A 2 PROMOTING INNOVATION MORE EFFICIENTLY THROUGH THE EVALUATION OF POLICY MEASURES

Political action serves to achieve specific goals. In addition to ensuring the effectiveness of a policy measure, it is also important to ensure the cost efficiency of allocated resources. Evaluations of individual policy measures can help determine how a measure must be designed in order to achieve the desired effect with a reasonable use of resources.

This also is of central importance in the area of research and innovation policy. The key is to know what type of funding is economically sensible. Not only the type of support, but also the potential recipient and the amount of support are decisive for the success of a measure. By means of a valid evaluation of funding measures, it can be established which of the allocated resources are most effective, how big the impact of funding is, and how measures can be improved further. Evaluations thus facilitate the competition of ideas in the search for the best long-term results in promoting innovation. If an evaluation is integrated into the project planning from the very start, additional costs arising from the evaluation will be relatively modest. In contrast, if evaluations are not employed, this can lead to efficiency losses and thus significantly higher costs. This is especially true if, as a result, other innovative ideas do not receive sufficient financial support.

To date, the practice of evaluation in Germany has shown a mixed picture: the need for evaluations has been recognised by various stakeholders, and many policy measures are already being evaluated. Yet, the scientific quality of evaluations is an issue that remains to be addressed. In many ministries, support measures are mostly evaluated prior to the allocation of funds, i.e. ex-ante. However, this does not suffice for determining the actual achieved efficiency of support measures. Therefore, accompanying evaluations as well as ex-post evaluations are of central importance. Based on valid scientific methods, such evaluations determine the extent to which a measure has been effective, thereby providing a solid basis for policy decisions on the continuation of or necessary adjustments to policy measures.

Several attempts have been made to anchor the topic of evaluation in research and innovation funding. In mid-2011, the Federal Ministry for Economic Affairs and Energy set up a specialist staff unit *(Aufbaustab Födercontrolling/Evaluation)* responsible for assessing and controlling support measures and the evaluation thereof. The ministry's subject-specific divisions are required to involve the staff unit in the early stages of tendering and awarding evaluations, as well as in the reporting process. In the Federal Ministry of Education and Research (BMBF), the ministry's subject-specific divisions are currently in charge of evaluating the effects of support measures. In addition, the BMBF is planning to establish an additional expert unit, which shall be responsible for fundamental and strategic issues relating to evaluations.

Furthermore, on 23 January 2013, the State Secretary Committee decided that in future, any proposed legislation shall be evaluated ex-post if the ex-ante expected annual compliance costs exceed one million euro. Such ex-post evaluations shall be conducted three to five years after the launch of a measure. Yet, the committee does not elaborate on the methodological standards to be applied.[18]

Germany's evaluation practice is most developed in the field of labour market policy, which is also the case in other countries.[19] The evaluation practice in this field offers impressive evidence of the importance of impact analyses for decision-making in economic policy. Based on evaluations, job creation measures that had been introduced in the 1990s were abolished in 2012 as their effectiveness could not be demonstrated. In the course of advancing the evaluation practice further, scientists have also been granted access to research data on major labour market programmes. Above all, this has ensured that competing studies could be produced on the same research issues. This has led to improvements in the overall quality of studies. The success of this change in practice illustrates the enormous potential of evidence-based economic policy. Other policy areas, however, are still far from employing such

evaluation practices, in spite of promising approaches. A view of the bigger picture shows that Germany could hugely benefit from experience gained in other countries and other policy areas.

