

Leibniz leadership lecture 2023

Documentation

"Reform of the scientific performance evaluation"

October 6/7, 2023, Head Office of the Leibniz Association, Berlin

Participants with introductory talks

<i>Mine</i>	<i>Altinli</i>	<i>Bernhard Nocht Institute for Tropical Medicine (BNITM)</i>
<i>Matthias</i>	<i>Beller</i>	<i>Leibniz Institute for Catalysis (LIKAT)</i>
<i>Sophie</i>	<i>Biesenbender</i>	<i>Commission for Research Information in Germany (KFiD)</i>
<i>Luc</i>	<i>De Meester</i>	<i>Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)</i>
<i>Jean-Emmanuel</i>	<i>Faure</i>	<i>European Commission, DG Research and Innovation</i>
<i>Gregor</i>	<i>Kalinkat</i>	<i>Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)</i>
<i>Matthias</i>	<i>Kiesselbach</i>	<i>DFG Head Office</i>
<i>Stefan</i>	<i>Kuhlmann</i>	<i>University of Twente, Netherlands</i>
<i>Anne</i>	<i>Krüger</i>	<i>Weizenbaum Institute</i>
<i>Ulf</i>	<i>Müller-Ladner</i>	<i>Justus Liebig University Giessen</i>
<i>Isabella</i>	<i>Peters</i>	<i>ZBW - Leibniz Information Center for Economics</i>
<i>Iris</i>	<i>Pigeot</i>	<i>Leibniz Institute for Prevention Research and Epidemiology - BIPS</i>

Participants without introductory talks

<i>Artemis</i>	<i>Alexiadou</i>	<i>Leibniz-Center General Linguistics (ZAS)</i>
<i>Rainer</i>	<i>Danielzyk</i>	<i>ARL - Academy for Territorial Development in the Leibniz Association</i>
<i>Frank</i>	<i>Ewert</i>	<i>Leibniz Center for Agricultural Landscape Research (ZALF)</i>
<i>Florian</i>	<i>Heider</i>	<i>Leibniz Institute for Financial Research SAFE</i>
<i>Katharina</i>	<i>Helming</i>	<i>Leibniz Center for Agricultural Landscape Research (ZALF)</i>
<i>Michael</i>	<i>Hintermüller</i>	<i>Weierstrass Institute for Applied Analysis and Stochastics (WIAS)</i>
<i>Alice</i>	<i>Hohn</i>	<i>Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB)</i>
<i>Oliver</i>	<i>Ibert</i>	<i>Leibniz Institute for Research on Society and Space (IRS)</i>
<i>Sunhild</i>	<i>Kleingärtner</i>	<i>Deutsches Bergbau-Museum Bochum – Leibniz Research Museum for Geo-resources (DBM)</i>
<i>Klaus</i>	<i>Lieb</i>	<i>Leibniz Institute for Resilience Research (LIR)</i>
<i>Jürgen</i>	<i>May</i>	<i>Bernhard Nocht Institute for Tropical Medicine (BNITM)</i>
<i>Jörg</i>	<i>Overmann</i>	<i>Leibniz-Institute DSMZ–German Collection of Microorganisms and Cell Cultures</i>
<i>Kathrin</i>	<i>Rosing</i>	<i>Leibniz-Institut für Wissensmedien (IWM)</i>
<i>Anna-Maria</i>	<i>Schielicke</i>	<i>Leibniz Institute of Ecological Urban and Regional Development (IOER)</i>
<i>Raimund</i>	<i>Seidel</i>	<i>Leibniz Center for Informatics (LZI)</i>
<i>Heike</i>	<i>Solga</i>	<i>WZB Berlin Social Science Center</i>
<i>Veronika</i>	<i>Somoza</i>	<i>Leibniz Institute for Food Systems Biology (LSB)</i>
<i>Silke</i>	<i>Voigt-Heucke</i>	<i>Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science (MfN)</i>
<i>Christof</i>	<i>Wolf</i>	<i>GESIS – Leibniz Institute for the Social Sciences</i>

Moderation

<i>Bettina</i>	<i>Böhm</i>	<i>Secretary General of the Leibniz Association</i>
<i>Claudia</i>	<i>Labisch</i>	<i>Head of EU Office</i>
<i>Lisa</i>	<i>Kressin</i>	<i>Division Scientific Affairs</i>

