

SKEP ERA-NET Work Package 3

**Approaches and practices in the evaluation
of environmental research programmes**

**Paula Kivimaa, Hanna Mela & Eeva Furman
Finnish Environment Institute (SYKE)**

ENVIRONMENTAL
PROTECTION



SKEP ERA-NET Work Package 3

**Approaches and practices in the evaluation
of environmental research programmes**

Paula Kivimaa, Hanna Mela & Eeva Furman

Helsinki 2008

FINNISH ENVIRONMENT INSTITUTE



THE FINNISH ENVIRONMENT 13 | 2008
Finnish Environment Institute

Front cover: Marja Vierimaa
Page layout: Marja Vierimaa

The publication is only available in the Internet.
<http://www.environment.fi/publications>

ISBN 978-952-11-3078-6 (PDF)
ISSN 1796-1637 (verkkokj.)

CONTENTS

Executive summary	5
Introduction	9
Scope, material and methods of the study	12
Part I – Background on evaluating research	14
1.1 Evaluation of public interventions	14
1.2 Evaluating research and research funding	15
1.3 Defining evaluation of individual research funding programmes	17
1.4 Case: PROTEE as a mid-term or continuous evaluation tool	19
1.5 Linking research and policy through evaluation	21
Part 2 – A study on the experiences and practices of mid-term and ex-post evaluation of environmental research programmes	25
2.1 Planning programme evaluation	25
2.2 Choosing the type of programme evaluation	27
2.3 Choosing the focus(es) for the programme evaluation	30
2.4 Choosing the method(s) for the evaluation	33
2.5 Evaluating the environmental dimension	37
2.6 Utilisation of evaluation results	39
2.7 Synthesis	40
Part 3 – Evaluations linked with supra-national collaboration	42
3.1 Evaluating at the EU-level	42
3.2 Examples of evaluation in other ERA-Nets	43
3.3 Outcomes from the workshop – How to evaluate a jointly funded research programme	45
3.4 Synthesis and recommendations on evaluations linked with supra-national collaboration	46
Acknowledgements	48
References	49
APPENDICES	51
DOCUMENTATION PAGES	54

Executive summary

The SKEP ERA-NET project is one of the EU-funded partnerships for strengthening the coordination and cooperation of research activities (www.skep-era.net). Work Package 3 of SKEP ERA-NET aims at collecting good practices on the management and evaluation of environment-related research funding programmes and at providing jointly developed recommendations for the SKEP member organisations. This report, which concentrates on programme evaluation, is one of the contributions of Work Package 3, led by the Finnish Environment Institute (SYKE) and the Finnish Ministry of the Environment (FiMoE). A team of SYKE researchers from the Research Programme for Environmental Policy has produced the report.

Evaluating research programmes: Background & definitions

Publicly funded research programmes can be defined as public interventions that aim to produce knowledge for political decision-making - in the case of an environment-related research programme especially for environmental policy. Evaluation research, largely concentrating on assessing the effectiveness of public interventions, has become an independent research field. The literature on environmental policy evaluation and on the evaluation of research are both relevant to evaluations of environment-related research programmes.

Evaluation of environmental policies or programmes has special characteristics compared to other types of evaluation due to the complexity of environmental problems. Complex causal chains, insufficient knowledge about many environmental issues, complex geographical patterns as well as time-scale effects create special challenges for evaluating public interventions intended to solve environmental problems.

The key concepts of evaluation research include the inputs, outputs and outcomes of public interventions. In a research funding programme, inputs are typically human resources and money invested in the research, while outputs are, for instance, calls for proposals, decisions to fund projects, and seminars organised by the programme management. These take place at different stages of the programme cycle. A research programme may have different kinds of outcomes. Immediate outcomes are often seen right after the programme has ended. They are the results of the research inputs and outputs, for example, in the form of publications or presentations. Final outcomes, in turn, may occur long after the programme has ended. They may be, for instance, innovations or new policies and practices that improve the state of the environment. Evaluating final outcomes is more challenging than evaluating immediate outcomes because they are usually caused by several other factors in addition to the research programme. The methods of the evaluation and the interpretation of evaluation

results, thus, should take into account the other contributory factors as well as the setting of the evaluation. It may take several years for all the final outcomes of a programme to occur, and some of them remain undetectable.

Linking research and policy through evaluation

Through discussions on science-policy interface, "policy-relevance" has become an increasingly important criterion in public research funding. Yet policy-relevance is a context-dependent concept that is difficult to define and measure on a general level. The links between research and political decision-making are complex, and policy-relevance needs to be defined case-specifically for each research programme or project.

To produce knowledge that is useful for political decision-making, research should have enough continuity. Research may have long-term relevance for policymaking that cannot be detected immediately after the programme has ended. Tacit knowledge created through long-term research enables researchers to respond quickly to new problem situations which may arise in society and in policymaking. Although challenging, this should be taken into account when evaluating research programmes. In addition to more traditional projects, research programmes should fund learning-based projects focused on exploration and risk taking. These may be vital to innovation and to policies created in response to unexpected environmental problems.

Planning the evaluation

Planning evaluations is an important part of programme management. The context of the programme, partly arising during the implementation of the programme, defines the focus and methods that would be beneficial in evaluation. However, certain areas, such as the general need for the evaluation, likewise the budget and the main purpose can already be decided upon at the programme initiation stage.

Several actors may take part in planning the evaluation, including the funding agency, the programme preparation committee, external experts and programme management. In some programmes, evaluators are given an active role in planning the evaluation, while in others the funding agency may impose strict requirements on an evaluation undertaken by external evaluators.

It is often difficult to estimate to what extent the programme has influenced a given outcome. This is referred to as the attribution problem. An assessment of the operational context of the programme could be included in the planning phase of the evaluation. This would help to recognise which changes can be attributed to the programme and which are due to changes in the operational context.

Type of the evaluation

Programme evaluations can be divided into four categories: programme evaluations, thematic evaluations, and evaluations of knowledge systems and of research fields. Programme evaluation is the most common type. Mid-term and continuous evaluations belong typically to this group, as they concentrate on improving a certain ongoing research programme.

Thematic evaluations and evaluations of knowledge systems can be seen as alternative or complementary types of evaluation. Thematic evaluations examine a group of programmes from the point of view of a given goal or outcome, while

evaluations of knowledge systems concentrate on the social configurations around a research programme, emphasizing the idea that scientific knowledge is produced as a result of interaction and cooperation between several actors.

Focus of the evaluation

Programme evaluations may focus on different issues. The most typical focuses are on evaluating processes within the programme, the programme's effectiveness and its outcomes. Outcomes can be divided into immediate and final outcomes. Evaluations tend to focus on the immediate, often scientific outcomes of the programme, while it is not as common to evaluate outcomes that are sometimes seen long after the programme has ended, e.g. impacts on policy, environment, society or economy.

Evaluations may also combine several focuses, as the funding agency may require different tasks of the evaluator. For instance, process evaluation is often combined with the evaluation of programme outcomes.

Evaluation methods

Methods for evaluating research programmes can be categorised on the basis of the evaluator or the techniques used for carrying out the evaluation. Evaluation may be carried out, for example, by a panel of experts, external consultants or other professional evaluators, or it can be done internally by the funding agency. Self-evaluation, in turn, means that people participating in the programme evaluate themselves. These methods are useful for different purposes. Peer panels are suitable especially for evaluating the scientific results and the quality of the programme, while professional evaluators have more competence in evaluating the management or other than scientific outcomes of the programme. Internal evaluations may be especially suitable for undertaking mid-term evaluations, as they require commitment from the programme owners.

Different techniques for carrying out evaluations include, for instance, bibliometric methods, stakeholder surveys, economic evaluation and benchmarking. One method is rarely used alone, but rather supplemented by other methods.

Evaluating the environmental dimension

Evaluating the environmental outcomes of a research programme is challenging because several societal and other factors influence the actual uptake of the results or the recommendations of a research programme. Thus, a programme may have potential impacts on the environment, but their realisation may depend on factors outside the programme. One way to address this problem is to evaluate the impact potential of the research programme separately from its actual impacts. Also an analysis of the operational context of the programme could be included in the evaluation. This would help to identify actors and activities, which are influenced by the programme and are relevant from the environmental perspective.

Utilisation of evaluation results

Utilisation of evaluation results is very important from the point of view of improving learning and research funding programmes. It concerns several groups, such as

programme and project managers, researchers, the funding agency, policymakers and the private sector, who can all learn from the evaluation and use its results for different purposes. Some factors that may hinder the use of evaluation results by the programme's funding agency are that evaluations can be experienced as criticism instead of a possibility to learn and improve future programmes. There may also be a lack of time and resources for going through and taking onboard the recommendations of the evaluation, especially after the programme is over. This is why mid-term evaluations are important.

