WR

WISSENSCHAFTSRAT

Drs. 1656-11
Halle, 11 November 2011

Recommendations on the Assessment and Management of Research Performance

Preliminary remarks **5**

A. Introduction **7**

A.I Background and subject matter **7**

A.II Our position **10**

A.III Objective of the recommendations **13**

B. Observations **15**
I.1 Assessment by peers 16
I.2 Assessment with the aid of quantitative indicators 18
I.3 Evaluation methods 21
I.4 Ratings/rankings 23
I.5 Management via the allocation of resources 24
I.6 Standards of good scientific practice 30
I.7 Consequences for teaching 31
I.8 Science policy framework 31

C. Recommendations **33**

C.I Higher-level recommendations **33**

C.II Individual recommendations **35**
II.1 The use of complementary management approaches 35
II.2 The use of peer review 37
II.3 Indicators of research performance 38
II.4 Evaluation methods 42
II.5 Ratings/rankings 43
II.6 Resource allocation methods 44
II.7 The documentation system 47

Preliminary remarks

Since the 1980s, the German Council of Science and Humanities has repeatedly and in various contexts expressed its opinion concerning issues such as performance-based differentiation, competition, and methods of assessing and managing research performance. At the same time, through evaluation and accreditation processes, its research rating and the excellence initiative, it plays an essential part in shaping the scientific system.

However, the Council is aware that at the present time, many scientists view current methods of measuring performance and ways of allocating funding with dissatisfaction, and are increasingly sceptical about the benefit of these processes for science. The Council therefore considers it necessary to engage in these debates, to identify the opportunities and risks associated with recent developments, and to offer recommendations that can fulfil the paramount aim of facilitating good research.

The Council recognises that changes in research practice, which are often attributed to methods of managing and assessing research performance, may have a wide range of causes, especially systemic ones. For example: the growth in size of the scientific system, increasing external demands on the scientific community, the international nature of research, the challenges of federalism, the complex funding channels, the increasing reliance on external funds, the fact that science policy initiatives are not always sustained, and the structure of the system, with a university and non-university sector. These conditions constitute the background against which the Council gives its recommendations.

To prepare its recommendations, the Council set up a working group which began its work in September 2010. Participants in this working group also include experts who are not members of the Council. The Council owes them a particular debt of thanks. In order to obtain a highly detailed and balanced picture of the current situation, the working group held discussions with representatives of the science ministries in selected *Länder*, with large scientific organisations, and with young scientists concerning the fundamental problems

and the effectiveness of existing management and assessment methods. The Council wishes to thank everyone involved for their support.

The Council adopted these recommendations for assessing and managing research performance on 11 November 2011 in Halle an der Saale, Germany.

A. Introduction

A.I BACKGROUND AND SUBJECT MATTER

For a number of years, the measurement and assessment of research performance – and resulting attempts to manage research and its impacts – have been the subject of intense discussion in the scientific community and in science policy. Proponents of new management approaches, who seek to enhance the ability of universities and non-university research institutions to act strategically, are opposed by advocates of traditional academic organisational principles, who see a threat to the role of the individual as the nucleus of research. Put simply, and ignoring more nuanced tones, these contrary positions can be described as follows:

Advocates of new management methods in research emphasise that the distribution of scarce public resources requires comprehensible, transparent justification with respect to policy makers and the scientists concerned. Accordingly, extensive investments in science are best legitimised by outstanding and useful research results. Moreover, this position assumes that there are measurable differences in performance between research institutions and between researchers. It calls for an increase in performance and efficiency through the targeted use of resources, and supports the further differentiation of the institutional landscape. To promote competition, the state should withdraw from the detailed management of scientific institutions and instead create targeted incentives, on an informed basis, from a higher-level perspective. The "costs" for the gain in autonomy on the part of science consist in the requirement for greater transparency and accountability. According to this view, it is primarily autonomous institutions which have the ability to act strategically. To increase competition, a range of target and performance-based measures should be implemented to make differences in performance transparent, create incentives for particular behaviour, and ultimately also reward or sanction the stakeholders. Supporters see the prospect of there also being material rewards for "good" scientific performance as being a motivational force, provided that the targeted institutional distribution of resources is possible based on valid criteria. This view has become widely

established in science policy to the extent that it currently shapes the framework within which the scientific system operates. Keywords associated with this view are: institutional autonomy, competition, performance-based resource allocation, new public management, efficiency, evaluation, transparency.

Critics, on the other hand, believe that the introduction of new assessment and management instruments calls into question the classical ideal of the self-determined researcher acting on his or her own initiative. In this view, the individual researcher fundamentally requires neither supervision nor disciplining; he or she performs their best when given trust along with sufficient freedom and resources to conduct their research activities. Trust, time freedom and intrinsic motivation are the keys to high-performance, creative research. Competition in the scientific system primarily consists in competition to deliver new research results and for recognition in the scientific community. Additional external, particularly monetary, competitive incentives which promise a higher personal income or a better equipped working environment to more prominent researchers are not required in order to increase research performance – in fact they may even be counter-productive and weaken intrinsic motivation or destroy its requisite social conditions. As well as effects on individual scientists, there is a fear of negative effects on the scientific system as a whole which may only become apparent in the long term, for example the neglect of research quality in favour of research quantity, the tendency to carry out low-risk mainstream research, declining diversity among researchers and in research topics, and the increasingly fragmented nature of published research results. According to this view, therefore, a preferable state of affairs is one in which appropriate conditions are created under which researchers can autonomously develop their potential. The demand from funders for accountability, transparency and evaluation is set off against the unintended drawbacks which these hold: an increase in competition to the point of mistrust, demotivation of the "losers", and an unreasonable increase in the amount of time spent writing applications and conducting evaluations. Attempts by science policy makers and the management of scientific institutions to intervene in the management of the scientific process are therefore rejected as being alien to science. Keywords associated with this view are: individual autonomy, trust, intrinsic motivation, research freedom, creativity, persistence.

This is a simplified portrayal of the debate surrounding the assessment and management of research performance and its effects, a debate which is taking place just as much in specialist publications and at conferences as in the press. The positions taken result from fundamental differences in assessments and perceptions of the current scientific system, which can be associated with the keywords "new public management" on one side and "Humboldt ideal" on the

other. Both poles of this debate first need to be placed in context in order subsequently to identify the tensions between goals which result from the different ways of measuring performance and assessing research.

Substantial changes in the framework within which scientific work is performed provide the starting point for the debates outlined above. For the German scientific system since the 1980s, the following developments have been the most influential: at the international level, the introduction of "new public management" in science policy in the English-speaking world; at the national level the impression that Germany is falling behind other countries in terms of research performance in many disciplines, despite two decades of continuous expansion of the scientific system. Moreover, core funding for universities stopped keeping pace with increases in the demands made of them. As compensation for a lack of core funding, and as a performance and quality indicator, external funding gained increasing significance in all areas of the German scientific system. Finally, the international and interdisciplinary spread of indicator-based rankings and formalised evaluation methods should be mentioned. Both are related to increased expectations for management and accountability standards from science policy makers and the public, and the stepping up of competition in the scientific system. |[1] However, emphasising competitiveness did not itself introduce a new principle; competition for scientific discovery, for professorial positions, for the formation of schools, for publications and for reputation had always been an effective mechanism in research. There had always been a differentiation of scientists by their performance, yet this was almost exclusively communicated to an internal audience in the subject field. What is new is that science policy uses competition within a subject in publicly financed research for external management measures and hence attempts to bring about a disclosure of performance differences beyond the subject field.

The assessment and management methods introduced in recent years, which form the subject of consideration below, serve partly to allocate resources on the basis of indicators, and partly to reward special achievements. A fundamental distinction can be made here between methods of direct management and methods of indirect management. While the former aim to achieve particular effects directly by imposing corresponding requirements (e.g. performance-based resource allocation by the *Länder* to universities), the latter aim to motivate stakeholders to act independently for the sake of quality

|[1] Cf. Wissenschaftsrat: Empfehlungen zu Forschungsberichterstattung der Hochschulen, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 1980, Cologne 1980, pp. 39-45; Wissenschaftsrat: Empfehlungen zum Wettbewerb im deutschen Hochschulsystem, Cologne 1985.

improvement or the purposeful use of resources by providing information and creating greater transparency (e.g. rankings or ratings). The most important incentives which can be employed here for effective management, both at the institutional and the individual level, are money, reputation, organisational freedom, and time. Most methods used in Germany for managing research performance, which are considered here, rely on assessment processes because they work via rewards and there needs to be a comprehensible basis for giving these rewards.

Assessment and management methods have been implemented at various levels of action. Essentially four levels can be distinguished: the funding bodies for universities and non-university institutions (German federal government and *Länder*), the institutions which conduct research (universities and non-university research institutions), the intermediary organisational units (faculty, institutes or similar) and the scientists within these institutions. Of course the stakeholders at these four levels create incentives and are influenced by these incentives to different degrees.

A.II OUR POSITION

The dichotomy between the positions described above, which formed as a reaction to the changes in the framework conditions for research, makes it difficult to reach a consensus. In the following, the Council assumes a position which recognises the incompatibility of the different points of view and attempts to deal with the situation in a way which demands concessions from both sides.

Good research as a goal of management

There is no value in assessing and managing research performance for its own sake; the purpose is to facilitate and encourage good research. Although it is not easy to define what quality research is, regardless of the specifics of particular disciplines there are clearly a number of principles of "good" research:

1 – Research is a practice of its own kind, a practice of discovery, which firstly follows the logic of the search for truth. Those who practice research have to be primarily inspired by curiosity and working on unsolved problems, and guided by the joy of discovering the new.

2 – In addition to working on given subjects and problems in the context of familiar theories and methods, good research is also characterised by originality and relevance. Furthermore, it requires a willingness to call oneself, one's own assumptions, theories, methods and data into question and constantly re-examine their significance.

