EUROPEAN
SCIENCE
FOUNDATION

Evaluation in Research and Research Funding Organisations: European Practices

A report by the ESF Member Organisation Forum on Evaluation of Publicly Funded Research

European Science Foundation (ESF)

The European Science Foundation (ESF) is an independent, non-governmental organisation, the members of which are 72 national funding agencies, research performing agencies and academies from 30 countries.

The strength of ESF lies in its influential membership and in its ability to bring together the different domains of European science in order to meet the challenges of the future.

Since its establishment in 1974, ESF, which has its headquarters in Strasbourg with offices in Brussels and Ostend, has assembled a host of organisations that span all disciplines of science, to create a common platform for cross-border cooperation in Europe.

ESF is dedicated to promoting collaboration in scientific research and in funding of research and science policy across Europe. Through its activities and instruments, ESF has made major contributions to science in a global context. ESF covers the following scientific domains:

• Humanities
• Life, Earth and Environmental Sciences
• Medical Sciences
• Physical and Engineering Sciences
• Social Sciences
• Marine Sciences
• Materials Science and Engineering
• Nuclear Physics
• Polar Sciences
• Radio Astronomy
• Space Sciences

www.esf.org

Member Organisation Fora

An ESF Member Organisation Forum is an output-oriented, issue-related venue for the Member Organisations, involving other organisations as appropriate, to exchange information and experiences and develop joint actions in science policy. Typical subjects areas discussed in the Fora are related to:

• Joint strategy development and strategic cooperation with regard to research issues of a European nature.
• Development of best practices and exchange of practices on science management, to benefit all European organisations and especially newly established research organisations.
• Harmonisation of coordination by MOs of national programmes and policies in a European context.

www.esf.org/mo-fora

Contents

1.
Introduction

● ● ●

Funding agencies and research organisations today need to demonstrate that their funding goes to excellent scientists who produce excellent results. Ex-post evaluation is one means to approach this challenge.

The evaluation of research programmes and funding schemes is a relatively new field. However, it has gained increasing attention within a short time and is rapidly developing. European research organisations are looking for more efficient, robust and meaningful approaches to find out which conditions and funding procedures are conducive to good research, to account for the public money spent, and to show the scientific as well as economic and societal impact delivered from these investments.

The success of evaluation depends on the acceptance of the results among decision-makers and research communities being evaluated. Therefore, evaluations must meet the highest standards. Funding and research performing organisations are striving to improve measures of research progress, productivity and quality, and test new organisational set ups and methods to this effect. The ESF and EUROHORCs Road Map states: "Research evaluation is a small, highly specialised field, and on a national level, there are only a limited number of actors involved in it. Therefore, the exchange of knowledge and sharing of experiences of evaluation strategies on an international level is especially valuable and is the prerequisite of benchmarking purposes."

The Forum on Evaluation of Publicly Funded Research launched in 2010 addresses Action number 6 of the 'EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions' to "develop common approaches to ex-post evaluation". While there are good reasons for variation in the evaluation approaches that national organisations take, the Forum defined some basic understandings and mechanisms that all European research and research funding organisations can agree on. This is the bottom line for any joint activity.

The aim of this report is to give insight into the practices of ex-post evaluation from the perspective of European research and funding organisations and to explore the opportunities and challenges of evaluation. It cannot substitute for 'toolbox' documents, which focus on methodologies and practical issues to be approached before commencing an evaluation project. There are plenty of excellent guides to this[1]. Instead, the report aims to be helpful in developing a research organisation's evaluation policy and therefore the organisation's strategy itself. The recommendations draw on the experience of a wide spectrum of research organisations – funding and research performing – in Europe: small and big, with long or brief tradition, from diverse geographical locations and with different missions. It gives an overview of evaluation principles accepted among the ESF Member Organisations and includes examples and good practices from those organisations.

The report builds on the work of three working groups set up within the Forum:

1. **Quality assurance and evaluation guidelines** (chaired by Gro Helgesen/Research Council of Norway)
2. **Impact assessment on science and society** (chaired by Per Janson/Swedish Research Council)
3. **Classification systems and categorisation of output data** (chaired by Ian Viney/UK Medical Research Council)

---

1. European Foundation Centre 2010, EU Commission 2004, and many others.

The Forum's aim was to improve evaluation studies of funding schemes by developing guidelines, to learn about best practices of impact assessment and to identify challenges in conducting transnational comparative research portfolio evaluation.

Each of the working groups has published a report about its results. The report at hand is a distillation of these results. Beyond that it addresses overarching issues that have emerged during the discussions of the Forum.

Rapid advancements in the field of research evaluation are reflected in a number of actors and networks in the field. They include the Nordic Network on Research Evaluation, the G8 HORCS Working Group on Research Assessment, the RTD Evaluation Network of the European Commission, the European Foundation Centre and many others. The Forum has established connections with these, profited from joint discussions, and has acted as the voice of national funding agencies and research performing organisations and their diverse experiences and interests.

Data-based assessments of the capacity of national and supranational systems will shape future thinking about research investments. It is therefore vital to continue the discussion on how to improve those evaluations and the data collection procedures that go with them. The interaction that has been established among national research funding agencies and national and European partner institutions in this Forum will help bring this discussion to the next level.

How to use this guide

This report explains why and in which cases organisations make use of ex-post evaluation (chapter 2). It shows some of the recent developments in the field (chapter 3) and describes which methods are used and which challenges arise (chapter 4). Furthermore, the specifics of diverse European national set-ups for research evaluation are considered (chapter 5). The report concludes with recommendations and policy advice. The ambition is to contribute to the improvement of evaluation strategies and studies.

2.
Why Research Evaluation?
● ● ●

Why is there research evaluation? Research and research funding organisations need evaluation to[2]:

- provide an evidence base for strategy development,
- document funding practices and thereby establish transparency about taxpayers' money,
- decide on the allocation of resources,
- support internal processes for learning about the research system and funding activities which may result in the adaptation of funding programmes or research fields,
- demonstrate that research performing and research funding organisations are accountable and are concerned with quality assurance,
- sharpen concepts: for example what is understood by internationalisation[3], interdisciplinarity or impact of science[4],
- establish a direct channel of communication with stakeholders, to communicate the impact and results of research funding to government or to allow grantees (scientists) to articulate their opinions about the funding system, application procedures and research conditions (for example during site visits, interviews, surveys).