Headquarters of the Leibniz Association

<i>Karin</i>	<i>Effertz</i>	<i>Head of Division - Leibniz Competition Procedure</i>
<i>Anja</i>	<i>Görnitz</i>	<i>Deputy Head of Division - Leibniz Competition Procedure</i>
<i>Peter</i>	<i>Heil</i>	<i>Head of Division Evaluation</i>
<i>Britta</i>	<i>Horstmann</i>	<i>Head of Division - Scientific Affairs</i>
<i>Maria</i>	<i>Migawa</i>	<i>Division Integrity, Committees and GWK</i>
<i>Gabriel</i>	<i>Sollberger</i>	<i>Division Leibniz Competition Procedure</i>
<i>Veronica</i>	<i>Thanner</i>	<i>Head of President's Office</i>
<i>Sylvaine</i>	<i>from Franqué</i>	<i>Administration</i>
<i>Christine</i>	<i>Wennrich</i>	<i>Head of Division Leibniz Transfer</i>

Program

Friday, October 6, 2023

13:00 **Lunch**

14:00 **Greeting**

Bettina Böhm

14:15 **External perspective: Scientific performance assessment in research funding**

Matthias Kiesselbach (DFG Head Office): On the Reform of Research Assessment: Current Initiatives and Perspectives in DFG and CoARA

Jean-Emmanuel Faure (EU Commission): Reforming research assessment: rationale, progress made and role of the European Commission

Discussion

15:45 **Internal perspective I: Procedures for scientific performance assessment in the Leibniz Association**

Matthias Beller (LIKAT, SAW) Research evaluation in internal organizational competition

Ulf Müller-Ladner (Justus-Liebig University Giessen, SAE): On the evaluation procedure of the Senate of the Leibniz Association

Discussion

16:45 **Break**

17:15 **Internal perspective II: Impulses for change from the Leibniz Association**

Leibniz-StG "Scientific Publishing": Iris Pigeot (BIPS), Luc De Meester (IGB): Suggestions on the principles of publication-based research assessment within and outside of Leibniz

Mine Altinli (BNITM), Christian Nehls (FZB), Gregor Kalinkat (IGB): Exploring Challenges in Science Evaluation: Perspective from Early Career Researchers

Discussion in groups

19:30 **Dinner**

Pascarella

Chausseestraße 30, 10115 Berlin

Sa-
turday, **October 7, 2023**

9:00 **Reflection on the previous day**

All participants

10:00 **International science or national science systems? Performance as-
sessment in international comparison**

Stefan Kuhlmann (Prof. Emeritus, University of Twente): Recognition and rewards systems for academics: current transformations in the Netherlands

Isabella Peters (ZBW, G6 Task Force on Research Assessment): National perspectives on research evaluation

Discussion

11:00 **Break with photo opportunity**

11:30 **How we want to evaluate: (Digital) structures and basics of perfor-
mance recording**

**Sophie Biesenbender (Commission for Research Information in Ger-
many):** The KDSF as a science-led standard for the collection and use of research information

Anne K. Krüger (Weizenbaum Institute): Commercial and open infra-
structures and their consequences for scientific performance measure-
ment

Discussion

12:30 **Conclusion and outlook**

13:00 **Lunch**


External perspective: Scientific performance assessment in research funding

Matthias Kiesselbach (DFG Head Office): *On the Reform of Research Assessment: Current Initiatives and Perspectives in DFG and CoARA*

The DFG is actively involved in CoARA's international reform process, both because it agrees with CoARA's goals and because it wants to help shape the reform process. The DFG supports the main goals of CoARA - the strengthening of qualitative review, the rejection of the inappropriate and unreflected use of quantitative proxies in research evaluation, especially in project proposals, and the consideration of a broader range of scientifically valuable contributions and practices in the evaluation of researchers. With regard to its own funding activities, the DFG believes it is already on the right track. Even before its involvement in CoARA, the DFG had already undertaken reform efforts. For example, the number of own publications that can be listed in DFG proposals has long been limited, and reviewers have been asked to give more weight to the quality of preliminary work than to the quantity of publications. More recently, CV forms with optional narrative elements and the possibility to cite previous work beyond traditional journal articles have been introduced. In addition, applicants are now expected not only to mention previous work, but also to relate it to the content of the current proposal. Overall, it is an important concern of the DFG that the evaluation of research is scientifically appropriate and does not create scientifically problematic incentives. Critics of the reform process see a danger in the diversification of the criteria. They fear that new criteria could override the existing primary criteria in the evaluation and thus dilute the understanding of excellence, or that additional criteria could lead to additional burdens. The DFG does not share these concerns. Instead of diluting the definition of excellence, CoARA is interested in identifying excellent research more precisely through qualitative assessment. Furthermore, some German science stakeholders fear a loss of autonomy for science through political intervention. The DFG does not see this danger either. The EU Commission, for example, is only one of many members of CoARA and cooperates with the other participating organizations on an equal footing.