International best practice examples

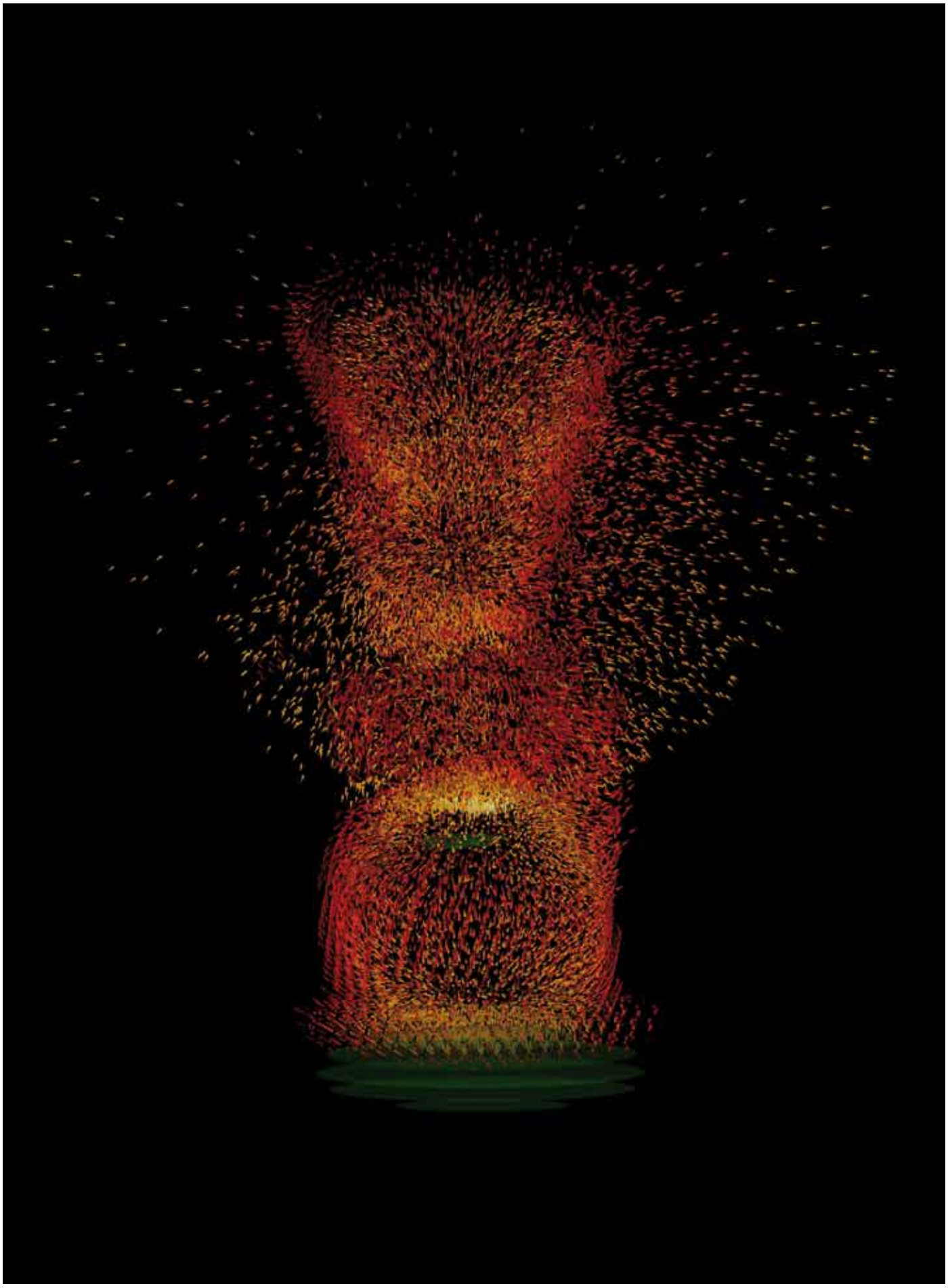
In a number of policy areas, the United States have a long tradition of evaluating government programmes based on scientific procedures. Relevant examples include the Tennessee STAR experiment,[20] which focussed on measuring the effect of class size on the learning outcomes of students, as well as the RAND experiment[21] on the effects of health insurance.[22]

