The Use of Evaluations in Europe
Report from The European RTD Evaluation Network Meeting

**The Danish Institute for Studies
in Research and Research Policy
2003/2**

The Use of Evaluations in Europe
Report from The European RTD Evaluation Network Meeting
The Danish Institute for Studies in Research and Research Policy 2003/2

The Use of Evaluations in Europe

**Report from The European RTD Evaluation Network Meeting
Denmark November 2002**

Introduction

The European RTD Evaluation Network was created in 1997 by DG Research with the purpose of establishing a dialogue between users and producers of evaluations of activities within research and development. The background was the need for open discussions of evaluation methods as applied across Europe.

The network was from the beginning established with one representative from each of the member states; gradually the network expanded and today five years after it was started the network has members from the fifteen member states as well as members from accession countries.

The network meets twice a year, always in the country having the presidency of the European Union. In the fall of 2002 the network held its meeting in Denmark as a part of the Danish presidency activities. The list of participants is presented in appendix A.

The theme chosen for the network meeting was " The use of evaluation", and as background for the discussion at the meeting, planned for November 2002, we at The Danish Institute for Studies in Research and Research Policy took the initiative to collect in a systematic way information about the initiation , construction and use of evaluations concerning public research activity in Europe.

As part of the preparations for the network meeting Fil.dr. Evanthia Kalpazidou Schmidt produced and presented the paper: "Evaluations and Science Policy", which is included in this report. The power point used by Evanthia Kalpazidou Schmidt for her presentation at the network meeting can be found in Appendix B.

On the basis of the survey and data collected we produced and presented:" The Use of R&D Evaluations in European Science Policy", which illustrates the comparative perspective of the use of evaluations of R&D. The data collected in the survey was administered by research assistant Per Stig Lauridsen, and all the information is attached in tables in Appendix C.

This report focuses on evaluations of public R&D, and it is based on the information gathered among the members of the network, but at the meeting held in November 2002 in Denmark, other issues were discussed also such as the use of evaluations internally in the Commission and the use of evaluations within the European Space Agency based on presentations from invited speakers.

Karen Siune,
Director, dr.scient.pol.
Member of the European RTD Evaluation Network

Table of content

EVALUATION AND SCIENCE POLICY

Presented at The European RTD Evaluation Network Meeting
Molskroen (DK) 27-28 November 2002

Evanthia Kalpazidou Schmidt
The Danish Institute for Studies in Research and Research Policy

INTRODUCTION

This paper discusses the use of evaluation in science policy focusing first and foremost on the theoretical and conceptual apparatus of evaluation. Moreover the paper focuses on the adoption, construction, implementation and use of evaluation in science policy in Denmark in a historical perspective and outlines some perspectives for the use of evaluation in science policy in general.

There are many good reasons to pay attention to the theory and the concept of evaluation. Firstly, a well-described evaluation theory can be as useful for policymaking as larger empirical descriptions. Secondly, more research in evaluation, both theoretical and empirical, is certainly needed to develop concepts and standards in order to understand processes and be able to support practitioners and science policy actors. This may also help to reduce the delivery gap between evaluators, policymakers and other stakeholders. Finally, more research in the implementation, impact and utilisation of evaluations at macro-level is necessary in order to study the interaction between science policy and the effects of it on different socio-economic systems (Siune, K & Kalpazidou Schmidt, E. 2002).

The perception of evaluation may be very wide. The different perceptions of the evaluation concept may have implications on efforts to map out the evaluation activities in an area, a country or - even more complicated - in a European perspective.

What is evaluation then? Albæk & Rieper (2001) give a wide definition of the evaluation concept: A systematic collection of knowledge about implementation and effects of public efforts based on explicit evaluation criteria and having utilisation as a goal. Efforts are defined as all the activities, in public and semi-public regime (obviously even activities in private organisations may be object for evaluation) such as public organisations, management, projects, programmes, policy areas, development activities, innovation systems etc. Implementation in this definition comprises the processes and the outcome. Effects consist of the immediate effects of the evaluated efforts as well as the effects expected in the long run. This definition of

evaluation opens up (besides research-based evaluations) the inclusion in the concept of other activities and practitioners too. Even practitioners without any training in evaluation methods can accordingly be included. The only precondition is that their work is based on a systematic collection of data.

The evaluation literature in general often considers the different forms of evaluations in relation to adoption, construction (methods and standards used) and in relation to implementation issues. Consequently the evaluation is usually discussed in relation to the elements included in the process. The question of what is evaluation can therefore not be answered in general, according to many evaluation researchers. A definition of evaluation in this sense is not possible without regarding the context for evaluation. Evaluation is a social process after all. As Guba & Lincoln (1989) argue "…there is no right way to define evaluation, a way that, if it could be found, would forever put an end to argumentation about how evaluation is to proceed and what its purposes are".

EVALUATION –
SOME FUNDAMENTAL QUESTIONS

In the evaluation process and practice the following fundamental questions have to be considered:

A.   OBJECTIVES AND MOTIVES

Why evaluate? This question is essential and should be answered in the beginning of the evaluation process since the objectives and motives are going to affect almost every aspect of the evaluation effort.

Motives and objectives can vary greatly. A distinction could be made between evaluation that aims at accountability or control (confirms or strengthens hierarchical control), and societal relevance; and evaluation that focuses on quality of research or on organisational structures and institutional efficiency. Initiation of evaluation could have as a motive an attempt to measure how well the actual accomplishment of an activity or programme matches the anticipated accomplishment. These are though some examples of possible motives for evaluation. Different motives may be combined.

The objectives for evaluation of research, which is in focus in this presentation, can vary to a great extent. In the literature of science policy the following reasons are described as motives for initiations of evaluations (Georghiou 2001):

- Accountability (financial or stakeholder)
- Resource allocation or reallocation of resources (stop/go and selectivity/prioritisation)
- Learning tool (at policy level or operational level)
- Justification and legitimation

These objectives can be combined. Same objectives may apply to a variety of schemes ranging from resource allocation to a variety of schemes such as to promote industry and university links.

B. OBJECT

What is to be evaluated? The definition of the evaluation object is the second essential step. Clearly defined and hierarchical formulated objectives can facilitate the identification of the level of aggregation on the one hand and on the other the clarification of which standards and methods that are adequate in a certain evaluation situation.

An evaluation can focus on either one or more of the functions of an activity (partial evaluation) or on all of them (comprehensive evaluation). Partial evaluation is by far the most used evaluation type. Even though partial evaluation is easier to practice, justifiable in many cases and useful as well, it is obvious that some of the key problems of the evaluated activities, organisations or programmes have often been found in the interaction between the different functions of these activities. This interaction should be addressed when designing a comprehensive evaluation.

Regarding the evaluation of research, the object for evaluation can be identified on different levels of aggregation. The object for the evaluation can consequently be one of the following:
- The individual researcher
- Appraisal of a project
- Programme analysis (National and/or European programmes)
- Institutional assessment (universities and/or other public research institutions)
- Fields of science
- System level/National systems
- Policy instruments
- Multi-level

Multi-level evaluations often combine activities at meso- and macro-levels.

C.   INITIATION AND TIMING

Who takes the initiative to undertake evaluation and when? This is an important issue that is closely related to the motives, the object and the organisation of the evaluation. Moreover the initiation and timing are connected to implementation and the potential of the impact of the evaluation.

In the case of research evaluation, the implementation and impact of the evaluation is to a significant extent determined according to whether this is a top-down initiative (National Ministries/research councils or EU initiatives) or a bottom-up initiative (The units that are subject to evaluation are the initiators) or a combination of these (Universities, ministries and research councils take the initiatives together). The combination of initiatives may establish a truce between the stakeholders and, despite eventual structural disagreements on evaluation in general, facilitate the process and the implementation of evaluation. And what's more may intensify the impact of evaluation.

D.   ORGANISATION

Who should perform the evaluation? Is it an external or an internal evaluation? What is external respectively internal evaluation in a specific evaluation process? Is an evaluation that relies on second order material produced by the evaluated unit/institution and performed by external evaluators a purely external evaluation? There are many problems related to the organisation and implementation of evaluations that should be addressed in the beginning of the process.

There are also many different forms of organising evaluations (a more detailed description of these follows in section 2). The use of various forms of peer review as evaluators is a common feature. Advantages and disadvantages with respect to classical peer review, the modified and informed peer review are often discussed in the literature of evaluation. Other evaluation forms often used are bibliometric analyses (publication or citation analyses) and activity, capacity and users evaluations. Management and impact evaluations are also developing rapidly. Evaluation of strategies of programmes and innovation systems have also been intensively used in recent years. However, when designing an evaluation it is important to look at the activity as a part of a wider setting e.g. in a context.

E.   CRITERIA

What criteria should be used in an evaluation?

Evaluations may derive criteria from different sources such as authoritative goal statements, interest group statements or have theories as their point of departure.

The first types of sources, the authoritative goal statements have been the most used. If no formal statements exist, the point of departure for the evaluation could be some kind of a common understanding of the goals of an activity. However, authoritative goal statements are not operational in connection with some types of evaluations. In these cases criteria are usually derived from various interest groups statements. Such criteria can though be vague and conflicting since different interest groups often have different knowledge needs and motives for initiation and adoption of evaluation. Attention to interest groups and stakeholders has increased recently.

Another source of criteria could be the theory. The disadvantages when looking for criteria derived from theory are that existing theories are often disputed. More research in and development of theories is therefore essential for future evaluation practice and consequently for science policy based on it.

F.   METHODS AND STANDARDS

What methods and standards should be used? This depends on the evaluation motives, the object, the criteria used etc.

Evaluation literature in general is very much focused on methodology issues and especially the issue of quantitative versus qualitative approaches. Quantitative indicators often used in science policymaking can not be used in each evaluation situation but need to be complemented by other methods.

Comprehensive evaluations require a combination of quantitative and qualitative methods. Comprehensive evaluations are though - being expensive - a rather rare phenomenon.

Another basic distinction often made is that between summative and formative evaluation. The first one is concerned with the output and the outcome of activities (and therefore uses more often quantitative methods) while the second is concerned with processes (and uses more qualitative approaches). Monitoring is another evaluation form used regularly, often as an ongoing control of activities and focusing on outputs, performance and efficiency.

Focusing on evaluation of research we can conclude that there is an intensified interest in the building-up and developing of more accurate and effective instrumentation, especially at the macro-level. Research is characterised by unpredictability, fragmentation and non-uniformity. This implies that the adoption of primarily formalised, quantitative methods or the adoption of performance indicators in every evaluation situation should be used with caution. However, there is an increasing tendency to use performance indicators. The main problem with such indicators is that they often measure what is measurable rather than what is needed. The basic requirement for the use of performance indicators is though the understanding of context, objectives and relationships between objectives and effects (Ziman 1985).

G. DISSEMINATION AND AVAILABILITY

Who is the user of the evaluation? This question is related to the motives and criteria of the evaluation. Another aspect of it is related to the availability and the effective dissemination of the results of the evaluation.

In the literature many argue that there is a "utilisation crisis" and that there is need for "utilisation-focused" evaluations. Others argue for an integration of users in the evaluation process at a very early stage.

With respect to evaluation of research we may conclude that the fragmentation, non-uniformity and abstract nature of research in some cases put limits to the dissemination of results (there are cases where dissemination of results is possible only to peers). However, audiences for evaluation of research can be different depending on the level of analysis and the differentiation of the level of the evaluation issues. The audiences or users can accordingly be:

- The society in general
- Policymakers (politicians and other officials)
- Programme/institutional/research managers (or other managers)
- The participants in the evaluation (researchers)
- Other stakeholders/actors or agents (these have a growing role to play as research engages more and more with socio-economic issues)
- Multiple audience or dual function

In some European countries the availability of research evaluations is limited however and in some others the results are not publicly available at all.

H.   IMPACTS

What impacts may be expected? There are in general two different types of impacts: The expected impacts that should correspond to the motives and other "hidden" impacts (which are not part of the explicit motives and objectives).

The usual types of "hidden" impacts are: The Matthew effect and the fostering of conflicts (usually when evaluating individual performance or institutional accountability). Another hidden impact is - according to Goodhart`s law - that the use of evaluation indicators distort behaviour. Evaluations may in some cases be counter-productive and not necessarily lead to expected impacts such as improved performance.

Another distinction is between evaluation impacts at the micro-, meso-level and macro-level. It is also vital to take note of the time-scale over which uptake can reasonably be expected. Research that is used in policymaking may in some cases take many years to achieve impacts. It is consequently necessary to make a third distinction between short-term and longrun impacts.

However, we have to conclude that more research on evaluation impacts is necessary. In general little is known about the impacts of evaluations at the different levels of implementation.

I.   IMPLEMENTATION OF RESULTS

Evaluation is a social process undertaken within a context, which as a consequence creates barriers that complicate the implementation of the results and recommendations of the evaluation. In the literature of science policy three factors are described as the main barriers to implementation of results namely the issues of absorbability, credibility and steerability (Georghiou 2001).