Jean-Emmanuel Faure (EU Commission): *Reforming research assessment: rationale, progress made and role of the European Commission*

Making the research system attractive to the younger generation requires a fundamental cultural and systemic change. To achieve this, all stakeholders - researchers, universities, policy makers, etc. - must be brought on board. Since the adoption of the DORA Declaration in 2012, progress has been relatively slow. However, the work of the former Open Science Policy Platform, initiated with the support of the EU Member States, shows the importance of reforming research evaluation in order to generalize the practice of open science. More than 600 research organizations have agreed on principles for improved research assessment and a common vision. The signatories have formed the CoARA Coalition. CoARA should act as a learning platform with the aim of collaborating, identifying best practices and respecting the autonomy of research organizations. A critical mass of research organizations is important for this.



Regarding the relationship between qualitative and quantitative assessment, it is important to find the right balance. The aim of reform should not be to exclude metrics, but to use them responsibly. The risk of subjective bias should not be ignored. With regard to the evaluation burden on scientists, new selection procedures or a broader distribution, e.g. including young scientists, could provide relief. The development of the importance of artificial intelligence in scientific application and evaluation procedures should be taken into account. There is no EU position on this yet, but the EU Commission has started some work on this.

Internal perspective I: Procedures for scientific performance assessment in the Leibniz Association

Matthias Beller (LIKAT, SAW): *Research evaluation in internal organizational competition*

Three theses on research assessment in general and on the Leibniz competition in particular are formulated: 1) performance assessment is a central necessity in science, 2) it differs depending on the object of assessment, and 3) quantitative assessment is not per se inferior to qualitative assessment - ideally, they would complement each other.

The SAW process already takes into account a variety of possible performance criteria, including differences between programs. There are no decisions based solely on quantitative or qualitative criteria. For example, personality plays an important role in the Professorial Fellowship Program and the Junior Research Group funding line. There can be no one hundred percent objective assessment. For all, but especially for rejected applicants, the procedure must be presented with the greatest possible transparency in order to create trust. The strong emphasis on quantitative indicators, such as the h-index, is more pronounced at universities than at non-university research institutions. It could be problematic if these standards were applied externally. Overall, evaluation in the Leibniz competition is a dynamic process that needs to be adapted regularly.

Ulf Müller-Ladner (Justus-Liebig University Giessen, SAE): *On the evaluation procedure of the Senate of the Leibniz Association*

The evaluation assesses the Leibniz institutions with regard to five central "subject areas". The focus is on the overall concept of an institute and the services based on it. In the overall concept, the institutions weight the importance of the three central types of tasks: research infrastructures, research, and knowledge transfer. The internal logic and consistency of this weighting is then assessed.

Institutions were asked to describe their performance in all three areas. While the qualitative and quantitative information on research is generally well structured and therefore assessable, the information on research infrastructure and transfer services could often be improved. To facilitate the qualitative assessment, the SAE and the Senate have asked the institutes to provide "highlights" of their achievements over the past five years. This approach has proven successful. However, it remains a challenge for the Leibniz Association to define standards and new performance parameters for FI and transfer services across all institutes.

In the ensuing discussion, the diversity and interdisciplinarity of the assessment groups was addressed. It was pointed out that the members of the SAE, who were responsible for the composition of the evaluation groups and relied, among other things, on suggestions from the institutes, sought to achieve the greatest possible diversity of personnel. In the 7-year cycle of evaluations of all institutes (2016-2023), 40% of the experts worked abroad and 40% were also women. The disciplinary composition corresponds to the range of expertise of each Institute. The academic age of the experts is not recorded, but experts with management experience tend to be more involved.