3 – In its findings and in public communication, good research is bound by its own theoretical and methodical quality criteria, but it is open to examination by third parties; it documents and presents results which are comprehensible to and reproducible by other researchers, and in its conclusions is transparent and responsible towards the public.

4 – Truthfulness and mutual trust are fundamentally important for good research, which is not compatible either with deception or tolerated carelessness. Acting as the first line of control here is the community of scientists. In addition, justification towards the public is also a self-evident duty. The central regulatory principle is the quest for objectivity of results. In addition to discipline-specific criteria for theoretical and methodical working, this implies scientific ethics which are first of all bound to the rules of good scientific practice.

Conditions for good research

Good research not only requires motivated, creative, risk-taking researchers who are theoretically and methodologically competent and reflective, it also requires specific conditions: a scientific community that is not only able to do its work and offer criticism, but which is also willing to offer criticism, and an environment which recognises and facilitates good research. Securing good research in the long term is linked to certain conditions:

1 – It is essential to recognise the particular logic of research. This needs to be safeguarded not only in terms of political policy but also institutionally so that legitimate expectations and limits, rights and duties, practices and structures are transparent and comprehensible not only for science but also for policy makers and society, for cultural and business sectors; good research needs trust from its environment, freedom in its day-to-day work, and the opportunity to practice its own work and time patterns.

2 – Recognition and trust should be reflected in the organisation of the framework: in the autonomy of the institution, also as performance expectations for everyone involved, e.g. for the forms and duration of employment for scientists; in legal frameworks, via which the results of research in publication and use are protected; in financial resources which both cover continuous basic needs in universities and support their willingness to take risks, and which promote diversity in the practice of discovery.

3 – A broad foundation of high quality scientific work is indispensable and provides the basis for first-class performance (“excellence”) in the scientific system. The difference between pure research, applied research and research-oriented services does not describe scientific worlds of different value. Rather, it

reflects dimensions of scientific practice as a whole which cannot be sharply distinguished from one another.

4 – Scientists should be willing not only to use the resources provided to them by the state and society responsibly and effectively, they should also seek out the risks of working on open and unsolved problems, and communicate these to the public. This requires that scientists should not only work along theoretically or methodologically familiar lines and promise success, but also that they should seek to refute assumptions and findings and forcefully advocate the view that a lack of success, and failure – just like errors in one's own work – are a part of research practice.

External expectations of good research and scientific practice

1 – Autonomy of research does not mean autarky or a sense that research is located in worlds separate from social, cultural, economic and political reality and financial feasibility. Rather, science always requires open dialogue with public, political, professional and social expectations – and not just in return for the fact that science depends on external resources. If the scientific community remains aware of the fact that it is bound by social structures and expectations, this does not mean subjugation to the market for products and goods, nor does it mean the economisation of scientific work. Expectations for the efficiency and effectiveness of the use of resources are therefore just as legitimate as the expectation of the relevance of subject matter, and the hope for success in scientific work and for the opening up of new, promising research fields.

2 – Outside of the renowned ivory tower, science finds necessary and inevitable challenges to its creativity, advisory ability and problem-solving capacity. Accountability to the public and not just within the scientific community is therefore not alien to scientific practice but inherent to it. External expectations with regard to the subjects and problems to which science should turn its attention are not illegitimate but necessary and desirable provided the intrinsic logic of research is recognised.

3 – Good research benefits from a full willingness to engage in teaching, not just out of self-interest, because good research cannot survive without good young scientists, or because of society's need for training and qualification, but also for reasons lying within research itself, to strengthen criticism and self-criticism. Teaching in higher education always also presents the challenge of facing the impartial scrutiny of new generations of scientists and of demonstrating and substantiating the validity and relevance of one's own findings, including in respect of external expectations, in such a way that societies remain able to learn from generation to generation.

The Council recognises the co-existence as outlined above of legitimate claims to autonomy by science and inescapable external expectations of science. As a result, it calls upon universities and non-university institutions, upon science policy makers in the German federal government and in the *Länder*, upon the public and upon individual scientists to make their arguments about the assessment and management of research more impartial and to strive for balance between the different interests. The goal of these recommendations is to make suggestions as to how this can be achieved.

To this end, the Council adopts the following method: as a first step, observations are collected which point to possible links between practised assessment methods and management instruments, and positive as well as negative changes in research. As second step, while existing conflicts of interests are recognised, these observations are used to develop suggestions as to how positive effects can be strengthened and negative impacts as far as possible avoided. Some initial leads have already been provided elsewhere and are included here. |[2] The goal is to design assessment methods so that they satisfy the demands of policy makers and society for accountability at the same time as safeguarding the necessary freedoms for research and achieving a reasonable balance between expenditure and returns. The tension between having the most detailed and informative research assessments on the one hand, and the lowest possible expense and few unwanted effects on the other, is by no means the only tension between goals to which there is no easy solution. Indeed, such tensions also arise between the given complexity of research as a subject and the necessary simplicity of the methods used to assess and manage it, and between the autonomy of researchers and the legitimate demand of funding bodies for accountability for the use of public resources. It is therefore a further goal of the Council's recommendations first of all to raise awareness of these tensions between goals in the application of management instruments, and to propose corrections where necessary and possible. At the same time, the Council is aware that it is itself a stakeholder in the assessment and

|[2] Cf. e.g. Jansen, D. (ed.): Forschungspolitische Thesen der Forschergruppe "Governance der Forschung" – Rahmenbedingungen für eine leistungsfähige öffentlich finanzierte Forschung anlässlich der Tagung "Neue Governance für die Forschung". Berlin, 14 and 15 March 2007, German Research Institute for Public Administration, Speyer 2007; university chancellors' study group on performance-related resource allocation and target agreements (Arbeitskreis der deutschen Universitätskanzler(innen) "Leistungsorientierte Mittelvergabe und Zielvereinbarungen"): Empfehlungen zur Gestaltung von Steuerungssystemen auf der Ebene Land/Hochschule, Giessen 2009; Robert Bosch Stiftung: Thesenpapier der Gäste des 4. Berliner Wissenschaftsgesprächs der Robert Bosch-Stiftung, Stuttgart 2011.

management of research performance and therefore shares responsibility for the achievement of these goals.

The recommendations focus on publicly funded research, regardless of its respective form of institutionalisation or organisation. In the following, the Council addresses the German federal government and *Länder* as the funding bodies for universities and non-university institutions. Furthermore, it addresses the management staff of universities and non-university institutions at the various levels within these institutions, and moreover also the specialist groups and bodies representing the interests of specific disciplines, which are organised in various ways. Not least, scientists themselves are also addressed.

B. Observations

Changes in the scientific system over the last two decades, which also include the introduction of instruments for assessing and managing research performance, have contributed to making differences in research performance more transparent and comprehensible. Whereas performance differences that existed were previously known for the most part only to an initiated specialist community, an explicit clarification of the respective concept of quality in research became necessary in the process of developing assessment methods. Differences in performance were to be revealed to a broader public. An assessment of performance which is based on transparent criteria is comparatively less susceptible to an abuse of power; it is structurally more open and more dynamic, and it encourages competition between institutions and individuals. A further effect can be seen in that many disciplines – as a result of changed publication strategies, the acquisition of important externally funded projects, working on socially relevant issues also and the formation of coalitions of researchers – have become significantly more visible over the last two decades and can hold their own in international competition. For research institutions in Germany, not only greater visibility but also increased attractiveness can be observed as a result of their profile enhancement activities and targeted funding programmes. However, there are fears that these positive developments could change and that unwelcome effects could increasingly occur.

No consensus exists as to whether the scientific system in Germany has now reached a point where things have gone too far and effects which may have been positive initially will be undone as a result of negative side-effects. No valid empirical basis exists which would enable a comparison of the performance of the scientific system before the introduction of these instruments with its performance following the establishment of assessment and management methods, and which would also exclude other changes in the same period of consideration such as the growth in size of the scientific system, the relative cutback in core funding, and the increasing demands placed on science. To date only isolated studies have been produced, but these have not been able to demonstrate any causal relationships. Taking a comparative look at other

countries is also only helpful to a limited extent because of the very different conditions, and there too one finds only sporadic studies on the effects of assessment methods and management in research. |[3] Hence the examination below must confine itself to observations and plausible assumptions concerning possible effects and reactions in the scientific system resulting from the methods for assessing and managing research performance which have been introduced in Germany since the 1990s.

I.1 Assessment by peers

In order to assess scientific performance, a large number of different reviews and reports are produced in the scientific system by peers, i.e. by relevant experts in the scientific field. With regard to research, these include: reviews of journal articles, assessments of project applications and of scientists in procedures for the appointment of professors, and reviews of institutions by means of evaluations. They are complemented by processes relating to teaching (especially accreditation processes). |[4]

In a comparison of assessment methods, this performance assessment by peers – in what are referred to as peer review processes – best allows an in-depth quality assessment of research and institutionalised (self-)control within the scientific community. It best does justice to the diversity and complexity of research. Peer reviews are also used increasingly frequently because they underpin and legitimise science policy decisions. Although peer reviews are often demanded and performed on account of their clear advantages as assessment methods, some disadvantages are also evident: |[5] in a system of permanent assessment processes with short cycle times, the review system reaches its limits because of the high workload for everyone involved:

_ At the institutional level, the workload involved in conducting reviews is perceived to be high due to the poorly coordinated timing of the various processes and the frequently too short intervals between the individual assessment processes. Moreover it is problematic that the internal and

|[3] An overview of studies in the international field is provided, for example, by Butler, L.: Impacts of Performance-based Research Funding Systems: A review of the concerns and the evidence, in: OECD: Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings, Paris 2010, pp. 127-165.

|[4] It is anticipated that in January 2012 the Council will deliver recommendations on accreditations as an element of quality assurance in the scientific system.

|[5] For criticism of the peer review cf. e.g.: Osterloh, M.; Frey, B.: Das Peer-Review-System auf dem ökonomischen Prüfstand, in: Kaube, J. (ed.): Die Illusion der Exzellenz: Lebenslügen der Wissenschaftspolitik, Berlin 2009, pp. 65-73.

external data requests which are necessary to support peer reviews are no more standardised than the data held by the institutions.