Ex-post evaluation of research helps to identify conditions that are conducive to science and to assess the contribution of funding organisations. The insights, sharpened definitions and bigger empirical bases, all gained by evaluation, build up and feed into later studies and into the research funding process itself.

The added value and potential learning effect from evaluation is therefore huge. To profit from these insights and information, organisations have established different ways to feed the results into funding schemes and policies[5].

Ex-post evaluation of research can have different objects. In the 2009 ESF report 'Evaluation in National Research Funding Organisations' five levels of evaluation are explored in detail. These are:

1. the evaluation of the research funding agency,
2. the evaluation of funding policies or particular strategic issues,
3. the evaluation of research fields or scientific disciplines,
4. the evaluation of funding schemes and
5. the evaluation of research grants.

---

2. As Hanne Foss Hansen points out (2009), these goals of evaluation can sometimes be ambiguous.
3. See also the report of the ESF MO Forum on Indicators of Internationalisation (ESF 2012a).
4. Some examples are laid out in detail in the Report of Evaluation Forum WG 3-Output Data: 'Research Funders and Research Output Collection' (ESF 2012d).

---

5. See the report of Evaluation Forum WG 1-Guidelines: 'A Guide to Evaluation Activities in Funding Agencies' (ESF 2012b).

3.
Developments in Research Evaluation

● ● ●

Research evaluation has rapidly developed within a short period of time. It has grown in importance with the advancement of New Public Management[6] when policy-makers identified it as a means to determine the effect of their actions. Its development also accelerated from within research funding organisations: from new ideas about the way research funding works, from an increased professionalisation of the field, from an improved database and from advanced methodologies which increase the range of feasible evaluation questions.

Evaluation models of the effects of research funding gain in sophistication

For every evaluation, it is necessary to have a basic assumption of the mechanism by which the activity, also called "intervention" (here: funding research), contributes to the desired outcome. The classic linear model of evaluation looks at the chain of input, output, outcomes and impacts. Inputs are the funds that funding agencies provide. Outputs are usually publications or patents. Outcomes can take the form of a new process. An impact might be a shift in a scientific paradigm.

This linear model has been changed into a more complex understanding of the interaction of inputs, outputs, outcomes and impacts. Depictions of this are the so-called Payback Framework (Buxton and Hanney 1996) or Logic Models[7].

New models of evaluation take account of a complex reality. However, they are also difficult to translate into an actual evaluation design. To draw valid conclusions from these models requires advanced methodologies and well-informed and vigilant peers.

Electronic gathering and storage of data creates a much better basis for evaluation

The advent of databases on research funding and on research outputs was and still is a strong driving force in evaluation. Only now, relatively cheaply and with little effort, is it possible to analyse data from a large number of research projects. Unlike paper documents, electronic data and documents can be easily extracted, shared and compared. Also, by now time series data can be analysed. This is important to draw conclusions on developments and to establish a connection between input and output – which in research takes some time. By building up new databases and data acquisition tools, research and research funding organisations will give this development a further push.

Methodologies advance rapidly

With more effort given to evaluation and more data available the methodologies in research evaluation gain sophistication. The power of statistical analysis progresses, new qualitative techniques are borrowed from the field of psychology, empirical social research etc. and the concepts and methodologies of impact studies progress rapidly.

---

6. The market oriented management of the public sector.
7. A logic model graphically shows how a programme is intended to work to achieve its objectives.

Increasing professionalisation in the field

Research evaluation is becoming more widespread, and, as a result, it is becoming more professionalised in several aspects.

First, the research organisations are getting more experience in organising (e.g. commissioning) evaluation projects. Many funding organisations and research performing organisations have reacted to the increasing demand by developing their evaluation capacities, either by establishing in-house evaluation units, by offering training to officers in charge of ex-post evaluation, by establishing evaluation routines and monitoring systems or by commissioning evaluation studies to external experts. The funding agency holds the full responsibility for an evaluation. Therefore, even if the analysis is to be conducted by external parties, it has proved helpful to build up internal expertise to assess whether the chosen methodology is appropriate and whether the analysis is correct.

Second, the scientists, who are – directly or indirectly – the object of evaluation have more experience. They learn to accept evaluation as part of the system (and sometimes learn to "play the system"). Studies indicate that younger generations of scientists are accustomed to and much more willing to accept a performance-oriented and indicator-based approach to research[8] and evaluation procedures.

Third, the evaluators are becoming more professional. If peers (scientists) are consulted, they have often already been involved in an evaluation process as either evaluators or as the objects of evaluation. For evaluation conducted externally funding organisations often commission professional evaluation institutes with long experience in evaluation methodologies, additional experience in the specifics of research evaluation being an asset. The choice of institutes may be limited to those with access to data or databases, as in the case of bibliometric studies. This leads to a market "niche" for providers of these services, with whom funding organisations have sometimes developed long-term partnerships, e.g. in framework agreements.

Nevertheless, the capacity worldwide, and the methods for carrying out evaluation are still poor compared with the amount spent on research and development. While there is the will to spend money to better understand the link between research and impact, there is not enough research, too few researchers and too few evaluation institutes to take up the questions. This rarely happens in any other

---

8. Böhmer et al. 2011.

research field. Capacity development and new ideas are required. With more evaluating institutes, competition will further spur development.

> Evaluation can be organised in different ways. The final report of the ESF MO Forum on ex-post-evaluation of Funding Schemes and Research Programmes 'Evaluation in National Research Funding Agencies: approaches, experiences and case studies' (2009) offers an overview. It shows that at the national level there is a variety of evaluation cultures which are linked with the orientation and direction of the funding organisation. Also, the requirements for research evaluation differ according to national priorities, institutional set-up, availability of data and expertise, and existing resources. Whether an evaluation is conducted internally, externally or in any form in between depends a lot on its purpose and the tradition of the organisation that is in charge of the evaluation.
> The report of WG 1 *'A Guide to Evaluation Activities in Funding Agencies'* (ESF 2012b) describes in detail the different variations in the organisational set-up.