Absorbability is primarily connected to how evaluators can make the audience pay attention to evaluation results and recommendations. As evaluators often put it, the expected implementation in many cases does not take place. Awareness of the audience is for that reason necessary. The timing of the evaluation is another significant factor. The evaluation should be undertaken in time for follow-on decisions, as it may later need validation. Furthermore a differentiation between intermediate and ultimate objectives is necessary. The level of recommendations has to be adjusted to the level of the project. The recommendations may therefore neither be too specific nor too general so that options are made unavailable. Finally, the report may be formed and structured in a way that is easily accessible to stakeholders.

Credibility is connected to the issue of trusting the results of the evaluation and depends on the credibility of the evaluators and the quality of the findings/evidence. The credibility of evaluators is connected to the competence, the standing, view and the independence of the evaluators. The quality of the findings/evidence depends on the breadth and depth of the coverage of the evaluation, the standards and methods used and the appeal of the data. In the case of use of quantitative data in particular, considerations on further need for supporting evidence (such as case studies or comparative approaches) may be necessary.

Steerability focuses on whether the results of the evaluation can be implemented. Evaluation and evaluators after all are only one among several influences and actors in a certain policymaking process. On the other side policy instruments themselves have limitations. It is for that reason not always possible to carry out recommendations and decisions. Furthermore, it can take time to implement the results of evaluations and assess the impacts of activities.

THE FRAMEWORK FOR DIFFERENT FORMS OF EVALUATIONS

In conclusion, evaluations differ with respect to object that these focus on, the aim, the organization and the evaluator undertaking these. The different forms of evaluation, their purpose and carrier are described in figure 1.

Figure 1. Characteristics of methods used in evaluation of research
        (based on Foss Hansen 2001)

Methods for evaluation	Focus on	Purpose	Evaluator
Survey descriptions	Issues/problems Currently being researched	Overview of the work in the field	Researchers from the field (not necessarily peers)
Peer review	Quality of production	Decisions about publications and appointments Research policy advice	Peers
Bibliometric analyses: Publication analyses Citation analyses	Quantity of publications Quantity of citations	Assessment of quantity Assessment of effect and exposure of excellent research teams	Analysts, statisticians (sometimes researchers)
Users evaluation	Utility experienced by users (short run perspective)	Assessment of utility and exposure of need for research	Representation of users or analysts through interviews with users
Historical evaluation	Development and utility (long run perspective)	Assessment of utility (long run perspective)	Experts in the history of science and the sociology of science
Activity evaluation	Quantity and quality of activities	Assessment of quantity and quality of activities in the research organisation	Experts in the sociology of science or organisation theory (sometimes peers)
Capacity evaluation (management evaluation)	Analyses of production quality and qualifications	Assessment of the capacity and potential of the research organisation	Experts in the sociology of science or organisation theory (sometimes peers)

The above-described methods are systematised as evaluations of structure, evaluations of process and evaluations of outcome. Evaluation of process focuses on research activities, evaluation of structure focuses on the capacity of production and evaluation of outcome focuses first and foremost on the production of research (Foss Hansen 2001).

Systematising the methods of evaluation of research activities we could conclude that the most used methods in relation to structure, process and outcome of an organisation are:

  a. Methods within evaluation of structure (assessment of performance in the future)
      Capacity evaluation
  b. Methods within evaluation of process (focus on present performance)
      Activity evaluation
  c. Methods within evaluation of outcome (assessment of performance in the past)
      Peer review
      Bibliometric analysis
      Users evaluation
      Historical evaluation

Evaluation of structure focuses on assessments of the organisational capacity. Examples of such evaluations are analyses of the research qualifications (capabilities, research experiences), analyses of the technological capacity (access to instruments and other facilities, grants for travelling etc.) and analyses of development within disciplines or fields. Finally, there are analyses of the capability of an organisation to engage in new scientific fields and renew research activities by recruiting highly qualified researchers, inviting visiting professors, training researchers and attracting grants.

Evaluation of process focuses on the activity e.g. the range and quality of the activity. Examples of such evaluations are analyses of research processes, communication, research environment and culture, research traditions and conditions but also analyses of other activities related to research such as supervising, teaching etc.

Evaluation of outcome is an assessment of products by focusing on output, effect and utility. The output evaluation is an assessment of achievement. Examples of output evaluations are measurements of the published units per researcher, per unit of time or measurements of internationally published articles per researcher, per unit of time. However, evaluations of output can be assessments of the scientific level of an organisation in general or assessments of the quality of articles, books and qualifications of researchers (Foss Hansen 2001, Kalpazidou Schmidt 2002).

Effect evaluation is an assessment of the effect of accomplished achievements. Examples of this type of evaluation are measurements of production in form of doctoral dissertations, or analyses of patterns of citations or number of patents.

Utility evaluations focus on the utility/value of past achievements and have therefore a historical perspective. Assessment of the use of research production by interest

groups and stakeholders is central in utility evaluations. Interest groups can though have varied, even conflicting interests and evaluate the same organisation having different motives and using different criteria. The results of this type of evaluations can consequently be diverse and in some cases, even contradictory.

Input evaluation (usually an integral part of other forms for evaluation) of different organisations or programmes can constitute a fourth type of evaluations. In these cases the input is not evaluated in relation to the outcome or any other assessment of effectiveness.

WHAT IS A STRONG EVALUATION?

There are strong and weak evaluations in all types of evaluation practice. A weak evaluation is a systematic, however usually a non-comparative evaluation. It is a kind of an ex-ray illustration of the selected dimensions of an activity. This issue is though closely related to the objectives of the evaluation in question and the need for knowledge that the initiators have.

A strong evaluation is a systematic, comparative evaluation. It is an assessment of value based on a comparison of something outside the evaluated organisation, system or process. It is a sort of record viewing present stock and indirectly pointing to the potentially "empty shelves". A strong evaluation is an assessment of several dimensions and expresses an opinion on the effectiveness of the process, the organisation and the activity (Christiansen et al. 1989)

Strong evaluations raise the questions of which other similar organisations are to compare with. Research activities, organisational structures and the interests surrounding the activities vary. The identification of similar organisations with respect to activities, structures, research fields, actors/stakeholders and surroundings is highly significant in comparative evaluations. Attempts to compare very differentiated organisations within a country (an issue particularly relevant in small countries with few research organisations as in Scandinavia) should be used cautiously since these comparisons may in some cases not be measuring the same dimensions (Kalpazidou Schmidt 1996).

Evaluation of research can be done at micro-, meso- and macro-level. At micro-level, evaluation first and foremost takes the form of the classical peer review that is based on the traditional recognition that only scientists are able to evaluate research outcomes. Evaluation at meso-level focuses on evaluation of disciplines, fields, programmes, organisations, institutions/departments and research groups. Evaluation at macro-level focuses on evaluation of R&D in different socio-economic systems. It is

especially at this level that comparative evaluations could be used to a much greater degree than is the case today.

Another concept - discussed intensively in recent years in an attempt to organise and implement stronger evaluations - is that of interest evaluation. The interest model has as its intention the involvement of all actors and stakeholders that have an interest in the evaluation outcome. According to the interest model, all actors and stakeholders should have the opportunity to make their voice heard already at the beginning of the process, in connection with the identification and formulation of the evaluation object and standards. This implies the assumption that processes, implementations and impacts must be regarded from different points of view and that these should be decided in a pluralistic way that involves all stakeholders. This model is therefore based on the idea of consensus building. In cases where consensus is not possible to achieve, the model cannot give any answers to problems. Evaluators in the interest model function as mediators between different interests aiming at consensus. Evaluators focus at last only on issues where consensus building has been achieved.

THE ROLE OF THE EVALUATOR

There are many actors with divergent interests involved in the evaluation process. Different agents/actors have different knowledge needs and therefore different motives, objectives and interests in the evaluation process. These interests depend on the varied ways that activities are controlled. In the case of research there are many control mechanisms that are active in the process, such as the political system, the market, the organisation, the professionals and the users. It is obvious that some designs and methods of evaluation are more appropriate for some agents/actors and their objectives, motives and need for knowledge than for others. The perception of the role of the evaluator is consequently also differentiated. There are accordingly several evaluator roles. The most usual evaluator roles are outlined below (Albæk 2001).

- Evaluator as a neutral, problem solving *social engineer* in the political-administrative decision making process
- Evaluator as a *controller* in an attempt to keep implementation bodies responsible for their dispositions
- Evaluator as an *adviser* in praxis, in an attempt to build up or adjust public initiatives or activities
- Evaluator as a *mediator* between divergent knowledge interests
- Evaluator as a *midwife* or therapist for disadvantaged groups in the society (often used in action research).

The different roles of evaluator may be combined. An evaluator can function as a mediator and adviser at the same time. It is though not possible to outline the pros and cons of the different roles without knowing the context for the evaluation. It is important on the other hand to underline that the evaluator has to be aware of his/her role and having this as a starting point choose the appropriate design and methodology. Moreover, the evaluator must be aware of the strengths and weaknesses of the methods and instruments that he/she uses in a particular case.

EVALUATION OF RESEARCH AND RESEARCH CONTROL – AN ORGANISATIONAL PERSPECTIVE

Research planning and budgeting, resource allocation, legitimation of activities, initiation of research programmes, changing of the rules concerning the structure of appointments, administrative legislation, establishing of ethical and scientific norms and initiation of research evaluation are some of the instruments that constitute political control of research activities.

In the literature, a differentiation between several methods of control is made namely this between concepts as bureaucracy, market, meritocracy, policy and dialogue. Bureaucracy is perceived as control that the state uses through a hierarchical system by means of planning, budgeting and implementation of legislation and, other rules and guidelines. Market is perceived as control through supply and demand on the research market, and meritocracy is control of research through an academic elite that functions as gatekeeper to the system. Policy is control of research through persuasion, negotiations and compromises between the interest parts. Dialogue on the other side is perceived as control through conviction e.g. exchange of information within the organisational network (Kalpazidou Schmidt 1996 & 2002).

Some specific forms of evaluations are identified within each of the above-mentioned models of control. In bureaucracy the evaluation usually focuses on current evaluation of outcome. This type of evaluation is often connected to resource allocation, distribution of rewards and sanctions, and aims at resource control based on "value for money" demands.

Evaluation in the market model is more of a prognostic character and operates with concepts such as expected utility. The aim of the evaluation is to identify the best among competing researchers.

Evaluation in the meritocracy model is carried out by colleagues and focuses on quality and outcomes. The aim of the evaluation is to secure the quality of the

production as well as to regulate access to the scientific community. Evaluators are accordingly functioning as gatekeepers to the scientific community.

In the political model all forms of evaluations are possible. The aim of the evaluations is first and foremost to legitimise activities for distribution of growth or on the contrary to suggest cutbacks.

In the dialogue model evaluation is current and prognostic. It focuses primarily on structures and processes but also on outcome, quality and utility issues. The aim is to develop qualifications, networks and research cultures but also to achieve renewal of research fields and innovation.

One of the most developed and significant evaluation theories has been presented by Vedung (1997). This theory describes different organisation models of evaluation as follows:

- The efficiency model, where focus is on cost-benefit or cost-effectiveness analyses.
- The productivity model: evaluations focus on the relationship between input and output.
- The effectiveness model, where the organisational principle is the effort and the objective of the activity. The question to be answered is: is the objective achieved and moreover, is this a result of the evaluated effort?
- The effect model, where the focus is on the effects that the effort have, both intended and unintended.
- The process model, where the evaluation focuses on the implementation process.
- The utility model, where focus is on the users assessment of the effort.
- he interest model, where focus is on the different agents/interests assessments of the effort.
- The peer model, which implies that experts within the same field evaluate other experts.

The different models may be combined depending on what the aim of the evaluation is and also on the need for knowledge that different actors have in a certain process.

EVALUATION AND POLICY MAKING - AN ANALYTICAL PERSPECTIVE

There are, as discussed above, many ways of designing, organising and implementing evaluations. The different evaluation models presented in the literature can be used as frameworks for the discussion of the use of evaluation in policy making. These can highlight differences and changes in the evaluation practice and probably suggest alternate adequate forms.

Having as point of departure an institutional perspective based on theories of organisation, Foss Hansen (2001) discusses the development of evaluation in relation to the preconditions that influence this in a historical analysis. According to this perception, the concept of evaluation perceived as an organisational form element in a policy field can be analysed in three types of processes: *adoption, construction* and *implementation* (see also Røvik 1998). It is of significance to consider, in a science policy frame, the conditions under which these processes take place.

Adoption of evaluation is related to the process of taking up and discussing evaluation as a concept in policymaking (as one among other possibilities). Construction is related to the process of development of evaluation models and standards. Finally, implementation is related to the process of using one or more of the standards, practicing the evaluation and reporting of the results.

The three types of processes can function as different phases in a specific evaluation situation but this is not always the case. The processes can even function simultaneously. Moreover, some standards can be in action and others only discussed and not implemented; e.g. adoption of evaluation does not necessarily imply construction of standards, not implementation either. A process can therefore be limited to adoption of evaluation and not continue to the next phase, the implementation of evaluation (Brunsson 1989).

Adoption, construction and implementation processes may have as a starting point *top-down* initiatives (when ministries/authorities initiate evaluations) or *bottom-up* initiatives (when the organisations themselves initiate evaluations).