_ At the individual level, accusations are levelled of "evaluitis" in a perpetually accelerating system. Reasons here include uncoordinated evaluation processes which overlap in time and content, data surveys for ratings and rankings, enquiries from science research, demands of various internal and external reporting systems, and internal quality assurance measures. For individual researchers, another factor contributing to a high workload is that sufficient support is not always provided by a service-oriented administration with close links to the sciences. For review activities, the work seems to be unevenly distributed, i.e. there is a small group of reviewers who have a particularly high workload resulting from reviews of projects, institutions, subjects or articles. |[6]

A further criticism of peer review is that the opinions of peers are sometimes contradictory or inconsistent, that the processes of group dynamics can lead to homogenisation and hence to judgements which are anti-innovation, that opinions rely too much on the reputation of persons and institutions, and that reviewers sometimes tend to allow their own preferred subjects and methods to influence their opinion. Hence there is a warning that peer review processes primarily promote the scientific mainstream.

One common way of dealing with these problems is the "informed peer review", in which the opinion of peers is supported by a certain set of indicators. This informed peer review is preferable to both a purely reputation-based assessment and an automated indicator-based assessment, partly because the indicator basis puts purely subjective opinions (or prejudices) to the test, and partly because the indicators, which are not always sufficiently informative in themselves, are classified and possibly relativised by experts.

One particularly important peer review process is the procedure for the appointment of professors, given that it is the key hurdle on the route to being awarded a professorship. Within the framework of statutory regulations, its form is largely left to the departments and faculties, although the university boards have an increasingly strong influence. In the review, suitability for the faculty and for the focus of the professorship plays an important role alongside

|6 For a recent discussion of the extent of review activities based on a representative survey, see: Böhmer, S.; Neufeld, J.; Hinze, S. et al.: Wissenschaftler-Befragung 2010: Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten, Institute for Research Information and Quality Assurance (iFQ) working paper no. 8, Bonn 2011, pp. 131-135. Only around one-third of respondents in this case reported having a medium or high review workload. However, the group of highly active reviewers now spend an average of 13% of their working time on review activities.

scientific quality. Apart from the lecture which is customary in most cases, quantitative indicators – mostly the number of publications, but occasionally also bibliometric instruments such as the "impact factor" |7 of journals in which candidates are published or the Hirsch index, named after its inventor |8 – are included in the quality assessment, especially in the pre-selection of candidates.

Variously voiced criticism of the appraisal process within the procedure for the appointment of professors relates to the narrowing of performance requirements, lack of transparency in decision-making with respect to candidates, insufficient consideration of a lack of research time due to family commitments, and in Germany the extremely long duration of the process. |9 Consequently – according to critics – it is not always guaranteed that the best qualified candidate with the greatest potential is appointed in an impartial process which is transparent with respect to those concerned. Weaknesses in the appointment procedure for professors should not be neglected, particularly because a professorial appointment is a decision which sets the course not only for the institution but also for the respective specialist community concerned, it has a far-reaching and long-lasting impact on the scientific system as a whole.

I.2 Assessment with the aid of quantitative indicators

Quantitative indicators are used in different assessment and management processes to measure performance in order either to generate information which is directly relevant to a decision or to provide the basis for informed peer review processes. The most popular indicators for measuring research performance are: acquisitions of external funds, numbers of doctorates, publication figures, and numbers of citations. As well as these more simple quantitative indicators, there is also increased demand for complex indicators relating to publications and citations (bibliometric indicators). Examples of

|7 A journal's impact factor states how often articles in the journal are cited on average in a particular period of time. It is considered a measure of the standing and importance of a journal.

|8 The h-index, Hirsch index or Hirsch number is a citation-based indicator which is primarily aimed at defining the research impact as a quality indicator, for which it considers the number of citations across the entire published output of a scientist; cf. Hirsch, J. E.: An index to quantify an individual's scientific research output, pp. 16569-16572, in: PNAS, 15 November 2005, vol. 102, no. 46. Interpreting and assessing this indicator is more demanding than a simple number would suggest, as it requires, for example, knowledge of the specialist context and age of the scientist.

|9 In 2005 the Council issued recommendations aimed at shortening and implementing quality assurance in the process, Wissenschaftsrat: Empfehlungen zur Ausgestaltung von Berufungsverfahren, Cologne 2005; cf. also on current criticism: Junge Akademie: Qualität statt Quantität – auch in Berufungsverfahren, Berlin 2010.

these include citations per publication compared to the average for the subject, or the Hirsch index. In addition, further indicators such as patents, licensing revenues, prizes, lectures or contributions to the organisation of conferences are used.

The use of quantitative indicators in the evaluation of research activities primarily serves to reduce complexity and enable comparability. Quantitative indicators meet the need for measurability and are relatively easy to operationalise. Furthermore, a formula-based comparison of quantitative indicators reduces the workload for reviewers. However, the supposed simplicity and clarity of these indicators may result in them being used too readily in the management of research. For example, common research assessment methods do not take sufficient account of self-reinforcing processes, e.g. with regard to citation figures, and hence contribute to an intensification of the "Matthew effect". |[10] The use of indicators is susceptible to manipulation and can lead to goal displacement, ultimately with the result that the goal is no longer to improve research performance but to achieve a quantitative increase with respect to certain indicators.

Overemphasising quantity over quality in assessment processes encourages a focusing on those activities which are measurable in one way or another: on high acquisitions of external funds, on a large number of publications, on numerous doctorates, etc. Up to a certain point, an increase in these measured variables can actually indicate research achievements and it is not proven that a focus on quantities necessarily entails a loss of quality. However, there are justified fears that a performance increase in measurable variables will reach (subject-specific) limits, and saturation effects will occur. The level of external funding, for example, may exceed limits above which the expense involved in acquiring and using the external funds is greater than the possible gain for research – studies for a number of subdisciplines already show this. |[11] With regard to faculty-student ratios, it is equally plausible that once a certain (subject-specific) number of doctoral students per professor is exceeded, it is no

|10 The Matthew effect described in the sociology of science by Robert K. Merton takes its name from a line in the Gospel of Matthew: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." (Matthew 25:29). Merton uses this to describe an effect whereby scientific publications by known authors are cited more frequently than publications by unknown authors. The term is now also used for other areas of the scientific system. A recent example is the German excellence initiative, which in the eyes of some of its critics particularly (and unfairly) encourages a small group of universities to dissociate themselves even further from the other universities as a result of these reinforcing processes.

|11 Cf. Jansen, D.; Wald, A.; Franke, K. et al.: Drittmittel als Performanzindikator der wissenschaftlichen Forschung, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, year 59, vol. 1, 2007, pp. 125-149.

longer possible to give appropriate support and supervision to the next generation of scientists.

A publication culture that is solely geared to publication figures can also generate problems. It is true that, *prima facie*, an increase in publication activity is not negative as it can demonstrate the growth and differentiation of science and be the expression of a useful strategy of imparting knowledge effectively. |[12] However, in so far as the number of publications and citations is crucial in determining positioning in performance-based resource allocation within universities, in procedures for the appointment of professors, in project applications, in rankings and in evaluation processes, a strategic publication practice is implicitly encouraged which aims to utilise a finding in as many individual publications as possible. In some disciplines, an increase in insufficiently quality-controlled anthology publications can be observed. Mainly they document the outcomes of externally funded projects and conferences, and this increase is encouraged by the fact that the external funders explicitly attach value to such publications. It can be assumed that such changes in publication practice are further supported by the fact that scientific publishers have an economic interest in publishing as much as possible.

Since the frequency with which publications are cited often plays a role alongside the simple number of publications, a tendency can be seen in a number of disciplines for their publication activities to be oriented towards journals which have a high impact factor. As a result of such adjustments, which take on different subject-specific forms, goal displacement can occur: publication strategies are no longer optimised on a subject-specific basis to achieve the most useful transmission of findings and search results, but to meet certain criteria in the performance assessment. The German Research Foundation (DFG) has attempted to counterbalance publication cultures which are too strongly focused on quantities by permitting only a limited number of selected publications in its application procedure as of 1 July 2010, instead of publication lists that are as extensive as possible. It remains to be seen what effect this will have on publication behaviour, which is also shaped by international practice.

|[12] In the name of scientific progress it may make perfect sense to publish as early as possible and successively regarding a research project; it can also be appropriate to increase publication output in order to raise awareness of a research field or increase the visibility of an entire discipline. It may also be useful not to include more than a key finding in a publication so that it can be better communicated and noticed.

Evaluations in research are generally geared to research projects, research units or whole institutions and may evaluate an institution's performance and achievement of objectives after the event (*ex post*) or with regard to the future (*ex ante*). |[13] They serve both to disclose and assess performance and to identify development prospects and improvement potentials. In Germany, evaluations are usually conducted in a peer review process, i.e. an appraisal is carried out by scientists who are relevant to the institution or research project. The disadvantages of peer reviews described above, especially the high workload and need for coordination, therefore also apply to evaluation processes.

At universities, evaluations of research and/or teaching form an integral part of internal quality management and are an instrument of quality assurance. The exact structure of the evaluation (intervals, method, etc.) is not specified in the individual Higher Education Acts (*Hochschulgesetze*) in Germany, and is therefore left to the universities. In some cases, the effect of evaluations tends to be to exercise indirect control and initiate a process of self-reassurance, which may result in learning effects for the development and improvement of institutions. Critics object that not all evaluation processes actually have consequences and that in some cases, particularly at universities, they are still purely a ritual whose only purpose is the formal fulfilment of quality assurance requirements. Far-reaching direct consequences such as the closure of university institutions or individual departments due to poor research performance are generally not to be feared owing to the importance of the respective institutions for teaching, while significant increases in funding are not possible due to financial restrictions. The extent to which research-related evaluation processes at universities are actually directly relevant to decisions or at least used as an instrument for indirect management cannot be clearly determined and probably varies between universities and *Länder*.