Evaluations linked with supra-national collaboration

The report sought to discover previous experiences in evaluating research programmes at the EU level, both the European Commission's framework programmes and ERA-NET activities, to identify specific elements related to jointly funded research programmes at supra-national level. The previous experiences of evaluations of supra-national programmes and the opinions in the workshop discussions did not present greatly different elements from mid-term and ex-post evaluation more generally. Yet some important elements to consider in evaluating jointly funded research programmes can be identified:

- Evaluations need to take into account the settings in the various funding countries. At the same time, a common evaluation protocol or uniform terms are needed to some extent in order to guarantee fair and consistent evaluations.
- The composition of the evaluation panels, if used, is important. There needs to be a balance in panel members, so that each funding country or organisation is adequately represented.
- Modified peer panels including scientific and other experts are often regarded as important, because policy and business level outcomes also need to be evaluated.
- Evaluations of programmes with supra-national collaboration should focus especially on the added value of joint funding. Regional and cultural differences and conflicting national interests are also important focuses of evaluations of supra-national programmes.

As ERA-NET is a new kind of networking instrument, new evaluation approaches may also be needed. Some innovative tools for mid-term or continuous evaluation have been developed, such as PROTEE described in Chapter 1.4 but their use for evaluating jointly funded programmes has not been tested. Descriptions of evaluation processes and lessons learned from ongoing and finalised funding programmes of various ERA-Nets should be collected, analysed and synthesised. The pilot research funding programmes of ERA-Nets could also be important contexts for developing and testing evaluation processes for supra-national research programmes.

Introduction

There are two basic types of formal institutions for research and technological development: organisations and programmes (Stein, 2004). Programme management as a discipline has been defined as “the integration and management of a group of related projects with the intent of achieving benefits that would not be realised if they were managed independently” (Lycett et al., 2004). Following from this, a public research funding programme is a directed aggregate of research projects, but also a tool for the development of research, science policy, research funding and cooperation of different actors. It is usually confined in topic or by problem definition and has a fixed period (Furman et al., 2006).

Evaluation research since the 1960s has focused on assessing the effectiveness of public interventions. An environmental research funding programme, when financed by public actors, is a public intervention, aiming to generate knowledge that improves the links between research and environmental policies or other environmental protection activities. Its effectiveness, equally to other public interventions, should be evaluated to improve the management of research funding programmes by learning from previous experiences. In addition knowledge about the programme’s outcomes is important to know about the implications of research.

The SKEP ERA-NET project is one of the EU-funded partnerships for strengthening coordination and cooperation of research activities (www.skep-era.net). SKEP includes seventeen governmental ministries and agencies responsible for funding environmental research in thirteen European countries. The project aims at improving the coordination of environmental research in Europe in order to enhance the cost-efficiency of research, to encourage innovation through more efficient use of research funding, and to further the environmental protection capability by setting down foundations for coordinating research programmes. Work Package 3 “Best practice in research management” of the SKEP ERA-NET aims at collecting good existing practices on the management and evaluation of environment-related research funding programmes (www.skep-era.net/site/79.asp). Figure 1 illustrates the life cycle of a single research programme.

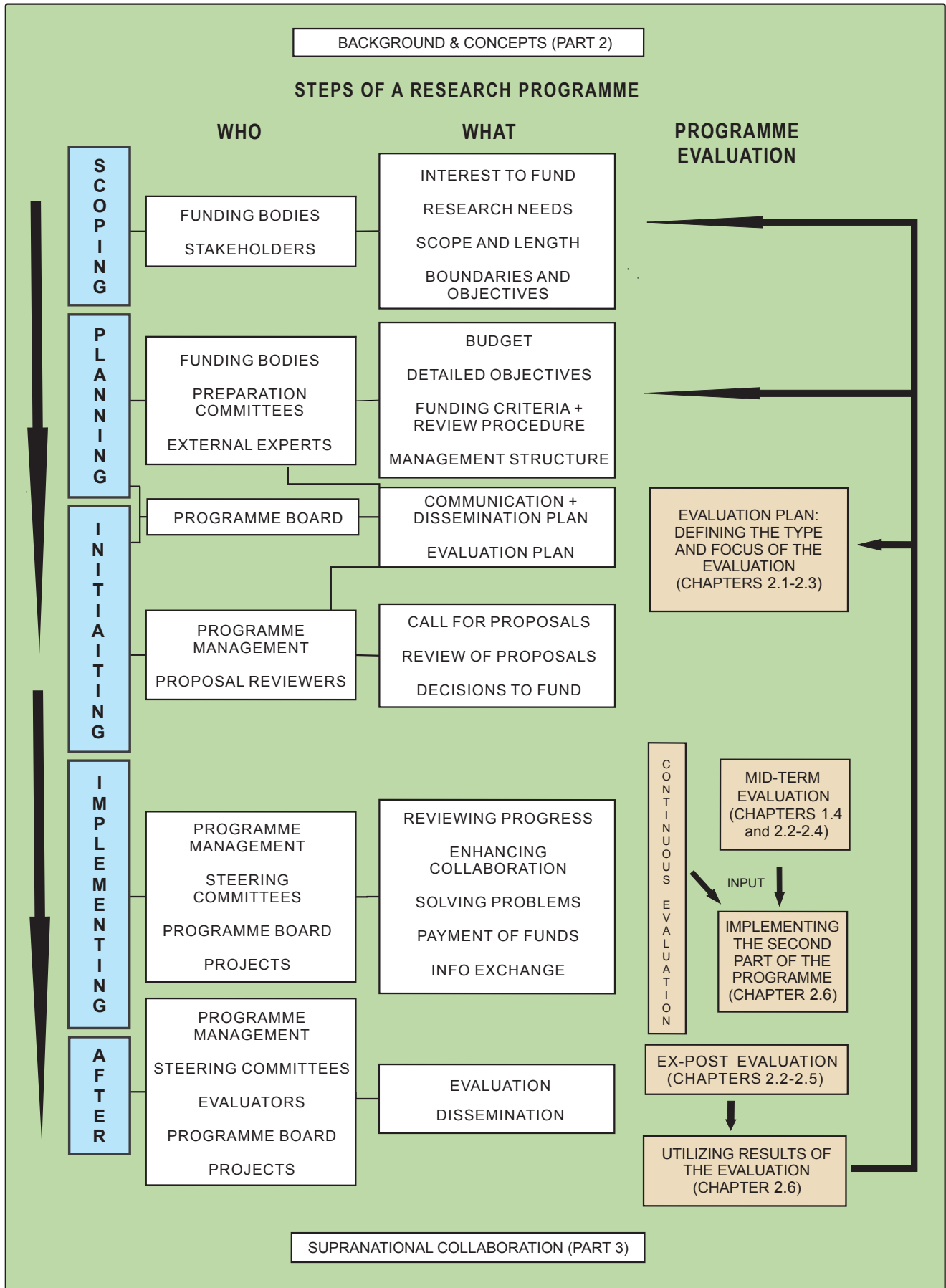


Figure I. Evaluation as an element of research programme management.

This report is one of the contributions of Work Package 3, led by the Finnish Environment Institute (SYKE) and the Finnish Ministry of the Environment (FiMoE) and it was produced by a team of researchers from SYKE's Research Programme for Environmental Policy. The report examines the approaches and practices of planning and carrying out research programme evaluations in the European Union and highlights some of the good practices available in environment-related spheres. It is intended for SKEP member organisations and for other funders of research. It also serves as a review for others who are interested in research funding programmes and their evaluation.

The report is divided into three parts:

The first part presents the background to evaluating research, summarises some evaluation literature, and defines the concepts and terminology of programme evaluation. Specific attention is paid to the issue of linking research and policy. Thus, a section in the first part concerns the interface of research and policy in relation to evaluation.

The second part of the report presents the empirical findings based on the practices of different research funding agencies financing environmental research. It goes through the programme evaluation cycle from planning and undertaking evaluation to the utilisation of its results. Feedback received from the participants of a workshop on evaluation held for SKEP members (described below) is also used in this section. As SKEP is oriented towards "scientific knowledge for environmental protection", a chapter in the second part deals with the special characteristics of evaluating environmentally related research programmes.

The third part of the report concentrates on evaluations of programmes built on international cooperation. The evaluation practices on the EU level and in two ERA-Nets are described. Further experiences of and proposals for evaluations of jointly funded programmes emerging in the workshop are also presented.