Already, the increasing professionalism has resulted in the establishment of standards and the endorsement of codes of conduct. These are very widespread on the part of evaluators, scientists or market research associations. As a result, evaluation studies are now an accepted part of research management in most European research and research funding organisations.

> As the report of WG 1 *'A Guide to Evaluation Activities in Funding Agencies'* (ESF 2012b) shows, evaluation standards are increasingly developed or endorsed by research organisations carrying out or commissioning evaluations. The Austrian FTEval (Forschungs- und Technologieevaluierung) Standards have been a long-standing example how this practice serves all involved parties. The "Golden Rules" are a proposition by the ESF MO Forum to summarise the commitments.

4.
Methods, Approaches and Challenges for Evaluation

● ● ●

Turning to the actual evaluation project and its prerequisites, the following aspects should be considered. They should also inform evaluation strategies.

A commitment to quality

Evaluation will only convince and bear fruit if it meets the highest standards. A database of the highest quality, the employment of appropriate, state-of-the-art methodology and sound processes are cornerstones of a useful evaluation study. If research evaluation is to serve as a base for decisions on funding schemes or research fields, it can have important implications and its results are therefore likely to be contested. This means that the process, the quality of the database and the analysis and conclusions must be able to stand the most stringent test.

The importance of evaluation questions

We need to move away from symbolic or routine-based evaluation. Evaluation is most legitimate when it addresses a specific problem and can offer advice on decision-making. Evaluation exercises should have a specific goal and address a real problem.

Therefore, at the beginning of every evaluation project, there is a precise question to be answered. One such question could be whether a funding scheme meets its objective in taking young researchers more swiftly to a degree[9]. Another question might

be whether the additional (third-party) funding of research gives added value compared to purely intramural funding. Yet another question could be how a scientific discipline in a specific country fares in worldwide comparison[10]. It is very important to set specific questions at the very beginning of an evaluation project and to involve the relevant stakeholders in the discussion. The more precise the question, the more likely it is to get an answer. Still, not all questions can be answered. The information needs to be accessible, data available and the methodology appropriate. After the purpose of the evaluation has been considered, then structured steps to answer the evaluation questions are put in place[11].

Cost, timing and follow-up

In the conception phase of an evaluation, costs and benefits of different evaluation designs are always weighed against each other. This of course also applies to the evaluation project as such. The resources that flow into an evaluation project must not exceed the expected value. This calculation should take into account personnel costs for the internal supervision of the evaluation as well as costs to other stakeholders. For example, one might refrain from approaching grantees with a survey if the additional insights do not justifying the time spent on answering them. As for the direct cost, some organisations try to ensure this by setting a fixed amount of the programme budget apart

9. For example this has been analysed in the Marie Curie Impact Studies, see Report Evaluation Forum WG 2: 'The Challenges of Impact Assessment' (ESF 2012c).

10. For example: Evaluation of Mathematical Studies, Research Council of Norway (2011).
11. This is the bottom line of most logic models/evaluation frameworks.

for evaluation purposes and therefore to cover the financial cost[12].

Timing of the evaluation itself needs some consideration. The time when evaluation results are needed might not be the ideal time to conduct an evaluation. This is perhaps because the research that is to be evaluated has just started and results cannot be expected yet, or because the research was completed long ago and it is hard to distinguish the outcome from other research of the scientists, or get hold of the grantees[13].

Finally, the follow-up of the evaluation, which again requires careful consideration and generates costs, must not be neglected.

<div style="border: 2px solid red;">

RECOMMENDATION 1
A systematic approach to evaluation ensures its usefulness.

- It is advisable to develop a strategy to deal with evaluation which encompasses the needs of the organisation, the interests of the stakeholders, and the processes for handling data collection and evaluation studies and which will inform the organisation's strategy.
- The internal organisation of evaluation determines the quality of the outcome. It is important to build up internal expertise, for example to write terms of reference or to structure the discussion of evaluation results.
- Every process of evaluation should be planned carefully from the design of the study to the discussion of the results.
- Evaluations should have a specific goal and address a real problem. Therefore, they cannot be conducted in a standardised or mechanical way. Evaluation must allow for different needs of different organisations, in different situations and at different times.
- After the evaluation study has been completed and recommendations have been developed, it is useful to check the progress of implementation at intervals or on a continuous basis (monitoring).

</div>

Methodologies used in research evaluation

The evaluation design and the evaluation questions determine the choice of methods. Research evaluation profits from a long tradition of peer review processes to assess scientific merit[14], from experi-

---

12. Example see WG 1 survey in 'A Guide to Evaluation Activities in Funding Agencies', p.19 (ESF 2012b).
13. See reports of WG 2 and 3: 'The Challenges of Impact Assessment' and 'Research Funders and Research Output Collection' (ESF 2012c/d).
14. See also the European Peer Review Guide (ESF 2011).

ences of evaluation studies in other fields and from efficiency gains in data collection and improvements in analysis techniques.

Ex-post research evaluation often makes use of additional methodologies, mainly drawn from the field of social science and science metrics. Methods include interviews, site visits, bibliometric and scientometric analysis, surveys, panel studies, different forms of statistical analyses, document analyses etc. It is of utmost importance to use the right methodology for the respective evaluation question.

With every evaluation study there is the opportunity to expand the methodology. While it is sometimes useful to take the "tried and tested" approach, at other times new pathways allow new insights. Here, the boundaries to science studies are fluid. It is therefore useful to be in touch with the scientific community in the field. Also, with developing methodology and more information electronically available, the opportunities for analysis are increasing rapidly.

The power of indicators

Research evaluation makes use of different indicators that directly or indirectly try to capture aspects of research production, progress and impact. The intention of quantitative approaches is to overcome the dangers of subjectivity and dependence upon the judgemental power of a limited number of experts. However, quantitative information can be interpreted in different ways. Indicators, which are increasingly used and asked for in evaluation, are very dependent on the conditions (of a field, of a country) and cannot be interpreted out of context. Although there is substantial agreement between organisations over established metrics such as publications and patents, the types of output data to be collected provide an area of active discussion. There is a need for more effort to capture outputs that are more difficult to quantify, but give a more holistic picture of research output (such as influences on policy and practice). Since this is a new field, there are few validated metrics to use.