At non-university institutions, evaluations are widely used. The methods vary depending on the funding organisation. The member institutions of the Leibniz Association (WGL) are evaluated at least every seven years by the Senate Evaluation Committee (SAE). In addition to this, the institutions' scientific advisory boards conduct audits for continuous quality assurance. The Max Planck Society (MPG) has its institutes assessed every two years by its scientific advisory boards, which are composed of external appraisers. In addition, as part

|[13] An example of an *ex-post* evaluation would be the evaluations of non-university research institutions conducted by the German Council of Science and Humanities; an example of an *ex-ante* evaluation is the assessment procedures which were performed by DFG and the Council as part of the German excellence initiative.

of an extended, strategic evaluation, institutes which deal with related subjects are combined in one research field and compared with each other every six years. The MPG also emphasises the evaluative character of its *ex-ante* appraisal in its selection of scientists, |[14] which is similar to the appointment procedure for professors at universities. In the Helmholtz Association (HGF), the research programmes are reviewed every five years across the Helmholtz Centres by external experts. In addition, the centres themselves are evaluated. The Fraunhofer-Gesellschaft (FhG) evaluates its institutes approximately every five years in what it calls technology audits. External representatives from the business and science communities are included in the appraisal process. The institutes are continuously measured in terms of the project funding acquired from the business community and from public research funding; internally this has a direct effect on the amount of core funding that is allocated. In addition, the institutes undergo an external evaluation of their performance on the market.

In the non-university sector, evaluations can often have many kinds of immediate and far-reaching effects both in financial respects and with regard to the further development or indeed the future of an institution (in terms of subject areas, staffing, and structurally). In the non-university sector, the introduction of evaluation methods as a quality assurance measure is mostly seen as being positive. However, just like in the university sector, the associated cost is criticised.

| [14] This is referred to as the “Harnack principle”, according to which institutes are founded on the basis of extraordinary scientists rather than a particular subject.

I.4 Ratings/rankings

Ratings/rankings are methods for the comparative assessment and disclosure of performance in research. As instruments of indirect management, unlike for example the earlier Research Assessment Exercise (RAE) in the United Kingdom, in Germany they are not linked to the allocation of funding. Since the 1980s various different rankings have become established in Germany also. They include research and university rankings of different providers in individual subjects or as a performance comparison between whole institutions (e.g. university and research rankings by the Centre for Higher Education [CHE], Handelsblatt ranking). In response to existing methods, in 2004 the Council itself issued recommendations for a research rating and since then has trialled such a method in several subjects in a pilot study. |[15] The methodology of comparative assessment varies. Generally indicators are compared to establish a ranking; in the Council's research rating, quality and quantity-based indicators are assessed by peers ("informed peer review"). |[16] Interdisciplinary rankings, in which entire universities are compared with each other based on a small number of indicators – such as, in particular, publication figures, grant money, or, if applicable, major prizes – and ranked in order of merit, are not carried out in Germany. However, they play an important role in the international arena (especially the Academic Ranking of World Universities, also known as the Shanghai ranking, and the World University Rankings), and also include German universities and in some cases also the German non-university research institutions. Due to a lack of sophistication in the criteria used and problematic bases for comparison, these rankings should be viewed critically.

As a method for the comparative surveying and assessment of research performance, ratings and rankings have the positive effect of revealing differences in performance and providing a basis for comparison across institutions and individuals. This provides a reference point for research institutions and specialist groups regarding their positioning in the national and where applicable also the international arena. Sophisticated, scientifically based methods therefore serve as guidance. They can have a competitive effect

|15 Cf. Wissenschaftsrat: Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften (Drs. 10039-10), Cologne 2010; Wissenschaftsrat: Pilotstudie Forschungsrating, Empfehlungen und Dokumentation, Cologne 2008; Wissenschaftsrat: Empfehlungen zu Rankings im Wissenschaftssystem – Teil 1: Forschung, in: Wissenschaftsrat, Empfehlungen und Stellungnahmen 2004, vol. I, Cologne 2004, pp. 159-220.

|16 Cf. Wissenschaftsrat: Empfehlungen zu Rankings im Wissenschaftssystem – Teil 1: Forschung, in: Wissenschaftsrat, Empfehlungen und Stellungnahmen 2004, vol. I, Cologne 2004, pp. 159-220; IREG Observatory on Academic Ranking and Excellence: Berlin Principles on Ranking of Higher Education Institutions, Berlin 2006.

which can also influence the strategies of whole *Länder*, individual universities, and individuals. However, it is problematic that with regard to their informative power, ratings/rankings are not always sufficiently detailed. This can lead to premature judgements concerning national and international competitiveness. For this reason, as in the case of direct assessment using quantitative indicators, goal displacement can occur. In the extreme case, the interest in the content and quality of research is supplanted by the interest in a position in the ranking.

I.5 Management via the allocation of resources

Performance-based allocation of Länder funding to the universities | [17]

Methods of performance-based resource allocation (PBRA) for the university budgets of the *Länder* to their universities serve to support the change from detailed ministerial management to greater autonomy of the individual institutions, while simultaneously placing the emphasis on quality and competition. Since the mid-1990s, a majority of the *Länder* have introduced different PBRA models and developed them further in subsequent years. There were various intentions behind the introduction of PBRA systems at the level of the *Länder*: (1) elimination of the arbitrary features of a financing system which had developed historically; (2) resource allocation based on clear criteria and science policy goals; (3) financial recognition of performance and workload differences in research and teaching; (4) accountability towards the *Länder* parliaments; (5) appropriate use of scarce budgetary resources.

Currently, in the individual *Länder*, very different proportions of resources are allocated based on teaching and research-related indicators. Until recently, in most *Länder*, it was mainly material resources that were allocated via PBRA models. Now, however – also as a result of the introduction of global budgets – human resource funding is increasingly being allocated on the basis of indicators. The figures range from less than five to nearly 100% of total higher education resources, with limits generally being imposed on any redistribution effects. | [18] This cap on PBRA funding, and the fact that a *Land*'s total budget for

| [17] In the following, no reference is made to the performance-based allocation of resources to non-university institutions because the individual organisations each have specific resource allocation mechanisms which would require a more detailed individual consideration.

| [18] A comparative study of PBRA systems in the *Länder* was last presented by the Higher Education Information System (HIS) in 2004: Leszczensky, M.; Orr, D.: Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung: Dokumentation und Analyse der Verfahren in 11 Bundesländern, HIS-Kurzinformation A 2/2004, Hanover 2004; an update of the study has been announced for the end of 2011.

its universities is usually limited and hence the indicators only govern allocation between universities, may lead to individual universities having to accept cuts despite having increased their performance, if other universities show even bigger increases. The successes of one university are therefore relativised by the successes of another university.

The indicators on which PBRA is based reflect different science policy goals, of which increasing research performance is only one. The most important factors are usually teaching-related indicators such as the number of students who are expected to complete their studies within the normal course duration, the number of new students, or the number of graduates. This is because the *Länder* are required to finance the universities according to the demand from students. As a result, their scope for management via further incentives is considerably limited. The main indicators of research performance are research-related external funding (acquired or awarded; in some cases weighted by origin) and doctorate numbers. |[19] Depending on the science policy goals, indicators relating to equal opportunities and international diversity are also often used. Initially more complex allocation formulas have been successively simplified in most *Länder*.

Critics of PBRA methods complain that performance is not always given priority as an allocation criterion, and that instead the *Länder* are often primarily concerned with having a comprehensible basis for resourcing their universities according to their respective teaching and research workloads in order to maintain their ability to function. These intentions have not always been clearly stated. Indeed, the introduction of PBRA systems has primarily been justified in terms of competition and meritocratic principles. There is also criticism of the fact that the choice of indicators on which resource allocation is based primarily depends on the availability of corresponding quantitative data. In contrast, important considerations as to the advantages and disadvantages that particular indicators have – whether they are at all suited to giving an adequate representation of the performance profile and to exercising management control in the sense that universities are actually able to influence them – often remain secondary.

A significant positive and intended effect of PBRA is that it limits the scope of the *Länder* for discretionary influence on resource allocation, resulting in an increase in rationality and equality. It seems plausible that universities will

|[19] Research-based resource allocation is often determined using only these two indicators, or where applicable with the addition of the number of postdoctoral degrees. In some cases further indicators are included, such as success in competitive research funding programmes. More rarely, publication figures, patents, industry partnerships and transfer activities are used.

orientate themselves to the given indicators and that incentives for being performance-oriented are therefore created, particularly since, looking to other countries, it appears that adjustment reactions can occur even where small proportions of funding are involved. |[20] However, strategic decisions by policy makers which are not related to performance continue to have the effect of compensating for funding differences between universities which PBRA is actually supposed to encourage. Hence, particularly with regard to increasing research performance, there remain doubts as to the management effectiveness of performance-based resource allocation in its current form.