In the following we shall discuss adoption of evaluation used as an instrument in policymaking, the conditions that influence the construction of evaluation standards in such a context and finally the conditions that influence the intensity of the evaluation practice (Foss Hansen 2001).

Adoption of evaluation as a science policy instrument is possible under the following conditions:

- Evaluation is adopted when central actors have a need for control (summative evaluation) and/or learning and development (formative evaluation) and if the evaluation is perceived as an instrument to reach these.
- Evaluation is adopted when concrete problems need solving and the evaluation is perceived as an instrument to uncover responsibilities for these problems and find solutions.
- Evaluation is adopted when there is a perception of a need for evaluation. Evaluation can be used as an activity which signals that something is being done e.g. as a justification or legitimation.
- Evaluation is adopted only if evaluation matches the existing repertoire/agenda in a policy area. Evaluation is thus not used in policy areas that are not familiar with it. It has to be an "antenna" out there to receive the signals of the process.
- Evaluation is adopted when it is perceived as a strategy that can be used in a political setting.
- Evaluation is adopted when there is a need for opinion making and development of an identity in a policy area. In this case the evaluation gives a possibility for interaction and dialogue.

Focusing on *the construction of standards* for the evaluation, the following conditions within a policy area are influencing the process.

- Evaluations standards are constructed in a way that these match the aim of the evaluation, in summative or formative terms.
- Evaluation standards are constructed in a way that these match the problems that are the cause for the evaluation.
- Evaluation standards are constructed in a way that these reflect central institutionalised values in a policy area.
- Evaluation standards are constructed in a way that these reflect the profile of the dominant actors who are adopting the evaluation.
- In highly politicised areas there is an ongoing struggle about the construction of evaluation standards.
- Evaluation standards are constructed in a way that these reflect the dominant interests and their agenda.

Focusing on *the intensity of the implementation* of evaluation in a policy area, it is possible to provide a description of the conditions that influence these as follows:

- The intensity of implementation of evaluation reflects the need for control and learning. (Evaluation practice is more intensive in areas that show high or increasing decentralisation of responsibilities and therefore increasing need for control as well as in areas with high theoretical uncertainty and consequently increased need for learning).
- The intensity of implementation of evaluation in a policy area reflects the range of the problems in the agenda.
- The intensity of implementation of evaluation in a policy area reflects the legitimation that is related to the evaluation process. Evaluation can develop to an institutionalised value in itself (a symbol for good administration and management).
- The intensity of implementation of evaluation in a policy area reflects the range of the evaluation competence in the area. The standards are the projection of the competence.
- The intensity of implementation of evaluation in a policy area in general reflects the political conflict level in the area. In highly politicised areas the intensity of the implementation of the evaluation is high and vice versa.
- The intensity of implementation of evaluation in a policy area reflects the degree of identity uncertainty in the area.

The processes described above can be systematised based on the rationale the different actors make use of in relation to adoption, construction and implementation of evaluations (see Albæk 2001, Foss Hansen 2001).
The first rationale has a target-oriented approach. It is a *target-instrumentation rationale* that perceives evaluation as a control and/or as a learning mechanism. The evaluation standards used and the implementation of evaluation reflect the need for control and/or learning.

The problem-responsibility rationale is driven by problem solving, a moving away from problems by placing responsibility, a kind of "muddling through" (Hirschman & Lindblom 1969). Standards are constructed based on the characteristics that the problems in focus show. Implementation illustrates consequently the range of the problems in a certain policy area.

The value-based rationale perceives evaluation in a policy area as an appropriate activity and the use of standards reflects central, institutionalised values. The intensity of implementation is a reflection of the legitimacy of the evaluation.

The cognitive rationale perceives actions as determined by actors thinking and cognitive processes. Actors have some kind of a "repertoire" or agenda (Weick 1969). Adoption of evaluation is possible only if this matches the existing expertise repertoire, and the standards used are a reflection of the expertise repertoire. The intensity of the implementation is a reflection of the familiarity of the evaluation to this repertoire but also a reflection of the existing evaluation competence in the area.

The dialogical rationale perceives evaluation as a dialogue process that can be used as opinion and identity development instrument in a specific policy area. Standards are constructed in order to form opinions and achieve the identity development of the area. The intensity of the implementation reflects the identity uncertainty in a policy area or the need for identity change.

The political rationale perceives action as determined by stakeholders interests, dialogue processes and willingness to compromise (Mintzberg 1983). Evaluation is adopted if central actors perceive it as an activity to promote their interests and/or as a negotiation/dialogue in order to achieve problem solving. Standards are constructed in a political process and the intensity of the implementation of evaluation in a policy area reflects the politicisation of the area. Policy is the product of a negotiation process where several actors with different interests and demands use their resources – in this case evaluations - trying to influence the decision-making according to their interests. In the political arena knowledge gained through evaluation undertaken by an independent instance is a valuable resource to make use of in the argumentation process. Accordingly, actors use evaluations as political ammunition to achieve their objectives.

The above-described rationales may be combined in complicated processes and in some cases even in contradictory processes. The rationales can be used as a framework for the understanding of the evaluation process. Policy areas and their characteristics, interests, values, expertise and competence influence evaluation processes and implementations. Evaluations can consequently function as significant contributions to decision-making and science policy making. Evaluation is nonetheless only one of several instruments used for this purpose. But when evaluations are used as inputs in science policy, this is often made in an indirect and not always obvious way since evaluations are one of many different components used by decision-makers in the synthesis of policymaking.

EVALUATION AND POLICYMAKING IN DENMARK

In many OECD countries unprecedented economic growth in the 1960s made it possible to channel resources from the private to the public sector and to hereby provide welfare services. These services were often the result of deliberate political choices. This process has its origin primarily in the US where the welfare state expanded so fast that a considerable knowledge vacuum appeared on the welfare states function and services. This took place simultaneously with a growth in the public sector resulting in a desire to hold the administration accountable for its activities and its financial spending. The institutionalisation of evaluation so started in the US in the 1960s, although evaluation as such has historical roots that go further back in time.

The concept of evaluation developed considerably during the last three decades in most OECD countries, though with noticeable differences between the countries. The problems addressed by evaluation research and practice emerged in connection with efforts to form, consolidate and reform the welfare state. The development of evaluation research and practice was consequently to a large extent related to developments in the welfare state's provisions and services (Albæk 1998).

Denmark - despite being a welfare state - was a late starter with regard to adoption, construction and implementation of evaluation. Not until the 1980´s and the 1990´s did evaluation become a common feature. The institutionalisation of evaluation during this period has been intensive and is still a topic on the policy agenda. The adoption of evaluation was - as in other countries - a result of the development on the economic and political scene and the development in the Danish administrative apparatus.

What was it then that generated this need for adoption of evaluation procedures in Denmark? **Firstly** a policy of "value for money" was introduced with high demands for increased productivity and documentation. Evaluation was perceived as an instrument to reach this goal. **Secondly** this period was characterised by an unstable political situation with minority governments constituted of many parties. This resulted in political compromises to a much higher extent than before. Such political compromises were often followed by agreements that the implementation of specific activities should be subject to evaluation after a few years. **Thirdly** the 1980s and 1990s were characterised by changes in the political administrative system. New organisational structures and management styles were introduced. During this period organisations gained a higher degree of autonomy simultaneously with an increased public demand for greater transparency and central control. At the same time the profile of the administrative personnel changed, from being an administration dominated by lawers to one dominated by personnel with a background within economics and politics. At **last** evaluation developed - especially during the 1990s - to

a research field in its own within social science. A capacity of evaluation researchers was build up. This assisted in reforming the welfare state (Foss Hansen & Borum 1999).

Focusing on **adoption** issues Denmark was - compared to other Scandinavian countries - late to adopt meso- and macro-level research evaluation. Not until the end of 1970s were the first meso-level evaluations introduced. Despite this Denmark adopted in early cases a non-traditional and experimental way of organising evaluations with results and experiences that, from a scientific point of view, were not completely convincing.

Focusing on the **construction of standards**, a more internationally oriented approach was introduced, the *modified peer review*. This standard is a peer review approach where the researchers perform the evaluation. The task and focus of the modified peer review are different from the classical peer review. Even the processes and the expected results are different. The context for decision-making is though more vague in the modified peer review approach than in the classical peer review, as this standard usually has to relay on second order material (such as annual reports, publication lists etc.). This standard was introduced as the subject for evaluation was expanded enormously at the same time. The evaluation process in modified peer review hence focused on activities different from the classical peer review. Both classical and modified peer review specify though the evaluation process, but the criteria used in the evaluation are usually not specified. Such criteria and definitions of quality rely on the evaluators' competences, the context and the processes that they are involved in (Foss Hansen & Borum 1999).

Modified peer review was introduced into the science policy debate in Denmark in 1983 by the Council for Research Policy and Planning in a policy paper where the Council – in contrast to previous recommendations – explicitly mentions the need for adoption of modified peer review. This evaluation form was conducted and organised along the lines of the Swedish model used by the Swedish Natural Science Council since 1977. The Council for Research Policy and Planning once again in 1984 and later in 1987 (supported by an OECD review of Danish science policy in which it recommended concentration of resources on specific fields) stressed the need for further developing and implementing the modified peer review evaluation standard. The Council suggested that the responsibility of the evaluation itself should be with the Council as well as selection of evaluation subjects. Furthermore the Council recommended introduction of international peers as evaluators. These evaluations were supposed to help the Council in giving advice primarily on resource allocation. This implies that evaluation policy in Denmark in its early days focused primarily on allocation of resources.

The initial phase of the evaluation **implementation** process in Denmark showed low intensity, however with increases over the years. Traditionally the politician's set the agenda for research evaluation and the research councils acted as mediators and performers of operations. The universities on the other hand demonstrated a shifting role over time from a passive and reactive to a more proactive role. The implementation of evaluations in Denmark can thus be categorised in three different phases. Phase one is mainly initiated by governmental bodies, science policy advisory councils and research councils (the 1980s). Phase two is characterised by initiatives taken by actors at all levels (the beginning of the 1990s). Phase three is dominated by initiators at university level (1993-1996). The tendency seems to be that the later phase is continuing as the universities take more of the initiatives to evaluation practice at the same time as evaluation as a policy instrument is high on the political agenda.

Ministries have though traditionally been the initiators for another highly intensive activity e.g. evaluations performed at other public sector institutions than universities.

Figure 2 outlines the use of modified peer review evaluations of research at universities in Denmark 1985-1996. Evaluations initiated by international research policy organisations are not included in this presentation (Foss Hansen & Borum 1999).

Figure 2. Research evaluations in the university sector in Denmark 1985-1996
(Foss Hansen & Borum 1999)

Initiator	Parliament, Government, Science policy advisory bodies	Research councils	Universities	Total
Initiated year				
Phase I: 1985-89	3	2	1	6
Phase II: 1990-92	3	6	5	14
Phase III: 1993-96	5	8	31	44
Total	11	16	37	64

The institutionalisation of research evaluation in Denmark started with certain coercive pressure that later on was supplemented by imitative and normative pressure. 31 of the 37 university evaluations initiated by the institutions are results of evaluation programmes at the Copenhagen Business School and Aalborg University. The answer to the question why these universities became bottom-up initiators is probably that these are young universities experiencing stronger legitimation demands. Imitation processes seem therefore to be more important than normative pressure in these cases. These universities developed at the same time a considerable expertise in the field.

The standard used in research evaluation in Denmark has been the modified peer review as a flexible and broad standard that leaves space for varied definitions of quality in disciplines, programmes and institutions. This could be an important reason for the adoption of this standard. Studies show that evaluations using modified peer review, conducted within the same framework and supervised by the same evaluation committee, vary significantly with respect to processes and results (Foss Hansen & Borum 1999).

However, research evaluation is still not a shared norm at Danish universities as it has been at the Danish public sector research institutions. The later are independent research institutions under the auspices of the relevant line ministries. Figure 3 illustrates the evaluations conducted at these institutions (an area with high evaluation intensity during the last years) initiated by different ministries 1989-1998.

Figure 3. Research evaluations in public sector research institutions in Denmark 1989-1998

	Public sector research institutions
Phase I (1989-93):	12
Phase II (1994-98):	210
Total	222

The high intensity of evaluations in this area is due to the fact that the Danish ministries use evaluations as ongoing control mechanism. The majority of these evaluations were performed on an ad hoc basis. Expert panels constructed the standards and performed the evaluations.

Two additional standards for research evaluation have been considered in science policy in Denmark, *the performance indicator standard* and *the informed peer review standard*. These standards were adopted and constructed but have not been implemented in a wider scale. Figure 4 summarises the main differences between the different standards.

Figure 4. Characteristics of standards for research evaluation in Denmark
(adopted from Foss Hansen & Borum 1999).

	Classical peer review	Modified peer review	Performance indicators	Informed peer review
Year of construction in Denmark		1983	1985	1993
Process	Reading first order material Assessing Rec. decision	Reading second order material Site visiting Assessing	Collecting data Calculating Performance indicators	Reading second order material Rating
Organisation	Several indiv. Assessments or committee	Committee	Calculator	Committee
Outcome	Yes/No Decision	Qualitative assessment	Quantified assessment Ranking	Quantified assessment Rating

The performance indicator standard, based on an economic model, was introduced in 1985 by two economists who ranked departments of economics anonymously, applying Data Envelopment Analysis (DEA, Jennergren & Obel 1985-86). The argument for using the DEA approach was to distinguish between departments in a research field when reallocating resources.