It is the responsibility of the parties involved (evaluators, research organisations) to communicate the limitations of the respective indicators. The way forward is to try to find remedies. This will require long term effort and cooperation among research funders and the research community.

<div style="border: 2px solid orange;">

The ESF MO Forum on Indicators of Internationalisation has gone a long way to discuss the promises and benefits of indicators and suggest solutions.

</div>

Data requirements

Evaluation is an empirical exercise and requires the existence of useful data. In evaluations that are based on peer review, these data might stem from the publication base of the funded researchers, project reviews or site visits. Other evaluation methodologies analyse existing databases, such as bibliometric databases or process-produced data from the proposal administration process. Yet other evaluations need to generate new data, e.g. in surveys or interviews. For every evaluation project the quality of the database is crucial. It needs a lot of effort to ensure that data are complete or soundly sampled, correct and capable of analysis.

While research funding agencies usually have high expertise and long experience in handling grant proposals and the data generated in the application process, there has not been the same progress in the field of output data.

> The ESF MO Forum on Evaluation of Publicly Funded Research undertook an international survey of research organisations to discover current practices. The aim was to share this information in order to help organisations develop their evaluation approaches (see report of WG 3, ESF 2012d).
>
> The outputs examined in further detail include:
> - Mobility of researchers
> - Collaboration
> - Intellectual property and routes to commercialisation
> - Publications
> - Influence on policy
> - Development of new products/interventions

Some organisations are planning and implementing different practices for the collection of this information and develop systems to monitor research grants and outputs. There are few established processes for the collection of output information. If data for evaluation and monitoring are collected and stored, it must be clear where the responsibility for a data collection system lies. This responsibility includes:
- data security,
- quality assurance,
- access to the data,
- privacy issues.

Having dealt with these challenges, research organisations have a powerful instrument at hand that allows much more information on research to be gathered and analysed much faster. This benefits both organisations and researchers: research and funding organisations must, and are committed to, alleviate the burden on researchers by carefully examining data requirements, avoiding duplication of data gathering and working on mechanisms to facilitate data entry.

Classification and standardisation of data

Furthermore, a stable, consistent and appropriately detailed classification system for analysis of research portfolio information is a solid foundation for the evaluation of research progress, productivity and quality[15]. If datasets more broadly comply with data standards such as CERIF[16] then a larger range of data might be mapped and exchanged among institutions. Sharing data and joint analysis can also establish useful benchmarks and help determine what "good" performance is.

Besides capturing the data, one of the greatest challenges in research evaluation is to connect information to a researcher, a grant, an output. One of the issues where funding agencies can play a major role in the future is therefore about linkage between datasets.

> There are initiatives that are aimed at reducing the burden on researchers. These include the monitoring systems of the UK Medical Research Council or the Swiss National Science Foundation, which allow the direct extraction of data from publication databases instead of typing them into the system. Often, electronic and structured versions of final reports serve as a data source, such as at the Austrian Science Fund. The ERC is planning to establish a research information system that is based solely on information available through the process-produced data of the handling of application and through public information. The German Research Foundation (DFG) plans to expand its existing research information system (Gepris/online database on DFG-funded research projects) to include output, personnel and collaboration information extracted from final reports of research grants.

Initiatives like ORCID or the commercial ResearcherID[17] strive for an unambiguous identification of researchers and therefore allow linkage between datasets (mainly publication data and funding data) based on the identity of a researcher. This

---

15. Different approaches to disciplinary classifications and possible harmonisations are presented in the Working Paper WG 3: 'The classification of research portfolios' (ESF 2012d).
16. CERIF is a data model put forward by the European Organisation for International Research Information (Eurocris).
17. ResearcherID is an identifying system for scientific authors introduced in January 2008 by Thomson Reuters. ORCID (Open Researcher and Contributor ID) is a proposed nonproprietary code that can uniquely identify scientific authors.

would allow researchers to be properly acknowledged for their work, and also reduce the need to re-enter data into multiple systems. It would also allow funding organisations to more accurately depict what they have funded. Funding agencies, research performing organisations and individual researchers have an interest in the improvement of data quality, because external research assessment will increasingly use databases. The correct coverage of entries (no over-reporting and no under-reporting) is important for that.

RECOMMENDATION 2
Data collections for evaluation and monitoring purposes are a valuable source of information.

- The purpose of any data collection should be clear. Responsible handling of scarce resources means that data should not be collected for its own sake – that is, if it is not being used. If an evaluation produces a lot of data, for example about a research field or information from grantee surveys, it is helpful to share the data in the form of scientific use files, which can be published or put into a data repository.

- Existing information and data (e.g. from the proposal processing and final reports but also external sources like publication databases etc.) should be exploited. Funding agencies need to try to limit the number of surveys and avoid duplication of effort, e.g. reporting obligations and other forms of data acquisition.

- To make information comparable will reduce the burden on researchers. Data that follow certain standards or are harmonised and agreed upon by several institutions can be more easily used in other contexts as well.

- The use of common classification systems even as a secondary level might also help to make data more easily compared and exchanged between funders and countries.

- Joint evaluation efforts require a common agreement on the data used and a common standard of the quality of the data. There needs to be more effort to develop these standards.

- Standardisation of data collection is an ambitious but much needed objective. Research Organisations should therefore support standardisation initiatives like CERIF, ORCID and implement those standards in their data collection procedures.

Demonstrating impact

The effect of research programmes, funding schemes or any other intervention is the fundamental question of every evaluation study. "Impact studies" aim to take a broader perspective and encompass unintended consequences of funding and conducting research projects, side effects and the influences of the research system on the outcome of research. There are not many examples of true impact studies and they face some difficulties. Any attempt of an impact study depends crucially on a solid database and a sensible methodology to establish hints for causal relationships.

The examples of impact studies show the diversity and breadth of the methodology employed. For example, the Impact Evaluation of the Academy of Finland of Finnish Programmes of Centres of Excellence used case studies, interviews, document analysis and other means; along with integrated different stakeholder perspectives, and funded and non-funded researchers; and made comparisons with other countries' excellence programmes.

