
Evaluation of research and innovation policies: a discussion of trends with examples from Germany

Stefan Kuhlmann

Fraunhofer Institute for Systems and Innovation Research,
Breslauer Str. 48, 76139 Karlsruhe, Germany
E-mail: Stefan.Kuhlmann@isi.fhg.de

Abstract: Recent changes in the field of evaluation refer to new demands by politics, economics and society to extend the subject of evaluation processes to cross-sectoral research promotion programs and research institutions, and new developments within the research of evaluation itself. The paper presents an overview of these trends and consequences for the function and methods of evaluation of research and innovation policies against the background of recent German experiences.

Keywords: Self-referential evaluation; German research system; evaluation of innovation programs; summative and formative evaluation.

Reference to this paper should be made as follows: Kuhlmann, S. (2003) 'Evaluation of research and innovation policies: a discussion of trends with examples from Germany', *Int. J. Technology Management*, Vol. 26, Nos. 2/3/4, pp.131-149.

Biographical notes: Stefan Kuhlmann, is head of the Technology Analysis and Innovation Strategies department of the Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe, Germany. He is also Professor of Innovation Policy Analysis at Utrecht University, Copernicus Institute for Sustainable Development and Innovation, Dept. of Innovation Studies. As a political scientist Kuhlmann focuses on public policy analysis, in particular the analysis, planning and evaluation of research, technology and innovation policies (national, international, regional). He is on the editorial advisory board of *Evaluation* (Sage) and author of numerous books and articles on research and evaluation policy.

1 Introduction

Institutions of scientific research and communication among themselves can be described as a self-referential system, whose performance-orientation and quality criteria are sharply differentiated from those of other social subsystems [1]. Part of the evaluation procedures used today – above all the assessment of scientific performance, projects and publications by colleagues (*peer review*) – is rooted in this self-reference. But it has its limits: more than a third of all science and research conducted (in Germany) is paid for with public funds, and some two-thirds are financed by industrial companies [2]. Politics, economics and society are demanding – and increasingly so since the 1990s – evidence of the performance, quality and benefits of government-promoted science and research, partly because public funds are scarce and should be employed as 'effectively' as

possible, and partly because policy makers and the public cherish a certain scepticism regarding the self-centred workings of the science system. Large research projects, cross-sectoral research promotion programs and research institutions increasingly became the subject of evaluation processes: the equally increasing relevance of industrial evaluation practice can only be mentioned in passing in the present paper.

Parallel to the spread and increasing differentiation of these attempts at assessing success, research into evaluation has also developed. It analyses the conditions for application as well as promises for success of evaluation in the area of research and innovation promotion, and develops further the concepts and methods used. This paper sketches an overview of this development, taking the case of Germany as one example.

The reader will find firstly a short description of trends of government research and technology promotion (Section 2), then (Section 3) an overview of the most significant concepts, application areas and methods of evaluative procedures in this policy field, then (Section 4) a sketch of the continuous structural change in research and innovation systems, as well as (Section 5) a discussion of the consequences for the utilisation of evaluation procedures in the process of political decision making; an outlook for the future concludes the paper (Section 6).

2 Trends in public research and innovation policies

Government interventions in research, technology and innovation have a long tradition and today indisputably form an independent field of state policy, (for the German case, see e.g. [3–5]). When research and innovation policy is referred to in the following, we mean primarily all the efforts on the part of the political system to shape the ‘research and innovation system’, here understood as the ‘landscape’ of research and technology-developing, innovation-facilitating institutions, including their regulations (see Section 4 below). At the latest from the 1970s, public initiatives have been aiming at stimulating the innovative behaviour of industrial enterprises [6–9]. It must be emphasised that research and technology policies show significant interfaces to other policy fields, especially education policy (education, vocational training, further education), economic policy (structural change, trade policy), legal and domestic policy, environment and transport policy, etc. These policy areas influence either decisive frame conditions for research and innovation on the supply side (infrastructure, qualified personnel, etc.) or operate from the demand side (such as transport infrastructure, industrial safety, environmental regulations).

In an international comparison, the German research and innovation system can be ranked as relatively well developed: in 2000 approx. 480,000 people (full-time equivalents) were employed in research and development (R&D); the total expenditures for R&D amounted to approx. 50,1 bn Euro, which corresponds to 2.45 % of the GDP [10]. The research infrastructure can be characterised as comparatively differentiated:

- *Industry* carries out the greatest part of research and development (R&D) in Germany (306,700 employees in R&D): in 1999 industry invested ca. 33,6 bn. Euro, for the most part in applied research and experimental development. Only a very few large multinational enterprises, especially in the chemical and electrical engineering industry, carry out long-term application-oriented basic research themselves.

- The second largest share of expenditure on research (8.1 bn Euro, mainly provided by the federal states) falls to the *higher education institutions (HEIs)* (345 national or officially recognised HEIs, including 116 universities and comparable institutions; 101,500 employees in R&D). They concentrate on basic research and long-term application-oriented research, for the most part financed by the federal states and the German Research Association (DFG), a government-funded but largely independent and self-organised body in research promotion. Since the 1980s the share of industrial research contracts in the research budgets of individual universities, especially technical ones, increased significantly (to 11.3% = 910 million Euro in 2000).
- The 15 ‘national research centres’ of the federal *Helmholtz Society* (21,500 R&D employees; 2.1 bn. Euro), principally above all long-term oriented research, which is considered too risky, or entails high costs (plants/facilities) and large research teams. In the past years the Helmholtz Centres have transformed their fields of activity significantly, ranging today from high-energy physics to space technology, medicine, biotechnology, applied mathematics and software development up to environmental technology. Moreover, various federal ministries maintain so-called *departmental research institutions* to execute their scientific-technical tasks.
- The institutes of the *Max Planck Society* (9,200 R&D employees, 1.0 bn Euro), a research organisation, which has its origins in the ‘Kaiser-Wilhelm-Gesellschaft’ established in 1911, concentrates on selected fields of basic research in the natural sciences and humanities. It focuses above all on those research areas in which a significant knowledge and development potential is assumed, while not yet anchored in university research, either because of their interdisciplinary character or because of the resources required.
- The institutes of the *Fraunhofer Society* (ca. 9,000 R&D employees, 730 million Euro) aim to promote the practical utilisation of scientific knowledge via long-term application-oriented and applied research. The Fraunhofer Society principally carries out contract research, which is financed partly by industry and partly by government bodies. The society regards itself as an ‘interface’ between science and industry in Germany. It was established in 1949; no other research organisation in Germany has grown so rapidly in the past 25 years.
- The institutes of the *Science Association G.W. Leibniz (WGL)* (formerly the ‘Blue List’) form a last category, which has in common that they are supported by the federal government and the federal states (ca. 10,000 employees, 811 million Euro). After the fusion of the East and West German research systems in 1990, many East German institutions found their home base here.
- Finally, the research institutions of the *Confederation of Industrial Research Associations (AiF)* principally carry out applied research and experimental development for the sector-specific needs of industrial enterprises. Their research palette, which is partly financed by public means and partly by industry, is geared especially to small and medium-sized enterprises, which are organised in industry-sectoral research associations.