Performance-based resource allocation within the universities

Most universities in Germany use performance-based resource allocation methods internally as well. This PBRA within universities is the competence and responsibility of university and departmental management. The proportion of the budget, which is allocated by university management to the intermediate level (faculties/departments/subjects) and from these in turn to the institutes/chairs/professorships, varies. Different approaches can be observed, ranging from internal target agreements to fine-tuning via highly sophisticated indicator systems. Experience suggests that universities are wholly or at least predominantly guided by the PBRA model in the respective *Land*, adopting its performance indicators and not differentiating sufficiently between subjects. |[21] These predominantly quantitative criteria are generally not assessed by specialist colleagues (peers). Even if relatively less funding is allocated via PBRA

|[20] Australian universities were allocated approximately 7.9% of their income via an indicator-based funding model, which is nearly all their core funding. Although this proportion is not high, there were additional incentive effects because this form of resource allocation seems easier for the universities to influence than the portion of funding which does not come from core funding, and because the gains in reputation are high and the university's research performance is a crucial factor when students decide where to study. Since the reward model gave a weighting of more than 50% to the external funding indicator, specific reactions could be observed on the part of the universities. Effects on the system as a whole could be demonstrated in Australia, especially a concentration of research funding, changes in the research portfolio of individual researchers, and a gradual separation of research and teaching. There are various studies on the Australian system, cf. e.g. Gläser, J.; Lange, S.; Laudel, G.; Schimank, U.: Evaluationsbasierte Forschungsfinanzierung und ihre Folgen, in: Mayntz, R.; Neidhardt, F.; Weingart, P. et al. (eds.): Wissensproduktion und Wissenstransfer: Wissen im Spannungsfeld von Wissenschaft, Politik und Öffentlichkeit, Bielefeld 2008, pp. 145-170.

|[21] On this point cf. also the following study: Leszczensky, M.; Orr, D.: Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung. Dokumentation und Analyse der Verfahren in 11 Bundesländern, HIS-Kurzinformation A 2/2004, Hanover 2004: of the 43 universities which were based in *Länder* having an official formula model, 13 universities (30%) had to a large extent followed the model used by the *Land* when they developed corresponding internal methods in the university. This was at least partly the case for a further 24 universities (56%).

at university level than at the *Land* level, more complex indicator systems are now increasingly being introduced. Attempting to carry out a more sophisticated allocation of resources based on individual research performance, these systems follow the respective university's structural and development plan. This trend for more sophisticated models is contrary to the PBRA models at the *Land* level, which have been incrementally simplified. A fundamental problem is that while different quality criteria and requirements of various subjects can be taken into account in a more complex model, it is necessary to reduce complexity in order to formulate allocation keys, for example by assigning a number of points to the specific performance criteria. Consequently, it is questionable whether, ultimately, the amount of work required to allocate the resources is reasonable, especially when only relatively small amounts of resources are involved.

If universities follow the models of the *Länder*, which primarily reward growth in the number of graduates and doctorates and in external funding, there is the danger within universities of giving preferential treatment to faculties which have large numbers of graduates and large amounts of external funding, without this actually revealing performance differences between subjects. The effect of PBRA within universities is difficult to assess across universities and subjects. |[22] Particularly strong effects are attributed to PBRA methods in medicine; this could be due to the special funding situation and organisational structures in medicine, and to the fact that PBRA methods were introduced very early on in this field. |[23] Even aside from the special case of this discipline it can be assumed that scientists react to incentives created at management level and, for example, make efforts to increase doctorate numbers or acquire more external funding. However, it remains questionable whether increasing acquisitions of external funding are primarily due to the monetary incentive created by PBRA systems, or whether instead to the financing needs of

|[22] In 2006, the university chancellors' study group on performance-based resource allocation and target agreements (Arbeitskreis der Universitätskanzler(innen) "Leistungsorientierte Mittelverteilung und Zielvereinbarungen") came to the conclusion that the management effect of PBRA within universities was limited due to its generally low budget relevance. Cf. Arbeitskreis der deutschen Universitätskanzler(innen) "Leistungsorientierte Mittelvergabe und Zielvereinbarungen": Hochschulinterne ziel- und leistungsorientierte Mittelvergabe. Eine Handreichung, Oldenburg 2006, p. 14; cf. also: Jaeger, M.: Wie wirksam sind leistungsorientierte Budgetierungsverfahren in deutschen Hochschulen?, in: Nickel, S.; Ziegele, F. (eds.): Bilanz und Perspektiven der leistungsorientierten Mittelverteilung. Analysen zur finanziellen Hochschulsteuerung, CHE working paper no. 111, Gütersloh 2008, pp. 36-50, here p. 45.

|[23] A study called "Governance der Hochschulmedizin – Intendierte und nicht intendierte Effekte dezentraler Anreizsysteme am Beispiel der fakultätsinternen leistungsorientierten Mittelvergabe (LOM) in der Medizin" by the Institute for Research Information and Quality Assurance (iFQ) is scheduled for completion and publication in 2012.

particular research projects, or whether also to the increasing importance of external funding for the attribution of reputation.

Salary scale W

At the individual level, new performance incentives have existed since 2004 in the form of salary scale W (*Besoldungsordnung W*). While scientists' basic pay is significantly lower than under the old salary scale C, now for example they can receive "special payments" for particular achievements in research, teaching and other areas of activity. Corresponding provisions for the granting of these payments have been included in the ordinances of the individual *Länder* and in the respective universities' regulations concerning the award of special payments. With regard to research activities, criteria include, for example, results of research evaluations or quantitative parameters such as external funding, publications and patents. Universities have a certain degree of freedom in selecting the criteria and can therefore grant additional pay in accordance with their respective strategic goals. Hence they are able to use salary scale W for management purposes.

Basically, salary scale W is a complex body of regulations on several levels – the German federal government, *Länder* and universities – which is still being developed and tested, including in the courts, making a comprehensive assessment difficult. Criticism of the salary scale W regulations for rewarding special achievements in research revolves around the fundamental question of whether individual monetary incentives are a way of achieving sustainable management of research performance, and is also directed at the specific, often too detailed formulation of criteria for awarding the relevant additional pay.

External funding in universities

Compared to other countries, Germany has a very diverse funding landscape. The most important funding bodies are the German Research Foundation (DFG), the German Federal Ministry of Education and Research (BMBF), the European Union (EU), a wide variety of different national foundations, and industry. The wide range of funding instruments allows individual research projects to be carried out – for example under the DFG's "standard procedure" – as well as large cooperative joint projects such as those funded by the BMBF and EU. Moreover, particularly risky projects can also be implemented, for example under the DFG's Reinhart Koselleck programme. Due to the scope and diversity of available funding instruments, completely different research questions arising from debates in the scientific community, with different requirements in terms of methods, time demands and subject focuses, can be investigated with the aid of financial resources provided by third parties.

In view of the fact that core funding for research at universities has stagnated for many years, external funding plays an increasingly important role. |[24] As reliance on external funding increases – albeit to varying degrees in different subjects – the allocation practices associated with it have a stronger impact. This is intensified by the fact that external funding is referred to as an indicator of quality in evaluation processes and also in most PBRA systems at the *Land* and university levels. Hence additional importance is attached to it for the allocation of core funding. Conceivably the worst case scenario that could develop from this situation would be that acquisitions of external funding no longer serve to fund specific research questions and instead their primary purpose is to be used as a central performance indicator in performance-based resource allocation. |[25]

For all the diversity that exists in the external funding landscape in Germany, research using external funding is broadly characterised by certain specific features which, given the growing importance of external funding, may have a disproportionate effect on the research system as a whole. |[26] Critics point out that short-term external funding creates opportunity structures which in the long term may lead to a preference for narrow research questions. Faced with a limited duration of funding, it would seem obvious to deal mainly with research questions for which data or studies already exist (secondary analysis research). Alternatively, the time pressure of project research – possibly combined with the development of ever faster measurement techniques – could encourage a development whereby, although primary data for research questions is collected and published, there is no longer sufficient analysis of the data acquired with the goal of theory formation.

Critics particularly often argue that a consequence of the increased importance of external funding may be that it leads to a concentration on subjects which

|[24] On the changed relationship between core and external funding at German universities, see the recent report by the Chairman of the Council on current trends in the scientific system: Neuere Entwicklungen der Hochschulfinanzierung in Deutschland, July 2011: while core funding for universities in Germany increased by only 6% between 1995 and 2008, universities' expenditure of external funding more than doubled. The share of external funding in the total budget rose from 11% to nearly 20% between 1995 and 2008. If one looks only at research funding, it is clear that as a result there has been a significant shift in the relationship between core funding and external funding.

|[25] In a survey recently conducted by iFQ, a majority of scientists themselves stated that the possibility of answering research questions and the positive effect on reputation were the most important motivating factors for becoming active in the external funding market, not the use of the external funding indicator in performance-oriented resource allocation methods. Cf. Böhmer, S.; Neufeld, J.; Hinze, S. *et al.*: Wissenschaftler-Befragung 2010: Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten, iFQ working paper no. 8, Bonn 2011, pp. 85-89.

|[26] On this point, cf. e.g. Torka, M.: Die Projektförmigkeit der Forschung, Baden-Baden 2009.

can be considered to belong to the scientific mainstream or which happen to be in fashion. This is based on the assumption that in the application and review processes associated with external funding, the reviewers are more likely to consider these subjects worth working on, and hence there will be a strong orientation to the subject focuses of extensive research funding programmes. Finally it is problematic if scientists, regardless of their own research interest, deal with research questions that promise a large acquisition of external funding with little outlay in respect of the application process. This can lead to a preference for projects and subjects which are sought after primarily for political reasons and therefore receive strong financial support. On top of that, while the external funding landscape in Germany does allow a wide range of individual questions to be worked on by individual researchers, particularly under the DFG's "standard procedure", successes in individual funding increasingly receive less institutional recognition than successes in the acquisition of larger, more visible and cooperatively implemented joint projects, which in addition usually have a larger volume of external funding. Critics also argue that a concentration of external funding at a few locations can occur, resulting in medium to long-term changes in the institutional landscape.

However much criticism there is of the possible negative impacts of excessive reliance by research on external funding, competitively allocated external funding still is an essential and indispensable element of quality assurance through scientific competition and can – up to a certain point – have a positive influence on research performance.

I.6 Standards of good scientific practice

The assessment and management of research performance serves to ensure that standards of good scientific practice are upheld. Owing to the great importance of trust as a basis for scientific communication processes, any neglect of the standards of good scientific practice threatens the success of science. Given the growth of the scientific system and the intense scrutiny of scientific practice by the media and by policy makers, it is not surprising that scientific misconduct is more frequently mentioned as a problem in the public arena today than in the past. However, this misconduct does not always take the form of scandalous plagiarism or serious falsification of data. Dishonest conduct in the research process is often difficult to detect and may find expression, for example, in methodologically sloppy work that fails to meet established standards, the neglect of sources that are hard to obtain, improper authorship, neglect of citation standards, quoting unread references, and citing potential reviewers just to win favour. In many respects there is a prevailing impression that such practices have become noticeably more widespread as assessment and management systems have intensified time pressure and competitive pressure. It seems reasonable to suppose that the inherent high expectations of quality

and originality in research combined with high time pressure would encourage such misconduct. Increasing competitive pressure in an internationalised scientific landscape, in which an ever greater number of researchers from Germany and other countries are in competition for reputation and employment opportunities, may also be a cause of such misconduct.