Following the implementation of the American Recovery and Reinvestment Act of 2009, the United States have also been placing greater emphasis on employing impact analyses in the field of economic policy measures. The Office of Management and Budget (OMB) plays a central role in putting these objectives into practice.[23] According to the OMB guidelines, evaluations shall constitute an integral part when allocating resources. The objective is to award the bulk of funding to support measures with proven effectiveness. Another portion of funding is earmarked for programmes that are endorsed by positive initial evaluations, but have not been conclusively assessed. Resources are allocated to these programmes under the condition that further evaluations will be implemented. The remaining portion of funds will then be invested in innovative and potentially successful programmes that are yet to be evaluated.[24]

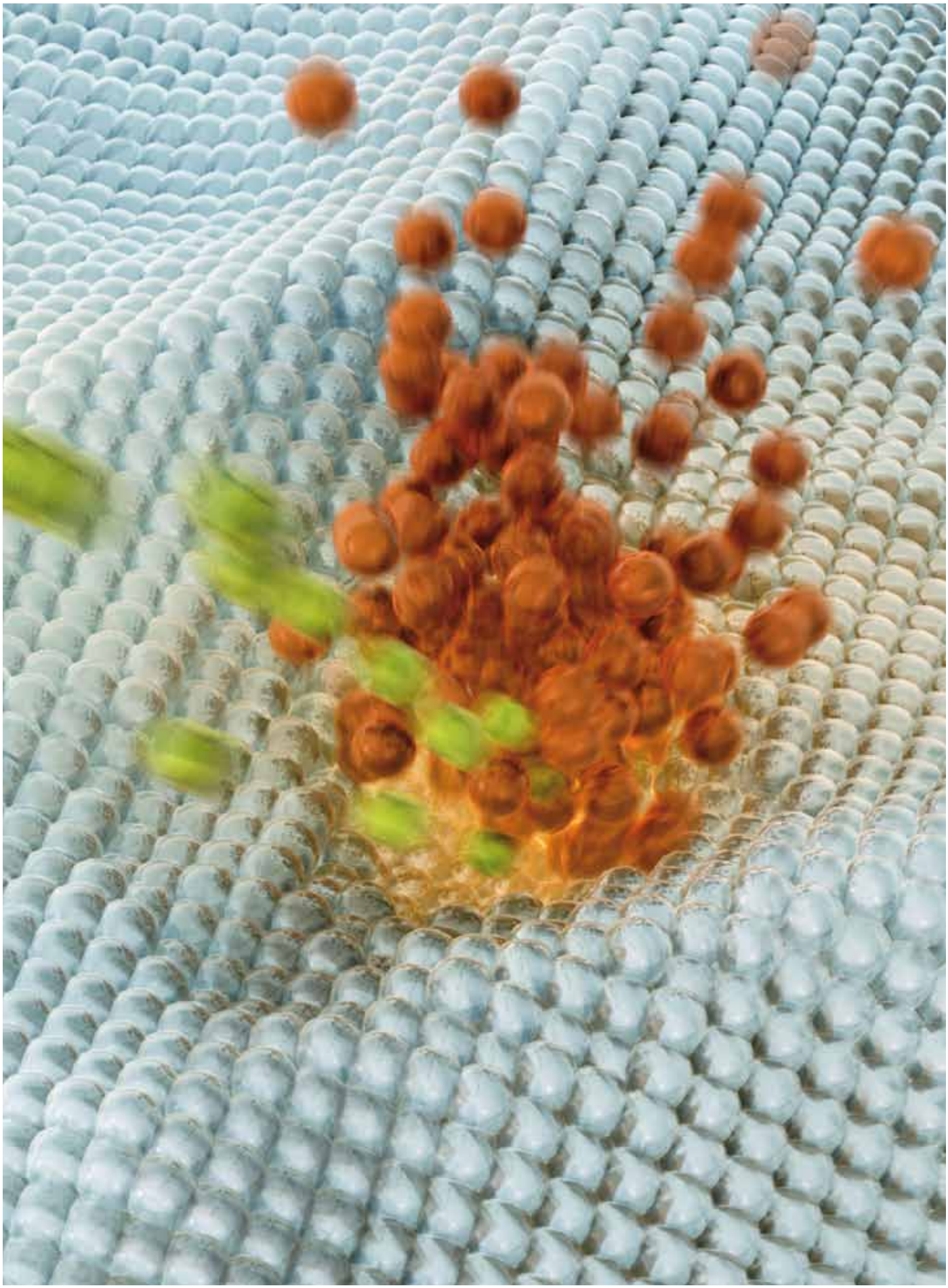
To date, policy evaluations have largely focussed on social programmes. To solve the United States' most pressing social problems, promising concepts are being identified, and pilot projects are being implemented and post-evaluated. If an evaluation proves positive, measures are widely implemented.[25] But even after implementation, measures are continuously monitored with regard to their effectiveness. In this process, it is important to regard impact analyses as an integral part of the decision-making process from the outset of a project. This will guarantee comprehensive evaluation opportunities, allowing stakeholders to factor in data requirements during a measure's planning phase. Furthermore, it will also ensure the cost-effectiveness of an evaluation, since the additional expenses incurred by evaluations tend to be very modest in relation to a measure's overall costs when planned in advance.