Attempts to introduce the performance indicator as a standard for evaluation and research resource allocation in Denmark are closely connected to the debate on allocation of resources for higher education. The introduction of performance indicators based budgeting system for higher education in 1981 started the debate on the principles of allocation of resources to research activities in general. The question raised was why research in a field should be coupled with the number of students in the field, as was the case in Denmark. This debate is still ongoing. The Ministry of Education opposed the use of performance indicators for research. However, the ministry adopted the performance indicator standard in 1989. In the beginning of the

1990s, The Research Policy Council made two attempts to introduce performance indicators based review projects, but failed due to lack of support from the ministry and the organisation of vice-chancellors at the universities. Moreover, the Council was unable to get agents to take managerial responsibility for these projects. The Ministry of Education – after stressing the problem of the principles for resource allocation – introduced in 1993 its own performance indicator review project to be used as a supplementary tool to peer review. As a result the *informed peer review standard* was launched (Foss Hansen & Borum 1999).

The informed peer review standard is a combination of the modified peer review and the performance indicator standard and was introduced in an attempt to tighten peer review evaluations and make them more transparent. The reaction to it from researchers was negative pointing out - based first and foremost on the assumption of the great differentiation that characterise research activities - that this was an attempt to compare the incomparable. However the ministry proceeded in 1995 to find a system for resource allocation and recommended the development of a system similar to the British model: a specific informed peer review standard. Strong reactions against the proposal from both political and administrative sources, academics and the media resulted in failure of adoption of the specific informed peer review standard once again. Alternate proposals on this issue have not been implemented either. After more than 20 years this issue is still on the agenda.

In 1998 the responsibility for the universities was transferred to the Ministry of Research. A new government transferred responsibility again in 2001 to the Ministry of Education. The Ministry of Research established during this period the concept of contracts in the university sector. The Ministry of Science, Technology and Innovation (as the Ministry of Research was renamed in 2001) has the responsibility for RTD and the universities. New ministerial initiatives have recently proposed increased external control of the universities involving more interests and actors.

In conclusion, a systematic and overall analysis of evaluation practice, its institutionalisation, consequences and impacts for policymaking has so far not taken place in Denmark. A project financed by the Council for Social Sciences is ongoing. The project focuses on the use of evaluation in the Danish state political-administrative system. Concluding results have though not been presented yet (Foss Hansen & Borum 1999).

EVALUATION PROCESS AND POLICYMAKING

The typical problems in the relationship between evaluators and policymakers are discussed in the literature of evaluation as the delivery gap on the one hand and the customer/user gap on the other. The delivery gap can in brief be described as the gap between what policymakers expect from evaluators and how evaluators perceive a certain evaluation process. The customer gap refers to what evaluators wish and demand during the process and how policymakers respond to this. The delivery and customer gap can so be summarized in the following (Georghiou 2001):

The delivery gap

What policymakers want	What evaluators say
Information in time for resource allocation decisions	Research may take years to have effects
Clear attribution of effects to investments	Linear model is a rare case and additionally is complex to assess
Independent evidence of research excellence	Peers defend their subject field and international colleagues
Key indicators to monitor and benchmark	Crude organization distorts performance and can be manipulated

The customer gap on the other side between how evaluators perceive an evaluation process and how policymakers usually respond to this is outlined in the following:

The customer gap

What evaluators want	What policymakers say
Clearly defined and hierarchical objectives	Programmes are a compromise involving multiple and conflicting objectives and interests
Guaranteed independence and transparency	Recommendations must be within realistic policy constrains
Time and resources for the task	We need the results in x months and the resources are limited
Full access to information and stakeholders/actors	Actors are overworked and busy

The question is consequently how we can overcome the customer and delivery gap by offering tools that enable **policymakers to understand the evaluation process** on the one hand and increase **evaluators awareness of the conditions for decision and policymaking** on the other. This is an important issue that has to be addressed in the context of the European science evaluation policy.

CONCLUSIONS AND PERSPECTIVES

The subject for this conference is how we can make use of evaluation practice as a strategic science policy tool in order to build the European Research Area. In this purpose some perspectives of significance for the development of this instrument are outlined in the following. It is obvious that we need to focus on:

- The 'state of the art' (needs at a European level, activities, competences etc.)
- Access to accumulated know-how and data (e.g. possible need for an internet portal)
- Targeting specific issues and problems
- Right timing in adapting processes (including quick identification of trans-border emerging issues)
- Higher involvement of policymakers and stakeholders in the specification of the evaluation from the beginning
- Involvement of both evaluators and key stakeholders in an open and interactive process (it may contribute in raising awareness, cooperation and participation and facilitate the implementation of evaluation results)

Furthermore, a number of good practice recommendations are outlined below:
- Evaluation instruments recognised and targeted on needs at European level have to be developed
- Evaluators have to be international, independent and adequately resourced
- Improvement of communication between different policy levels/The results of evaluations have to be disseminated at several levels
- Evaluations should be followed up by studies to determine the impacts
- Societal relevance of activities has to be addressed/Development of an evaluation system that can even measure social and not only economic benefits
- Multidisciplinary approaches should be used to achieve the above-mentioned
- Trans-national comparative evaluations are needed
- Identification and addressing of issues that are inherently trans-border/European and/or particularly complex and which need a higher level of aggregation
- Evaluations should serve as learning experiences to identify and develop adequate forms of organisation/best practice at European level

In conclusion, some general remarks may be of significance to the discussion of the use of evaluations in science policy in a European context:

- Evaluation provides scientific support to policymaking
- Prioritisation and concentration of resources are significant in policymaking and so is evaluation in order to achieve it
- Evaluation is only one of several instruments used in science policy
- Policy instruments have limitations
- The value of evaluation is both in the process and in the products generated
- Evaluation results are not always implemented or have expected impacts
- Utilisation and impacts of evaluations have to be addressed
- Stakeholders and different actors are not always aware of the impacts
- The more transparent the processes the better implementation chances
- Quantitative indicators must rely upon an approach that relates inputs to effects and outputs
- Indicators are useful but need to be complemented by formative/learning approaches and case-studies
- Differentiation of research activities is essential
- Contextual issues need to be addressed
- Coordination and systematisation of evaluation across Europe is needed/activities to tackle European-wide issues
- Development of a more strategic and coordinated evaluation approach in Europe: The European Research Evaluation Area?

Evaluations may of course not replace politics but EU policymaking may be based on systematic approaches. Scientific support to policies is part of the 6th framework programme launched recently. Further development of evaluation instruments and practices at a European level is vital to the use of evaluation as scientific support to European policymaking.

REFERENCES

Albæk E. (1998), Knowledge interests and the many meanings of evaluation: A
    development perspective, *Scandinavian Journal of Social Welfare*, vol. 7,
    pp. 94-98.

Albæk E. (2001), Vidensinteresser og de mange betydninger af evaluering–et
    udviklingsperspektiv, in Dahler-Larsen P. & Krogstrup H. K. (red): *Tendenser i
    evaluering*. Odense Universitetsforlag.

Albæk & Rieper (2001), Evaluering i Danmark: Effectevaluering, monitorering og
    formative evaluering, Dahler-Larsen P. & Krogstrup H. K. (red): *Tendenser i
    evaluering*. Odense Universitetsforlag.

Brunsson N. et. al. (1989), *Standarder*. Nerenius & Santerus Forlag.

Christiansen J. et al. (1989), Research on Research - Evaluation of Evaluation in the
    Nordic Countries. *COS-forskningsrapport nr 3*. Copenhagen.

Foss Hansen H. & Borum F. (1999), The Construction and Standardization of
    Evaluation. The Case of the Danish University Sector, *Evaluation*, vol 5, no 3, pp.
    309-329.

Foss Hansen H. (2001), Rejsende praksis: Hvilke forhold påvirker udviklingen af
    evalueringspraksis på forskellige politikområder? In Dahler-Larsen P. & Krogstrup
    H. K. (red): *Tendenser i evaluering*. Odense Universitetsforlag.

Georghiou L. (2001), Evaluation in the Balance: Matching Methods and Policy Needs,
    paper presented at The Evaluation Conference, Bruxelles, Nov. 10-12, 2001.
    http://les.man.ac.uk/PREST

Graversen E. K., Kalpazidou Schmidt E., Langberg K. (2002), Innovation and
    Dynamics in Public Research Environments in Denmark – a research policy
    perspective. *The Danish Institute for Studies in Research and Research Policy,
    paper nr 10.*

Guba E. G. & Lincoln Y. S. (1989), *Forth Generation Evaluation.* Sage Publications.

Jennergren L. P. & Obel B. (1985-86), Forskningsevaluering - eksemplificeret ved 22
    økonomiske Institutter, *Økonomi og Politik*, vol. 59, no 2.

Hansen A. & Foss Hansen H. (2000), Evaluering i Danmark. *Nordisk Administrativt
    Tidsskrift* 2/2000, 81.årgang.

Hirschman & Lindblom (1969), Economic Development, Research and Development,
    Policy Making: Some Converging Views in Emery F. E. (red): System Thinking.
    Harmondsworth.

Kalpazidou Schmidt E. (1996), *Research Environments in a Nordic Perspective. A
    Comparative Study in Ecology and Scientific Productivity*. Acta Universitatis
    Upsaliensis. Uppsala Studies in Education 67, Uppsala.

Kalpazidou Schmidt E. (2002), *Organisation and Management of Research
    Environments*. The Danish Institute for Studies in Research and Research Policy,
    paper nr 11.

Siune K & Kalpazidou Schmidt E. (2002), The Use of R&D Evaluations in European Science Policy, Presented at the RTD European Evaluation Network Meeting, Molskroen (DK) 27-28 Nov. 2002.

Mintzberg H. (1983), *Structures in Fives. Designing Effective Organizations*. Englewood Cliffs, N. J. : Prenticehall.

Røvik K. A. (1998), *Moderne Organisasjoner-Trender i organisasjonstekningen ved tusenårsskiftet* Fakbokforlaget.

Ziman J. (1985), *Criteria for National Priorities in Research*. London: Imperial College.

Vedung E. (1997), *Public Policy and Program Evaluation.* Transaction Publishers. New Brunswick. London.

Weick K. E. (1969), *The Social Psychology of Organizing*, Reading: Addison-Wesley.

THE USE OF R&D EVALUATIONS IN EUROPEAN SCIENCE POLICY

A SURVEY

Karen Siune
Evanthia Kalpazidou Schmidt

The Danish Institute for Studies in Research and Research Policy

Abstract

The theme for this report is the outlining of the landscape for R&D evaluation activities in Europe and the use of these in European science policy. The report is based on a study that aimed to increase insight into national and European evaluation practice as well as focus on the implementation and utilisation of evaluation results in policymaking. The rationale for the study is accordingly to outline the concepts and types of evaluations of public R&D used in different countries, how these have been utilised in science policy and what impact, both at micro- and macro-level these have had. This is made in order to illustrate comparatively the R&D evaluation landscape in Europe with regard to adoption, construction, implementation and use of the results in science policy. Furthermore, the study aimed to point out the future needs with respect to evaluations of public R&D, both in terms of how these should be conducted and what perspectives these have in a European science policy framework.

The study of the landscape for R&D evaluation practice and use in science policy is of great significance for the future instrumentation of science policy in Europe and consequently also for the building of the European Research Area.

1. Introduction

Evaluation of R&D activities across Europe has been gaining importance throughout recent decades and is becoming a significant element in policymaking in different contexts. At the same time evaluation seems to be undergoing an institutionalisation and standardisation process in several European countries, in which more and more research organisations and research communities - but also other actors and interests implicated in the research process - engage. The tendency is that evaluation processes - observed at national or European level - are likely to increase over the coming years. Evaluation is consequently perceived in many countries as one of the strategic instruments used in science policy in efforts to promote socio-economic development.

Evaluation landscapes in European countries vary in terms of context, history and paradigms. What is more, evaluation practices in Europe vary with respect to involved actors and interests and the types of issues and problems that the countries are facing and dealing with. In consequence, initiations, objectives, processes and their outcomes, interpretations, impacts, use and learning may be quite different across the continent. A number of countries make systematic use of evaluation and have for this purpose established national evaluation centres and constructed specific standards. Others initiate evaluations and develop standards on an ad hoc basis. Some countries use top-down models, other bottom up or combinations of these two. However, it is a fact that we have limited knowledge of the R&D evaluation landscape in Europe with regard to adoption, construction, implementation and use of results. These points out the need for a mapping of the evaluation landscape and of the evaluation concepts as these are used in European public R&D. A more systematic approach to the use of R&D evaluations in science policy in a broader European context is for that reason essential.

In the project Managing with Uncertainty in Science Policy at the MUSCIPOLI workshop no 3 (Athens, October 2002), where the focus was on building European Research Capacity, it has been concluded among others that there is a need for a more systematic approach to the instrumentation used in building and developing the European Research Area. The uncertainty is high with respect to the instruments that the European countries use in science policy. The uncertainty is even greater with respect to the impacts of the different instruments used. One of the instruments used in science policy is evaluation. The intensity of the use of this instrument in science policy in Europe has not been subject for methodical studies so far.