Scope, material and methods of the study

The report focuses on the mid-term and ex-post evaluation of research funding programmes, while the evaluation of project proposals has been covered in an earlier report of the SKEP ERA-NET project (Furman et al., 2006). The idea is to share experiences of evaluation between research funding agencies. The study aims to present different ways and methods to evaluate research programmes, not to give a comprehensive picture of how research programmes are evaluated in the specific countries and research funding agencies in question. Some emphasis will be placed on the evaluation of jointly-funded programmes to support the joint activities of the SKEP member organisations.

The report synthesises the experiences of evaluations of research funding programmes by exploring the terms of reference for evaluation, case studies and the outcomes of a workshop. Therefore, each chapter will include a description of current approaches and practices related to programme evaluation. The report contributes also to Work Package 5 of SKEP ERA-NET; management of joint calls under SKEP. Approaches for evaluation identified in this study can be applied to the joint and individual research funding activities of the SKEP members. The report also provides information and tools for the management and evaluation of European joint calls in the sphere of environmental research.

The study is based on literature, evaluation-related documentation from eighteen research funding organisations in eleven EU Member States, two detailed case studies of evaluation activities by two funding organisations, the outcomes of an expert workshop, and further fact finding through contacting experts by email and telephone.

Literature. As a starting point of the study, a review of the academic literature was carried out. The literature review covered scientific evaluation literature in general and that related to the evaluation of research and of environmental policies in particular.

Written documents. At the end of 2006, the SKEP member organisations were asked to provide by e-mail and by post material concerning programme evaluation in their own countries and/or institutions. No strict instructions were given. The participants were asked to provide any relevant documents concerning programme evaluation, such as mid-term or final evaluation reports, guidelines for evaluation, and evaluation forms. As a result, eighteen institutions from eleven EU Member States provided documents on evaluation of research that they had funded or coordinated (Austria, Belgium, Finland, France, Italy, Ireland, the Netherlands, Norway, Poland, Sweden and the United Kingdom). The organisations providing material for the study are listed in Appendix 1. The material received was somewhat heterogeneous and consisted of various documents related to evaluation, such as final or mid-term evaluation reports, evaluation guidelines, frameworks and forms, invitations to tender and preparatory documents for planning programme evaluation. A few of

the documents provided concerned evaluating research proposals and were thus outside the scope of this study.

An advantage of such heterogeneous material is that it gives an insight into the large scope of the research programme evaluation field and into the diversity of evaluation activities. A disadvantage is that it is difficult to compare procedures between different organisations as the descriptions are not standardized and concern different issues. Expert interviews were used to complement the written material. In addition, experts in programme management of the European Commission and ERA-NET management were contacted to obtain personal comments.

Case studies. To examine the evaluation procedures in different organisations in more detail two evaluation case studies were carried out by combining interviews and written material. The interviewees represented different actors in evaluation: funding agency staff, evaluators, and project managers in projects funded by a programme. The case studies include a thematic evaluation of the climate change impacts of a selected group of programmes funded by the Finnish Funding Agency for Technology and Innovation (Tekes) and external and internal evaluation processes of Irish ERTDI programme (Environmental Research, Technology, Development and Innovation) funded by the Department of the Environment, Heritage and Local Government (DEHLG) and managed by the Environmental Protection Agency (EPA). In addition two case studies on evaluation in other ERA-Nets are presented in the third part of the report: Bonus and Vision ERA-NET.

Workshop. A workshop gathering experts from the SKEP member organisations and some other experts in the field of research evaluation was held on 18th April 2007 in the Åland Islands. A list of the workshop participants can be found in Appendix 2, and the workshop programme in Appendix 3. The workshop included discussions in small groups on how to evaluate jointly funded research programmes. The groups sought answers to the following questions:

What kind of characteristics should the evaluation of a joint research programme have in relation to:

- *Planning an evaluation?*
- *Type of evaluation?*
- *Focus of the evaluation?*
- *Carrying out the evaluation?*

Other questions discussed related to the special requirements for evaluation arising from the environmental context of a research programme and to the organisation of evaluations of joint programmes from the point of view of effective utilisation of their results. The questions also dealt with how evaluation and its results should be communicated, what the target groups of the evaluation are and how the results are going to be used and by whom. Finally, one question concerned the linking of research and policy through the evaluation of a jointly funded EU programme.

Outcomes of the group discussions are presented in Chapter 3.3. The participants were offered an opportunity to comment the contents of the draft of this report before and after the workshop. A few participants used the opportunity, and their comments have been added in the report.

Part 1 – Background on evaluating research

1.1

Evaluation of public interventions

Since the late 1960s, programme evaluation has been a major part of tracking the effectiveness of policy programmes and interventions in the public sector (Solomon, 1998, 45). Evaluation is also strongly linked to the political process of improving regulation, as a mechanism of development, learning and accountability (Mickwitz, 2006a). The term programme evaluation does not refer to research funding programmes as such, but public interventions in general. Nevertheless, the same ideas and needs apply to improving publicly coordinated research funding activities as for policies in general.

Vedung (1997, 3) has defined evaluation as “careful retrospective assessment of the merit, worth, and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations”. Later his definition was extended to include ex ante evaluation of programmes and policies (e.g. Mickwitz, 2006b). Scriven (1991, 285) has described programme evaluation to be the “largest area of evaluation to which a self-conscious specialty has been devoted... Program evaluation has a long history of practice, but only became a recognised specialty in the 1960s. ...three most active sub-areas are education, health and ‘criminal justice’”. In this description, the word programme refers to all public interventions, i.e. policies

Evaluation has developed into a distinct research field, and several books on the theory, concepts, and practice of evaluation have been published since the 1990s (e.g. Chen, 1990; Scriven, 1991; Shadish, et al., 1991; Pawson & Tilley, 1997; Vedung, 1997). Different concepts and practices of evaluation have been adopted and developed in different policy fields, such as technology policy (e.g. Georghiou, 1999; Georghiou and Roessner, 2000) and environmental policy (e.g. Mickwitz, 2006b). In technology policy, the demand for evaluation has been fuelled by the desire to understand the effects of technology policies and programmes, to learn from the past and, more instrumentally, to justify the continuation of those policies to a sometimes sceptical audience (Georghiou and Roessner, 2000, 657). In environmental policy, evaluation has also been highlighted for reasons of learning, development and accountability (Mickwitz, 2006b). For the evaluation of environment-related research programmes, two fields of literature can be seen as relevant: literature on environmental policy evaluation and literature on the evaluation of research and research funding.

According to Knaap and Kim (1998, 5), methods in policy evaluation are numerous and can be grouped into three categories of focus: process, impacts, and efficiency. Vedung (1997) groups different models of public policy and programme evaluation as effectiveness models (e.g. goal-attainment and side-effects, goal-free, and client-oriented models), economic models (cost-efficiency and cost-effectiveness), and

professional models (peer review). Owen and Rogers (1999, 40) propose yet another way of grouping evaluations, including proactive, clarificative, interactive, monitoring and impact evaluation. The following examines in more detail the concepts of environmental policy evaluation and research evaluation that could be applied to evaluating environmental research funding programmes.

The evaluation of environmental policies has been claimed to have specific characteristics in contrast to policy evaluation in general due to the nature of environmental problems and the complex causal connections related to the causes and implications of environmental problems (Mickwitz, 2006b). For instance, Lafferty and Meadowcroft (1996) have identified specific features characterising environmental problems and the complexity surrounding them, including: the knowledge deficit; complex geographical patterns of impact and causation; redistribution of losses and gains, and time-scale effects. These features complicate the evaluations focusing on the impacts of specific policy measures targeted to reduce environmental problems.

Methods utilising especially the side-effects model have been developed and applied for evaluating environmental policies (e.g. Hildén et al., 2002; Mickwitz, 2006b). They have shown how policies may create both anticipated and unanticipated side-effects, which should be taken into account in policy planning and implementation as well as possible. The context specificity of policy interventions has also been described in evaluation studies focusing on environmental policies.

1.2

Evaluating research and research funding

The need to evaluate the effects of scientific activity and research has increased due to increases in public spending on research (Callon et al., 1997). The five purposes for evaluating research, according to Coryn (2007), include accountability and efficiency; resource allocation; improvement; synthesis; and decision-making. Thus, the improvement of research programme management is merely one of several purposes in evaluating research. The literature on evaluating research and research funding is mostly focused on system-level outlooks of science and education and on programmes funding technology and innovation (Callon et al., 1997; Capron & van Pottelsberghe, 1999; Luukkonen, 1997; Georghiou and Roessner, 2000; Perrin, 2002). The literature focusing specifically on evaluating the more natural or social science based research programmes is rather rare. Thus, the literature review mainly focused on evaluating technologically-oriented research funding, the elements of which are largely relevant to science-based research funding.

According to Kanninen and Lemola (2006) the main challenges in evaluating technologically-oriented and to some extent also other basic research relate to:

- Time lag – when should impacts of basic research be assessed?
- Attribution – what is the significance of the research efforts for the observed impacts?
- Appropriability – who benefits from the research efforts?
- Complexities – what are the mechanisms through which impacts are generated?