I.7 Consequences for teaching

Research is embedded in a broad network of relationships, with the result that changes in the research funding framework will also impact on other areas of the scientific system. For example, teaching may benefit from good research if the acquisition of additional research funding leads to an increase in the resources available for teaching. Alongside this, however, there may be negative consequences for teaching particularly if incentives for research are created at the expense of teaching. This is the case, for example, when research freedom is used as a reward and a complete withdrawal from teaching occurs, and where, conversely, rewards are not given for greater involvement in teaching (for example in the form of better resources or "sabbatical leave for teaching enhancement"). |[27]

Finally it should also be considered that in the value system of science, research with its national and international prominence enjoys a higher status than teaching with its locally limited effects. For as long as a greater reputation can be acquired through research and this generates higher regard, teaching will assume a subordinate role. Enshrined in the scientific system itself, this reputation differential makes it difficult to establish effective compensatory incentives for teaching.

I.8 Science policy framework

In recent decades, universities in Germany have been faced with numerous new requirements imposed by science policy in respect of course and governance structures, the legal framework, and target-setting. Frequently only a short time was given in which to implement the requirements. It was and is hardly possible to reflect on the initiated processes when new targets are set within short periods of time and management measures affecting the ability of the universities to act strategically are introduced and changed by science policy makers on short terms. The implementation pressure has an effect at the levels

|[27] The purpose of a teaching semester is to allow professors to give sustained consideration to methodological and didactic issues relating to teaching and learning, with the aim of improving their teaching skills or developing new teaching and learning concepts, cf. Wissenschaftsrat: Empfehlungen zur Qualitätsverbesserung von Lehre und Studium, Cologne 2008.

of faculties, departments and, finally, chairs. Often, implementing new political goals as rapidly as possible takes precedence over the cautious introduction of innovations. Correspondingly, it is difficult for there to be a longer-term adjustment to changed conditions, which could also serve as guidance and have an incentive effect. In the non-university sector also, extensive and lasting reforms have taken place, |[28] but the conditions here are comparatively more stable.

|[28] For example the introduction of programme-oriented funding in the Helmholtz Association and the establishment of the Leibniz Association's evaluation system.

C. Recommendations

The observations of possible impacts of the assessment and management of research performance which are described above bring to light various tensions between goals which need to be taken into account in the formulation of recommendations: tension between goals for example between the priorities within science for the selection of subjects and the need to take into account relevant factors outside the field of science, or between specific and general funding, and above all between reducing the workload involved in the assessment process and recognising the complexity of the subject matter, and between the demands of the scientific community for research freedom and the public demand for accountability and efficiency.

As described above, such goals in tension continually engender controversy. These recommendations follow in response to this controversy. The aim should be to find a pragmatic way of dealing with unresolvable tensions between goals. Even if no permanently applicable courses of action can be identified, at least extreme positions and strong swings towards one or the other pole should be avoided.

C.I HIGHER-LEVEL RECOMMENDATIONS

The Council makes the following higher-level recommendations as guidelines for the development of recommendations for action (C.II):

1 – Self-reflexivity: organisations in the scientific system need institutional mechanisms (e.g. reporting systems, benchmarking processes) in order to be extensively informed and able to provide information about themselves. Externally, this is necessary for the sake of accountability and disclosure of their performance, and in the internal perspective in order to enable strategic action, decision-making, learning and development. In particular this also includes the development of assessment competence within the non-university research institutions and universities.

2 – Procedural review: assessment processes and management instruments must not become an end in themselves. Their form and structure absolutely require reflection as to their function, their desired and undesired effects, their reach and their impacts on other areas. Regular reviews of the processes and their effectiveness – where possible with the involvement of science research – reduce the risk that they will take on a life of their own or come to be used for another purpose.

3 – Limitation of workload: however great the demands for complexity, diversity and reflexivity in the processes, efforts to limit workload must take priority in order to reduce the burden on everyone involved. Workload can be limited via various specific measures (for example, coordinating different processes, standardisation of data collection); it does not necessarily mean an inappropriate reduction in complexity.

4 – Appropriate timing: balance between the unpredictabilities of the research process and demands for regular reporting and predictability can be achieved by slowing down assessment process cycles and/or bringing them into an appropriate rhythm.

5 – Avoiding a "tonnage ideology": in the assessment and management of research performance, an increase in quantity does not necessarily mean an increase in quality. Hence the corresponding processes must recognise and take into account that it is primarily the quality of research that counts.

6 – Suitability: differences in the "missions" of universities and the various non-university research institutions must be taken into account in the development of methods for the assessment and management of research performance. These differences may mean that there is necessarily a different research orientation. The missions shape the task profile of an institution, which may for example include tasks relating to teaching, the provision of services for science, and consultation, in addition to research.

7 – Inclusion of all stakeholder levels: the inclusion of all relevant stakeholders and improved communication between the levels is necessary for the development of appropriate and accepted methods for the assessment and management of research performance. In particular, open participation by specialist groups in the further development of these methods is essential. Only in this way can subject-specific special features and demands be incorporated efficiently and effectively.

8 – Financial requirements: core funding for universities and non-university research institutions must enable them to perform their core duties in research and teaching, fostering young talent, and – depending on their task profile – the provision of services. Only appropriate core funding can guarantee security and a long-term approach, thus also safeguarding "risky", unconventional and time-consuming research and enabling researchers to change between lines of research.

9 – Ensure attractiveness: management systems should attract, support, and facilitate the independence of young scientists, offering them attractive prospects and working conditions at universities and non-university institutions.

10 – Take teaching into consideration: to counteract the preference for research over teaching in the scientific system, effective incentives for teaching and management mechanisms which aim for a balance between incentives for teaching and research are urgently required at the individual level and at the level of the universities.

11 – More diverse approaches to management: to avoid goal displacement and other unwanted developments, management in research should draw on the widest possible repertoire of instruments and not give priority to monetary incentives and quantitative performance measurement.

C.II INDIVIDUAL RECOMMENDATIONS

II.1 The use of complementary management approaches

To prevent individual management approaches – particularly monetary ones – from causing undesired adjustments, the Council recommends the use of diverse management instruments, particularly a guarantee of appropriate scope for autonomous self-management and time flexibility.

In the science field, autonomy and time flexibility are valuable assets. So far their effectiveness as an incentive and management instrument, especially for scientists, has been neglected. The Council recommends that researchers should be granted an appropriate degree of autonomy depending on their career stage and individual requirements: scientists should be allowed the opportunity to focus flexibly on research or teaching activities for limited periods of time; to prevent teaching and research becoming detached, research should always be conducted to a certain extent even if there is a focus on teaching, and vice

versa. |[29] The only exceptions should be shorter phases in the form of sabbatical leave for research or teaching enhancement. The Council therefore recommends that in addition to the established instrument of sabbatical leave for research, universities should also make increased use of the instrument of sabbatical leave for teaching enhancement – not only for the purpose of flexibilisation but also to compensate for the generally stronger incentive in research. |[30] In addition, universities should enable researchers to attend institutes for advanced studies more frequently than at present for limited periods of time. This kind of flexible career model could be an advantage as universities compete (internationally) for scientists.

Since procedures for the appointment of professors control entry into the scientific system, with long-term consequences, university management and the *Länder* should devise these procedures transparently with suitable standards so that different dimensions of performance (research, teaching, transfer activities), including diversity and equal opportunities aspects, are given appropriate consideration (see also C.II.3). |[31]

Target agreements can also be a useful tool for results-oriented quality management. They can be used at different levels (the individual, the intermediary level, the institution, the funding body). For them to be used effectively, a high level of management competence is required for the formulation of suitable targets which are not too detailed or one-sidedly focused on quantities.

Furthermore, the Council recommends the targeted use of monitoring of research activities and instruments for indirect management at the various levels of the scientific system. The disclosure of evaluation results and performance-related data should be considered in order to leave reputation-based management in large part to the scientific system itself. Ratings and

|[29] Cf. also Wissenschaftsrat: Empfehlungen zu einer lehrorientierten Reform der Personalstruktur an Universitäten, Cologne 2007.

|[30] Sabbatical leave for teaching enhancement has been awarded at Technische Universität München, supported by funding from the "Excellent Teaching" competition organised by Stifterverband and the Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder* in the Federal Republic of Germany (KMK), cf. http://portal.mytum.de/studium-und-lehre/lehrpreise/freisemester_fuer_lehre.html as at 9 October 2011. Another strong incentive for greater involvement in teaching is offered by the German federal government and *Länder* with their funding programme for quality in teaching, cf. http://www.bmbf.de/foerderungen/15440.php as of 7 October 2011 or http://www.gwk-bonn.de/index.php?id=269 as of 10 November 2011.

|[31] Cf. Wissenschaftsrat: Empfehlungen zur Ausgestaltung von Berufungsverfahren, Cologne 2005; criticism of current practice has also been expressed, for example, by Junge Akademie: Qualität statt Quantität – auch in Berufungsverfahren, Berlin 2010.

rankings which satisfy scientific requirements (cf. C.II.5) can support this at the institutional level. In order to compensate for the disadvantages of competition for reputation, particularly the preference for experienced scientists, young researchers should at an early stage carry out research projects independently and on their own initiative, they should bear responsibility for the quality of the projects, and should be able to publish the results themselves. In keeping with the rules of good scientific practice, clear indication of authorship by young scientists is absolutely essential here.

Scientific misconduct and inadequate performance should in the first instance be dealt with by the scientific community itself. If scientific misconduct is proven in the context of externally funded projects, the funding institutions should also examine possible sanctions. This has been the practice of DFG, for example, for some time. As a preventive measure, possible impacts on the upkeep of standards of good scientific practice should be considered when assessment and management processes are devised. In the context of evaluation processes, it should also be possible at universities for action to be taken at the individual level in the case of poor research performance.