Methodological considerations concern the time between the intervention (funding of a research project) and the analysis, the question "what would have happened if the funding did not exist?" (counterfactual), and comparative reference. This can include funded vs. non-funded research, research from different fields, other funders or comparable research organisations from other countries. The need for a comparative group is a powerful argument for harmonising data internationally and for sharing knowledge about similar funding schemes at an international level.

As with other forms of social sciences or social research, it will remain very difficult for research evaluation to establish causal mechanisms. The research system is "alive" and influenced by many different factors that cannot be held constant or blanked out. The difficulty of impact studies lies in attribution. A grant given to a researcher hopefully enables him or her to conduct the research project envisaged. However, on a micro-level it is hardly possible to distinguish which research result stems from which source of funding, even less to establish a direct and linear chain of knowledge expansion. Research is an additive process to which many individuals contribute. Basic research yields important results, but they are always mediated by other spheres, such as technology or public opinion. It is unlikely to be able to detect the one single contribution of a funded project. This is why it is more honest to talk about contribution instead of attribution and to look at a macro instead of a micro level.

The emphasis on the impact of science and therefore the demand for impact studies mirror a change in the perception of the purpose of research. However, impact studies should be handled with caution, not only because the methodological challenges but also because of their tendency to focus on short- and medium-term effects.

RECOMMENDATION 3

The use of appropriate methodologies and indicators needs to be given special consideration.

- Indicators are a prerequisite for quantifying research (input, output, outcome, impact). The development of robust and meaningful indicators is a continuous task for organisations involved in research evaluation.

- Consider the three main methodological challenges to assessments of impact a) how to determine attribution of an action to the effects of that action, b) how to deal with time-lags between research and consequences of research, and c) how to determine what would have happened without the action taken (the counterfactual position).

- As academics from the field of sociology of science, science studies or bibliometrics advance methodologies and data analysis and establish new indicators, it is important to work closely with this research community.

- Diverse approaches to evaluation spur development. It is useful to commission to new institutes to help build up knowledge in a broad field and to prevent one-sided market power of the established companies.

5.
Diversity and Harmonisation
●●●

The Member Organisation Forum has succeeded in bringing together national research and research funding organisations from Europe to discuss and document good practice and challenges for research evaluation. The discussions have revealed a great amount of agreement, a potential for cooperative approaches to evaluation, but also different practices and priorities. There are merits to both diversity and harmonisation in research evaluation.

The case for diversity lies in the heterogeneity of the actors of research evaluation in Europe. These national research and research funding organisations deal with many languages, different legal conditions, different processes and different missions. Conditions on added value, which are assessed in evaluation studies, might be different depending on research groups, research institutions or the research system. A uniform approach to evaluation is therefore neither feasible nor desirable.

Also, the research and research funding organisations themselves and their uses for evaluation studies can be very different. While some use them for the distribution of resources, others see them purely as a learning tool. While some organisations conduct evaluations because of national legislation or the demands of their funders, others see them as a voluntary exercise. National priorities have to be taken into account. While the reach of funding programmes to researchers of diverse ethnicities might be an important objective for one organisation, knowledge transfer might be a bigger issue for another. All these differences require a different evaluation design. That is why one approach to evaluation cannot be simply translated to another setting.

Another argument for a pluralist approach to evaluation is that different methodologies should be explored. This requires different organisations taking a different approach on evaluation. However, this is also a strong argument for harmonisation, cooperation and knowledge exchange: while every organisation conducts only a limited number of evaluation studies, to develop evaluation methodology and experiences in process, the evaluators need to look further than their own organisation. This makes bridges between funding agencies and their evaluation efforts all the more important, if the plurality of approaches is to grow into a rich and robust set of methods. This report, and the recommendations in the following section, is intended to contribute to such a development.

No research or research funding organisation is the first or the only one to deal with the many questions to be considered when setting up an evaluation project. If issues like data acquisition, standardisation and harmonisation of data, adequate indicators, or quality standards and processes can be addressed collectively, there is much scope for synergies in the development of procedures and technical support infrastructure.

Also, in most organisations the critical mass of people concerned with evaluation is very small. To keep up with the international developments in the field of evaluation and monitoring, in terms of methodology, indicators and technology, it is therefore crucial to exchange and cooperate in multinational projects.

The Nordic countries have for a long time cooperated in sharing data resources. They also face challenges in exchanging data, and have compiled information on political, legal, ethical, organisational, technical and financial challenges in a recent report (Sandberg 2012). ●●●

•••
A good example of an international cooperation has been in the area of classification with the HRCS (UK's Health Research Classification System). Pilot projects in the Research Council of Norway, the Swedish Research Council and the German Research Foundation demonstrate how a translation of subject classification can be achieved.

This does not require uniform approaches to evaluation. But it requires an exchange about the prerequisites for successful evaluation activities among actors from different nationalities, different roles in the research system and from experts and practitioners. This provides the basis for the development of a shared understanding about the right level of harmonisation.

RECOMMENDATION 4
Alignment of evaluation activities enables Research Organisations to position themselves on the European and global research landscape.

- Harmonisation and standardisation of data and evaluation processes are the prerequisites for comparison and joint activities.

- It is advisable to jointly establish strategic alignments with key actors, such as publishers of publication databases. This might include the harmonisation of acknowledgements, the use of unique IDs (Researcher's ID, Grant ID, Institutional ID) and the capacity to analyse bibliometric information on a comparative basis.

- Whenever evaluations are concluded, as much information as possible about the process, about the data used and about the conclusions should be made available for partner organisations so that it can be used for wider learning and for benchmarking purposes.

- Heads of Research Organisations have a pivotal role in ensuring that exchange and networking is possible and enforced. This requires the allocation of resources for these activities on an organisational level and for organisations to support joint activities.

6.
Recommendations
● ● ●

The debates among the members of the MO Forum have resulted in a set of recommendations.

1. A systematic approach to evaluation ensures its usefulness

- It is advisable to develop a strategy to deal with evaluation which encompasses the needs of the organisation, the interests of the stakeholders, and the processes for handling data collection and evaluation studies.
- The internal organisation of evaluation determines the quality of the outcome. It is important to build up internal expertise, for example to write terms of reference or to structure the discussion of evaluation results.
- Every process of an evaluation should be planned carefully from the design of the study to the discussion of the results.
- Evaluations should have a specific goal and address a real problem. Therefore, they cannot be conducted in a standardised or mechanical way. Evaluation must allow for different needs of different organisations, in different situations and at different times.
- After the evaluation study has been completed and recommendations have been developed, it is useful to check the progress of implementation at intervals, or on a continuous basis (monitoring).