Interestingly these challenges are similar to those related to the environmental problems identified by Lafferty and Meadowcroft (time-scale, complex patterns of impact and causation, redistribution of losses and gains). The attribution problem, sometimes referred to as the impact problem, is typically identified in the evaluation literature (e.g. Vedung, 1997; Similä, 2007). Thus, the methods of the evaluation and

the interpretation of evaluation results need to take into account the other factors that impact on the outcomes of research.

Georghiou and Roessner (2000, 658) have argued that there is no universally applicable method for evaluation and that it is usually necessary to understand the setting of the evaluation and the discourse in which its results are located before the choice of approach can be fully appreciated. Kanninen and Lemola (2006) present a similar argument. Thus, several different models of evaluation exist solely for the purposes of evaluating research on technology and innovation. A classification between particular evaluation tools differentiates between qualitative methods (e.g. peer review), semi-quantitative methods (e.g. the historical tracing of scientific events) and quantitative methods (e.g. econometric and cost/benefit analyses) (Capron & van Pottelsberghe 1999). Kanninen and Lemola (2006) have presented an extensive list of methods. used in the evaluation of basic research that serves as a basis in this report for analysing the evaluations of environment-related research funding programmes (see Part 2 of the report).

The challenge of evaluation research is to find suitable approaches and tools for evaluating the final, more indirect outcomes of research. The links between the different components of research-funding programmes as policy interventions are illustrated in Figure 2, which is based on descriptions of the policy cycle by Hildén et al. (2002) and Kivimaa and Mickwitz (2006).

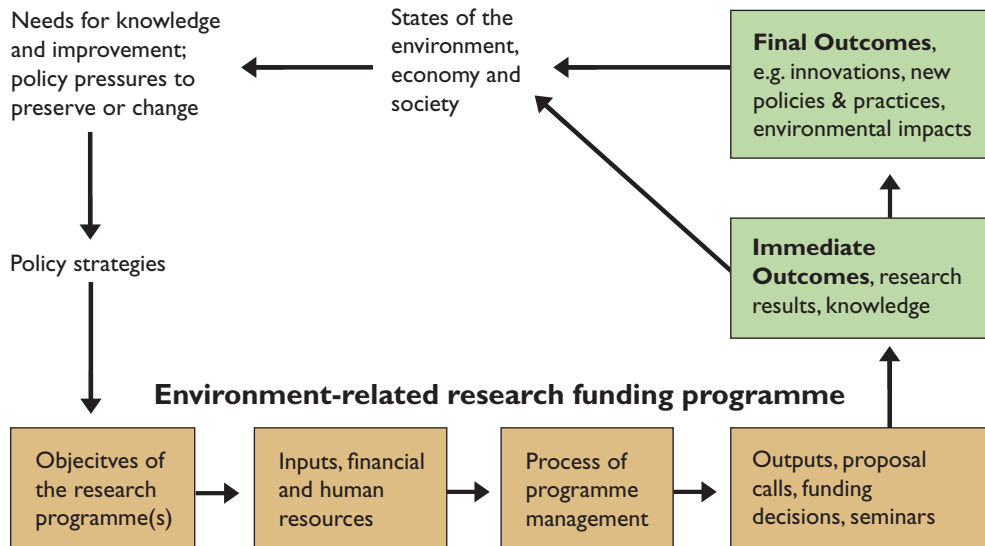


Figure 2. The policy cycle of environment-related research funding programmes from the evaluation perspective (inspired by Hildén et al., 2002 and Kivimaa and Mickwitz, 2006).

Defining evaluation of individual research funding programmes

Evaluation is seen as a process that puts the research programme management on the right track, keeps the management on course and helps the management and other stakeholders to learn from previous experiences. This kind of learning is important to improve the research management for future programmes and to know about the implications of research. Therefore, evaluation should be embedded into programme management. Evaluation is essentially perceived as decision-making support (Callon et al., 1997), while it is also crucial in demonstrating the legitimacy and effectiveness of the use of public resources to various stakeholders.

While the context of evaluating public interventions differs widely, some common terminology has been applied in evaluation research. The key concepts of evaluation include inputs, outputs and outcomes of the policy or programme being evaluated, while target groups are those whose actions the policy or programme is intended to influence and the conversion process transforms inputs into outputs. The inputs, outputs and outcomes of evaluation may be very different depending on the policy being evaluated. Table 1 provides a general definition of the key concepts and also provides a definition of the key concepts in the context of evaluating research funding programmes.

Table 1. The key concepts of evaluation and their application in the context of the evaluation of research funding programmes

Key concepts of evaluation	General description (Mickwitz and Kivimaa, 2007)	Description in the context of evaluating research funding programmes
Target groups	The actors, i.e. decision-making entities, such as companies, organisations and individuals whose actions the intervention is intended to influence.	Researchers in the organisations funded by the programme; other researchers, policymakers and stakeholders that the dissemination of programme results is intended to reach.
Inputs	What is used to produce outputs. Resources, such as personnel and finance, but also matters coming from the target groups that the agencies take into account or respond to.	Money invested in research. Human resources used for planning, managing, advising and evaluating the programme. Management guidelines and research ideas and proposals from the different target groups.
Administration or process	The conversion of inputs, by an agency or a network, into outputs.	The conversion of money, personnel effort, management guidelines and research ideas into the outputs defined below.
Outputs	What the administration produces and the target groups are provided with or expected to respond to.	Calls for research proposals, decisions on funded projects, conferences and seminars organised by the programme management and disseminating programme results, evaluation reports.
Outcomes (immediate)	The actions and the consequences of the actions taken by the target groups due to responding to the outputs.	Research results (publications, presentations, models & tools), knowledge of researchers and seminar participants, and new networks of people.
Outcomes (intermediate, final)	Hardly any outcome is the results of some policy outputs alone, but instead is affected by a variety of factors.	New policies, innovations (technologies, products, services, practices), new business models, patents, environmental improvements, public knowledge, etc.

Research programme inputs typically include human resources and the money invested in the research projects funded and in the programme management and administration. Provided the programme is well planned in advance, general management, dissemination and evaluation guidelines may also be considered as inputs into the intervention process. The outputs of research funding programmes are produced in different phases: calls for research proposals are the output of the preliminary planning and scoping exercises, decisions for funded projects follow the assessment of research proposals, and conferences and seminars organised by the programme management are the output of synthesising the research carried out in the projects funded by the programme. The outcomes of research funding programmes also occur in different phases and are varied in nature. Immediate outcomes are often detectable after the programme has ended, consisting of research results in the form of publications, presentations, models and tools. Table 2 shows an example based on the empirical material collected for this study of the variety of outcomes a single research funding programme may produce.

Table 2. Examples of outcomes from programmes funded by the Ministry of Ecology and Sustainable Development (MEDD), France (unpublished material)

Teaching and training	Public decision-making	Supporting economic activity	Public awareness
<ul style="list-style-type: none"> • Recruitment of young researchers by Water Agencies • Use of program booklets for teaching/training 	<ul style="list-style-type: none"> • Dialogue between researchers and decision makers – new networks • Decision-making tools • Use of scientific results in new legislation • GIS • New set of maps 	<ul style="list-style-type: none"> • Methods for evaluating fish populations • Managing a natural reserve, an eco-tourism facility 	<ul style="list-style-type: none"> • Exhibitions • Films

The final outcomes of a research programme are harder to evaluate, because they are often a result of several other factors in addition to the research programme and occur long after the implementation of the programme. In the context of environment-related research programmes, these include innovations, policies and practices that reduce the negative environmental impacts or improve the protection of natural resources, thus improving the state of the environment. According to a SKEP member, some outcomes of a research programme may also be seen as having a “conceptual” character as opposed to an “instrumental” one. Conceptual outcomes may include, for example, seeing things differently as a result of the research programme, understanding the problem better or realising that less is known about the problem than was previously thought. Evaluation of this kind of outcomes is more problematic than evaluating outcomes of a more concrete character.

Inputs, outputs and outcomes as terms are also related to the evaluation process. From this perspective evaluation as such can be seen as an intervention with implications for research funding and for the conditions that the research affects. Based on the model of the Center for Disease Control and Prevention, Jordan (2007) defines the inputs to evaluation, for example, as data, tools, analytic expertise, collaborative partnerships and a culture of evaluation. The inputs are transformed first into outputs and then into outcomes through the evaluation process. The outputs and outcomes of an evaluation can be useful from the point of view of accountability, improving the programme or improving practices of future evaluations (Jordan, 2007).