In 2000 the federal government spent a total of 8.4 bn Euro on research and development. The spectrum of instruments available to German research and technology policy is widely differentiated meanwhile (see Figure 1), ranging from *institutional* support for research facilities (roughly half of the federal expenses), over various forms of financial incentives (programs; roughly the other half of federal expenses) for research and experimental development in public or industrial research laboratories, to the creation of an ‘*innovation-oriented*’ *infrastructure*, including the institutions and mechanisms of technology transfer. These instruments have characterised, by and large, the practice of research and innovation policy in the Federal Republic of Germany since the 1970s.

Figure 1 Instruments of government research and innovation policy

<i>Instruments in the narrower sense</i>	<i>Instruments in the wider sense</i>
1 Institutional support <ul style="list-style-type: none"> • National science centres • Max Planck Society • Fraunhofer Society • Higher education institutes • Other institutions 	4 Education, training and further education <ul style="list-style-type: none"> • Schools, HEIs, enterprises
2 Financial research and innovation incentives <ul style="list-style-type: none"> • Research programs and cooperative projects • Innovation programs (indirect support) • Venture capital 	5 ‘Discursive’ measures <ul style="list-style-type: none"> • Evaluation of innovation policy • Technology assessment • Long-term visions • <i>Awareness</i> measures
3 Other infrastructure and technology transfer <ul style="list-style-type: none"> • Information and advice for SMEs • ‘Demonstration centres’ • ‘Technology centres’ • Cooperation, networks 	6 Public demand
	7 Related policy fields <ul style="list-style-type: none"> • Industry and competition policy • Regulatory policy, eg influencing private demand • Social policy

Source: Meyer-Krahmer and Kuntze [11]

3 Evaluation processes

3.1 Concepts and areas of application

In Germany, as in the other industrialised countries, two heterogeneous lines of development of evaluation procedures in the area of research and innovation policy can be determined in a historical perspective [12] They can also be depicted as a *shell model*

with three layers (see, for overviews in similar ‘evaluation cultures’ in other industrialised countries, [13–16]):

First shell – individual research performance: the ‘core’ was formed by *peer reviews* and later by additional procedures to measure the research performance of individual researchers and groups (bibliometrics etc.) as *internal, scientific* instruments for deciding on the allocation of promotional funds to research institutions [17–19]. *Peer review* procedures are in widespread use in the German research system, especially in the ex ante evaluation of projects in basic and long-term application-oriented research [20]. Peer review is the predominant evaluation instrument of the German Research Association (DFG). The DFG plays a central role in the promotion of basic research in universities, principally by granting individual researchers funds on application (so-called standard procedure). Applications for grants are assessed by *peers*, who are elected every four years by the entire scientific community [8]. Each expert is advised to judge the application on the basis of its scientific quality alone.

Second shell – programs: around this core of peer review procedures a ‘shell’ was formed, which consists of evaluation studies in the sense of *impact analyses* of research and innovation policy programs. The studies carried out in Germany up till now mainly follow the conceptional tradition of *impact analysis*, the basis for which is formed by the manifold and traditional work above all of US impact research, together with approaches from *policy analysis*, which has developed into an instrument of policy advice used in many political fields (especially social politics) (on the ‘policy-analytic movement’: see [21], for overviews [22–25]). This evaluation concept can be characterised as external to science, because it is mainly set in motion by actors from the policy administrative system, with the intention of checking the attainment of politically set scientific, technological, economic or societal goals. Impact analyses have gained acceptance in Germany since the 1970s in many political fields with the spread of program policy [26–29]. Since this time approx. half of the public research budget has been directed towards project or program promotion, the other half towards support for institutions: in contrast to institutional promotion, programs are designed to pursue and control specific political aims, e.g. in the area of promoting ‘key technologies’ or stimulating innovation activities in medium-sized firms (*Mittelstand*). Political claims for control call for an efficiency review: program evaluation and impact analysis have experienced a considerable upswing since then [30]. The spread of evaluation is closely bound up with the increase in strategic programs (initiated by the EU Commission) to promote science and technology. In Germany the Federal Ministry for Research (then BMFT) was particularly active; between 1985 and 1993 some 50 large evaluation studies were commissioned [31], and in the years following the efforts did not slacken. As a rule, independent research institutes acted as evaluators on behalf of research policy administrators who are in charge of running the evaluated programs (i.e. the neutrality of evaluation may be questioned). The experience of over two decades of program evaluation led to the establishment of a certain ‘evaluation scene’ in the German-speaking area, consisting of a group of experts and institutes from the field of economics and social sciences, who use a relatively broad spectrum of evaluation concepts, methods and instruments [32] and who have been organised professionally in a ‘German Society for Evaluation (DeGEval)’ since 1998.

Third shell – institutions: here the performance of entire research institutions is dealt with. In Germany, evaluations by the ‘Wissenschaftsrat’ (Science Council) have been

playing an important role for a long time; in the restructuring of the ‘research landscape’ of East Germany after reunification they even assumed a shaping function [33]. Since the 1990s, evaluations of institutions have been carried out with greater frequency [34]. In spring 1999 an international commission completed a ‘system evaluation’ of the German Research Association (DFG) and the Max Planck Society (commissioned by the heads of federal and state (Länder) governments, [35]). At the same time, a system evaluation of the Fraunhofer Society was carried out [36]. An evaluation of the institutions of the ‘Science Community G.W. Leibniz (WGL)’ was completed in 2000 [37,38]; also a system evaluation of the national science centres in the ‘Helmholtz Society’ [39] as well as a strategic evaluation of the ‘joint industrial research’ mechanisms (see contributions by Blum and Lorenzen in this issue). Parallel to the growing interest in the performance of publicly promoted research programs and institutions, private enterprise is paying increased attention to the efficiency, effectiveness and strategic slant of its laboratories [40–42].