This is the background for the initiative taken as a part of the preparations for the biannual meeting of the European RTD Evaluation Network, which was held in Denmark in November 2002, focusing on ex-post evaluation practices at national and European level of public R&D activities.

This report is based on empirical data gained through a questionnaire answered by different agents involved in R&D activities and policymaking from 22 European countries, including EU Member States, accession and associate states. The study is made in an attempt to contribute to increase insight into national and European evaluation practice as well as the implementation and utilisation of the results in European policymaking. The purpose of the study was accordingly to map out, in a comparative perspective, the concepts and types of evaluation of public R&D used in the European countries, how these have been used in science policy and what impact, both at micro- and macro-level these have had. This, also in order to contribute to a better understanding of the variations in evaluations of public R&D, both in terms of how these are conducted and what perspectives these have in a European science policy framework. The focus is accordingly on evaluation of R&D activities and not on evaluation of R&D policies.

The study, this report is based on, ought to be seen as a first step, which outlines the contours for the topic in focus. More extensive research is however required in order to study the complete range of evaluations and their use in European science policy.

The central questions in this report are consequently:

- How does the conceptualisation of evaluations look like in a European perspective?
- What are the objectives and motives for public R&D evaluations in the European countries? Who takes the initiatives? Who are the evaluators?
- What are the similarities and the differences with regard to context, adoption, construction, implementation and learning of R&D evaluation procedures in science policy in the European countries?
- What characterises the evaluation standards and methods used in R&D within Europe?
- What is the use and what are the impacts of R&D evaluations in science policy?
- Is there a need for increased coordination and systematisation of evaluation activities in science policy in Europe?
- Can we in addition - based on the results of this survey - talk about plausible regional evaluation practice patterns or patterns that differentiate EU members from accession countries and non-member states? And moreover, can we point out any patterns in R&D evaluation practice that distinguish north from south and/or east from west Europe?

The answers to the questions presented here are of great significance for the future instrumentation of science policy in Europe and consequently also for the building of the European Research Area.

2. Types of evaluations related to public R&D used by European countries – Object and level of aggregation

In order to map out the landscape of evaluation in Europe we collected data on the different types of evaluations and objects that these focus on as well as data on which level of aggregation this is conducted. The respondents were for that reason asked to give information on the percentage breakdowns of evaluations related to public R&D, making a distinction between evaluations at universities, evaluations at other types of public research institutions, evaluations of national or European programmes and/or other types of evaluations used. The results of the study reveal that there is great differentiation with respect to the types of evaluations related to public R&D that European countries make use of.

According to the revealed results 7 countries (Austria, Belgium, Bulgaria, Denmark, Finland, France and Spain) perform *evaluations of R&D at the universities* in less than 25% of the total evaluation efforts, while for 7 other countries (Germany, Greece, the Czech Republic, Ireland, Norway, Poland and the UK) the figure is between 26-50%. 3 countries (Estonia, Hungary and Latvia) apply this type of evaluation in 51-75% of all the evaluations related to public R&D. Lithuania, the Netherlands, Slovakia and Slovenia make use of evaluations of R&D at the universities more frequently, namely in more than 76% of all evaluations related to public R&D.

Focusing on the *evaluations of R&D at other types of public research institutions*, the results of the questionnaire show that 5 countries (Hungary, Lithuania, the Netherlands, Slovakia and Slovenia) make intensive use of evaluations of R&D at other types of public research institutions, with a frequency that is between 76-100% of the total evaluations, while in 4 other countries (Bulgaria, Estonia, France and Latvia) evaluation of such research institutions engages between 51-75% of the activities. 7 countries (Belgium, the Czech Republic, Finland, Greece, Ireland, Spain and Sweden) use this type of evaluation in less than 25% of the cases, while 6 other countries (Austria, Denmark, Germany, Norway, Poland and the UK) make use of it in between 26-50% of all types of evaluations that are the subject for this analysis.

Looking at *evaluations of national programmes* (including programmes at universities and/or other public research institutions), we can conclude that 5 countries show an intensity that is higher than 75% namely Hungary, Latvia, Lithuania, the Netherlands and Slovenia. On the other side, 7 of the responded countries use evaluation of national programmes in less than 25% of the total evaluation activity related to public R&D. 9 of the countries in focus demonstrate such activity in an extent that is between 26% and 75% (with the Czech Republic, Estonia, Finland, Germany, Greece and

Ireland showing the next most intensive activity e.g. between 51-75%). Austria, Sweden and the UK have an activity that considers 26-50% of the total evaluations related to public R&D.

Moreover, according to data received through the questionnaire, 13 countries *evaluate European programmes* (including programmes at universities and/or other public research institutions) with an intensity that is less than 25% of the evaluations carried out, while in 4 countries (Estonia, Germany, Greece and Hungary) such activity covers between 51-75% of all the evaluations carried out while in Latvia it covers more than 76% of the total public R&D evaluations. Evaluations of European programmes are not relevant in the Czech Republic and the Netherlands.

In addition, it is important to emphasise that other types of evaluations such as *evaluations of innovation systems* in Finland and Sweden, *system level evaluations of all major R&D institutions* in Germany and *evaluation of cooperation between public institutes and industry* taking place in Slovakia and finally evaluation of projects carried out by The Lithuanian State Science and Studies Foundation, are other frequently used instruments.

3.   The initiators of R&D evaluations – Adoption

The question of initiation and adoption of evaluation is essential first and foremost with respect to implementation and impact of the activity as it is closely connected to the motives and objectives of the evaluation. The respondents were therefore asked to give information about the initiators of the majority of the public R&D evaluations in their country.

According to the results of the study, *the ministries (national or regional) that are responsible for R&D activities* in the European countries are taking the initiatives for public R&D evaluations in 76-100% of the cases in Estonia, Greece, Latvia, Lithuania and Slovenia. The ministries are highly initiating actors also in Austria, Hungary, Ireland, Poland and Slovakia (initiated between 51-75% of the entire evaluation activity). The top-down evaluation model is consequently the most usual in the above-mentioned countries.

In Belgium, Denmark and Germany, ministries are initiators of evaluations in more than 26% but less than 50% of the cases. The ministries are low profile initiators responsible for R&D activities in Bulgaria, the Czech Republic, Finland, France, the Netherlands, Norway, Spain and Sweden. These countries demonstrate a ministerial initiation percent that at highest does not exceed 25% of all the public R&D evaluations.

Focusing on other actors taking initiative to evaluations in Europe namely *the research councils,* we can conclude that high profile initiators (initiated more than 76% of the carried out evaluations) are the research councils in Estonia, France, Latvia, Norway, Slovakia and the UK. Even research councils in Bulgaria, Poland and Slovenia show a high degree of involvement, as these are the initiation agents for evaluations of R&D in between 51-75% of the cases. In Denmark, Sweden and the Netherlands, research councils function as initiators of public R&D in between 26 and 50% of the cases. In Austria, Belgium, the Czech Republic, Germany, Ireland and Lithuania on the other side, research councils are low profile actors with respect to initiation and adoption of R&D evaluations, which means that these are initiators in less than 26% of the evaluations carried out in these countries.

There has to be pointed out that initiators of a certain evaluation process can represent two or more of the stakeholders. This is the case for instance in Estonia, Poland and Denmark where the ministries assisted by the research councils adopted and carried out evaluations, or in Slovakia where the ministries and the research councils in cooperation with the universities initiated and adopted evaluations. In some countries (Greece, Hungary and Spain) the research councils do not have the responsibility of functioning as initiators to R&D evaluations.

Focusing on the bottom-up evaluation model where *the universities* take the initiative to carry out evaluations we can differentiate between countries as the Czech Republic, Greece and the Netherlands, where a majority of the evaluations are initiated by the universities e.g. between 51-75% of the total public evaluations (as mentioned-above, often in cooperation with the research councils). The universities in Slovakia are initiators of more than 76% of the evaluations. In Austria, Denmark, Finland, Germany, Hungary and Lithuania the universities are low profile initiators. In Slovenia and Spain on the other side 26 -50% of the initiations are bottom-up undertaken by the universities. Universities as initiators of R&D evaluations are not relevant in Sweden.

The Czech Republic, Bulgaria and Finland show however a different pattern since it is often *the unit for evaluation* that initiates and decide the adoption of evaluations. According to the responses this is the case in more than 50% of the initiations to evaluation processes in the later countries and in more than 76% of the cases in the Czech Republic. This implies that the Czech Republic uses a clear bottom-up evaluation model where the university *or the unit for evaluation decides internally* on the adoption and implementation of the evaluation process. Greece and Ireland use a bottom-up model in 26 - 50% of the evaluation cases. The unit for evaluation is the initiator of R&D evaluations also in Lithuania, Poland, Slovenia and Spain, though in less than 26% of the evaluation activities.

EU is functioning as an initiator (together with national/regional ministries and other R&D agents or actors) usually in Hungary and Greece, and in a minor degree (less than 25% of the adoptions) in Austria, Belgium, Germany, Ireland and Lithuania. The EU as initiator of R&D evaluations is not relevant in countries such as the Czech Republic, Spain, Sweden and the UK.

Other R&D agents and actors (with exception of the above-mentioned) are initiators in 51-75% of the cases in Hungary and in 26 - 50% of the evaluations in Sweden. Even in Austria, Finland, Germany, Lithuania, Slovenia and the UK other R&D agents are initiating evaluations, though in less than 25% of the adoptions in question are initiated by other agents than ministries, research councils, universities and research units, and the EU.

In Germany the initiators of the evaluations were an interesting combination of ministries, universities and EU initiatives and *other R&D agents and actors*. A closer look at the German situation that is characterised by a great decentralisation might be of great interest, even from a methodological point of view.

It is of great significance for the use of evaluation in science policy in Europe in general (and in particular for countries as Hungary, Sweden, Germany, Austria, Finland, Lithuania, Slovenia and the UK, where other R&D agents are actively engaged in evaluation) to get more information on who these R&D agents or actors are and in addition get more information on what characterises their motives, objectives and need for knowledge. The survey points out though that such agents and actors that initiate evaluations are usually R&D funding agencies (such as VINNOVA in Sweden) or major research institutions in internal collaboration (as the case is in Germany).

4. The context for R&D evaluations

With respect to the overall context for evaluations related to public R&D, the results of the study reveal that 12 countries use evaluations with regular interval as a *part of a routine*, not specified in a contract, in more than 50% of the cases, 8 of them in more than 75% of all the cases. Almost all the EU accession countries make intensive use of evaluations with regular intervals (Hungary and the Czech Republic though in an extent no more than in 50% respectively 25% of the cases). Besides the accession countries, France, the Netherlands and Norway make intensive use of evaluations as part of a routine with regular intervals (in 76-100% of the cases) while Greece and Sweden carry out evaluations as a part of a routine as well in 51-75% of the cases. Even Germany and the UK use evaluations with regular intervals, though in an extent that is calculated to be between 26-50% of the public R&D evaluations taking place in the country.

On the contrary, countries as Austria, Belgium, the Czech Republic, Finland, Ireland and Spain use evaluation as a routine in less than 26% of the total evaluations of public R&D. Ad hoc evaluations are as a result carried out more frequently in most of these countries.

Evaluation as *a follow up of a contract*, known to take place from the beginning of the activity, is often used in the Czech Republic, Finland and Slovakia (in 76 – 100% of the cases). In Hungary and the UK, evaluation is part of a contract in between 51-75% of the cases and is known to be an integrated part of future contracts for all public institutions in Denmark. This feature is used in a limited scale in Austria, Belgium, Bulgaria, France, Germany, Ireland, Lithuania, the Netherlands, Norway and Sweden. This implies that 10 countries do not make use of evaluations in connection with contracts in a noteworthy degree. In Greece evaluation is an integrated part of a contract, known to take place from the beginning of the activity in 26 - 50% of the cases. Evaluations as part of a contract are not relevant in Spain.

In contrast to the routine or contract based evaluations, there are countries that decide on implementations of evaluations primarily on an ad hoc basis. According to the results of the study, 4 countries (Austria, Belgium, Bulgaria and Germany) use evaluation on an *ad hoc basis* in more than 50% of the cases while Greece and Ireland make use of it in more than 75% of the evaluation activity in the country. 11 countries (Estonia, Finland, France, Hungary, Latvia, Lithuania, Norway, Poland, Slovenia, Sweden and the UK) use evaluation on an ad hoc basis in less than 25% of the activity. In Spain, ad hoc evaluations are used in 26 - 50% of the total evaluations carried out in the country. In the Czech Republic and the Netherlands ad hoc evaluations are not relevant in a science policy context.

5. Reasons for ad hoc evaluations of R&D – Why evaluate? – Objectives and Motives

The ranking of the main objectives and motives for using evaluation as an instrument for R&D assessment does not show any significant differentiation between the European countries as the majority point out *the quality of research* as the main reason for the activity.