A conceptual distinction can be made between measurement and evaluation. Jordan (2007) characterises evaluation as being a more thorough process in comparison to measurement. Evaluation tries to find reasons for why objectives have or have not been achieved. The characteristics of measurement and evaluation are summarised in

Table 3. Evaluation is typically undertaken at a certain phase of a programme, while measurement is a continuous, more routine-based activity concentrating mainly on the outputs and immediate outcomes of the programme. Measurement is often used as a complementary element for programme evaluation. (Jordan, 2007).

Table 3. Characteristics of measurement and evaluation (Jordan, 2007).

Measurement/Metrics	Evaluation
Quick	Slow
Continuous	Periodic
Inexpensive	Costly
Primarily outputs and outcomes	Processes, outputs, outcomes, impacts
Answers "What?" (data)	Answers "Why?" (causality)

1.4

Case: PROTEE as a mid-term or continuous evaluation tool

In addition to more general evaluation concepts, the literature also provides examples of learning-based evaluation-related methods, such as PROTEE and SOCROBUST. Here, PROTEE is described as an example.

A tool for the evaluation and management of radical innovation projects, PROTEE, was developed under the European Commission's Framework Programme IV by a consortium of academic and industrial partners (PROTEE Final Report 1999). It is a management tool to evaluate the learning curves of projects and can be used as a mid-term or continuous evaluation tool in all kinds of programmes and projects.

PROTEE can be applied in the evaluation of projects and programmes which aim at change and introduction of novelty, including both technical and "social" innovation (McNally and Valve 2007). Importantly, the "methodology is apt for the evaluation of innovation projects not innovations per se" (Valve et al., 2007). Projects are perceived as "a learning process aiming at the progressive discovery...of what the states of the world are" (PROTEE Final Report 1999).

The basis of PROTEE lies in the idea that research and innovation, implicitly or explicitly, can survive only if they can mobilise or enlist human and non-human actors in specific ways.¹ PROTEE was designed to address the paradox that, even though proposals, interim reports and evaluations often look promising and favourable, many promises of innovation projects and programmes fail to be realised (Mc Nally and Valve 2007). In a sense it could be seen as responding to the critique presented in the literature that research evaluation often fails to acknowledge that innovative ideas are born out of learning from failures and imprecise research outcomes, and also "unproductive" research is needed for extraordinary accomplishments (see Chapter 1.5). However, although also applicable for looking back, the idea of PROTEE is to look forward, recognise the uncertainties in the projects and programmes, and attempt to mobilise allies supporting the innovation. The idea of PROTEE is to encourage innovators to make rich and risky, non-smooth descriptions of their projects as a requirement for continued support (McNally and Valve 2007).

This new evaluation tool is not intended to substitute traditional evaluation and management tools; rather it is seen to complement them. The design of PROTEE is pragmatic with the intention to accommodate the practical demands of project or

¹ This means that research participates in the production of social order. It creates and modifies reality through the assumptions it makes and through the knowledge it spreads (e.g. Calàs and Smircich, 1999).

programme management. Its core is a structured dialogue between the funders and the project or programme team, taking place in real time at project or programme appraisal meetings. Based on the dialogue and qualitative indicators, the project or programme is re-described in ways that enhance learning and project or programme descriptions. (McNally and Valve 2007) The participants to the PROTEE evaluation should examine what the project or programme under evaluation “expects” (Valve et al., 2007). The researchers, thus, have an active role in the evaluation through dialogue with the evaluators. Moreover, PROTEE is based on an assumption that the representatives of the research project or programme and the evaluators both agree to enter into a learning pact (PROTEE Final Report 1999). This means that changed actions and practices are more likely to follow from PROTEE than the more standard evaluations.

According to the PROTEE Final Report (1999), “research in history, management and sociology of technology has shown that innovations fail to explore their environment in a way that ensures a positive learning process for four basic reasons: lack of realism; lack of strategy; lack of falsifiability; and lack of innovativeness.” Based on these reasons, four classes of indicators have been created for PROTEE. The classes of indicators are described in the Manual for the PROTEE instrument (1999) to include the following questions:

I. Realism

- a. How rich is the innovators’ description of the project/programme?
- b. How diverse is the range of elements they describe?
- c. To what extent do the innovators distribute levels of uncertainty throughout the project/programme description?
- d. Is the innovators’ description of the project/programme contingent or inevitable?

II. Strategy

- a. How much opposition do the innovators take into account?
- b. How coherently do the innovators describe the project/programme from the point of view of the antiprogrammes?
- c. Are the project’s/programme’s central and peripheral aspects clearly distinguished?
- d. Can the project/programme sacrifice elements to adapt to criticism by opponents, as part of a negotiation process?

III. Falsifiability

- a. How specific are the trials to test various aspects of the project/programme?
- b. How diverse and independent are the experts and methods to evaluate the project/programme?
- c. How well do the innovators justify decisions about the project/programme with respect to losses and gains in information?
- d. How relevant/critical are the trials for the project/programme?

IV. Innovativeness

- a. Is the project/programme learning from its environment?
- b. Can the project/programme reconcile its contradictory supports?
- c. Is the project/programme a “hopeful monster”, a “hopeless pet” or a “mature beast”?
- d. What is the project’s/programme’s degree of innovativeness?

Five phases are included in a PROTEE approach in each meeting between the Evaluator (anyone in a position to offer advice on the continuation of the research project) and the Innovator (a representative of the research project). In the first phase, (1) the Evaluator listens to the Innovator's account of the research project or programme, its aims, and the chronology of events. After listening and recording the Innovator's story of the project or programme, (2) the Evaluator retells it to the Innovator. This constitutes the second phase of the meeting. The idea behind this approach is to build "a common basis for further work". The historical analysis of the project or programme is in effect tied to the processes ongoing in society. (PROTEE Final Report 1999; Valve et al., 2007) In a subsequent phase, (3) "the Evaluator and the Innovator enter into an analytic dialogue structured by questions formulated by the Evaluator to encourage the Innovator to make risky descriptions of the project". In the fourth phase, (4) a re-description of the project or programme is made. Finally, (5) an evaluation is produced based on differences in project descriptions made in consecutive meetings. (PROTEE Final Report 1999). This can be interpreted as a sign of the learning that has taken place.

PROTEE has been applied, for example, to the evaluation of a set of transport infrastructure projects and to that of a risk assessment framework (McNally and Valve 2007, Valve et al., 2007).

1.5

Linking research and policy through evaluation

As the conceptions about the role of science in policymaking have changed (Guimarães Pereira et al., 2006), increasingly calls have simultaneously been made for more "policy-relevant" research in public research funding programmes. This raises questions along the lines of "what does policy-relevant research mean?" and "how should policy-relevance be measured?". Despite the increasing importance of basic and applied research to the economy and increased calls for accountability and demonstration of results, particularly for publicly funded research, there is no agreement on how to describe or measure the outcomes of scientific advances in general (Jordan, 2006). Moreover, the literature does not suggest direct criteria for evaluating the policy-relevance of knowledge (Sørensen, 2002), although there is a long tradition of evaluation research. This is partly explained by the fact that both "policy" and "relevance" are ambiguous concepts and that there is no recipe allowing the translation from research into practice (Sørensen, 2002). According to Sørensen (2002), "relevance" is not a simple property of knowledge or research that can be claimed, rather it is related to several properties, including: (i) availability of knowledge/research results, (ii) applicability of knowledge/research results to a given policy problem, (iii) robustness of research, and (iv) acceptability/assumed reliability of research results.

The routes between research and policy are complex, having interaction with many other elements, while at the same time the boundary between the two is indistinct and overlap occurs. Studies on the use of social science reveal that knowledge is used for many purposes, not restricted to problem-solving and decision-making, demonstrating that the model of rational, linear and deductive knowledge transfer from science to policy does not reflect real life (Sørensen, 2002). There are, however, many studies indicating that opening the research process in an interactive way towards stakeholders leads to social robustness of the knowledge gained, which is attractive for decision-making (Sørensen, 2002). What is important from the evaluation perspective is that research may be useful even when it demonstrates no direct policy outcomes at a given point in time.

Research funding should encourage exploration and risk taking (Jordan, 2006), because the more innovative ideas are often born out of learning from failures and imprecise research outcomes at the beginning of research. According to Perrin (2002), this means that:

- Evaluation approaches largely based upon assessing the extent to which programmes have achieved pre-determined objectives ipso facto are not open to double-loop learning, and may penalise programmes that go beyond or demonstrate limitations in these objectives.
- Evaluations based upon mean scores rather than upon the recognition of the few but extraordinary accomplishments, penalise innovation and those who explore the unknown.
- Evaluators could put greater emphasis on identifying positive examples rather than 'averages', even if they are small in number; additionally, other lessons that might arise from 'failures' as much as from 'successes' could be emphasised.
- Evaluations of innovative projects and programmes should identify the extent to which there has been any attempt to learn from failures, to identify implications for the future, and the extent to which action has been taken based upon what has been learned.
- The criteria for success should not be whether the project succeeded or failed in what it was trying to do, but rather should be the extent to which it truly explored something new, identified what can be learned and acted upon these.