Internal perspective II: Impulses for change from the Leibniz Association

Iris Pigeot (BIPS), Luc De Meester (IGB) (Leibniz-Steering Group "Scientific Publishing"): *Suggestions on the principles of publication-based research assessment within and outside of Leibniz*

Two principles of scientific performance assessment seem to be central and consensual: 1) a balanced use of quantitative and qualitative indicators, and 2) the recognition of a variety of scientific activities as performance.

The discussion made it clear that both blind reliance on metrics and blind reliance on narrative should be avoided. Finding the right balance remains a major challenge, e.g. with respect to the demand for the most objective procedures possible, scarce time resources, and diverse career paths. It was agreed that the quality of publications should be given priority over their quantity. A pre-defined checklist of questions to be asked could provide guidance on the content of the qualitative assessment. It is not easy to adequately consider the quality of journals if one does not want to blindly follow the impact factors of commercial journals. Subject-specific or institution-specific positive lists of high-quality journals within a discipline could provide support, but should not replace an assessment of the quality of each individual publication. In addition to experience as readers and publishers, experience as editors and reviewers could also be taken into account, with access figures, for example, being an indicator of the quality of peer-reviewed journals. An alternative to positive lists could be a graduated categorization of journals (i.e. different lists with a certain quality differentiation). Such positive lists of journals could be created or approved with the support of the scientific advisory board of the respective institute and would have to be kept up to date. However, such positive lists are difficult to establish for highly interdisciplinary institutes. Any reform should take into account that the Leibniz Association must remain in line with international practice, as it cannot function as an island.

With regard to the recognition of a variety of scientific activities and not only high-ranking scientific publications as an achievement, it is important to note that scientists cannot be expected to be able to cover the entire range of scientific achievements. In addition, scientific publications should be given special weight in the future.

Within Leibniz, a group could come together to formulate a consensus list of "possible scientific achievements" with illustrative examples. Such a list could be continuously expanded as a "living document".

Ultimately, however, it is the responsibility of the institute's leadership to promote the cultural change associated with the points mentioned above.

Mine Altinli (BNITM), Christian Nehls (FZB), Gregor Kalinkat (IGB): *Exploring Challenges in Science Evaluation: Perspective from Early Career Researchers*

For many Early Career Researchers (ECRs), the performance expectations they have to meet are partly based on anachronistic notions of science and the person doing it. On the one hand, this contradiction with the actual practice of ECRs creates questionable incentives, e.g. a focus on individual rather than collaborative performance, and on the other hand, the actual achievements of ECRs, e.g. in the context of doctoral supervision, are not sufficiently recognized. In addition, there is a lack of transparency and consistency regarding the performance required to remain successful in academia. Against the backdrop of the current debate on the German law on fixed-term contracts for academics (WissZeitVG), a reform of performance evaluation must also take these aspects into account. In addition to the changing requirements for ECR, the top management level should also be evaluated more strongly with regard to leadership and mentoring skills.

The ensuing discussion highlighted the different expectations of academics, who are expected to be both exceptionally creative and willing to take on a heavy workload. ECRs must be given the opportunity to develop and demonstrate their independence. However, this requires appropriate relief from additional tasks and the opportunity to access budget resources beyond project funding. Moreover, taking greater account of the diversity of people working in science and multiplying the services taken into consideration cannot solve the fundamental problem of the few permanent positions in the system. Management services in the area of mentoring could therefore also include realistic advice on alternative career paths. In the evaluation of the Leibniz Institutes, the ECRs have so far had the opportunity to comment on the situation at their institutions in the "headless discussion". This could be expanded to include the possibility of anonymous feedback.

International science or national science systems? Performance assessment in international comparison

Stefan Kuhlmann (Prof. Emeritus, University of Twente): *Recognition and rewards systems for academics: current transformations in the Netherlands*