2. Data collections for evaluation and monitoring purposes are a valuable source of information

- The purpose of any data collection should be clear. Responsible handling of scarce resources means that data should not be collected for its own sake – that is, if it is not being used. If an evaluation produces a lot of data, for example about a research field or information from grantee surveys, it is helpful to share the data in the form of scientific use files, which can be published or put into a data repository.
- Existing information and data (e.g. from the proposal processing and final reports but also external sources like publication databases etc.) should be exploited. Funding agencies need to try to limit the number of surveys and avoid duplication of effort, e.g. reporting obligations and other forms of data acquisition.
- To make information comparable will reduce the burden on researchers. Data that follows certain standards or is harmonised and agreed upon by several institutions can be more easily used in other contexts as well.
- The use of common classification systems even as a secondary system might also help to make data more easily compared and exchanged between funders and countries.
- Joint evaluation efforts require a common agreement on the data used and a common standard for the quality of the data. There needs to be more effort to develop these standards.
- Standardisation of data collection is an ambitious but much needed objective. Research Organisations should therefore support standardisation initiatives like CERIF, ORCID and implement those standards in their data collection procedures.

3. The use of appropriate methodologies and indicators needs to be given special consideration

- Indicators are a prerequisite for quantifying research (input, output, outcome, impact). The development of robust and meaningful indicators is a continuous task for organisations involved in research evaluation.
- Consider the three main methodological challenges to assessments of impact a) how to determine attribution of an action to the effects of that action, b) how to deal with time-lags between research and consequences of research, and c) how to determine what would have happened without the action taken (the counterfactual position).
- As academics from the field of sociology of science, science studies or bibliometrics advance methodologies and data analysis and establish new indicators, it is important to work closely with this research community.
- Diverse approaches to evaluation spur development. It is useful to commission to new institutes to help build up knowledge in a broad field and to prevent one-sided market power of the established companies.

4. Alignment of evaluation activities enables Research Organisations to position themselves on the European and global research landscape

- Harmonisation and standardisation of data and evaluation processes are the prerequisites for comparison and joint activities.
- It is advisable to jointly establish strategic alignments with key actors, such as publishers of publication databases. This might include the harmonisation of acknowledgements, the use of unique IDs (Researcher's ID, Grant ID, Institutions) and the availability of information to analyse bibliometric information on a comparative basis.
- Whenever evaluation is concluded, as much information as possible about the process, about the data used and about the conclusions should be made available for partner organisations for wider learning and for benchmarking purposes.
- Heads of Research Organisations have a pivotal role in ensuring that exchange and networking is possible and enforced. This requires the allocation of resources for these activities on an organisational level and for organisations to support joint activities.

7.
Outlook

● ● ●

With its recommendations, the MO Forum aims to contribute to increased quality and usefulness of evaluation studies. Still, there is more to be done to achieve alignment between the evaluation activities of national and European organisations to make research landscapes and results of funding more comparable. Future activities could unfold in the following directions.

Agreeing on standards

- To allow for international cooperation and comparisons of research portfolio, funding agencies and research performing organisations should agree on data standards and quality requirements of monitoring data. This includes translation of subject classification and pilot use of international classification systems.
- Funding Agencies and Research Performing Organisations should develop guidelines for researchers, e.g. how to acknowledge the provision of research funding. Vis-à-vis publishing houses, funding organisations should support the inclusion of the name of the funding organisation and the grant number in publication databases so that any publication that is funded externally can be found. This necessitates further improvement of the quality of the databases, e.g. in the field of acknowledgements.
- To make international collaboration in evaluation projects, e.g. evaluation of research fields or funding schemes, possible, it is necessary to agree on a common set of quality standards for evaluation procedures. A helpful way to implement this is to follow the so called "Golden Rules on evaluation projects" developed within the framework of the Forum (ESF 2012b). We encourage the ESF Member Organisations to adopt them.

Learning from each other

- The exchange of evaluation criteria and questionnaires used in evaluation surveys are first steps to make information from evaluation exercises in different countries comparable. Also, a "meta-perspective" can be taken, for example by commissioning internationally comparative studies on brain drain, research areas, or research networks. Meta-studies of national evaluation studies and monitoring systems could also give insights into the implementation of evaluation procedures.
- Another way to build up knowledge within the organisation that conducts evaluation, among stakeholders and among a wider audience – for instance other organisations that plan evaluation studies – is to exchange evaluation studies or make them publicly available, preferably (at least the summary) in English. In this way the opportunities that evaluation studies provide can be widely exploited.
- Easy access to this knowledge could be provided by a knowledge base on evaluation, e.g. a collection of evaluation studies online, searchable by different criteria.
- The use of international experts for evaluation may help to harmonise procedures and standards. This includes invitations to panel deliberations and discussions with international academics and practitioners that bring expertise to the field.
- Joint evaluation projects allow actors to accumulate knowledge about research funding and to draw conclusions about national specifics. For example, the tools of impact assessment could be further developed by conducting an assessment

of impact that involves different countries and research funders and research performing organisations.

Network building and professionalisation

- To further the discussion on evaluation, there needs to be more contact between funding agencies and research performing organisations, academics in the field of research evaluation and researchers who are to be evaluated. This could include network meetings and/or topical workshops. The range of topics includes – amongst others – the evaluative use of final reports, data standards or indicators for impact.
- For funding agencies it is important to spur the development of and knowledge about evaluation. Therefore an important future field of activity is the transfer of best practice and mutual learning, for example by offering opportunities for staff exchange and training.

Voicing common interests

- National research organisations work in a field where many of their stakeholders and different political actors (governments, EU Commission, OECD, etc.) have an active interest in their evaluation activities. Therefore, research organisations should join forces and speak with one voice to policy players about strategic decisions such as indicators used, forms of data collection, use of evaluation studies, etc.
- This includes the development of common statements, e.g. about the ex-post-detection of research projects' impact, about data standards or about the (mis)use of monitoring data.