The continuation of research is also an important aspect in producing knowledge that is useful for policymakers. Only long-term research enables the maintenance of capabilities that are needed in the case of policy crises (Georghiou, 2006). Maintaining capabilities is crucial, because only a certain kind of knowledge is easily transferable to others. As scientific knowledge consists of skills, indirect knowledge and embodied knowledge in addition to the easily communicated written and spoken knowledge (Gallopín and Vessuri, 2006), research activities will build up the tacit knowledge of a researcher. Tacit knowledge is difficult to transfer to others, but enables fairly quick responses of that specific researcher to new, but related, problem situations. Thus, evaluations should also emphasise the long-term policy impact of research rather than the immediate impacts after the research programme has ended. The passing of time, however, complicates the evaluations of policy outcomes due to the attribution problem identified earlier.

An important thing to acknowledge is that research funding or its evaluation can never be totally impartial. This is because scientific knowledge is both socially produced (e.g. Pinch and Bijker, 1987) and participates in the creation of social order. It creates and modifies reality through the assumptions it makes and through the knowledge it spreads (e.g. Calàs and Smircich, 1999). Furthermore research is a strategic activity, which is influenced by decisions made by public authorities, researchers and others (e.g. Callon et al., 1997). This means that policy-relevance is defined subjectively by people with different interests.

The Science Meets Policy workshops in London (2005) and Helsinki (2006) tackled the question of how to link research and policy. The role of programme evaluation did not come up directly in the recorded discussions, but the general recommendations can be applied in that context. The London workshop also revealed further steps that are needed for effective science-policy interface (Table 4.).

Table 4. Summary recommendations and actions from Science Meets Policy 2005 (Scott, 2005) and their applicability for research programme evaluation developed for this report.

Recommendation	Action	Options research programme evaluation (developed for this report)
Overarching recommendations	Create incentives for researchers and policymakers to engage with each other	<ul style="list-style-type: none"> • plan the evaluation to explore the incentives and dialogue; include a mid-term evaluation, after which these can be improved; • compare engagement between policymakers and researchers in different programmes through thematic and knowledge system evaluations; • a comparison of programme goals and process regarding policy and actual results useful; • self-evaluation useful when evaluating incentives; • policy citation index useful when evaluating policy outputs; • add engagement through modified peer panel evaluations including policymakers
	Policymakers to seek dialogue with researchers throughout the policy cycle	
	Support training, education and secondments	<ul style="list-style-type: none"> • a comparison of programme goals regarding policy and actual results useful; • evaluate through peer panels with scientific experts
	Support interdisciplinary research	
	Maintain a longer-term perspective	<ul style="list-style-type: none"> • evaluations of a group of programmes and the development of research fields useful; • evaluate the projects' continuation before and after
Planning and managing research programmes	Identify clear policy needs for research	<ul style="list-style-type: none"> • evaluate the policy relevance and origin of programme goals and objectives • identify good examples of interaction between research and policy
	Project selection: scientific criteria need to be balanced by policy relevance criteria	<ul style="list-style-type: none"> • specify what is meant by policy relevance for the programme and its evaluation • take into account the long-term implications of research in policymaking
	Engagement and stakeholder dialogue	<ul style="list-style-type: none"> • compare engagement between the public, media and researchers in different programmes

The Science Meets Policy workshop in Helsinki (Häkkinen and Pyykkö, 2006) concentrated on the challenges of multi-disciplinarity and policy integration which arise in the pursuit of a stronger science-policy interface. Defining the research questions in an interactive and interdisciplinary mode was seen crucial. Also, providing opportunities to discuss theme-specific questions in continuing processes loaded with a learning environment was seen as important. From the programme evaluation perspective this requires evaluators or evaluator teams with a broad perspective. The planning of the programme focus and providing opportunities for setting research questions in project planning could serve as potential evaluation criteria.

SKEP WP4, which aims at sharing and identifying practices in dissemination, implementation and use of research, produced a report on the dissemination and implementation of research (Holmes 2007). One chapter of the report is devoted to evaluation. According to the report, few evaluations have included criteria for dissemination and implementation of research results. Criteria such as the effectiveness of the dissemination process and the impacts on stakeholders and policy development have been measured by methods such as peer scoring by experts and stakeholder surveys. In these cases the evaluations have been carried out on project level. They could also, however, be useful criteria on the programme level.

Linking research and policy through evaluation involves many unclear and yet to be resolved issues. As research, policy and the world around us are social constructions, there are no easy answers to tackle evaluation related issues, such as:

- The interface or continuum between science and policy
- Defining and assessing policy relevance
- Evaluation of long-term policy impacts
- Evaluation of results versus evaluation of learning and trial
- Guaranteeing even relative impartiality in evaluations

Part 2 – A study on the experiences and practices of mid-term and ex-post evaluation of environmental research programmes

2.1

Planning programme evaluation

Most of the research funding agencies and evaluators consulted for this study view evaluation planning as a part of the overall programme management. An interviewee took the view that evaluations should be linked to efficient feedback systems that inform the programme board and management. It is crucial that the programme board knows the implications of its decisions. The variety of different kinds of evaluations emphasise that the context of the research programme widely defines the methods and approaches that may be relevant for the evaluation of a particular research programme or a group of programmes. Yet it is possible to define and explicitly state the general need for, the budget of and the purpose of evaluation already at the research programme planning stage. The more detailed planning of the mid-term or ex-post evaluation of research funding programmes can be carried out later on in the process.

Evaluations are often planned to reflect the goals of the programmes. Nearly as often evaluations may be performed "goal-free" to evaluate in general the impacts of the programme, or sometimes to focus specifically on the side-effects. Process evaluations, examining how well programme management has supported the achievement of programme aims, are an important part of evaluations to improve research programmes in the future. Wider issues, such as policy-relevance or societal impacts of research programmes can be evaluated when this is deemed necessary, but such evaluations are not included as a standard procedure. The models and focuses of evaluations are explored in more detail in the subsequent chapters.

In addition to the funding agency, several other actors may take part in planning the evaluation, including the programme preparation committee, external experts and programme management. The role of the evaluator in planning the evaluation varies. When external evaluators are used, strict requirements may be imposed by the programme's funding or implementing agency (Box 1). On the other hand, evaluators may be invited to take an active part in planning the evaluation (see Box 2). Interviews involving the activities of one funding agency stated that evaluations and their planning activities have become more systematic in recent years, partly due to strict requirements imposed on the contents of invitations for tenders by public organisations. This has limited experimenting opportunities on the part of the evaluators. Yet tenders have typically left the methodological choices to the evaluator.

According to a SKEP member, one of the positive features of outsourcing the planning of evaluation is that it brings in people with expertise in evaluation processes. A threat of outsourcing, however, is that it may not make enough use of the first hand knowledge of those who are involved in the programme. The funding agency and the programme staff have valuable information for instance on stakeholders, and this information may be useful for the evaluation.

Box 1. External evaluation largely planned by the programme's implementing agency

An example of an external evaluation where the programme's implementing agency largely influences the content of the evaluation is provided by the Irish EPA and its ERTDI programme. The framework for the external final evaluation was to some extent defined in-house. It included the issues that the external evaluators were supposed to address in their evaluation and the tasks that they were to undertake. These requirements were presented in the invitation to tenders for conducting the evaluation, thus playing a major role in defining the way the evaluation was to be done. However, the process still left room for interaction, as recommendations from the external mid-term evaluation were utilised when developing the ex-post evaluation framework for the programme. Possible modifications to the framework will also be made based on the conclusions and recommendations from the final external benchmarking review. The evaluation framework that was developed in-house was also presented to the Irish Evaluation Network, which gave a very encouraging feedback on it. The evaluation network provides a forum for exchanging ideas and organises seminars and conferences for people involved and interested in evaluation.

Box 2. Planning thematic evaluations

The Finnish Funding Agency for Technology and Innovation (Tekes) plans and structures to a great extent the evaluations it commissions from professional evaluating agencies. The ex-post evaluation of the climate change impacts of a group of programmes funded by this agency and initiated in 2001 differed from their current practice by providing the consultancy with ample leeway to generate ideas on how to evaluate the impacts. The leeway was caused by two factors: the newness of the type of evaluation conducted and the point in time, where systematic evaluation procedures were less developed at the funding agency. Despite the flexibility in the process, the evaluation was planned in-house by discussing the evaluation with the relevant units of the organisation and the research programme managers to define the need for the evaluation and the main questions of the evaluation. The evaluations ended by having three different parts: 1) an evaluation of the climate change reducing potential of the agency's technology programmes, 2) an international comparison of the development and adoption of technologies beneficial for reducing climate change (based on case studies), and 3) a popular communication of the impacts of technology policies on climate change. After the evaluation the results were utilised in planning a programme focusing on generating business from climate change technologies. The full use and usability of the evaluation results will be found out when the subsequent programme is evaluated.