To sum up, German evaluation practice in the area of judging the performance of research and research institutions can be characterised as both strong and fragmented at the same time: strong because of the high degree of self-organisation, which encourages consensus and commitment among researchers; unsystematic and fragmented, because up to now few efforts have been made to coordinate and systemise evaluation practices.

3.2 *Potential impacts of research and innovation funding*

Since any public funding of research and innovation involves a variety of actors with differing interests and expectations – policy makers, program managers, participating researchers and the program’s clients – the potential variety of results, outcomes and impacts is huge. A well-designed evaluation has to take these different perspectives into account and to reconcile related desires for information with constraints on resources and the availability of information. Figure 2 shows different classes of impact (short or long-term; direct or indirect) in three different domains: the worlds of science, of the economy and society, and of policy making respectively.

Figure 2 Impact dimensions of public research spending

Main domains of impact of public RTD spending	Direct impacts		Indirect impacts	
	Short-term	Long-term	Short-term	Long-term
<i>Science</i> (‘Wissenschaft’) Typical impacts	scientific findings	knowledge	improved teaching	industrial spill-overs
Economy and society Typical impacts	improved technology	improved technical know-how	increased productivity	improved competitiveness
<i>Policy</i> Typical impacts	improved understanding	problem-solving	increased problem awareness	increased general satisfaction

Source: Airaghi *et al.* [43]

The ‘assessability’ of each of the cells in the matrix differs considerably: the darker the cell, the more difficult a direct attribution of policy input and potential impact, and the more challenging is the task of evaluation: capable and effective evaluation concepts and methodologies are needed.

3.3 Evaluation methods

Nowadays we have many *methods* to determine achieved or achievable impacts of research and innovation funding and related policies. The most important are before/after comparisons, the control or comparison group approach, as well as qualitative analyses (among others, plausibility checks, estimated judgements). These concepts can be carried out individually or in combination with different indicators (financial expenditure for research and development, patents, economic, social, technical indicators, publications, citations etc.), data collection methods (existing statistics, questionnaires, interviews, case studies, panels etc.) and data analysis methods (econometric models, cost/benefit analyses, other statistical methods, technometrics, bibliometrics, peer reviews) [30,13,44]. Despite all the efforts made to arrive at objectivising procedures and appropriate indicators, one must, however, warn against regarding quantitative indicators alone as adequate to evaluate research and innovation promotion [45]. The understandable desire for a standardised ‘indicator tool-box’ is not compatible with the tendency to pursue complex political goals [43,46].

A methodologically well-balanced concept of ex post evaluation of political programs, acceptable from the perspective of evaluation research, would contain the following fundamental questions [47,31,48,30,49,] (Figure 3): Does the evaluated policy measure make possible adequate solutions for the basic technical, scientific, economic or societal problem which sparked off the policy intervention in the first place? Are the assumptions on which the program concept is based correct as regards problem perception and causes; is the program in this sense ‘strategically efficient’? Are possible, noticeable structural changes in the promotional field effects of the measure, or have such structural changes for their part influenced the potential policy impacts? The attempt to answer these questions raises a whole series of methodical, conceptual and empirical problems, which must be solved in each case. In most cases they are only to be managed within well-defined conceptual boundaries (if at all). A crucial methodological problem has its roots in the practice of policy generation: only in exceptional cases are the research and technology policy targets formulated so explicitly and clearly that their attainment can be checked ([50] examined this question in several industrialised countries in detail; see also [51]).

Figure 3 Basic elements of an evaluation concept

1	Is the program appropriate? Are the basic assumptions right?
2	Is/was the target group reached?
3	Which direct and indirect impacts are there /are foreseeable?
4	Were the targets attained, or rather are they attainable?
5	Are the implementation and administration efficient?

The further questions of an evaluation concept concern the impacts and the conduct of the program: Was the target group reached? Which direct and indirect impacts can be attributed to the program? Were the program targets ‘achieved’? This question is by no means trivial, not only because of the vagueness of target-setting, but also because of the problem of attributing ascertainable changes to the stimuli given by the program. Besides the target attainment, questions should be asked about the ‘free rider effects’ and the appropriateness of the implementation and the administration of programs; these elements affect the operative efficiency of the measure being evaluated and belong to the methodically farthest developed in evaluation practice.

Objectivity and scope of evaluation processes

Even although evaluation processes are applied and accepted today in wide areas of public research and innovation policy, the ‘correctness’ and objectivity of the results and the recommendations derived from them are still doubted time and again with regard to inadequate evaluation methods. The methodological reliability of evaluation studies is doubted not only by policy makers but also by social scientists. Schön and Rein [21] identify the following problems:

- The evaluation criteria applied overlook/neglect the fact that most programs pursue ‘multiple, conflicting, and evolving purposes’.
- Program results are often evaluated without a proper understanding of the context of their origin (‘black box problem’).
- Evaluation has gained the reputation of being a ‘killer’ when it has proved the (partial) failure of a program, without concern for the possible causes.
- Many critics of program evaluations point out that these, as a rule, adopt the perspective of the political administration responsible for the program (or state supervisory bodies), but neglect the interests of other ‘concerned parties’ (e.g. citizens’ groups).

Besides these general problems of program evaluation, specific problems arise in the area of research and innovation policies [43]:

- 1 Research and innovation have various effects. To the typically short-term desired effects for the participants in promoted projects belong increases in turnover and enlarged market shares, improvements in know-how and new contacts. The effects of research, however, far exceed this: a project can have manifold impacts on actors who did not directly participate in it – such influences are difficult to measure.
- 2 Evaluation results are often needed early for policy decisions, sometimes even before the research work is completed and normally before the entire scope of possible socio-economic effects can have become apparent.
- 3 Many effects cannot be attributed unambiguously to one distinct project or program. They can be the result of a combination of influences, to which the processes of converting research results into practice also belong.