One of the exceptional countries Sweden ranks *the societal relevance of research* as the main reason for ad hoc initiatives. This has its natural explanation in the fact that the Swedish law for Higher Education and Research explicit mentions the societal relevance of research activities as one of the essential objectives of the Swedish higher educational system. Even Belgium and the Netherlands mention the societal

relevance as a significant reason for Ad hoc initiatives - second only to quality of research.

Germany ranks the *socio-economic/societal impact of R&D* as the main reason for ad hoc initiatives. Furthermore, Finland specifies that the main purpose for ad hoc initiatives is to find out the *societal impacts of R&D activities*. Austria uses ad hoc evaluation as a *learning tool in the process* of starting new programmes, while Ireland uses it in *assessments of impact and effectiveness of existing programmes*.

Assessment of *institutional efficiency* is the second reason of importance for ad hoc initiatives in European evaluation practice, as the results of the questionnaire reveal. This is used by Estonia, Greece, Ireland, Latvia, Lithuania, Norway, Poland, Slovenia and Spain. Hungary and Bulgaria however rank the issue of institutional efficiency as being the first reason for ad hoc evaluation while Austria, Finland, Germany, the Netherlands and Sweden rank it as the third reason for ad hoc evaluation activity.

Finally, *Cost evaluation* is not practised very often in Europe with an exceptional country Greece where this as the main reason for ad hoc initiatives.

6. Who is the evaluator of public R&D? – Organisation and implementation of evaluations

We have earlier in this presentation discussed the issue of the initiators of public R&D evaluations in Europe. The question put at this point is who are undertaking, organising and implementing the evaluations in the European countries.

The results of the study reveal that the majority of the European countries - that is to say 17 of the 22 countries that responded to the question - make first and foremost use of *ad hoc groups set up for the specific evaluation task*.

It has to be emphasised at this point that Estonia, the Netherlands and Sweden engage mainly *purely international peers* in the organisation and implementation of public R&D evaluations. Greece, Hungary, Latvia, Lithuania, Spain and the UK engage first and foremost *purely national peers* while Denmark, The Czech Republic, Germany and Norway engage a *combination of international and national peers* to carry out the evaluations. In several European countries there is a general request for the use of international evaluation panels.

Four countries namely Austria, France, Ireland and Lithuania have institutionalised research evaluation practice and make primarily use of *specific public institutions* for this purpose and do so in an even greater degree than countries such as Germany,

Greece, Hungary, Latvia, Poland and Spain that also engage specific public institutions, although second to Ad hoc groups.

Finland makes mainly use of *private evaluation contractors* and secondly of specifically organised ad hoc groups. Even Estonia, Belgium, Bulgaria, Ireland, Slovenia, Sweden and UK engage private evaluation contractors. Germany makes first and foremost use of ad hoc groups, secondly engages public research institutes and thirdly private research institutes as contractors to organise and implement evaluations of R&D.

7. The methods and standards used in R&D evaluations – Construction

The results of the study on the issue of construction of standards and methods in public R&D evaluations in Europe demonstrate that there is a range of similarities with regard to how the respondents rank the most frequently used standards in their countries.

Peer review is the dominating method used in the majority of evaluations in Europe. 15 countries rank peer review as the most frequently used standard.

However, Finland shows a different pattern by ranking *users evaluation* as the most frequently used method and *capacity evaluation* as the next most frequent standard used, followed by peer review evaluation on a third place. Users evaluation is the second regularly used R&D evaluation standard in Sweden (second only to peer review). Austria and Norway rank users evaluation as the third frequently used method. Capacity evaluation is also frequent in Bulgaria and Germany.

Survey descriptions are the most frequently used methods in evaluation of public R&D in Austria, Greece and Slovakia and the next most used standard in evaluation of R&D in Germany, Hungary, Poland, Slovenia and Spain. The Netherlands and Finland use survey descriptions too, though as a third respectively fourth option. However, the rest of the countries participating in this survey do not make use of this element in their evaluations.

Bibliometric analyses in form of *publication analysis* are frequently applied in Lithuania and the UK. Also Belgium, the Czech Republic, Estonia, Latvia and Slovenia make use of publication analysis. *Citation analysis* is the second frequently applied method in the Netherlands (peer review is the most frequent method used) and the UK. Citation analysis is also frequently used in Belgium, the Czech Republic, Estonia, Hungary, Latvia, Lithuania and Slovenia.

Activity evaluation is the second most used method in the majority of R&D evaluations in Denmark, France and Lithuania. The standard of *management evaluation* is most frequently used in Norway but also in Austria, Denmark, Germany, Greece and Sweden. Several responses indicate however that this element is going to be more prominent in future European evaluation activities.

Historical evaluation is very rare in European R&D evaluation practice.

Ireland uses structured interviews with programme participants, administrators and funders; and also questionnaires, peer reviews and comparative studies of similar programmes with other countries.

It has to be emphasised that *review of strategies and management* as well as *impact evaluations* are developing quickly, but also "*comparisons with similar programmes in other countries*" are coming up, as the comments to the questionnaire illustrate in the following: "Peer review and survey-based impact evaluations are predominating. If ´management evaluation´ means assessment of institutional performance, then there is an increase of such efforts since the late 1990s. Bibliometrics have gained more ground only recently, e.g. in the area of medical research and university hospital" and furthermore another respondent comments, "Many evaluations include review of programme strategy and expected impact".

However, the results of the questionnaire reveal moreover that it is difficult to make a clear distinction between the uses of different methods in a certain evaluation context since the methods are rarely used exclusively. Combinations of these - including many elements from different methods - are common standards. The following statements illustrate this: "These alternatives are difficult to use. Institutional evaluations usually include some peer-elements, but also user-surveys etc. Programme evaluations usually include all these elements. Evaluations concerning the innovation system are more policy-oriented. The main focus in all evaluations is in societal impact analysis: what difference these programmes or institutions etc. stand for?" and "The methods mostly are not used exclusively, but often in combinations, e.g. peer review supported by citation analysis. Users and management evaluations are now developing rapidly".

8. Availability and dissemination of R&D evaluation results in different European countries

There are different policies on the availability of public R&D evaluation reports to the public in general across Europe. The results of the study pointed out a differentiation with respect to this issue between four main groups of countries.

The first group, constituted by Denmark, Estonia, Finland and Norway, is characterised by *total openness and transparency* in evaluation processes, in the meaning that all public R&D evaluation reports and consequently results are available to the public.

The second group, which includes Germany, Greece, Hungary, Ireland, Latvia, Lithuania, Slovakia, Sweden, the Netherlands and the UK, is characterised by *great openness* since most of the evaluation reports and results are available to the public.

The third group of countries comprising Austria, Belgium, Bulgaria, the Czech Republic and France, is characterised by *some degree of openness* in view of the fact that only some of the public R&D evaluation reports are available to the public.

Finally, there are three countries, namely Poland, Slovenia and Spain, where *none of the R&D evaluation reports and results are available to the public*.

The format and language in which the evaluation reports are available to the public are different in the European countries. *Full text reports in the national language* are available in the Czech Republic, Germany, Greece, Lithuania, Norway and Sweden. *Full text reports in English* (usually when the ad hoc groups are constituted by international or a mixture of international and national evaluation experts) are available in the Czech Republic, Denmark, Estonia, the Netherlands, Norway, Sweden and the UK. *Summary of the evaluation reports in English* is available in the Czech Republic, Denmark, Hungary, Norway and Sweden. *Summary of the evaluation reports in the national language* is available to the public in Estonia, Greece, Latvia and Lithuania.

In conclusion, focusing on the issue of availability of the obtained public R&D evaluation results, it is obvious that we need some more information on the issue that could illustrate what the reasons for not being open are and how the dissemination of the results is made possible in such a context. On the opposite side what's more, we need to focus on the risks that are implicated in the policy of being fully open and transparent.

9. Impacts of R&D evaluations on meso-level - Utilisation

The impacts of public R&D evaluations vary, from immediate impacts on meso-level e.g. consequences for the evaluated unit or organisation, to impacts on macro-level e.g. national science policy initiatives.

The European countries were asked to rank the utilisation of R&D evaluations and mark the typical consequences of the evaluations for the unit in focus. The respondents were therefore asked to specify the consequences for the evaluated units as being "value for money"-based, based on the approach of exercising ongoing control of public research activities and finally based on efforts to adapt organisation structures or contents, which are of cognitive character.

The responses to the questionnaire show that *allocation of resources* and *determining of the future of the evaluated unit* are the most frequently estimated impacts of evaluations on meso-level since it is the first utilisation priority of 13 countries and the second of 9 countries. Allocation of resources is typically used first and foremost in Belgium, the Czech Republic, Estonia, Latvia, Lithuania, Poland, Slovakia, Slovenia, Sweden and the UK as well in Bulgaria and Greece, though in a minor degree in the last two countries than in the previous mentioned. The utilisation of evaluation in determining the future of the unit in focus is most recurrently the impact of the process in Bulgaria, France and Hungary. In other countries too namely, Denmark, Estonia, Lithuania, Poland, Slovenia, Sweden and the UK, the evaluation impact is determining the future of the unit, although this is the utility tool used as a second option.

It is however interesting to observe that *ongoing control* is perceived as the dominating utility tool in Denmark and the Netherlands, and the second frequently used in France, Hungary and Latvia.

Adaptation of organisation structure is used first and foremost in Norway and secondly in Germany and Spain.

Another interesting feature is that *adaptation of content or work programme* is the characteristic impact on meso-level mainly in Finland, Germany, Greece and Spain, and secondly in Ireland.

Changes in research design on the other hand are not a first priority utility tool in any country although Austria, the Czech Republic, Finland, the Netherlands and Norway make use of it as a secondary science policy tool.

It has to be emphasised that Austria shows a very different pattern as the evaluation in the first place is perceived as a *foundation for discussion*. Ireland on the other side, as mentioned earlier, uses evaluations in order to improve programme effectiveness and ensure "value for money".

In conclusion, the dominating impacts of evaluations on meso-level are related to the allocation or reallocation of resources and determine the future of the evaluated institutions or units. Even ongoing control activities are related to the "value for money" attitude.

10. Impacts of R&D evaluations on macro-level - Utilisation

We have detected how the evaluations are used on meso-level. The important question from a national as well as from a European evaluation perspective is how the majority of the public R&D evaluations are used on macro-level. The central question to answer in this context is consequently, whether the evaluations are the foundation for science policy at macro-level.

The results of the study reveal that R&D evaluations have significant impacts on science policy in the European countries with some remarkable exceptions. Evaluations as a tool in order to carry out *changes in science policy* as well as *changes in allocation of resources* and *changes in planning and budgeting* are used primarily in 17 of the 22 countries participating in this survey. Focusing on each of these categories, we obtain the following illustration.

Evaluations used for carrying out *changes in science policy* is an instrument used first and foremost in the Czech Republic, Finland, Slovakia, Spain and the Netherlands and secondly in Bulgaria, Denmark and Hungary. Even in Austria, Estonia, Germany, Greece and the UK evaluations are used as instruments in changing science policy, though in a more limited extend than in the countries mentioned above.

Evaluations as an instrument in order to achieve *reallocation of resources* are used firstly in Belgium, Bulgaria, Denmark, Greece, Latvia, Lithuania, Slovenia and the UK, and as a second instrument in the Czech Republic, Estonia, Germany, Norway, Sweden and the Netherlands.

Evaluations as a tool used in *planning and budgeting* of national research activities is common in Estonia, Germany, Hungary and Norway, and is used as a second priority instrument in Austria, Finland, Ireland, Greece, Latvia, Poland and the UK.

However, there are some countries that in general do *not make use of the majority of R&D evaluations* on macro-level namely Austria, France and Poland. The question is accordingly why. R&D evaluations in Austria are perceived as a basis for discussion. In France evaluations are used to achieve a general feeling on the quality of research while in the Netherlands evaluations are primarily used by university boards (are mainly used on meso- and not on macro-level).

Sweden has a different strategy that aims at *learning from evaluation experiences via questions and answers* in order to improve programme performance at the first place. Reallocation of resources is the second impact of R&D evaluations in Sweden. Ireland uses evaluations primarily in order to get value for money.

Evaluation as an instrument for carrying out *changes in the administrative legislation* or *changes in the rules concerning the structures of appointments* is rarely used in the European countries. Only in Lithuania, the Czech Republic and Slovenia is this the case - as far as the first instrumentation regards. Slovenia and Finland use evaluations also in order to restructure the rules for appointments.

11. Lessons learned and perspectives

The central question in a European perspective is evidently what the experiences in general are as well as the lessons learned of public R&D evaluation practice. According to the responses to the questionnaire, 10 European countries (Belgium, the Czech Republic, Estonia, Finland, Ireland, Lithuania, Poland, Slovenia, Sweden and the UK) are requesting *better evaluation methodologies and standards* while 3 countries (Estonia, Ireland and Slovakia) perceive the *existing evaluation methods as adequate* and sufficient.

Moreover, 14 countries (Austria, Bulgaria, the Czech Republic, Denmark, Estonia, Finland, Germany, Greece, Ireland, Latvia, Poland, Slovakia, Spain and the Netherlands) point out as lessons learned in general from evaluation practice a greater degree of *coordination and systematisation* of public R&D evaluation activities in Europe. The following comment illustrates the above-mentioned: "As of 2003, a new national approach will include managerial and strategic issues. Standardised methods of evaluation and collecting documentation from research institutes will bring the administrative burden for institutes and researchers down".