The institutionalisation of evaluations plays a key role in achieving the goal of an evidence-based allocation of resources. In the United States, specific institutions have been established for this purpose. In 2009, the Department of Labor set up its own Chief Evaluation Office (CEO), which serves as the central office for the evaluation of all labour market measures applied by the ministry. The CEO does not set its own goals, but rather aims to evaluate specific measures or arrange for their evaluation. 0.5 percent of each project's budget is allocated to the evaluation team (up to 1 percent from 2013).[26] The CEO's total budget for the planning and coordination of evaluations currently amounts to approximately 40 million US dollars. The impact analyses themselves are carried out by specialised institutes, and the main focus is placed on the efficiency and effectiveness of programmes. The CEO is responsible for identifying useful measures, for improving existing measures and uncovering new potential. For the purpose of quality assurance, the CEO also provides a publicly accessible data set of each of the CEO's quantitative evaluations, thereby allowing to reproduce the main findings of evaluations. Furthermore, all evaluations commissioned by the CEO are peer-reviewed. These elements of quality assurance are of great importance.

But also in other policy areas, the evaluation of measures is an integral part of political practice, education policy being a prime example. Here, the United States have set milestones in educational research, especially through the use of randomised experiments.[27] Since many of the programmes are launched at a local level, it is particularly important to coordinate and disseminate information efficiently. In the area of education, this is primarily done through the website WhatWorksClearinghouse (WWC). The WWC gathers and assesses studies on various educational programmes (cf. Box 1), which are then presented on a central website. For each measure, the WWC website provides a comprehensible overview of the contents, purpose and effects of a measure. To date, slightly less than 40 percent of the total of studies assessed have been classified as largely compliant with standards, and only about 20 percent have been classified as fully compliant with standards. From among those 328 (methodologically valid) results



Laser particle acceleration.



Ion beam.