The obvious limitations in the performance of the evaluation process can only be compensated for if its application context is deliberately kept in mind and its results are judged with caution:

- Evaluators and the users of evaluations must be clear about what they really want to know, which *questions* are to be posed in what breadth, depth and radicalism. The analytical scope of evaluation studies, however, always remains limited: if one assumes in principle an unlimited impact area of policy measures, in which *other effects* are also simultaneously at work, then the reconstruction of impact interactions causes more problems, the more comprehensive and generalised the definition of the input and output criteria.
- Evaluation research in the area of research and innovation policy assumes today that ‘realistic’ information about an object for evaluation can only be obtained by the combined application of various social science *methods* and *indicators*.
- Nevertheless, the ‘*objectivity*’ of evaluation results is *narrowly limited*. Taking into consideration the perspectives of various and different actors can enhance it.

Caution and care are called for. Innovation research knows today, that the connections between ‘investment’ in research and innovation and possible scientific, economic, social or ecological ‘returns’ are complex and cannot in any way be constructed as a simple input/output model [44]. This holds even more true since research and innovation systems are in transformation.

4 Research and innovation systems in transformation

The ‘research and innovation system’ of a society comprises, according to the internationally accepted understanding, the ‘cultural landscape’ of all those institutions, which are engaged in scientific research, which accumulate and disseminate knowledge, educate employees, develop technology, which create and distribute innovative products and processes; in this category also belong appropriate regulatory regimes (standards, norms, laws) as well as government investments in appropriate infrastructures [52–55]. The innovation system covers schools, universities, research institutes (education and science system), industrial enterprises (economic system), the policy-administration and intermediary instances (political system) as well as the formal and informal networks of the actors of these institutions. As a ‘hybrid system’ it represents a sector of society, which radiates far into other areas, via the education system, or via entrepreneurial innovation activities as well as their socio-economic effects [56]. No innovation system is identical to another, just as no society is identical to another [57]. Efficient innovation systems develop their special profiles and strengths only slowly, in the course of decades or even centuries. They are based on stable exchange relationships between the institutions of science and technology, industry and the political system.

Modern innovation systems were always, and at present are particularly under *pressure to adapt* to changed environments: globalised markets, revolutions in the geopolitical situation, as well as worldwide interconnected new communication infrastructures and technologies are transforming the innovation scene, which for its part also influences the international integration process through its own dynamic momentum:

- The development of high-tech products has *accelerated* significantly in the past years; the creation of a new idea and its commercialisation in the market follow each other today as rapidly as never before. Changed market requirements and the use of information and communication technology have contributed fundamentally to this.
- In many areas the ‘*science base*’ of innovation activities; the functions and procedures of ‘targeted’ basic research and its interfaces in the development of marketable products demand growing attention [58,59]. An often-cited example for this trend is the new significance of research in molecular biology for the pharmaceutical industry and practising medicine.
- Complex products and processes, however, require closer integration and the ‘fusion’ [60] of heterogeneous technology developments, by means of which traditional demarcations between knowledge and technology areas disappear and *interdisciplinary competencies* gain significance in the innovation process [61], new disciplinary clusters arise and old ones disappear.
- These mentioned trends reinforce as a whole the requirement to *cooperate* on the part of innovation actors: with increasing complexity of the necessary knowledge, isolated actors are less and less in a position to master this adequately without external support (see on this and the discussion about a transition of the production of scientific knowledge from Mode 1 to Mode 2 [62–66]). Innovation-oriented cooperation and the maintenance of corresponding networks [67] in the meantime belong to the day-to-day innovation activities – between enterprises as well as in the framework of joint research and development plans between enterprises and public research institutions, increasingly also in international networks [68].
- At the same time, we find a trend to establish especially strong, thematically focused innovation capacities in one particular location, which compete worldwide with only one or two other locations (e.g. vehicle construction in Germany and the USA; gene technological pharmaceuticals in the USA; home electronics and office technology in Japan; segments of the machine tool industry in Italy and so on). These ‘locations’ have also been graphically described as ‘*Industrial Hollywoods*’ [69]: whoever wants to innovate successfully must orient himself according to the criteria of these centres, which are an especially efficient part of regional, national or sectoral innovation systems, and cooperate with them if necessary.
- The relation of codified and non-codified knowledge in the innovation process, the extent to which stocks of knowledge are fixed and can be transferred, is also subject to new dynamics. Computer networks accelerate the circulation of codified knowledge, which is only then effectively utilisable, when it can be recontextualised, i.e. when the *learning capability of the innovation actors* grows [70].

Government policy in Germany has been reacting to these trends in science and innovation by seeking to bring about structural change in the ‘research and innovation landscape’ by means of promotional policy, and by exercising pressure to modernise on research institutions:

- So since the mid 1990s many research and innovation policy *programs* have been launched as competitions, which aim to bring about a structural change in science and the economy: consortia of candidates (usually institutions) should in a self-organised process elaborate joint project plans and detailed goals. Partnerships among the various institutions (research facilities, universities, private firms, technology centres, further education establishments etc.) aim with a bundle of coordinated measures to optimise entire innovation systems. Such *multi-actor and/or multi-measure programs* can apply either to regions and/or certain aims and/or certain technologies and branches.
- The *altered requirements of institutions* were summed up concisely by the above mentioned 'system evaluation' of the Max Planck Society and the German Research Association (DFG) [35]: an international group of evaluators called, among other things, for a loosening of the strictly disciplinary orientation, the development of flexible and efficient forms of organisation for a temporary cooperation of different disciplines and groups in problem-oriented research fields, efficacious procedures for quality assurance with external participation, an improved collaboration between universities and non-university research institutions, the promotion of institutions from cross-disciplinary research centres as well as an increasingly international orientation and networking among the institutions.

5 Performance measurement or learning medium: evaluation between two functional poles

The sketched changes of the functional conditions for research and innovation have led to a growing interest in evaluation since the 1990s and have provided an impetus for the application of relevant procedures [71,16]. The expectations of evaluation processes are divided between two functional poles: evaluation can in the first instance serve to *measure performance* and thus provide the legitimisation for promotional measures afterwards (*summative function*), or it can be utilised as a '*learning medium*', in which findings about cause and effect links of running or completed measures can be utilised as intelligent information for current or future initiatives (*formative function*).

The summative pole is nurtured above all by the evaluation practice of Anglo-US countries: here in the framework of efforts to reform and cut costs in the public sector ('New Public Management') performance measurement procedures also gained great influence in research and innovation policy [72,73]. The US government and a majority of the states are increasingly implementing 'performance-based management and budgeting systems' – not least in research and innovation promotion [74]. Promoter and promoted are under growing pressure to prove the advantages of their activities, financed as they are by taxpayers' money. This is not just because of relevant new legal requirements – such as the 'Government Performance and Results Act (GPRA)' – or tight public budgets, but also because of an intensive public debate about justification, direction and advantages of public investments in research and innovation. An example of this is the 'Advanced Technology Program (ATP)' – a hotly debated government program in support of cooperative research and innovation projects between science and industry in risky high-tech areas, which aims in the long term at far-reaching diffusion effects; but the program is always being confronted with expectations of short-term,

measurable (quantifiable) impacts. Ruegg [75] observed that the program “as met nearly continuous demand for measures of impact of the program since the day it was established”.