Summary of recommendations concerning the use of complementary management approaches:

1 – Grant a greater degree of individual autonomy, allowing flexibility in the setting of priorities over the course of a scientist's career.

2 – Devise procedures for the appointment of professors, which are a key quality control mechanism at the point of entry, based on a broad concept of performance (research, teaching and transfer activities) and comprehensible criteria.

3 – Examine the use of instruments for indirect management in the competition for research reputation by universities, the German federal government, and the *Länder*.

II.2 The use of peer review

Peer review, as a science-led assessment method with a primarily quality orientation, is indispensable in the assessment and management of research performance. At the same time, however, peer review processes involve a lot of work and tie up a large amount of resources, which is why there should always be good reasons for their use. Like all other evaluation methods, peer review processes are also susceptible to errors, and undesirable effects cannot be fully ruled out. Quality assurance and reflection on the processes are therefore just as essential as compliance with central standards that guarantee the reliability and validity of the results of the assessment. In particular, this applies to the selection of reviewers; it is especially important that the composition of

reviewer groups is as diverse as possible and, depending on context, also international. Young scientists should be systematically included in these groups. Reviewer groups should be made aware of typical problems associated with peer review processes, such as a preference for the mainstream, processes of group dynamics, etc.

Summary of recommendations concerning the use of peer review:

1 – Justification for carrying out a work-intensive peer review process.

2 – Adherence to standards for such processes, particularly the careful selection of the reviewer group, ensuring a variety of different scientific perspectives and career stages.

3 – Reviewer group to be sufficiently informed about problems and possible risks of peer review processes.

II.3 Indicators of research performance

General recommendations

Quantitative indicators of research performance have gained enormously in importance over the last few years. Indicators that are based on external funding, and bibliometric indicators, are particularly important. For the development and use of indicators of research performance, the Council recommends:

The scientific communities should agree on and document quality standards within subjects so that these can inform the application, interpretation and further development of indicators. The concept of research activities should be defined as broadly as possible. In addition to knowledge and technology transfer activities it should also include services for research, such as services in connection with research infrastructures, |[32] edited works, committee activities and review work.

The Council points out that there is diminishing marginal utility as the quantity of research activity increases. It raises for consideration the possibility that after a certain point – which differs between subjects – a linear relationship between an increase in certain research indicators and the quality of research can no longer be expected. This needs to be taken into account in the assessment of research performance. It can be assumed in the case of

|[32] Wissenschaftsrat: Übergreifende Stellungnahme zu Informationsinfrastrukturen, in: Wissenschaftsrat: Empfehlungen zu Forschungsinfrastrukturen, Cologne 2011, pp. 5-70, here p. 58.

numbers of doctorates, for example, that as the quantity of doctorates being supervised increases, at a certain point the quality of the supervision starts declining. An increase in external funding also does not necessarily mean an increase in research quality.

Since indicators merely indicate that research activity is happening, without explaining what it is, interpretation by qualified persons is always required. Only then will indicators be a meaningful, widely accepted complement to qualitative information and subjective judgements. Hence supplementary information, especially methodological information, is required in addition to the indicators that are used. In addition to the reviewers, the audience or users of the assessment should also receive interpretation guidance.

At the individual level, attention should focus on qualitative information and content-related aspects (most important publications, research and teaching activities, knowledge and technology transfer, services for research) when evaluating and rewarding research performance – for example in connection with project reviews, in deciding to award additional pay under salary scale W, and particularly in the appointment procedure for professors. Decisions in the professorial appointment procedure should not be based solely or predominantly on bibliometric indicators (such as Hirsch indices). Assessment of individual research performance exclusively on the basis of quantitative indicators should be completely ruled out. In any kind of individual assessment, the conditions associated with different life phases and career stages should be given appropriate consideration.

Finally, when optimising the use of research indicators it should be taken into consideration that they conflict irreconcilably with data economy goals. With this in mind, limitation to a few robust indicators would seem to be necessary.

Recommendations concerning key indicators | [33]

_ Bibliometric indicators: these are perfectly appropriate for supporting informed peer review processes. Nevertheless, knowledge of methodology and critical reflection on the potentials and risks associated with bibliometrics are essential. The Council therefore recommends that when bibliometric indicators are used, reviewers' attention should be drawn to the methodology, or (at least selectively) persons with corresponding knowledge of the methodology should be involved in order to give greater attention to specific problems. The need to take subject-specific publication cultures, practices and strategies into account in the interpretation of bibliometric indicators can

| [33] Cf. Wissenschaftsrat: Pilotstudie Forschungsrating. Empfehlungen und Dokumentation, Cologne 2008.

only be satisfied by peers; hence these indicators should only ever be used in peer review processes.

  - Publication lists: publication lists provide information on the nature and content of publications and on the subject profile and publication strategy. They are a good indicator of well-founded qualitative judgements by peers. Publication lists can be restricted to a few selected publications.
  - Publication figures: this indicator mainly provides information about the quantity of publications and hence an indication of research productivity. As a performance indicator, publication figures are problematic as they create an incentive to publish more than is strictly necessary for the transmission of scientific knowledge ("salami tactics"). Depending on the publication practice in the subject, a distinction should be made according to publication types, for example; where possible the figures should be compared to the average for the subject. The indicator should be accompanied by other – preferably qualitative – indicators.
  - Citation figures: these provide information on the impact of a publication. Here too, it is necessary to compare the figures to the average for the subject. However, the indicator can be manipulated ("citation cartels") and says less about the quality of the cited publication than it does about the visibility and reputation of the authors.

- Amount of external funding: the external funding indicator (acquisitions or awards) provides information about the quality of applications and implicitly about the applicant's previous research performance, but not about the quality of the research currently being produced with the external funding. External funding should be weighted relative to the average for the subject to take account of the different acquisition possibilities and the subject-specific practice in respect of external funding. Moreover, external funding should be weighted based on the awarding practice: external funding which is acquired competitively in a qualitative appraisal process should be given a higher weighting than funding which is awarded in non-competitive procedures. Finally the Council points out that there is a need when assessing the level of external funding to take into account that after a certain point, which varies between subjects and depends on the task profile, there is no longer a linear relationship between quantity and quality.

- Lists of externally funded projects: lists of externally funded projects provide information about the type of projects carried out and the strategy for acquiring external funding. Such lists should be limited to the smallest possible number of selected projects. They can be incorporated into a qualitative opinion by peers and are a useful addition to quantitative indicators of the level of external funding.

_ Number of doctorates / postdoctoral degrees: these provide information about the extent to which young scientists are supported, not about the quality of support offered to young scientists. They are problematic as indicators for research performance because of possible goal displacement. Especially in automated processes (e.g. PBRA processes), it should therefore be taken into consideration that after a certain discipline-specific point no positive relationship can be expected between the number of doctorates and the quality of support provided. In peer review processes also, these figures should only be used as evidence of activities to support young scientists. |[34]

_ Research prizes, awards, keynote and plenary lectures: such lists provide information about the recognition and assessment of research performance by third parties and about the reputation and visibility of a scientist. It should not be ignored that the use of this indicator may create a preference for research into fashionable topics.

_ Number of patents and patent value: the number of patents indicator shows that innovative research with a view to applicability is taking place, but it should be weighed against information about the patent value and in itself does not say anything about the actual application and hence success of the patent. It is therefore best used in conjunction with information about the exploitation of the patents.

_ Exploitation of patents: income from patents is *one* indicator of the success of products developed on the basis of research and to this extent is therefore more likely to provide qualitative information about research performance. In view of the wide-ranging functions of patents, indicators based on them are not by themselves suitable for the assessment of the applicability of research.

Summary of recommendations concerning indicators of research performance:

1 – Agreement within the scientific community is required regarding quality standards as the basis for the development of indicators; a broad concept of performance should be used here which includes services for research.

2 – Consideration should be given to diminishing marginal utility in the assessment of quantitative indicators such as the level of external funding and number of doctorates.

|34 On this point, cf.: Anforderungen an die Qualitätssicherung der Promotion – Positionspapier des Wissenschaftsrates (Drs. 1704-11), Halle 2011.

3 – Indicators should be interpreted by qualified peers to ensure they are placed in a subject-specific context taking the specific function and methodological structure of the indicator into account. Support is to be provided for reviewer groups and for the target audience of the review in the form of interpretation guidance.

4 – Assessment of individual scientists based on quality-oriented methods.

II.4 Evaluation methods

Where there are regularly occurring institutional evaluations, regardless of whether this is in the non-university or university sector, these should, as a rule, only be conducted every five to ten years. When the intervals are decided, the term of appointment of management staff should be taken into account. Longer evaluation intervals will limit the workload for the persons and institutions being evaluated and conducting the evaluation. In addition, snapshots are avoided, medium-term trends become visible and changes which occur between two evaluations can be meaningfully assessed. The commissioners of evaluations (funding agencies in the German federal government and *Länder*, funding institutions) should adjust their evaluation practices accordingly, if this has not already happened.

For the sake of reducing the workload on the individuals and institutions concerned, authorities and institutions which commission and carry out evaluations have a responsibility to coordinate the diverse range of evaluations and their different cycles better than has so far been the case. This presents a particular challenge owing to the structural composition of the German scientific system (federal structure, diverse funding bodies).

Precisely because evaluations can be used to exercise a direct management effect and their results influence the development of institutions, policy programmes, and so forth, great importance has to be attached to the quality of the process and the reliability of the assessment that is carried out. |[35] Commissioners of evaluations should make criteria and methods transparent at an early stage and regularly subject them to critical examination, making adjustments where necessary, in respect of their decision usefulness, effectiveness, and unintended effects. At the level of university and institutional management, a high degree of management competence which also enables failures to be dealt with in a professional manner is required for the necessary

|[35] Cf. also the Council's "Grundsätze für erfolgreiche Evaluationen": Wissenschaftsrat: Aufgaben, Kriterien und Verfahren des Evaluationsausschusses des Wissenschaftsrates (Drs. 10296-10), Lübeck 2010, pp. 13-16.

implementation – which should happen as promptly as possible – of the courses of action which are derived from the evaluation results.