BOX 01

How to recognise good impact analyses

To assess the validity of an evaluation's results, it is first crucial that the study is implemented transparently in every respect. In particular, this includes a clear presentation of the implementation of the evaluation. As part of an experimental study, it must be documented that groups have been assigned prior to the experiment, how the grouping was conducted and whether the random assignment led to groups that are actually comparable. This is to prevent results that are based on group differences from being attributed to the support measure itself. Studies designed to produce valid results should have large sample sizes so as to rule out random effects. Furthermore, to isolate the cause of an observed change, evaluations should be conducted in an environment that is as normal as possible. Ideally, the evaluation of a measure should be conducted in more than one location to prevent the particular circumstances of a region from producing spurious results. Prior to an analysis, it should be established based on which criteria results are to be determined. These results should be presented in a clear-cut way to allow the assessment of the measure's effectiveness. If the desired evaluation methods cannot be implemented, or if the random assignment has proven to lead to incomparable groups, the most important factors influencing the evaluation's results should be controlled for. The results of studies that merely show a before-and-after comparison or those that do not sufficiently control for influencing factors cannot be regarded as valid results. In such cases, it cannot be ensured that results are actually a consequence of the support measure and are not due to other, unobserved factors. However, whether a support measure in whole can be regarded as scientifically well-founded does not only depend on the quality of the studies, but also on the quantity of evidence. Therefore, to produce reliable results, a high-quality study should be replicated in at least one additional, representative environment.

from 209 measures, only about half of the results have been positive or potentially positive.[28] A comparable project is currently being conducted in the area of combating crime. The US Department of Justice has launched the website CrimeSolutions.gov, a site that aims to present impact analyses on policy measures, such as measures to promote the reintegration of youths that have been convicted of criminal offences. What both websites have in common is that they do not only include domestic studies in their databases, but also take into account studies from other countries. Thus, in terms of transparent evaluations of policy measures, the United States are very well positioned.

Some of the European countries also present findings systematically. For example, the University of Aarhus operates the Danish Clearinghouse for Educational Research, which serves a similar purpose as the aforementioned websites.[29]

The Canadian Social Research and Demonstration Corporation (SRDC), founded in 1991, is another example of good practice in North America. The SRDC evaluates policy measures for a number of public authorities. The SRDC specialises in test projects that are evaluated with a number of different methods, depending on the project. The SRDC uses both quantitative and qualitative methods. From the outset, the SRDC makes use of randomised experiments to facilitate well-founded evaluations.

In 2010, Great Britain established its Behavioural Insights Team (BIT). This facility aims to identify ways of making the government work in a more efficient and target-oriented way. In cooperation with authorities, the BIT examines e.g. strategies for improving the services of employment agencies, and strategies for making traffic offenders pay their outstanding fines.[30] At the BIT, the collection of data required for a well-founded impact analysis is integrated into the planning phase of the project to ensure a cost-effective evaluation. Results are then managed by centralised evaluation teams, which combine the findings of various projects. Following a trial period, the BIT's funding was continued and the team was expanded due to its great success. The BIT is now to become partly privatised.[31]

Since 1999, Switzerland has analysed the effectiveness of public policies as part of its constitution. Article 170 of the Swiss Constitution states: "The Federal Assembly shall ensure that federal measures are evaluated with regard to their effectiveness." The Swiss Agency for Development and Cooperation, Switzerland's national development aid organisation, has derived its own evaluation standards from this constitutional mandate. According to these standards, 0.6 to 0.8 percent of project funds must be allocated to the impact analysis budget.

In the Netherlands, the institutionalisation of the evaluation process was introduced shortly after the Second World War. There, the Central Planning Bureau (CPB) is in charge of evaluating economic policy measures. Above all, this includes the ex-ante cost-benefit analysis of policy proposals. In recent years, however, the CPB has also been initiating the use of scientific ex-post evaluations, while making frequent use of both natural[32] and actual experiments. The issuing of innovation vouchers to small and medium-sized enterprises (SMEs) may serve as an example of actual experiments.[33] With the help of these vouchers, businesses were able to award small projects to research institutions. The aim of the experiment was to determine whether vouchers are an effective means of promoting the innovation activities of SMEs, and whether vouchers can strengthen relations between SMEs and scientific institutions. The efficiency of the measure could be shown easily and convincingly as a limited number of vouchers was randomly assigned to candidates in an oversubscribed lottery.[34] It is planned to increase the use of scientifically founded evaluation measures in the future.[35]

These examples demonstrate that the implementation of well-founded evaluations is both possible and desirable. Tests conducted in other countries have been extended without exception and often play an important role in political decision-making. Germany could also benefit from such developments.

