 372 units. By point score, the institutes on our list for comparison rank as follows; the numbers in brackets indicate the position in the overall ranking of the 372 evaluation units of the EM (see Figure 34).

1. Research Institute of Textile Machines (89th)
2. Institute of Czech literature (101st)
3. Transport Research Center (107th)
4. CERGE-EI (119th)
5. Institute for Research and Applications of Fuzzy Modelling (181th)

Of course, the number of points depends not only on the output and the productivity of an evaluation unit but also on its size in terms of research active staff. Therefore, it is not justified to read this list as a ranking of research quality.

The ranking would look differently if not the total point scores but various output categories were used for the rankings. One could, for instance, rank by points achieved through publications in international impacted journals (J imp):

1. CERGE-EI
2. Institute for Research and Applications of Fuzzy Modelling, University of Ostrava
3. Transport Research Center, Brno
4. Research Institute of Textile Machines (VUTS) and Institute of Czech literature (AS CR) (ex aequo)

By contrast, our IPR exercise, being formative, did not rank or grade the institutions. They were grouped into thematic categories, each group of institutions visited by one peer group of three international experts. As a general trend, however, the feedback received by the applied research institutes was more critical as compared to the others, while the EM rates VUTS in 89th position and the Transport Research Centre 107th, both clearly ahead of CERGE-EI and the Institute for Research and Applications of Fuzzy Modelling.

Technically speaking, we could stop the comparison here because, in essence, the two different exercises produce different kinds of results. Where the findings can be compared roughly, they are not in line, i.e. the more favourable ranking of the two applied institutes by the EM is not confirmed by the results of the peer reviews. Yet,

we take this analysis of actual EM scores a little bit further in order to illustrate some or the main issues we have presented in the SWOT analysis in a very condensed form.

The 2009 EM results rank VUTS ahead of CERGE-EI – but is VUTS really performing better than CERGE-EI or are there other reasons for their different scoring in the EM? The EM itself cannot answer this question. Nevertheless, it actually contains already some hints at where the answers might be found. The EM stops its evaluation at the ranking by overall point score, thus missing a lot of information contained in the output data. One option would be to have a closer look at the different output categories filled by the individual institutions, i.e. their "output profile". This, of course, should be done with caution, given the skewed incentive structure offered by the EM, but it is nonetheless instructive:

It becomes evident that different institutions have different "output profiles", i.e. not all types of outputs are equally frequently produced by different institutions (see Figure 34 for the figures). What are the reasons behind these differences? These institutes work in different disciplines, they have different missions and tasks and they address different users and target groups. Therefore, different outputs are of different importance to them. Some conspicuous examples:

- VUTS produces most of its outputs in categories close to application which is appropriate to its main orientation towards regional and national companies as users of results.
- The Institute of Czech Literature publishes mainly in SCOPUS or ERIH listed journals and books, appropriate publication channels for a philological institute.
- The Transport Research Center publishes in non-impacted journals and books and also reports a substantial share of output in applied categories, and they are the only organisations in this comparison that produces results used by the provider of the funding. Their target groups are located in the CR, therefore it is important for them to be present on the national stage.
- CERGE-EI produces mainly publications in (impacted) international papers and in books, which is in line with its mission to carry out scientific research of the highest international standards.
- The Institute of Research and Applications of Fuzzy Modelling Nearly produces two thirds of their reported output as articles in international journals, matching its focus on theoretical research of soft computing methods.

For a university-based research group in economics or many of the technical sciences, publishing in international impacted journals is the most acknowledged way of publishing results, while an applied technical research centre is rightfully judged (among other things) by demonstrating its ability to produce prototypes, patents and the like. In other words, CERGE-EI producing more articles in international journals than VUTS is not a mistake, on the contrary, it is an indicator that these institutions (by and large) are performing well, measured by what their mission and tasks are. The EM makes no difference between different "output profiles", even if these are justified by the mission and role of a research organisation.

This raises the next question: why should **one** such appropriate result produced by CERGE-EI be worth less than **one** appropriate result produced by VUTS? As we have mentioned elsewhere (see 8.2), this decision is quite arbitrary. On average, CERGE-EI received 26.3 points for each publication in a J-imp journal and VUTS received 40 points per S-type result. This difference is apparently not to due to a generally higher rating of applied results, given that the Institute for Research and Applications of Fuzzy Modelling, after all, received 42.8 point per J-imp publication. This demonstrates how arbitrary the assignment of points to results is. Hence, the EM fails to convey a clear message to the Czech research community.