The Dutch science system and the governance of its universities had been shaped by New Public Management (NPM) for ten years. As a result, universities were treated like businesses, and performance evaluation, for example by the Netherlands Research Council (NWO), was one-sided and focused on quantitative indicators. A few funding programs became a bottleneck in academic careers. For example, it is almost impossible to become a professor in the Netherlands without funding from the "Veni, Vidi, Vici" program. Criticism of this system culminated in a protest by students of the Young Academy (Royal Dutch Academy of Sciences) in 2015. It focuses on competition rather than collaboration and is therefore detrimental to the quality of research. It also fails to recognize the diversity of performance types and career paths. In connection with the implementation of Open Science, there has been a reorientation of performance assessment at Dutch universities in recent years. With the concept of "Room for everyone's talent", the major scientific organizations have developed a new guiding principle for performance assessment in science, which focuses equally on different dimensions (e.g. impact of research, teaching, ability to collaborate). Traditional research outputs remain important (including publications and citations), but the focus on quantitative metrics has diminished. In contrast, the narrative of impact has become important. However, this reorientation has also brought with it problems and criticism. For example, there is a danger that excellence is now expected in all dimensions of performance. It is also necessary to remain compatible with international standards. In the natural sciences in particular, change has been slow.

Isabella Peters (ZBW, G6 Task Force on Research Assessment): *National perspectives on research evaluation*

A comparative look at performance assessment in other European countries is worthwhile for a better classification of academic performance assessment in Germany and for an informed participation in the European CoARA process. For example, there are major differences between Spain, France, Italy and Germany with regard to the criteria, procedures and responsible actors in performance assessment as well as to the implementation of career support. Thus, on the one hand, the starting conditions for participation in CoARA are differing, and on the other hand, this makes it clear that the aim of this process can be increased coordination, but not a "one size fits all" approach. The motivation to participate in CoARA also varies with the degree of freedom of science in a country. The more the state currently "reigns in", the greater the hope of being able to act in a more science-oriented manner through CoARA. In Germany, on the other hand, the primacy of self-government and the freedom of research and teaching make political influence a greater concern.

How we want to evaluate: (Digital) structures and basics of performance recording

Sophie Biesenbender (Commission for Research Information in Germany): *The KDSF as a science-led standard for the collection and use of research information*

The discussion on the understanding of scientific performance and indicators should also take into account the processes, systems and digital infrastructures for collecting and processing research information. In many contexts, research information (such as information on publications, research data or patents as outputs of scientific performance processes) forms the basis for evaluation and assessment procedures in research. The collection and processing of research information on the basis of definitions is a prerequisite for the transparent, fair and appropriate use or non-use of this information in the context of performance evaluation.


Conscious and purposeful management of research information is therefore the basis for research evaluation and a strategic issue for every research institution.

A standard for research information and its processing developed by the scientific community is the KDSF standard ("Core Data Set for Research"), which takes into account different information requirements and use cases. In the context of the reform process, the KDSF supports responsible research evaluation by allowing for the consideration of diverse research activities and results as well as subject and structural specificities and by providing a transparent basis for the use of quantitative research information.

The KDSF thus provides an impetus for the efficient, professional and responsible handling of research information, the minimization of dependencies on third parties, and the strengthening of data competence and data sovereignty in the context of digital transformation processes in the scientific system.

Anne K. Krüger (Weizenbaum Institute): *Commercial and open infrastructures and their impact on scientific performance measurement*

Digital infrastructures for research information and evaluation can be used to collect a variety of new data on scientific practice and performance. However, there are a number of problems associated with this: First, these infrastructures are distributed by commercial vendors, which could lead to new financial dependencies in the use of these databases and software products. Second, there is the question of the quality of the data used for information and evaluation, as well as their appropriate use. Third, the evaluation options are primarily tailored to the practice of the natural sciences, which does not do justice to other subjects in the collection and presentation of performance. Fourth, the possibilities for evaluating the societal impact of research are constantly increasing, which increases its relevance compared to, for example, basic research. Fifth, it is becoming clear that this not only changes the evaluation criteria, but also who - beyond scientific peers - can independently generate information and thus evaluate research. It also strengthens quantitative performance measurement, including considerations of predicting research success



("predictive analytics"). It becomes clear that the digital infrastructures of research information and evaluation play a central role in who can evaluate science and how. This needs to be critically reflected upon.

Conclusion and outlook

This leadership lecture should be seen as the beginning of the internal Leibniz exchange on the reform of research assessment. It was also intended to provide an initial overview of possible topics that would be explored in greater depth in various exchange formats over the coming years. The aim was not to produce new guidelines or manifestos, but rather structured discussions that would advance Leibniz's internal evaluation procedures and contribute to a common position - with feedback to the CoARA process.

The Leibniz Strategy Forum on Research Evaluation will be an important forum for this in the future.

Group photo