Challenges in the evaluation process

Evaluations provide important information on the effectiveness of a measure in the observed context. Yet, when it comes to provisional measures that have been introduced for test purposes, it is important to assess to what extent their effects can be generalised once the measure is introduced on a broader basis. In particular, it is essential to examine whether the results from a specific testing environment are transferable to other parts of the population. Another important point in the evaluation process is that long-term or downstream effects cannot be conclusively assessed if the evaluation period is too short.[36]

The implementation of evaluations must also be designed in a way that allows for a relevant selection of quantitative and qualitative criteria that precisely suit the respective measures.[37]

Due to major advances in methodology, it is possible to draw valid conclusions even in cases where experimental evaluations are not feasible.[38] A suitable design of the implementation of the measure can then facilitate well-founded results that come close to experimental standards. Qualitative evaluations can also contribute greatly to the evaluation of economic policy measures. In any case, the introduction of scientifically sound evaluations represents a significant advance towards an evidence-based innovation policy.

Recommendations

The Expert Commission recommends the systematic use of scientifically well-founded impact analyses in research and innovation policy measures. Already in the planning of measures, concrete objectives should be set and, whenever possible, suitable indicators should be defined that will allow for a subsequent assessment of whether objectives have been achieved. It is essential to consider evaluations already in the project planning phase and to immediately record the data required.

Evaluations provide the basis for informed political decisions. As a key prerequisite, evaluations have to be conducted without prejudging the outcome, and both positive and negative results have to be appreciated as knowledge advancements. Evaluations should not be regarded as a potential career threat by those responsible for a measure. To keep those in charge from refusing impact analyses on these grounds, the success criterion of an evaluation should be knowledge advancement and the improvement recommendations derived from it rather than a positive assessment of a measure. In particular, political

stakeholders should create the opportunity to reallocate budgets and to further develop measures instead of completely terminating them.

The Expert Commission recommends the establishment of centralised evaluation offices at ministerial level to be in charge of implementing and assessing scientifically valid evaluations. First attempts – such as the establishment of a staff unit for assessing and controlling support measures at the Federal Ministry for Economic Affairs and Energy – should be systematically advanced. The Expert Commission particularly recommends the creation of a Chief Evaluation Office following that of the US Department of Labor in all those federal ministries that have major budgets for the promotion of research and innovation. These centralised bodies should operate independently and should define standards for evaluation. Furthermore, the Expert Commission recommends introducing research data centres in all of the ministries concerned. These centres should have the purpose of collecting and processing data of all projects that aim to promote research and innovation. Data should be made available to researchers to be used for evaluation purposes. To allow the commissioning of evaluations by the ministries, the centralised evaluation bodies should receive 0.5 percent of the programme budget of all ongoing projects. These funds shall be allocated to evaluations and their methodological advancement. When putting evaluations up for tender, special attention should be attached to the methodological expertise of the tendering parties. The Expert Commission recommends fostering quality-based competition and focussing on providing ongoing incentives for new, qualified tenderers. This can be achieved through suitable tendering procedures and by considering international bidding consortia.

To improve the transparency of evidence-based support measures, the Expert Commission recommends the launch of central websites by the respective ministries. This should be done by using the above best practice examples and possibly in cooperation with the respective institutions. These websites should collate all results of evaluations in the area of research and innovation that aim to increase the effectiveness of individual measures. The collated results should be assessed according to their methodological quality.

To ensure that evaluation standards are implemented in the ministries, the Expert Commission recommends the provision of systematic training for ministerial staff and the recruitment of new staff with strong methodological skills. An ongoing exchange between ministerial staff and external empirical researchers in the form of further training and workshops would further support the development of skills.

The Expert Commission finally recommends promoting the implementation of these efforts also at a European level. By launching suitable initiatives, the Federal Government could become a pioneer in the field of evidence-based innovation policy.