Cooperation among European research organisations is the key to develop research evaluation practices even further. The Member Organisations which took part in the ESF Forum on Evaluation of Publicly Funded Research are committed to collaborate in the future in this matter.

Literature & Annexes

Literature

- Academy of Finland (2009): Evaluation of the Academy of Finland of Finnish Programmes of Centres of Excellence 2000-2005 and 2002-2007. Available at: *http://www.aka.fi/Tiedostot/Tiedostot/Julkaisut/2_09%20CoE%20in%20Research.pdf*
- Böhmer, Susan et al. (2011): Wissenschaftler-Befragung 2010: Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten. iFQ Working Paper No. 8.
- Buxton, Martin, Hanney, Steve (1996): How can payback from health services research be assessed? J Health Serv Res Policy. 1996 Jan;1(1):35-43. Review.
- EU Commission (2004): Evaluating EU activities. A practical guide for Commission services. Available at: *http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_activities_en.pdf*
- EU Commission (2010): Ex-post Impact Assessment study concerning the 'Marie Curie Actions' under the Sixth Framework Programme. Available at: *http://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/ex-post_impact_assessment_of_marie_curie_actions_under_the_fp6.pdf*
- European Foundation Centre (2010): Research Forum Evaluation Guidelines 3: How to evaluate. Available at: *http://www.efc.be/programmes_services/resources/Documents/Research_Evaluation_Guidelines_3_How_to_Evaluate.pdf*
- European Science Foundation (2009): Ex-post Evaluation in National Research Funding Agencies: A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes. Available at: *http://www.esf.org/activities/mo-fora/publications.html*
- EUROHORCs and ESF (2009). EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions. Available at: *www.esf.org/publications/corporate-publications*
- European Science Foundation (2011): European Peer Review Guide: Integrating Policies and Practices into Coherent Procedures. A Report by the ESF Member Organisation Forum on Peer Review. Available at: *http://www.esf.org/activities/mo-fora/publications.html*
- European Science Foundation (2012a): Indicators of Internationalisation for Research Institutions: a new approach. A report by the ESF Member Organisation Forum on Evaluation: Indicators of Internationalisation. Available at: *http://www.esf.org/activities/mo-fora/publications.html*
- European Science Foundation (2012b): A Guide to Evaluation Activities in Funding Agencies. A report by the ESF MO Forum on Evaluation of Publicly Funded Research - Working Group 1: Quality Assurance and Evaluation Guidelines. Available at: *http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html*
- European Science Foundation (2012c): The Challenges of Impact Assessment. A report by the ESF MO Forum on Evaluation of Publicly Funded Research - Working Group 2: Impact Assessment. Available at: *http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html*
- European Science Foundation (2012d): Research Funders and Research Output Collection. A report by the ESF MO Forum on Evaluation of Publicly Funded Research - Working Group 3: Comparative Research Portfolios. Available at: *http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html*
- Hansen, Hanne Foss (2009): Research Evaluation: Methods, Practice, and Experience. Danish Agency for Science, Technology and Innovation. Copenhagen.
- Sandberg, Marie (2012): Reinforced Nordic collaboration on data resources. Challenges from six perspectives. Nordic Council of Ministers.
- The Research Council of Norway (2011): Research in Mathematics at Norwegian universities an evaluation. Available at: *http://www.forskningsradet.no/en/Newsarticle/Worldclass_Norwegian_mathematicians/1253976964167*

Annex A.1 MO Forum Mandate and Description

The ESF MO Forum on Evaluation of Publicly Funded Research brought together the expertise and experience of representatives from 33 Member Organisations and 7 Observer Organisations.

The aims of the MO Forum included:
- To exchange and document experiences with current practices.
- To facilitate networking of science officers engaged in evaluation.
- To regularly update, further elaborate and disseminate the report on evaluation practice.
- To explore needs and possibilities for collaboration in future evaluation exercises.
- To identify best practice examples in research evaluation that could possibly lead to a Guideline for Evaluation Processes.

The MO Forum set up three working groups:
1. **Quality assurance and evaluation guidelines** (Chaired by Gro Helgesen/Research Council of Norway).
   - Draw up guidelines for ex-post evaluation of funding schemes.
   - Do a survey among the members on evaluation organisation and practices.
   - Develop a website repository containing evaluation reports and guidelines.
2. **Impact assessment on science and society** (Chaired by Per Janson/Swedish Research Council).
   - Inform about existing impact studies.
   - Produce a template for categorising the impact studies.
   - Conduct an in-depth analysis of impact assessment practices (methodology, criteria, best practice).
3. **Classification systems and categorisation of output data** (Chaired by Ian Viney/UK Medical Research Council).
   - Examine feasibility of common systems to classify research portfolios.
   - Produce a short paper summarising approaches for the analysis of output data.

The Forum has held the following meetings and workshops:
- First meeting in Stockholm, May 2010.
- Second meeting in Smolenice, November 2010.
- Meeting of Working Group 3, December 2010.
- Third meeting in Paris, May 2011.
- Fourth meeting in Bern, 7-8 November 2011.
- Fifth and final meeting in Oslo, 7-8 May 2012.
- Meetings of Working Groups in between, if required.

It enjoyed support by National Seconded Expert Sarah Chen (CNR Italy) from April-August 2011.

Apart from this report the MO Forum also published this work in other ways and interacted with other bodies, having significant impact in the course of its two-year-life-cycle:
- The Working Groups have launched three online surveys, which have so far circulated among MOs and external organisations through Europe and worldwide with a satisfying participation rate.
- Two working documents 'The capture and analysis of research outputs' and 'The classification of research portfolios' have been published at *http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html*
- Discussion among the Working Group on the classification of research portfolios has stimulated the production of an EMRC Science Policy Briefing on the classification of research portfolios: 'Health Research Classification Systems – Current Approaches and Future Recommendations' at *http://www.esf.org/publications.html* (November 2011 and an update in April 2012).
- The Forum presented its work, amongst others, at the G8 Heads of Research Assessment meeting (November 2011), the ERC workshop on the evaluation of final reports (November 2011), and the ESF Science Policy Conference (November 2010).
- It jointly developed with the European Alliance on Research Career Development the International Workshop on "How to track Researchers' Careers", 9-10 February 2012, Luxembourg which has published a report and recommendations.
- *Research Europe* published an article on the present landscape of evaluation in Europe focusing on the Forum positioning on monitoring systems, interviewing its Chair, Anke Reinhardt ("The Public's Right to know", 26 May 2011).