As the complexity of research and innovation policy programs and also the tasks of institutions have grown rather than diminished, summative performance measurements soon reach their limits. *Formative, learning-oriented evaluation approaches* were therefore – partly in competition, partly as a complement to the summative – also further developed and applied. Proceeding from the lasting experience that evaluation results often produce only small impacts on political decision-making processes, and support only a few of the expectations and interests represented in a policy arena, evaluation experts (and increasingly also policy makers) tried to relax the boundaries between evaluation and decision-making processes, indeed even partly to integrate the two spheres. Guba and Lincoln [76] have sketched the problems concisely. They see a basic problem of previous evaluation concepts in their naive science-theoretical statement, that investigation results inform ‘how things really are and function’, as if they could not have been influenced by the evaluator, his/her commissioner and others.

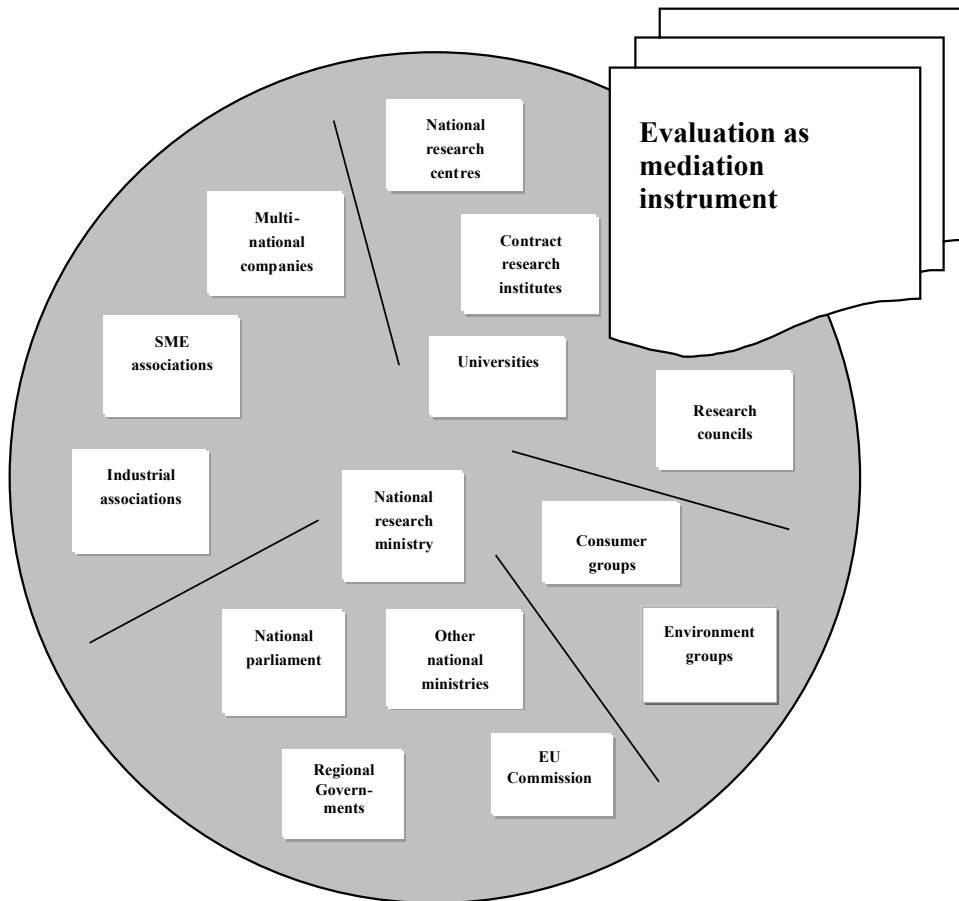
The key concept of the new, wide understanding of evaluation is ‘negotiation’ among the participating actors. The result of evaluations correspondingly designed, is, in contrast to conventional methodology, no longer “a set of conclusions, recommendations, or value judgements, but rather an *agenda for negotiation* of those claims, concerns, and issues that have not been resolved in the hermeneutic dialectic exchanges” [76]: decisions are made rather as a continuous process, in which competing actors achieve consensus interactively, or also not. Evaluation results are one piece of information among many. Here therefore the evaluation process, i.e. the communicating of the participating actors in its course, becomes crucial; the process is consciously designed to be ‘participative’ (‘Participatory Evaluation’; see [23,77]). The media aspect of the evaluation procedure takes centre stage. In particular, the following characteristics of the participatory evaluation approach can be further developed for use in research and innovation policy discussions (see also Figure 4):

- Evaluation is designed as a process of empirically, analytically prepared, structured presentation and confrontation of (partly conflicting) *actors’ perspectives*; the whole spectrum of evaluation methods (see Section 3) can be applied in this exercise.
- The evaluator acts as a *‘facilitator’*, supporting the *mediation of conflicts* in the negotiation arena through actors of the policy-administrative system. The mediation function of evaluation procedures is of particular importance in the case of experimental policies – at present a growing number of research and innovation policies [78].
- The target of the evaluation is not only an assessment of facts from the perspective of an individual actor (e.g. the policy-administrative system), or the examination of the appropriateness of a policy, but the stimulation of *learning processes* by breaking down rigid actor orientations, by investigating the multiple perspectives of involved actors on the role and effects of a policy measure – thereby exploiting the whole empirical tool box of social and economic sciences – the evaluation helps to illuminate the ‘expectable expectations’ of actors, thus providing *enlightenment and trust*. In this respect mediation-oriented evaluation concepts may be compared with the functioning of stock exchange and rating agencies for capital markets:

the trustworthiness of enterprises depends on ‘objectivised’ assessments of their expectable future performance, based on the expectations of other economic actors. There is nevertheless an important difference: finally, market information is measured in financial terms while research and innovation policy related evaluative information does not convert into one currency – on the contrary, it helps to articulate the diversity of ‘expectable expectations’ as a source of alignment and learning.