The evaluation literature and the material collected for this study show that an evaluation of research may take many different forms and it may be carried out for different purposes. The target audience or user of the evaluation and the purpose of the evaluation affect what kind of evaluation should be planned and carried out. There

are clear differences in the planning process and the structure of the evaluation if the evaluation is focused on the intermediate or final outcomes of the programme, e.g. impacts on policy, rather than on the immediate scientific results of the programme. In thematic evaluations, the choice of the evaluation theme is important, so that someone in the funding organisation has this theme as an area of responsibility to facilitate the use of results.

The attribution problem was identified as one of the most significant challenges of evaluation – how to define the significance of the research programme on the outcome studied, because outcomes are generally due to a variety of factors. According to an evaluator, the planning of the evaluation could include an assessment of the operational context, where the programme intervention is intended to take place – this kind of assessment helps the ex-post evaluation by identifying what changes are due to the programme and what changes result from a change in the operational context during the programme period. An assessment of the operational context could be made in synthesis with the general programme planning because the programme planning would also benefit from a stage-setting review.

Acknowledging the attribution problem can help to plan the evaluation in a way that gives better answers on the actual contribution of the programme to a given problem. Asking stakeholders about the impact of the programme and the extent to which it contributed to a certain problem may provide valuable information. It is, however, important that the questions are formulated in a way that enables the respondents to tell more about their understanding of the reasons underlying the successes or failures of the programme.

2.2

Choosing the type of programme evaluation

The evaluations addressing research funding programmes can be divided into three main types: 1) programme evaluations, concentrating on evaluating a single funding programme either mid-term or ex post; 2) thematic evaluations, focusing on the impacts or effectiveness of a group of programmes in relation to a specific goal or outcomes; and 3) evaluations of knowledge systems. This chapter briefly describes these three evaluation types. In addition, the report acknowledges evaluation of research fields.

Programme evaluation – mid term and ex post

Programme evaluation is the most typical and traditional type of evaluation of the process, the effectiveness or the impacts of a research funding programme. It consists of a variety of evaluation approaches (e.g. peer review, internal evaluation or evaluation by external experts). Common to all programme evaluations, as defined for the purposes of this report, is that the evaluation is focused on a single research programme. Mid-term evaluations are typically single programme evaluations, because their intention is to redirect or improve a particular on-going research programme during its course. Many ex-post evaluations are also focused on a single research programme.

Mid-term evaluations have been perceived as crucial to developing research programme design and management. According to one interviewee, mid-term evaluations are not focused on the realised impacts of the programme, but should rather answer to the question: What should be done to achieve the best possible results from the programme? The mid-term evaluations focus on whether the programme is researching the right things; on whether the world has changed in such a way

that there are new things that the research should focus on, and; on the ways in which to make sure that optimal use is made of the research results. Thus, mid-term evaluations are often designed to redirect the goals of the programme and improve the knowledge use processes.

Most of the evaluation material received for this study was based on the single research programme evaluation type. The subsequent sections, therefore, largely present alternatives for the focus and methods of evaluation from this perspective. Many of the methods, however, may also be applied when carrying out thematic or knowledge system evaluations, although the starting point of the evaluation is different in these cases.

Thematic evaluation

Thematic evaluation is a complementary approach to evaluating research funding programmes individually. It means that the impacts of a group of programmes on some specific outcome are evaluated simultaneously. For example, the Finnish Funding Agency for Technology and Innovation (Tekes) has commissioned ex post evaluations looking at the impacts of selected technology programmes for achieving climate change targets (Hjelt et al., 2003), innovation process changes (Valtakari et al., 2004) and internationalisation (Halme et al., 2004). The UK's Natural Environment Research Council (NERC) carried out an evaluation of the economic impacts of research it has funded beginning from 1985 until today (NERC, 2006). Examples of the results of this evaluation are provided in Table 5. One of the two deeper case studies presented in this report is focused on experiences from thematic evaluation.

Table 5. Economic benefits of NERC-funded research (Source: NERC, 2006)

Strategic benefits	Direct benefits
<ul style="list-style-type: none"> • Increased national reputation in some environmental science fields • Sustaining and improving the national science base • Improved policy development and implementation • New networks for collaboration • Increasing public engagement and awareness • Improved decision making and empowerment • Addressing market failure and encouraging innovation • Promoting competitiveness and knowledge transfer 	<ul style="list-style-type: none"> • Saved costs from avoiding environmental damage • More cost-efficient techniques for environmental management • Avoided casualties and deaths • Increased market shares for environmental technology

Thematic evaluation is deemed beneficial because some outcomes result from several subsequent (or simultaneous) research programmes, and continuity of research is the key to achieving wider changes towards sustainability. Moreover, thematic evaluations serve the general development of programme design and management. According to the interviewees, thematic evaluations may be challenging, because the target groups of evaluations are even larger due to several programmes being included and it is hard to find common measures for the impacts and conclusions of different programmes.

Evaluations of knowledge systems

Some evaluations focus on the social configurations and the knowledge system around the research programme rather than merely scientific outputs (e.g. Sandström, 2006). The evaluation is based on the view that scientific knowledge is socially embedded forming a collective of producers and users. Thus, the way in which the researchers interact with others affects the outcomes of the programme. This type of evaluation, if the resources are available for it, is a good complement for more traditional programme evaluation focused on direct programme outputs and outcomes.

Research evaluation methods using the historical tracing of events or case studies may be used to evaluate the knowledge systems around research programmes. For example, research adopting cases studies of technological innovations has (i) illustrated the significance of public research programmes for the emergence of innovations through providing funding and motivating the formation of new cooperative networks, and (ii) identified specific research programmes that have played a role in the innovation processes (e.g. Kivimaa and Mickwitz, 2004; Kivimaa et al., 2007).

While evaluating knowledge systems and social configurations sounds new compared to other forms of evaluation, the material used resemble those previously used in evaluating research programmes. However, slightly different questions are posed in examining the material (see Table 6).

Table 6. The proposal for evaluating the Swedish COPE Programme (Sandström, 2006)

Publication analyses	<p>a. Do programme members participate in international discussions? Is the group a leader or among the leaders in their research field? Are the channels used effective?</p> <p>b. Does the programme group appear to participate in national discussions? Are their channels effective?</p> <p>c. Are the popular publications – the grey publications – adequate for the purpose? Do the users use these grey publications?</p> <p>d. Popularisation – do the programme members participate in that? How significance is that in research groups cooperation?</p>
Human capital	<p>e. Is there personnel exchange between the research group and the related knowledge network?</p> <p>f. Are there PhD students who already have a career in the network?</p> <p>g. What happens to newly qualified doctors? Where in the network will they be placed?</p> <p>h. Is there post doc cooperation in the programme?</p>
Network cooperation	<p>i. Which actors in the network related to the programme topic are invited to conferences organised by the programme?</p> <p>j. How are programme's research activities made visible in national and international conferences</p> <p>k. To which conferences have programme participants been invited as speakers or commentators?</p> <p>l. With which international research networks does the programme have an active exchange? Are the exchange activities sufficient</p>
Expert competence	<p>m. In what way do the participants of the programme take part in practice-lead cooperation with users? How is this cooperation financed?</p>

Evaluations of research fields

Another evaluation type is evaluating the development and knowledge of a particular research field. Research funding programmes are included in these evaluations as a part of the whole development in the research field or area. Examples of research field evaluations often include national level or funding agency level assessments describing the current status and quality of a research area or field and the amount of public funding attributed to it, for example, in

- Energy research
- Biodiversity research
- Water research

Choosing the focus(es) for the programme evaluation

Process (management and research of the programme)

Process evaluation has been described as the assessment of everything occurring prior to the emergence of true outcomes - in most cases including the evaluation of the merit, worth, and/or significance of outputs, vision, design, planning, operation, justification, fidelity, management, activities, procedures, and so forth (Coryn, 2007). In the case of research funding programmes, such evaluations focus rather on how the planning, management and research have been organised in carrying out the programme than the results and outcomes of the programme.