The purpose of the EM is to allocate institutional funding. By EM logics, VUTS does qualify for more institutional research funding than CERGE-EI[122]. In other words, producing (sellable) output very close to application by regional and national companies in the CR is considered more deserving of public research funding than producing, mainly through basic research, research results published in international journals. It is, of course, possible to pursue such a funding policy but it should be done in a transparent way, based on a stakeholder dialogue. However, we cannot help noticing that internationally, public money is invested in basic research (including the training of young researchers) without direct market application rather than in R&D close to market, where the results of research are appropriable and hence there is a market for them. The general rule is that the more basic an institute's research orientation is, the higher the share of institutional funding.

Finally, the issue of productivity: the intention to increase the productivity of R&D performers has been one of the key objectives of Czech R&D policy making and, in particular, of developing and implementing the EM. However, the EM does not give any information about productivity as this would, in its simplest form, require a comparison of the results / points achieved per research capacity, e.g. measured by the number of FTE of scientists and researchers. As a quick and certainly not very precise proxy we have also listed the staff headcounts for 2009 as reported by the institutions in their self assessment reports of our IPR. Computing the ratio of total point score per capita would lead to another different ranking – again raising a whole range of issues, such as the composition of staff, the actual number of full time equivalents available for research, the quality of the research environment etc. Quite obviously, these issues are of high relevance for the development of research quality but they cannot be dealt with adequately by any formula but rely on peer assessment and feedback.

[122] At least the CERGE part of it, as EI is governed by the AS CR.

Figure 34 Evaluation Methodology 2009 – Results for research organisations participating in the informed peer review exercise of the Czech Audit*

Name (ID code)	Results	Points	Total staff 2009	Points per capita	Articles in impacted journals (Jimp)		Articles in non-impacted journals (Jneimp)		Articles in journals in SCOPUS or ERIH databases		Books and book chapters		Contributions in coference proceedings		Patents		Industrial design, registered design		G category (technically realized results, working example, pilot plant)		Results used by the provider of research financing		Certified methodics, medical procedures		Software		S category: prototype, methodology, working example		Small scale production plant, new technology, plant variety, or animal breed		T category: technically realized results, working example, pilot plant	
	Total No.	Total No.	Heads		No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points	No.	Points
Transport Research Center, Brno (44994575)	172,2	4402,4	125	35,2	6,2	340,2	63,2	294,7	1,2	9,7	21,6	501,8	2,0	16,0	0,0	0,0	3,0	120,0	0,0	0,0	2,0	80,0	12,0	480,0	2,0	80,0	57,0	2280,0	2,0	200,0	0,0	0,0
Research Institute of Textile Machines (VUTS), Liberec (46709002)	97,0	5968,0	167	35,7	0,0	0,0	0,0	0,0	2,0	16,0	0,0	0,0	4,0	32,0	9,0	1920,0	0,0	0,0	11,0	440,0	0,0	0,0	0,0	0,0	4,0	160,0	55,0	2200,0	1,0	100,0	11,0	1100,0
Institute of Czech literature of the AS-CR (68378068)	470,2	4868,3	80	60,9	0,0	0,0	28,0	280,0	230,8	2762,0	211,3	1826,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
CERGE-EI, joint workplace of the Centre for Economic Research and Graduate Education of Charles University and the Economics Institute of the AS-CR (216208 and 67985998)	327,4	4043,4	173	23,4	99,8	2623,2	157,0	628,0	15,0	120,0	49,5	496,2	2,0	16,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	4,0	160,0	0,0	0,0	0,0	0,0	0,0	0,0
Institute for Research and Applications of Fuzzy Modelling, University of Ostrava (61988987)	61,1	1833,0	14	130,9	39,3	1681,5	0,0	0,0	7,3	58,0	5,5	21,5	9,0	72,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Source: Technology Centre of the AS CR, research organisations' self assessment reports; computations by Technopolis.

* Please note: Output categories where none of the institutes listed has reported results are omitted in this table. CERGE-EI is a "synthetic" entity as compared to the original EM table, created by adding the results and scores from the Centre for Economic Research and Graduate Education of Charles University and the Economics Institute of the Academy of Science of the Czech Republic.

In Brighton, 30/09/2011

Erik Arnold
Technopolis Limited
Managing Director

technopolis |group|
JOANNEUM RESEARCH
The University of Manchester Manchester Business School
MANCHESTER 1824
chepls Center for Higher Education Policy Studies
Universiteit Leiden
PERITUS