- Trust is a precondition for involved actors to *mutually provide evaluative information* on their activities, performance etc. The willingness to exchange information is crucial since the demand for evaluative information – given the wide array of potential performance and impact dimensions as depicted in Figure 2 above – is principally unlimited, i.e. unmanageable.

Figure 4 Evaluation as mediation instrument in the research and innovation policy arena



These evaluation concepts aim primarily to facilitate a ‘reframing’ [21] of the orientations of corporatistic and policy-administration actors. In the context of the research and innovation system they can be regarded as an ‘*intelligent provider of strategies for negotiation and management*’, not only for the responsible political actors but also the interested public. Crucial preconditions of performing this intelligence function are:

- 1 obviously a high level of *expertise* of the involved evaluators, not only with respect to methodologies but also concerning the thematic content and the particular policy arenas of the funding area under evaluation
- 2 an *independent status* of evaluators vis-à-vis political decision makers – the latter condition not satisfyingly being realised yet in Germany.

‘Intelligent’ policy development processes in this sense, which are similar to ‘discourse’ or ‘mediation’ processes in the context of environmental and technology policy [79] can moreover be enriched by combinations with [80,81]

- ‘foresight processes’ (‘technology foresight’) [82,83], with the intention of delivering debatable ‘visions’ of more or less desirable future developments, and
- ‘technology assessment’ [84] as the attempt to anticipate the possible positive or negative effects of technological developments and to feed the information gleaned in such studies back into the process of technology generation.

This potential of methodically enriched actor- and process-oriented evaluation processes has encountered growing interest in the research, technology and innovation policy arena in Europe. Recent examples for such concepts can be found in measures of the Federal Research Ministry for the structural modernisation of interdisciplinary research at German university hospitals [85,86], or in the promotion of complex ‘nano-technology competence centres’ [87].

6 Outlook

Besides the insider peer review process to assess the scientific performance of colleagues, evaluation in the field of research and innovation promotion takes place principally in the form of a critical external examination: representatives of the political system instigate evaluations of the success of promotional measures regarding the scientific and innovation system. Such an efficiency review can be conceived as the question as to whether ‘return on investment’ of taxpayers’ money is justified – then however there is the danger that in this short-sighted view, the diversity of possible long-term and indirect, desirable but also undesirable effects of creative science and innovation activity in industry and society can be overlooked.

The use of evaluation processes as a mediation instrument, which does not deny the diverging perspectives of participating actors from science, industry and policy making, but on the contrary deliberately makes different interests the subject of competing success criteria, can on the other hand bring about a ‘reflexive turning point’ vis-à-vis older, often naively rationalistic concepts of policy initiative planning. Mediation of stakeholder perspectives by evaluation will not, of course, bring about revolutionary changes of direction in research and innovation policy (this can only can take place in the

polity sphere). But the practical implementation of such radical changes is greatly facilitated by mediation underpinning the learning ability of the actors.

References

- 1 Luhmann, N. (1990) *Die Wissenschaft der Gesellschaft*, Frankfurt/M.
- 2 European Commission (1999) *Second European Report on S&T Indicators. Key Figures*, Luxembourg.
- 3 Grimmer, K., Häusler, J., Kuhlmann, S. and Simonis, G. (Eds.) (1992) *Politische Techniksteuerung – Forschungsstand und Forschungsperspektiven*, Opladen.
- 4 Martinsen, R. and Simonis, G. (Eds.) (1995) *Paradigmenwechsel in der Technologiepolitik?* Opladen.
- 5 Krull, W. and Meyer-Krahmer, F. (1996) (Eds.) *Science and technology in Germany*, London.
- 6 Ostry, S. and Nelson, R.R. (1995) *Techno-Nationalism and Techno-Globalism. Conflict and Cooperation*, Washington.
- 7 Roobeek, A.J.M. (1990) *Beyond the Technology Race. An Analysis of Technology Policy in Seven Industrial Countries*, Amsterdam et al.
- 8 Neidhardt, F. (1988) *Selbststeuerung in der Forschungsförderung. Das Gutachterwesen der DFG*, Opladen.
- 9 Ergas, H. (1987) 'Does technology policy matter?', in B.R. Guile and H. Brooks (Eds.) *Technology and Global Industry. Companies and Nations in the World Economy*, Washington D.C., pp.191-245.
- 10 BMBF – German Federal Department for Education and Research (2002) *Faktenbericht Forschung 2002*, Bonn.
- 11 Meyer-Krahmer, F. and Kuntze, U. (1992) 'Bestandsaufnahme der Forschungs- und Technologiepolitik', in K. Grimmer, J. Häusler, S. Kuhlmann and G. Simonis (Eds.) *Politische Techniksteuerung – Forschungsstand und Forschungsperspektiven*, Opladen, pp.95-118.
- 12 Kuhlmann, S. (2000) 'Evaluation in der Forschungs- und Innovationspolitik', in R. Stockmann (Ed.) *Evaluationsforschung*, Leverkusen, pp.287-307.
- 13 Bozeman, B. and Melkers, J. (Eds.) (1993) *Evaluating R&D Impacts: Methods and Practice*, Boston et al.
- 14 Georghiou, L. (1995) 'Research evaluation in European national science and technology systems', *Research Evaluation*, Vol. 5, pp.3-10.
- 15 Georghiou, L. (1998) 'Issues in the evaluation of innovation and technology policy', *Evaluation*, Vol. 4, pp.37-51.
- 16 OECD – Organisation of Economic Cooperation and Development (Ed.) (1997) *Policy Evaluation in Innovation and Technology. Towards Best Practices*, Paris.
- 17 Daniel, H.-D. (1993) *Guardians of Science. Fairness and Reliability of Peer Review*, Weinheim et al.
- 18 Hornbostel, S. (1997) *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*, Opladen.
- 19 van Raan, A.F.J. (Ed.) (1988) *Handbook of Quantitative Studies of Science and Technology*, Amsterdam.
- 20 Campbell, D.F.J. and Felderer, B. (1997) *Evaluating Academic Research in Germany. Patterns and Policies*, Vienna.
- 21 Schön, D. and Rein, M. (1994) *Frame Reflection. Toward the Resolution of Intractable Policy Controversies*, New York.
- 22 Widmer, T. (1996) *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*, Bern et al.