Annex A.1 **MO Forum Mandate and Description**

- Last but not least, are the results of the Forum's Working Groups, in the three reports:
  – 'A Guide to Evaluation Activities in Funding Agencies'.
  – 'The Challenges of Impact Assessment'.
  – 'Research Funders and Research Output Collection'.

The Forum established interaction with other ESF Member Organisation Fora: Indicators for Internationalisation, Research Careers, Peer Review, Science in Society, Research Infrastructures.

MO Forum on Indicators of Internationalisation

This Forum was created in 2010 in order to undertake a pilot study aiming to design and to produce a set of indicators that could account for the internationalisation of European research activities and programmes and be useful for MOs' policy and piloting processes and in their relationships with the European Commission as well as their governments (for benchmarking and policy evaluation). The development of such a set of indicators is aimed at enhancing the development of a common strategic analysis of internationalisation among the institutions that are gathered within ESF.

The Forum has already included: I) an analysis of the literature, II) the design of a common framework to depict internationalisation objectives and activities of each organisation and III) the selection of a common set of indicators which would help MOs to position themselves within the R&D system at national and supra-national level. These indicators for funding and for performing organisations were chosen for their coherence with the framework and either drawn from existing sets of indicators or newly designed by the experts involved in the project.

All the findings were published in the report 'Indicators of Internationalisation for Research Institutions: a new approach'. A report by the ESF Member Organisation Forum on Evaluation: Indicators of Internationalisation.

- **ESF Member Organisations (MOs):** Refers to ESF member organisations which are Research Performing Organisations (RPOs) and Research Funding Organisations (RFOs).
- **ESF MO Forum:** An output-oriented, issue-related venue for the Member Organisations, involving other organisations as appropriate, to exchange information and experiences and develop joint actions in science policy.
- **Ex-Post Evaluation:** Evaluation is the description, analysis and assessment of projects, programme processes or organisational units. It is performed in the course of, or after, an intervention.
- **Funding Scheme:** Funding programmes or funding instruments distribute funding based on explicit requirements and often with an explicit objective, e.g. to promote scientific careers, to enable research collaboration.
- **Impact Studies:** Investigation of positive and negative, short- and long-term, intended and unintended effects of research funding.

- **Monitoring:** Monitoring is the systematic, recurrent collection of data to observe, track and record processes or activities.
- **Output Data:** Information on tangible and quantifiable research output (research findings) such as publications or patents.
- **Peer Review:** The process of evaluating research applications (proposals) by experts in the field of respected research.
- **Research:** The activity performed by researchers in all sciences.
- **Research Discipline:** Field of study, a branch of knowledge.
- **Research Field:** Research area or area of scientific study, often equivalent or closely linked to a research discipline.
- **Research Funding Organisation:** A governmental agency or private organisation which funds research.
- **Research Performing Organisation:** An institute or other organisation which is itself realising research and which employs active researchers.

Annex A.3 **List of Forum members**

Country	Organisation	Member
Austria	Austrian Science Fund (FWF)	Christian Fischer Falk J. Reckling
Belgium	Fund for Scientific Research (FNRS)	Raphael Beck Nadège Ricaud
	Research Foundation - Flanders (FWO)	Hans Willems
Czech Republic	Czech Science Foundation (GAČR)	Veronika Paleckova Radka Smrzova
	Academy of Sciences of the Czech Republic (ASCR)	Petr Ráb
Denmark	Ministry of Science, Technology and Innovation	Poul Schjørring
	Danish Agency for Science, Technology and Innovation (FIST)	Claus Beck-Tange
	Danish National Research Foundation (DNRF)	Marianne Gauffriau
Estonia	Estonian Research Council (ETAG)	Viktor Muuli
Finland	Academy of Finland (AKA)	Jaana Roos
France	French National Institute of Health and Medical Research (Inserm)	Isabelle Henry
	National Centre for Scientific Research (IPCMS/CNRS)	Pierre Gilliot
Germany	German Research Foundation (DFG)	Katharina Fuss Jürgen Güdler Anke Reinhardt (Chair)
Hungary	Hungarian Scientific Research Fund (OTKA)	Gyula Péter Szigeti
Ireland	Health Research Board (HRB)	Brendan Curran
	Science Foundation Ireland (SFI)	Helen O'Connor
Italy	National Research Council (CNR)	Sarah Chen
Luxembourg	National Research Fund (FNR)	Frank Bingen
Netherlands	Netherlands Organisation for Scientific Research (NWO)	Anko Wiegel
Norway	The Research Council of Norway (RCN)	Gro Helgesen (Co-Chair WG1) Ingrid Roxrud
Slovakia	Slovak Academy of Sciences (SAV)	Iveta Hermanovská
Spain	Council for Scientific Research (CSIC)	Juan José Damborenea González
Sweden	Swedish Council for Working Life and Social Research (FAS)	Inger Jonsson
	Riksbankens Jubileumsfond	Britta Lövgren
	Swedish Research Council (VR)	Per Janson (Co-Chair WG2) Jenny Nordquist
Switzerland	Swiss National Science Foundation (SNF)	Katrin Milzow
Turkey	Scientific and Technological Research Council of Turkey (TÜBITAK)	Mustafa Ay
United Kingdom	Medical Research Council (MRC) Head Office	Ian Viney (Co-Chair WG3)

Observers		
	All European Academies (ALLEA)	Rüdiger Klein
	European Commission	Peter Fisch
	European Research Council (ERC) Executive Agency	Alexis-Michel Mugabushaka
	Foundation for Polish Science	Marta Lazarowicz-Kowalik
	National Centre for Research and Development, Poland	Agnieszka Sosinska
	Wellcome Trust	Liz Allen
	Wellcome Trust	Briony Rayfield
Forum management		
	European Science Foundation (ESF)	Laura Marin Madelise Blumenroeder