This factor emphasises the need for further discussions on evaluation methodology and instrumentation issues in the context of a European network.

In addition, 7 countries (Bulgaria, Denmark, France, Lithuania, the Netherlands, Sweden and the UK) require *more issues to be included* in evaluations of public R&D.

The responses to the questionnaire have provided tangible suggestions of elements to be included in future R&D evaluation practice.

Issues as **socio-economic impact** assessments and assessments of technology and technology transfer, development of **common instruments for impact analysis** of R&D and research funding, evaluation of societal impacts and **relevance of research** activities, systematisation, coordination and **regularity** in evaluation processes are some of the issues to pay attention to in European R&D evaluation, raised in the responses to the questionnaire. The following statements illustrate this: "Evaluation is a strategic tool for development, not a control tool for management. Societal impacts of R&D and public funding, as well technology assessments are needed for improving the science and technology policy and strategy. We also need common European measures for impact analysis"

Reviewing of agency **strategies, objectives, management and instrumentation** are other issues of interest to focus on in evaluations, according to the respondents, as the following fragments demonstrate: "Evaluations to review agency strategy within an area (e.g. biotechnology), a tool (e.g. centres of excellence) etc." and likewise, "Although the existing methods seem quite adequate, the research community asks for more emphasis on the mission, leadership and strategy in the evaluation of research institutes. These changes also imply less emphasis on disciplinary evaluations".

Better coordination of evaluation, **adjustment to EU standards** and **more extensive use of evaluation results in R&D science policy** development are high on the agenda of the candidate countries as one respondent expresses it: "Better coordination with EU standards and more extensive use of evaluation results for S&T policy formulation".

The respondents requested moreover better evaluation methodologies such as improvement of **methods used at macro-level**, methods used in **assessments of societal impacts** and **relevance analysis** as well in **assessing interdisciplinary research** and investments in **basic research**. As some of the respondents put it: "Improvement of evaluation methodologies is necessary for projects on the boundary between basic and applied research and for interdisciplinary projects. Generally, an analysis of used methods and their adequacy should be done on macro-level", or "With the ever-growing demands from governments for proof of value-for money, improved techniques for assessing this are needed, particularly in the case of investments in basic research" and furthermore "there is a need to strengthen the measures of relevance".

Other issues asked for to be included in evaluation practice are **differentiation on evaluation practice** depending on the research field (with respect to basic research and interdisciplinary research in particular), as well as the tasks and the types of research institutions. The following statements illustrate this: "Issues to be included in evaluation studies vary according to the type of institution being evaluated, a fixed format can not always be used", and likewise "Evaluation should be more differential, paying attention to the fields of sciences and to the tasks of studies and research institutions".

In conclusion, the experiences of the evaluation practice in Europe show that there is a need for:

- **Further development of evaluation as a strategic tool in science policy**

- **Better evaluation methodologies, particularly on macro-level**

- **Better instrumentation on the socio-economic impacts of R&D activities**

- **Better evaluation methods and better instrumentation assessing the societal relevance of R&D activities**

- **The context for evaluation is also in focus in the European evaluation and science policy agenda calling for a differentiation of the evaluation processes depending on the context**

- **Increased coordination and systematisation of the evaluation activities across Europe.**

The latest points may however seem contradictory as coordination and systematisation asks for increased standardisation and it is recognised in general that the European countries are different both with respect to science policy and the level of economic development and consequently have a different point of departure in their investment in science and research.

Coordination and systematisation are moreover opposing to requests for the evaluation practice to make a differentiation of processes and methods depending on the context for the evaluation. Common standards and methods may not be possible to apply in all European countries. An open debate on these issues is important for the future development of evaluation practice and its use in science policy in Europe.

12. Concluding remarks

Based on the study of adoption, construction and implementation of R&D evaluation and the use of it in science policy we can conclude a greater than ever interest in and an intensification of the use of evaluation in science policy in Europe. Evaluation as a concept is undergoing a standardisation process not only in the EU but also in other European countries. The use of evaluations in assessments of public R&D activities with regular intervals in the majority of the accession countries and the focusing (of the majority of the European countries participating in this survey) on coordination and systematisation of the evaluation activities in general illustrates the intensity of the standardisation process that is ongoing in many European countries.

However, besides similarities with respect to evaluation practice, there has been observed a differentiation between the countries that participated in the survey regarding the types of evaluations, the instruments used and the utilisation of evaluation in science policy across Europe. The study outlines though a framework of general perspectives on the use of evaluation in a science policy perspective that could be useful as a basis for further efforts in a European context. The implications of these are calling for more issues to be included, in particular with focus on impact analyses, assessments of socio-economic impacts and societal relevance of R&D activities as well as better instrumentation on macro-level evaluations. Increased coordination and systematisation of evaluation of R&D activities is another potential that is worth to focusing on. There is an apparent request for further development of evaluation as a strategic tool in European R&D policymaking.

Comparative international evaluations, which make use of mainly international peers, may be one strong strategic instrument that can be used in European science policy.

The results of the comparative study presented here reveal though that it is not possible to make a distinction between certain patterns that differentiate EU member states from accession or associated states. Neither can we talk about regional patterns when focusing on the east-west or north-south dimension in the use of evaluation practice across Europe.

13. Further studies and analyses

In conclusion, there is in fact a need and a pronounced request for supplementary studies on and analyses of evaluation practice and its use as a strategic tool in science policy in Europe. The problems related to this issue are many and the questions far from answered. This study is just a prologue to the topic and may function as a starting point for discussion and inspiration for further analyses. These analyses have a lot to gain if complemented with qualitative approaches and case studies too.

Science policy has become a significant European policy issue. An overview of R&D activities and evaluation practice could be one of the important tools used in European science policy. It seems to be anticipated to focus on the European evaluation landscape, if we are going to use evaluations as an instrument in European science policy and if we intend to decrease the uncertainty with respect to the instrumentation and the impacts of science policy in Europe. Hence, this is one of the challenges that the European science policymaking is faced with.

REFERENCES

Albæk E. (1998), Knowledge interests and the many meanings of evaluation: A development perspective, *Scandinavian Journal of Social Welfare*, vol. 7, pp. 94-98.

Foss Hansen H. & Borum F. (1999), The Construction and Standardization of Evaluation. The Case of the Danish University Sector, *Evaluation*, vol 5, no 3, pp. 309-329.

Foss Hansen H. (2001), Rejsende praksis: Hvilke forhold påvirker udviklingen af evaluerings-praksis på forskellige politikområder? In Dahler-Larsen P. & Krogstrup H. K. (red): *Tendenser i evaluering*. Odense Universitetsforlag.

Georghiou L. (2001), Evaluation in the Balance: Matching Methods and Policy Needs, paper presented at The Evaluation Conference, Bruxelles, Nov. 10-12, 2001. http://les.man.ac.uk/PREST

Graversen E. K., Kalpazidou Schmidt E., Langberg K. (2002), *Innovation and dynamics in public Research environments in Denmark – a research policy perspective.* The Danish Institute for Studies in Research and Research Policy, paper nr 10.

Guba E. G. & Lincoln Y. S. (1989), *Forth Generation Evaluation.* Sage Publications.

Kalpazidou Schmidt E. (1996), *Research Environments in a Nordic Perspective. A Comparative Study in Ecology and Scientific Productivity*. Acta Universitatis Upsaliensis. Uppsala Studies in Education 67, Uppsala.

Kalpazidou Schmidt E. (2002), *Organisation and Management of Research Environments*. The Danish Institute for Studies in Research and Research Policy, paper nr 11.

Kalpazidou Schmidt E. (2002), Evaluation and Science Policy. Presented at the European RTD Evaluation Network Meeting, Molskroen (DK) 27-28 Nov. 2002.

Ziman J. (1985), *Criteria for National Priorities in Research*. London: Imperial College.

Vedung E. (1997), *Public Policy and Program Evaluation.* Transaction Publishers. New Brunswick. London.

Appendix A

List of participants in The European RTD Evaluation Network Meeting,
27-28 November 2002, Molskroen, Femmøller, Denmark.

Name	Organisation	Country
Klaus Zinöcker	Joanneum Research	Austria
Alberto Silvani	European Commission	Belgium
Birgit De-Boissezon	European Commission	Belgium
Françoise Thys-Clement	Université Libre de Bruxelles	Belgium
Olivier Rouland *	European Commission	Belgium
Dana Wagnerova	Academy of Sciences of the Czech Republic	Czech Republic
Anne-Mette Pedersen *	The Danish Institute for Studies in Research and Research Policy	Denmark
Evanthia K. Schmidt *	The Danish Institute for Studies in Research and Research Policy	Denmark
Karen Siune	The Danish Institute for Studies in Research and Research Policy	Denmark
Kaare Aagaard *	The Danish Institute for Studies in Research and Research Policy	Denmark
Helle Martinson	Estonian Science Foundation	Estonia
Robin Gustafsson	TEKES	Finland
Bastiaan de Laat *	Technopolis France	France
Stefan Kuhlmann	Fraunhofer Institute Systems and Innovation Research	Germany
Nick Constantopoulos	Ministry of Development	Greece
Zoltán Peredy	Hungarian Ministry of Education	Hungary
Bill Hogan	FORFAS, Science and Technology Division	Ireland
Anda Adamsone	Latvian Academy of Science	Latvia
Renata Bareikiene	Science Council of Lithuania	Lithuania
Arne Berge	The Research Council of Norway	Norway
Vojtech Rusin	Slovak Academy of Sciences	Slovakia
Edvard Kobal	The Slovenian Science Foundation	Slovenia
Javier Martinez Vassallo	Ministry of Science and Technology	Spain
Torbjörn Winqvist	Vinnova - Swedish Agency for Innovation Systems	Sweden
Niels Eilskov Jensen *	European Space Agency	Netherlands
Stephan van Galen	Association of Universities in the Netherlands	Netherlands
Margaret Dennis	Department of Trade and Industry	United Kingdom

* Invited speakers/participant but not a member of the network

Appendix B

Slides from

EVALUATION AND SCIENCE POLICY

Paper presented at The European RTD Evaluation Network Meeting
Molskroen (DK) 27-28 November 2002

Evanthia Kalpazidou Schmidt
The Danish Institute for Studies in Research and Research Policy

FUNDAMENTAL QUESTIONS

- Objectives and Motives
- Object
- Initiation and Timing
- Organisation
- Criteria
- Methods and Standards
- Dissemination
- Implementation of Results
- Impacts

THE ROLE OF THE EVALUATOR

- Evaluator as a *social engineer* in the decision-making process
- Evaluator as a *controller*
- Evaluator as an *adviser*
- Evaluator as a *mediator* between divergent knowledge interests
- Evaluator as a *midwife* or therapist

Different evaluation models (I)

- The **efficiency model:** focus is on cost-benefit or cost-effectiveness analyses
- The **productivity model:** evaluations focus on the relationship between input and output
- The **effectiveness model:** focus is on the effort and the objective of the activity
- The **effect model:** focus is on the effects (intended and unintended)

Different evaluation models (II)

- The **process model:** focus is on the implementation process
- The **utility model:** focus is on the users assessments of the effort
- The **interest model:** focus is on the different stakeholders/interests assessments of the effort
- The **peer review model:** implies that experts within the same field evaluate their colleagues.