- 23 Patton, M.Q. (1997) *Utilization-Focused Evaluation. The New Century Text*, Thousand Oaks *et al.*
- 24 Kuhlmann, S. (1998) *Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik*, Baden-Baden.
- 25 Wollmann, H. (1998) *Evaluation Research and Politics: Between a Science-Driven and a Pluralist Controversy-Responsive Policy-Making Model. Potential and limitations*. Rome.
- 26 Derlien, H-U. (1976) *Die Erfolgskontrolle staatlicher Planung. Eine empirische Untersuchung über Organisation, Methode und Politik der Programmevaluation*, Baden-Baden.
- 27 Mayntz, R. (1980) *Implementation politischer Programme. Empirische Forschungsberichte*, Königstein.
- 28 Mayntz, R. (1983) *Implementation politischer Programme II. Ansätze zur Theoriebildung*, Opladen.
- 29 Hellstern, G.-M. and Wollmann, H. (1984) *Handbuch zur Evaluierungsforschung*, Bd. 1, Opladen.
- 30 Meyer-Kraemer, F. (1989) *Der Einfluß staatlicher Technologiepolitik auf industrielle Innovationen*, Baden-Baden.
- 31 Kuhlmann, S. and Holland, D. (1995) *Evaluation von Technologiepolitik in Deutschland – Konzepte, Anwendung, Perspektiven*, Heidelberg.
- 32 Becher, G. and Kuhlmann, S. (1995) (Eds.) *Evaluation of Technology Policy Programmes in Germany*, Boston *et al.*
- 33 Block, H-J. and Krull, W. (1990) 'What are the consequences? Reflections on the impact of evaluations conducted by a science policy advisory body', *Scientometrics*, Vol. 19, pp.427-437.
- 34 Kuhlmann, S. and Holland, D. (1995) *Erfolgsfaktoren der wirtschaftsnahen Forschung*, Heidelberg.
- 35 BLK, Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (1999) *Bericht der internationalen Kommission zur Systemevaluation der DFG und der MPG*, <http://www.blk-bonn.de/papers/forschungsfoerderung.pdf>.
- 36 FhG – Fraunhofer-Gesellschaft (1998) *Systemevaluierung der Fraunhofer-Gesellschaft*, Bericht der Evaluierungskommission, Download ftp://192.76.176.135/eval_fhg.pdf
- 37 Wissenschaftsrat (2000) *Systemevaluation der Blauen Liste – Stellungnahme des Wissenschaftsrates zum Abschluss der Bewertung der Einrichtungen der Blauen Liste*, Köln.
- 38 Röbbcke, M. and Simon, D. (1999) *Zwischen Reputation und Markt. Ziele, Verfahren und Instrumente von (Selbst)Evaluierungen außeruniversitärer, öffentlicher Forschungseinrichtungen*, Berlin.
- 39 Wissenschaftsrat (2001) *Systemevaluation der HGF – Stellungnahme des Wissenschaftsrates zur Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren*, Köln.
- 40 Brockhoff, K. (1999) *Forschung und Entwicklung. Planung und Kontrolle*, 5th ed., München *et al.*
- 41 Gerpott, T.J. (1999) *Strategisches Technologie- und Innovationsmanagement. Eine konzentrierte Einführung*, Stuttgart.
- 42 Bürgel, H.D., Haller, C. and Binder, M. (1996) *F&E-Management*, München.
- 43 Airaghi, A., Busch, N.E., Georghiou, L., Kuhlmann, S., Ledoux, M.J., van Raan, A.F.J. and Viana Baptista, J. (1999) 'Options and limits for assessing the socio-economic impact of European RTD programmes', *Report to the European Commission, DG XII*, Evaluation Unit, Brussels/Luxembourg.
- 44 Grupp, H. (1998) *Foundations of the Economics of Innovation: Theory, Measurement and Practice*, Cheltenham.
- 45 Barré, R. (2001) 'Sense and nonsense of S&T productivity indicators', *Science and Public Policy*, Vol. 28, pp.1-8.

- 46 European Commission (Ed.) (2002) *RTD Evaluation Toolbox, a Report Prepared by the Thematic Network 'Socio – Economic Evaluation of Public RTD Policies (EPUB)*, forthcoming.
- 47 Kuhlmann, S. (1998) 'Moderation of policy-making? Science and technology policy evaluation beyond impact measurement: the case of Germany', *Evaluation*, Vol. 4, pp.130-148.
- 48 Callon, M., Larédo, P. and Mustar, P. (1995) *La gestion stratégique de la recherche et de la technologie. L'évaluation des programmes*, Paris.
- 49 Rip, A. (1990) 'Implementation and evaluation of science & technology priorities and programs', in S. Cozzens (Ed.) *The Research System in Transition*, Boston, pp.263-280.
- 50 Cunningham, P.N., Georghiou, L.G., Barker, K.E., Kuhlmann, S., Reger, G., Marciano da Silva, C. and Henriques, L. (1994) *Analysis of Experience in the Use of Verifiable Objectives*, Brussels/Luxembourg.
- 51 Mowery, D.C. (1994) *Science and Technology Policy in Interdependent Economies*, Boston *et al.*
- 52 Freeman, C. (1987) *Technology Policy and Economic Performance: Lessons from Japan*, London.
- 53 Lundvall, B.-Å. (Ed.) (1992) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London.
- 54 Nelson, R.R. (Ed.) (1993) *National Innovation Systems: A Comparative Analysis*, Oxford, New York.
- 55 Edquist, C. (Ed.) (1997) *Systems of Innovation. Technologies, Institutions and Organizations*, London, Washington.
- 56 Kuhlmann, S. and Meyer-Krahmer, F. (2001) 'Internationalisation of innovation, interdependence, and innovation policy for sustainable development', in G.P. Sweeney (Ed.) *Innovation, Economic Progress and Quality of Life*, Cheltenham, pp.86-110.
- 57 Amable, B., Barré, R. and Boyer, R. (1997) 'Diversity, coherence and transformations of innovation systems', in B. Barré, M. Gibbons, Sir J. Maddox, B. Marin and P. Papon (Eds.) *Science in Tomorrow's Europe*, Paris, pp.33-49.
- 58 Schmoch, U., Hinze, S., Jäckel, G., Kirsch, N., Meyer-Krahmer, F. and Münt, G. (1996) 'The role of the scientific community in the generation of technology', in G. Reger and U. Schmoch (Eds.) *Organisation of Science and Technology at the Watershed. The Academic and Industrial Perspective*, Heidelberg, pp.1-138.
- 59 Meyer-Krahmer, F. and Schmoch, U. (1998) 'Science-based technologies: university-industry interactions in four fields', *Research Policy*, Vol. 27, pp.835-851.
- 60 Kodama, F. (1995) *Emerging Patterns of Innovation. Sources of Japan's Technological Edge*, Boston.
- 61 Schmoch, U., Breiner, S., Cuhls, K., Hinze, S. and Münt, G. (1996) 'The organisation of interdisciplinarity – research structures in the areas of medical lasers and neural networks', in G. Reger and U. Schmoch. (Eds.) *Organisation of Science and Technology at the Watershed. The Academic and Industrial Perspective*, Heidelberg, pp.267-372.
- 62 van der Meulen, B. and Rip, A. (1994) *Research Institutes in Transition*, Delft.
- 63 Kuhlmann, S. and Reger, G. (1996) 'Technology-intensive SMEs: policies supporting the management of growing technological complexity', in W. Cannell and B. Dankbaar (Eds.) *Technology Management and Public Policy in the European Union*, Luxembourg, Oxford, pp.73-102.
- 64 Meyer-Krahmer, F. (1997) 'Science-based technologies and interdisciplinarity: challenges for firms and policy', in Ch. Edquist (Ed.) *Systems of Innovation. Technologies, Institutions and Organizations*, London, Washington, pp.298-317.