Evaluations should always be conducted in an informed peer review process based on a solid set of data, and include the reading of selected publications. In principle the recommendations made in section C.II.2 concerning the peer review also apply here. Just like for the reviewer group, the criteria for the evaluation should in each case be appropriate for the object of evaluation with its specific task profile, also aside from the core area of research, and therefore cannot be standardised. It is essential to have a balanced mix of quality and quantity-based indicators which, like the evaluation procedure, should be specified and published in advance. At the beginning of the process, the reviewers must be informed about the evaluation criteria and the function of the indicators that are used. As far as possible, all parties concerned should be given the opportunity to participate in the process. The results of the evaluation and the resulting courses of action should be published.

Summary of recommendations concerning evaluation processes:

1 – Institutional evaluations should be conducted at intervals of five to ten years.

2 – Improved coordination of timing when carrying out evaluation processes.

3 – Evaluations have consequences as an instrument of direct management; regular process review is required with regard to intended and unintended effects.

4 – Evaluations always to be conducted as an informed peer review. Adaptation of the assessment process (reviewer group, assessment criteria etc.) to the respective object of evaluation. Guarantee greatest possible process transparency and participation opportunities as basis for acceptance.

II.5 Ratings/rankings

Rankings can only begin to have a positive competitive effect through transparency and comparability if they meet certain requirements. Most well-known rankings are still some way from meeting these requirements. Hence on this subject the Council confirms the central recommendations which it made previously in the course of developing and testing its research rating: |[36]

[36] Cf. Wissenschaftsrat: Empfehlungen zu Rankings im Wissenschaftssystem – Teil 1: Forschung, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 2004, vol. 1, Cologne 2004, pp. 159-220; Wissenschaftsrat: Pilotstudie Forschungsrating. Empfehlungen und Dokumentation, Cologne 2008.

_ The various specialist groups should be involved in designing and implementing the processes as peers.
_ Assessment should always be in the form of an informed peer review and multidimensional, based on a comprehensive concept of research performance; methodological standards should be upheld and the data basis should be disclosed.

Following completion of the pilot phase of its rating – which is scheduled for 2013 – the Council will comment on rating processes and issue further recommendations.

II.6 Resource allocation methods

Continuity-oriented university funding

Universities should be funded so that they are able continuously to carry out good research and teaching, provide effective support to young scientists, and where applicable perform other core tasks effectively such as transfer and infrastructure activities and services to science, independently of particular projects and programmes. To this end, the Council recommends the continued development of existing funding modes. In the Council's opinion, a return to the *status quo ante* of input-based allocation of resources would be neither beneficial nor desirable. Instead, the funding system should:

_ be devised according to principles of transparency, equality and comprehensibility;

_ take into consideration universities' workload and needs in respect of performing their core tasks; and

_ meet universities' infrastructure requirements not solely according to student demand but also taking into account the principle of continuously safeguarding good research within universities.

If indicator-based resource allocation mechanisms are used for universities' core funding, the underlying science policy goals should be clearly stated and communicated. Indicators which are not performance-based should not be listed as such so as not to awaken the impression that merely ensuring the fulfilment of regular tasks in research and teaching is rewarded as a special achievement. To prevent undesirable effects, diminishing marginal utility should be taken into account with key indicators (such as the number of students, doctorates, and the level of external funding) (cf. C.II.3).

Furthermore, the Council recommends to the German federal government and *Länder* that in the medium term, externally funded projects should be funded in a way which takes account of the actual costs, i.e. the indirect costs as well. |[37] Since, under the existing practice, externally funded projects require partial co-financing from core funding (particularly infrastructure and administration), universities will not have the freedom they need for their strategic development until funding changes occur to cover such costs. At the same time, however, it must be ensured that this project funding which also covers indirect costs does not lead to a reduction in core funding.

Within universities also, enabling the continuous fulfilment of core tasks should determine the mode of resource allocation. At this level, in addition, there should be subject-specific differentiation in the allocation of resources. Particularly due to the requirement for subject-specific differentiation, it is essential for the further development of resource allocation methods within universities for there to be new dialog forums between the different levels that are involved or affected – university management, faculties and departments, dean's offices, and individual scientists.

Additional resources for strategic management

Aside from core funding, additional resources should be made available for furthering particular strategic goals both at the level of the *Länder* and at the level of the universities and faculties. As is already the practice in some *Länder*, these resources could be allocated by means of target agreements or on an application basis as part of targeted funding programmes. Additional resources of this kind, in contrast to core funding, could by all means be allocated on a discretionary basis, but not arbitrarily.

The goal of the incentive which is created in connection with these additional resources does not necessarily have to be to promote top-level research: while the *Länder* can give awards to their universities, for example for success in particular strategically determined areas, internally the universities can also set other management goals – such as profile enhancement or particularly unconventional, risky, innovative research – and further these goals by offering awards and risk capital. As a result of these additional resources, the faculties are also able to reward or specifically encourage particular activities.

|[37] On this point, cf. also Wissenschaftsrat: Empfehlungen zur künftigen Rolle der Universitäten, Cologne 2006, p. 55.

Based on the overarching research policy goals which are agreed with the funding agencies, the non-university research institutions should continue to be assured of a stable financial outlook and improved conditions enabling them to continue to fulfil their respective mission-specific tasks.

Assessment competence for salary scale W

The entire salary scale W system is currently undergoing dynamic development and legal review. Nevertheless, the Council already identifies a fundamental need in the universities to build up and develop internal assessment competence, which is essential for decisions with regard to evaluating performance and awarding special payments. In the assessment of individual achievements, the exclusive use of quantitative indicators should be avoided (cf. C.II.3). Instead, instruments such as qualitative self-evaluations should be used more than at present. It should be considered whether an assessment at greater intervals of time would also be appropriate in connection with salary scale W. The Council reserves the right to issue a more detailed statement concerning salary scale W at a suitable point in time.

In general terms, the Council points out that additional pay for special achievements in research cannot subsequently compensate for wrong decisions in the appointment procedure.

Performance incentives via external funding

External funding performs a key function in respect of competitive incentives and quality assurance in research, in both the university and non-university sectors. In terms of its significance for research funding, it should be reduced to a supplementary function again; it certainly should not have the function of supporting the core tasks of universities. The described negative consequences when research funding is too heavily reliant on external funds should be limited.

It is vital that the wide variety of external funding sources which exist in Germany should be preserved; this also includes special research funding from the *Länder* which is geared towards top-level research and allocated via an application procedure. A wide range of methodological and thematic approaches should continue to be supported via external funding in future also. Variable project durations should also be possible to a greater extent than at present, depending on needs.

Summary of recommendations concerning resource allocation methods:

1 – Core funding for universities to continuously facilitate good research, teaching, the development of new scientists, and infrastructure and other services where applicable.

2 – Determine infrastructure requirements according to universities' core tasks – also beyond the demand from students.

3 – Further development of indicator-based funding into a transparent and comprehensible system; avoid "false labelling" as performance orientation; introduction of corridors for the most important indicators.

4 – Externally funded projects which are supported by the German federal government / *Länder* to be financed according to the actual costs, i.e. including indirect costs.

5 – Provision of resources to support strategic goals, both at the level of the *Land* and within universities.

6 – Build up internal assessment competence in universities for salary scale W; avoid using purely quantitative indicators for the assessment of individuals.

7 – Maintenance and continued expansion of the wide variety of external funding sources by the German federal government and *Länder* as a competitive performance incentive.

II.7 The documentation system

Two kinds of measures can help reduce the workload associated with methods of assessing and managing research performance: firstly a reduction in the number of evaluations and similar processes, and secondly improvement and increased efficiency in the continuous documentation of research performance, which the Council therefore urgently recommends. In addition, such documentation provides an essential basis for institutionally safeguarding the scientific system's self-reflexivity and ability to learn.

If documentation of research activities is regular, comprehensive and at the same time easy to manage, it will pay dividends to all relevant stakeholders: scientists can fulfil reporting requirements on the part of institutional management without repeated data requests constantly generating a high workload. At the same time, as members of specialist groups, they gain a comprehensive picture of their own research activities. Scientific institutions are able to monitor their activities and take strategic decisions on an informed basis, where applicable also via comparative benchmarking, with regard to their continued development and profile enhancement. Finally, science policy makers receive detailed information which enables them to take decisions

concerning resource flows and science policy priorities, and gain insights into the effects of their measures. Not least, a comprehensive data basis allows science research to make more reliable statements concerning intended and unintended effects of performance assessment and management methods. Overall, the Council considers that there is still a considerable need for improvement in order to implement a good documentation system – also as the basis for a functioning monitoring system for the processes that are introduced.

The Council recommends that an understanding should be reached quickly – if possible between the *Länder* – concerning minimum standards for data collection along with harmonisation of the reporting systems at universities and non-university institutions, which in many cases are still in the development phase. A proposal for the standardisation of data for research assessment, which is currently being worked on by the Council's research rating steering group, should serve as a starting point for further efforts. |[38]

To ensure that improvements in reporting systems do not come at the expense of researchers, support from science administration is absolutely essential within the institutions and at the political level. In close consultation with scientists, science administrators should develop and maintain the reporting systems, be able to supply information for internal and external enquiries, and support the implementation of evaluation results. The direct and indirect costs for setting up and ongoing maintenance of the documentation systems will be compensated by the reduced data supply costs.

Summary of recommendations for the documentation system:

1 – Agreement on minimum standards for data collection and harmonisation of existing reporting systems and those under development at universities and non-university facilities.

2 – Reduce the workload for individual scientists relating to data storage, provision and maintenance via a service-oriented administration with close links to the sciences.

| 38 The publication of corresponding recommendations can be expected in early 2012.