Adoption of evaluation as a science policy instrument is possible under the following conditions (I):

- When central actors have a need for **control** (summative evaluation) and/or **learning** and development (formative evaluation)
- When concrete **problems need solving**
- When there is a need to use evaluation as a **justification or legitimation** instrument

Adoption of evaluation as a science policy instrument is possible (II):

- If evaluation matches the existing **repertoire/ agenda** in a policy area
- When evaluation is perceived as a strategic tool that can be used in a **political setting**
- When there is a need for **opinion making** and/or development of an **identity** in a policy area

Conditions within a policy area that are influencing *the construction of standards*

- Standards match the aim of the evaluation in summative or formative terms
- Standards reflect central institutionalised values in a policy area
- Standards reflect the profile of the dominant actors, their interests and agenda
- In highly politicised areas there is an ongoing struggle about the construction of evaluation standards

The intensity of the implementation of evaluation in a policy area (I)

- Reflects the need for control and/or learning
- Reflects the range of the problems on the political agenda
- Reflects the legitimation that is related to the evaluation process

The intensity of the implementation of evaluation in a policy area (II)

- Reflects the range of the evaluation competence in the area
- Reflects the conflict level in a policy area. In highly politicised areas - high intensity of implementation of evaluations
- Reflects the degree of identity uncertainty in the policy area

Rationales for use of evaluation in policymaking (I)

- The *target-instrumentation rationale:* evaluation is a control and/or a learning mechanism
- The *problem-responsibility rationale:* evaluation is driven by problem solving
- The *value-based rationale:* evaluation is a justification and legitimation activity

Rationales for use of evaluation in policymaking (II)

- The *cognitive rationale:* actions are determined by cognitive processes. Actors have some kind of a "repertoire" or agenda
- The *dialogical rationale:* evaluation is a dialogue process (opinion making and identity development)
- The *political rationale:* actions are determined by stakeholders interests, by dialogue processes and intentions to compromise

The use of evaluation in science policy in Denmark

- A policy of "**value for money**"
- An unstable political situation with **minority governments** constituted of **many parties**
- Changes in the **political administrative** system
- Evaluation developed - especially during the 1990s - to a **research field.**

Research evaluation in the university sector in Denmark 1985-1996

Initiator	Parliament, Government, Science policy advisory bodies	Research councils	Universities	Total
Initiated year				
Phase I: 1985-89	3	2	1	6
Phase II: 1990-92	3	6	5	14
Phase III: 1993-96	5	8	31	44
Total	11	16	37	64

Research evaluations in public sector research institutions in Denmark 1989-1998

	Public sector research institutions
Phase I (1989-93):	12
Phase II (1994-98):	210
Total	222

Perspectives

- **State of the art**
- **Access** to accumulated know-how and data
- **Targeting** specific issues and problems
- Right **timing** in adapting processes (quick identification of trans-border emerging issues)
- **Involvement** of policymakers/stakeholders in the specification of the evaluation
- Involvement of evaluators and key stakeholders in an **open** and **interactive process**

Good practice (I)

- **International, independent** and adequately resourced evaluators
- Improvement of **communication** between different policy levels
- Evaluations should be followed up by studies to determine the **impacts**
- **Societal relevance** of activities has to be addressed
- **Multidisciplinary** approaches
- Quantitative indicators must relay upon an approach that relates **inputs and goals to effects and outputs**

Good practice (II)

- Trans-national **comparative** evaluations
- Identification and addressing of issues that are inherently **trans-border/European** and/or particularly **complex** and which need a higher level of aggregation
- Evaluations should serve as learning experiences to identify and develop adequate forms of **organisation/best practice at a European level**

Concluding remarks (I)

- Evaluation is only one of several instruments
- The value of evaluation is both in the process and in the products generated
- Evaluation results are not always implemented or have expected impacts
- Utilisation/impacts of evaluations have to be addressed
- Stakeholders/actors are not always aware of the impacts
- The more transparent the processes the better implementation chances
- Indicators are useful but need to be complemented by formative/learning approaches and case-studies

Concluding remarks (II)

- Evaluation instruments recognised and targeted on needs at European level
- Coordination and systematisation of evaluation across Europe/activities to tackle European-wide issues
- Development of a strategic and coordinated evaluation approach in Europe: The European Research Evaluation Area?

How we can overcome the gap
between policymakers and
evaluators by offering tools that:

- *enable policymakers to understand the evaluation process*

- *increase evaluators awareness of the conditions for decision- and policymaking*

Appendix C

Results from a survey:
"The use of evaluations in Europe"

Data collected by:
Per Stig Lauridsen and Karen Siune
The Danish institute for Studies in Research and Research Policy
October-December 2002.

Prepared for the European RTD Evaluation Network Meeting 27th-28th November,
Denmark.

Tables:

Table 1: Percentage breakdowns of evaluations related to public R&D

	- 25%	26 - 50%	51 - 75%	76 - 100%	Not relevant
Evaluations of R&D at universities	Au, Be, Bu, DK, Fi, Fr, Sp	Cz, Ge, Gr, Ir, No, PL, UK	Es, Hu, Lv	Lt, NL, Sk, Si	
Evaluations of R&D at other types of public research institutions	Be, Cz, Fi, Gr, Ir, Sp, Se	Au, DK, Ge, No, PL, UK	Bu, Es, Fr, Lv	Hu, Lt, NL, Sk, Si	
Evaluations of national programmes	Be, DK, Fr, No, PL, Sk, Sp	Au, Se, UK	Cz, Es, Fi, Ge, Gr, Ir	Hu, Lv, Lt, NL, Si	
Evaluations of European programmes	Au, Be, DK, Fi, Fr, Ir, Lt, No, PL, Sk, Sp, Se, UK		Es, Ge, Gr, Hu	Lv	Cz, NL
Other types of evaluations	Au, Fi, Gr, Lt, No, Se	Sk, Sp		Ge	Cz, Hu

The abbreviations used in the table - and in the tables below - are: Au: Austria, Be: Belgium, Bu: Bulgaria, Cz: Czech Republic, DK: Denmark, Es: Estonia, Fi: Finland, Fr: France, Ge: Germany, Gr: Greece, Hu: Hungary, Ir: Ireland, Lv: Latvia, Lt: Lithuania, NL: Netherlands, No: Norway, PL: Poland, Sk: Slovakia, Si: Slovenia, Sp: Spain, Se: Sweden, UK: United Kingdom.

Table 2: The initiators of R&D evaluations

	- 25%	26 - 50%	51 - 75%	76 - 100%	Not relevant
Ministry	Bu, Cz, Fi, Fr, NL, No, Sp, Se	Be, DK, Ge	Au, Hu, Ir, PL, Sk	Es, Gr, Lv, Lt, Si	
University	Au, DK, Fi, Ge, Hu, Lt	Si, Sp	Cz, Gr, NL	Sk	Se
Research councils	Au, Be, Cz, Ge, Ir, Lt	DK, NL, Se	Bu, PL, Si	Es, Fr, Lv, No, Sk, UK	Gr, Hu, Sp
EU	Au, Be, Ge, Ir, Lt	Gr	Hu		Cz, Sp, Se, UK
The unit for evaluation decides internally	Lt, PL, Si, Sp	Gr, Ir	Bu, Fi	Cz	Hu, Se
Other agents - actors	Au, Fi, Ge, Lt, Si, UK	Se	Hu		Cz, Gr, Sp

Table 3: The relative weight of different types of contexts for evaluations of R&D

	- 25%	26 - 50%	51 - 75%	76 - 100%	Not relevant
Part of a routine, regular intervals	Au, Be, Cz, Fi, Ir, Sp	Ge, Hu, UK	Bu, Gr, PL, Se	Es, Fr, Lv, Lt, NL, No, Sk, Si	
Part of a contract, known to take place from the beginning of the activity	Au, Be, Bu, Fr, Ge, Ir, Lt, NL, No, Se	DK, Gr	Hu, UK	Cz, Fi, Si	Sp
Ad hoc	Es, Fi, Fr, Hu, Lv, Lt, No, PL, Si, Se, UK	Sp	Au, Be, Bu, Ge	Gr, Ir	Cz, NL

Table 4: The reasons for ad hoc initiatives

	Rank					
	1	2	3	4	5	6
To measure institutional efficiency	Bu, Hu	Es, Gr, Ir, Lv, Lt, No, PL, Si, Sp	Au, Fi, Ge, NL, Se			
To measure quality of research	Be, Es, Fr, Lv, Lt, NL, No, PL, Sk, Si, Sp	Au, Bu, Fi, Ge, Hu, Se	Gr, Ir			
To measure societal relevance of research	Se	Be, NL	Lt, Si, Sp	Ge, PL		
To measure cost evaluations	Gr		PL	Si		
To measure socio-economic/ societal impact of R&D	Fi, Ge		Es		Lt	
Other reasons	Au, Ir		Bu	Lt		

Table 5: The evaluators of public R&D

	Rank			
	1	2	3	4
Ad hoc groups, set up for the specific task	Be, Bu, Cz, DK, Es, Ge, Gr, Hu, Lv, NL, No, PL, Sk, Si, Sp, Se, UK	Au, Fi, Lt		
Specific public institutions	Au, Fr, Ir, Lt	Ge, Gr, Hu, Lv, PL, Sp		
Private evaluation contractors	Fi	Be, Bu, Es, Ir, Si, Se, UK	Au, Ge, Gr	
Others			Bu, Sp	

Table 6: Types of ad hoc groups, set up for the specific task

	Rank		
	1	2	3
Purely international	Es, NL, Se	Lt	Lv
Purely national	Gr, Hu, Lv, Lt, Sp, UK	Cz, Se	
Mixed	Cz, DK, Ge, No	Es, Gr, Lv, NL, Sp	Lt

Information about the types of ad hoc groups used as evaluators of public R&D was only obtained from 13 of the 20 countries that responded that they do make use of ad hoc groups as evaluators.

Table 7: The methods used in the majority of evaluations of R&D

	Ranking								
	1	2	3	4	5	6	7	8	9
Survey descriptions	Au, Gr, Sk	Ge, Hu, PL, Si, Sp	NL	Fi		Bu			
Peer review	Be, Bu, Cz, DK, Es, Fr, Ge, Hu, Lv, NL, No, PL, Si, Sp, Se	Au, Gr	Fi, UK	Lt					
Bibliometric analysis in form of publication analyses	Lt, UK	Be, Cz, Es, Lv	Si	Bu, Hu, No	DK, Ge		Au		
Bibliometric analysis in form of citation analyses		NL, UK	Be, Cz, Es, Hu, Lv, Lt, Si		Bu			Au	
Users evaluations	Fi	Se	Au, No	Cz, DK, Gr, NL, UK		Ge, Si			
Historical evaluation					UK				Au
Activity evaluation		DK, Fr, Lt	Bu, Sp	Be	Au	Si			
Capacity evaluation		Bu, Fi	Ge		Hu, Si	Au, UK			
Management evaluation		No	DK, Gr, Se	Au, Ge	Be, Fi	Hu			

Ireland uses: 1. Structured interviews with programme participants, administrators and funders. 2. Analysis of already available data on programmes. 3. Postal or telephone questionnaires. 4. Comparisons with similar programmes in other countries. 5. Peer reviews. 6. Bibliometrics (with main emphasis on citation analysis).

Table 8: The availability of the evaluation reports to the public in general

All of the evaluation reports are available	DK, Es, Fi, No
Most of the evaluation reports are available	Ge, Gr, Hu, Ir, Lv, Lt, NL, Sk, Se, UK
Some of the evaluation reports are available	Au, Be, Bu, Cz, Fr
None of the evaluation reports are available	PL, Si, Sp

Table 9: The formats in which the evaluation reports are available to the public in general

Full text, in the national language	Cz, Ge, Gr, Lt, No, Se
Full text, in English	Cz, DK, Es, NL, No, Se, UK
Summary, in the national language	Es, Gr, Lv, Lt
Summary, in English	Cz, DK, Hu, No, Se

Information about the formats in which the evaluation reports are available to the public was only obtained from 12 of the 19 countries that responded that some, most or all of their evaluation reports are available to the public.

Table 10: The typical consequences for the evaluated unit

	Rank						
	1	2	3	4	5	6	7
Allocation of resources	Be, Cz, Es, Lv, Lt, PL, Sk, Si, Se, UK	Bu, Gr	DK, Fr, Hu, No, Sp	Ge, NL	Au, Ir		
Determining the future of the evaluated unit	Bu, Fr, Hu	DK, Es, Lt, PL, Si, Se, UK	Ge, Ir, Lv, NL	Cz, No	Gr	Au	
Ongoing control	DK, NL	Fr, Hu, Lv	Si	Au, Gr, Ir, Sp	Lt		
Adapting organization structure	No	Ge, Sp	Bu, Cz, Es, Gr, Lt		Si		Au
Adapting content/work programme	Fi, Ge, Gr, Sp	Ir	Au, Se, UK			Lt	
Changes in research design		Au, Cz, Fi, NL, No		Bu, Lt, Si, Se		Gr	
Other	Au, Ir						

Table 11: The uses for the majority of evaluations (macro-level)

	Rank						
	1	2	3	4	5	6	7
Changes in the administrative legislation		Lt	Cz, Si	Au, Sp	Gr		
Change in the rules concerning the structure of appointments		Si	Fi	Gr, Hu, Lt	Au		
Changes in policy	Cz, Fi, NL, Sk, Sp	Bu, DK, Hu	Au, Es, Ge, Gr, UK	Ir, Si, Se	Lt		
Changes in allocation of resources	Be, Bu, DK, Gr, Lv, Lt, Si, UK	Cz, Es, Ge, NL, No, Se	Ir, Sp		Hu	Au	
Change in planning and budgeting	Es, Ge, Hu, No	Au, Fi, Gr, Ir, Lv, PL, UK	Lt, Se	Cz	Si, Sp		
Not used	Au, Fr, PL	Sp	Hu		Ir	Gr	
Other	Ir, Se						

Table 12: The lessons learned in general of R&D evaluations

Better evaluation methodologies must be created	Be, Cz, Es, Fi, Ir, Lt, PL, Si, Se, UK
More issues must be included in evaluations	Bu, DK, Fr, Lt, NL, Se, UK
The existing evaluation methods are adequate	Es, Ir, Sk
Coordination and systematisation	Au, Bu, Cz, DK, Es, Fi, Ge, Gr, Ir, Lv, NL, PL, Sk, Sp
Others	Au, Cz, Es, Ge, Gr, Hu, Lv, Lt, NL, No, Se

The Danish Institute for Studies in Research and Research Policy is a government research institute under The Danish Ministry of Science, Technology and Innovation.

The Institute was set up for the purpose of strengthening, through its own research and analysis, the foundations of the research advisory system and the basis for research policy decisions.

The Institute is responsible for carrying out both basic research and long-term competence-building, together with more practical-oriented analyses and investigations.