- 65 Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. and Trow, M. (1994) *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*, London *et al.*
- 66 Etzkowitz, H. and Leydesdorff, L. (2000) 'The dynamics of innovation: from National Systems and 'Mode 2' to a Triple Helix of university-industry-government relations', *Research Policy*, Vol. 29, pp.109-123.
- 67 Callon, M. (1992) 'The dynamics of techno-economic networks', in R. Coombs, P. Saviotti, and V. Walsh (Eds.) *Technological Change and Company Strategies: Economic and Sociological Perspectives*, London *et al.*, 72-102.
- 68 Niosi, J. (Ed.) (1999) 'The internationalization of industrial R&D', *Research Policy (Special Issue)*, Vol. 28, pp.107-336.
- 69 Meyer-Krahmer, F. (1999) 'Was bedeutet Globalisierung für Aufgaben und Handlungsspielräume nationaler Innovationspolitiken?', in K. Grimmer, S. Kuhlmann and F. Meyer-Krahmer (Eds.) *Innovationspolitik in globalisierten Arenen. Neue Aufgaben für Forschung und Lehre: Forschungs- und Technologiepolitik im Wandel*, Leverkusen, pp.35-65.
- 70 Lundvall, B.-Å. and Borrás, S. (1998) *The Globalising Learning Economy: Implications for Innovation Policy*, Luxembourg.
- 71 Shapira, P. and Kuhlmann, S. (Eds.) (2003) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Cheltenham, forthcoming.
- 72 Osborne, D. and Gaebler, T. (1993) *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*, New York.
- 73 Shapira, P., Kingsley, G. and Youtie, J. (1997) *Manufacturing Partnerships: Evaluation in the Context of Government Reform, Evaluation and Program Planning*, Vol. 2, pp.103-112.
- 74 Cozzens, S. and Melkers, J. (1997) 'Use and usefulness of performance measurement in state science and technology programs', *Policy Studies Journal*, Vol. 25, p.2.
- 75 Ruegg, R. (1998) 'Symposium overview. Symposium on evaluating a public private partnership: the advanced technology program', *The Journal of Technology Transfer*, Vol. 23, p.2.
- 76 Guba, E.G. and Lincoln, Y.S. (1989) *Fourth Generation Evaluation*, Newbury Park *et al.*
- 77 Worthen, B.R., Sanders, J.R. and Fitzpatrick, J.L. (1997) *Program Evaluation. Alternative Approaches and Practical Guidelines*, 2nd ed., White Plains, NY.
- 78 Smits, R. and Kuhlmann, S. (2002) *Strengthening Interfaces in Innovation Systems: Rationale, Concepts and (New) Instruments*, Brussels.
- 79 van den Daele, W. (1997) 'Risikodiskussionen am 'Runden Tisch'. Partizipative Technikfolgenabschätzung zu gentechnisch erzeugten herbizidresistenten Pflanzen', in R. Martinsen (Ed.) *Politik und Biotechnologie. Die Zumutung der Zukunft*, Baden-Baden, pp.281-301.
- 80 Kuhlmann, S., Boekholt, P., Georghiou, L., Guy, K., Héraud, J.-A., Laredo, Ph., Lemola, T., Loveridge, D., Luukkonen, T., Polt, W., Rip, A., Sanz-Menendez, L. and Smits, R. (1999) *Improving Distributed Intelligence in Complex Innovation Systems*, Brussels, Luxembourg, http://www.isi.fhg.de/abtlg/ti/pb_html/final.pdf.
- 81 Kuhlmann, S. (2001) *Management of Innovation Systems: The Role of Distributed Intelligence*, Apeldoorn, Antwerpen.
- 82 Grupp, H. (1998) (Ed.) 'Technological forecasting & social change', *Special Issue on Recent National Foresight Activities*, Vol. 57.
- 83 Cuhls, K. and Kuwahara, T. (1994) *Outlook for Japanese and German Future Technology. Comparing Technology Forecast Surveys*, Heidelberg.
- 84 Rip, A., Misa, Th. J. and Schot, J. (Eds.) (1995) *Managing Technology in Society. The Approach of Constructive Technology Assessment*, London, New York.

- 85 Braun, D., Hinze, S., Hüsing, B., Kuhlmann, S., Menrad, K. and Peter, V. (1997) *Interdisziplinäre Zentren für Klinische Forschung in der Frühphase der Bundesförderung – Vergleichende Analyse*, Stuttgart.
- 86 Bühner, S., Peter, V., Braun, D. and Kuhlmann, S. (1999) *Interdisziplinäre Zentren für Klinische Forschung – Kommunikation & Kooperation: Prozessevaluation der Fördermaßnahme 'Interdisziplinäre Zentren für Klinische Forschung an den Hochschulkliniken'*, Stuttgart.
- 87 Bühner, S., Bierhals, R., Heinze, Th., Hullmann, A., Studer, Th., Erlinghagen, R. and Lang, C. (2002) *Die Kompetenzzentren der Nanotechnologie in der Frühphase der Bundesförderung: Ein Bericht der begleitenden Evaluation*, Karlsruhe.