Explicit evidence of process evaluation was found at least in relation to the evaluation activities of nine research funding/coordination agencies providing material for the study. The process evaluation could, for instance, on a general level, ask how well the programme board, advisory groups, programme coordinator and funding organisations have cooperated and managed the programme. According to an interviewee, process evaluations require a reactive feedback system so that programme management can learn from the evaluations. Self-evaluation (see the Chapter 2.4 on methods) is a useful method to combine with other methods in evaluating processes.

Effectiveness (achievement of programme objectives)

Effectiveness evaluation typically focuses on how the programme has been able to achieve its goals. According to Vedung (1997), the two central questions for goal-attainment evaluation are: Are the results in accordance with the programme goals? Are the results produced by the programme? The effectiveness models of evaluation, however, are not always related to mere goals, but can also evaluate the results of the programme in general or focus on stakeholder concerns (e.g., Vedung, 1997). This is important, because the evaluation approaches largely based upon assessing the extent to which programmes have achieved pre-determined objectives are not open to double-loop learning and may penalise programmes that go beyond or demonstrate limitations in these objectives (Perrin, 2002).

Effectiveness was a standard focus for programme evaluations in the evaluation material received from research funding organisations. The questions asked related to the achievement of programme goals and to some specific criteria, such as scientific quality, innovativeness, industrial relevance, usefulness for policymaking, mobility and environmental utility. The specific criteria are dependent on programme goals and are also related to the evaluation of programme outcomes (below).

Cost-effectiveness

The purpose of cost-benefit analyses is to indicate the value of research as a proportion of the benefits to the costs of the research effort (Kanninen and Lemola, 2006). The costs of research programmes are easy to calculate, but putting a monetary value on research results and outcomes of research is difficult. Table 5 shows the kind of economic benefits that can result from publicly funded research, but it does not present figures for the benefits nor does it focus on the monetary value of other than economic benefits.

Immediate outcomes (scientific and other)

Most research programme evaluations focus, often exclusively, on the immediate scientific outcomes. The immediate, often scientific outcomes of research are commonly evaluated by examining publication quantities, the accumulation of references to publications and publication patterns (e.g. Oksanen, 2003; Kanninen and Lemola, 2006). Scientific outcomes may also extend to include presentations, services offered by researchers and the capabilities and skills of researchers (Kanninen and Lemola, 2006). Non-scientific immediate outcomes may, for example, include assistance to policymaking processes, appearances in the media and workshops disseminating practical knowledge to stakeholders.

An evaluation of research quality may be a part of evaluating scientific outcomes or the effectiveness of the programme. The Irish EPA has used several different indicators in evaluating the research quality criterion of the ERDTI Programme, including the quality of research inputs and outputs and reputational quality (Table 7). In addition to publication and other research output quantifications, scientific peer review panels (see Chapter 2.4 on evaluation methods) are used in evaluating scientific outcomes, focusing on the merit, worth and value of these outcomes.

Table 7. Indicators of research quality in the ERDTI programme (unpublished source)

<p>a. Quality of research inputs</p> <ul style="list-style-type: none"> • Number of Masters/Doctors/Post Docs funded (per thematic area) • Completion time for Masters/Doctoral students • Number of "newcomers" funded • Expected duration of projects vs. actual duration • Number of researchers funded & their gender ratio
<p>b. Quality of research outputs</p> <ul style="list-style-type: none"> • Number of publications (peer-reviewed & other) • Citations • Presentations of research findings (national and international) • Number of end-of-project reports • Number of reports to policymakers • Number of reports on integrated assessments of sectoral relevance • Number of firms adopting new practices reducing emissions/waste generation
<p>c. Reputational quality</p> <ul style="list-style-type: none"> • Contribution to national and international academic committees and working groups • Membership of national and international research committees • Leading role in scientific societies • Awards, prizes, honours

Intermediate and final outcomes

The intermediate and final outcomes of research programmes are fairly seldom evaluated. While the evaluation of scientific outcomes is fairly straightforward, the societal outcomes, such as impacts on policy or environmental or health impacts, are indirect, they occur after long periods of time and can be difficult to evaluate (e.g. Oksanen et al., 2003, Mickwitz, 2006b).

According to a questionnaire made in the first phase of this study in 2006, a majority of programmes have not evaluated impacts on policy, and socio-economic impacts have been evaluated even more seldom. In comparison, evaluating scientific results and immediate outcomes of research programmes has been much more common. This may indicate the challenges related to evaluating societal impacts.

Due to the long timescales of effects, the evaluation of societal impacts may focus on looking at the outcomes of a group of programmes thematically, such as outcomes related to climate change (e.g. Hjelt et al., 2003) after some time has elapsed.

Alternatively the intermediate outcomes could be approached bottom-up, where the significance of research programmes on selected cases of outcomes is evaluated.

In the evaluation framework developed for the Irish research programme ERTDI, societal outcomes have been divided into three categories:

- Policy impact
- Commercial impact
- Collaborative impact

These three categories are given a number of indicators each. When evaluating the policy impact, attention is paid to how the programme contributes to national and international policy development, whether the government has applied programme findings to policy development and whether the programme has contributed to work on advisory panels to industry or government agencies. Commercial impact evaluation looks at the generation of commercial potential, for instance, in the form of patent applications and a number of new operating practices that use less energy and materials or produce less emissions or waste. Collaborative impact evaluation examines the collaborative agreements established between different institutions as a result of the programme. Attention is also paid to linkages developed between research and industry.

As it often takes time for the policy impacts to take place, the ERTDI research programme will consider the re-evaluation of the policy impacts after three years from the first ex-post evaluation to see whether the policy impact of the evaluated projects is greater. This would give a more reliable picture of the effectiveness of the research.

The importance of evaluating societal outcomes depends on the type of the research funded. According to the evaluation report of an Austrian research programme KLF, evaluating societal benefits in addition to scientific quality is especially important for research aiming at sustainable handling of public goods and at supporting public interests. The programme KLF was identified to belong to this category and was thus evaluated from the point of view of societal impacts and scientific quality. Other types of research, such as free basic research or technically and economically oriented research, may have different needs for evaluation.

In the case of another funding agency, the societal outcomes have not been part of evaluation questions. However, the evaluation approaches utilising interviews of key stakeholders have nevertheless brought up societal effects, such as implications on regional development, to the awareness of the evaluators.

Regarding the evaluation of the outcomes of research programmes, a further aspect was raised by a workshop participant: the sustainability or longevity of the achieved outcomes. If one of the objectives of the programme was to create new interdisciplinary scientific communities, it might be worthwhile to evaluate whether these communities have the potential to continue to exist when the programme is over.

Side-effects

Side-effects evaluation may be used as a complement to other evaluation focuses. An evaluation focusing on side-effects examines the unintended consequences of research funding programmes in addition to the achievement of programme goals. The side-effects may be impacts that were anticipated while the programme was being planned or they may be unexpected occurrences. Public policy seldom turns out exactly as intended (Mickwitz, 2006b), thus in the evaluation of side-effects it is important to actually know what consequences the use of public resources have.

According to the empirical material, evaluating side-effects is rarely included in research programme evaluation as such. Addressing side-effects, however, can

sometimes be essential. As defining the scope of a research programme always entails selecting and excluding, the resources directed to a specific programme are not allocated to some other, possibly important and related areas. This can be considered a side-effect of a research programme. One way to address this issue is to undertake a gap analysis, as was done in the Irish ERTDI programme. A gap analysis tries to identify possible research areas that have not been addressed within the programme but which would be relevant for it.

A combination of different evaluation focuses

An evaluation may typically combine different focuses to examine both the process of the programme and its results. According to one evaluator, a funding agency often requires several different tasks of the evaluator, including:

- Evaluation of the programme strategy (i.e. Are the aims of the programme relevant with respect to the operating environment?)
- Evaluation of the management process (i.e. How well have the programme management and board supported the achievement of programme goals?)
- Evaluation of programme outcomes (i.e. an estimate of the expected and/or identification of realised outcomes)
- Evaluation of the added value from the programme
- An assessment of what the funding agency can learn

Combinations of different focuses may be useful to avoid some of the unintended side-effects of evaluation. When evaluations focus only on some aspects of programmes, the ignored aspects of evaluations may hide problems that would need attending to in the latter part of the evaluated programme and in future programmes. For example, if the evaluation concentrates on the immediate scientific outcomes of a programme and puts great emphasis on the number of publications, this may encourage the production of scientific publications at the expense of other activities that have not been the focus of evaluation.

2.4

Choosing the method(s) for the evaluation

Evaluation methods consist of two different but overlapping elements: the method from the perspective of selecting the evaluator (peer review, expert evaluation, self-evaluation and internal evaluation) and the method from the perspective of the techniques used to measure the process, effectiveness, side-effects or outcomes of the programme (e.g., case studies, bibliometrics, S&T indicators, cost-benefit analyses, stakeholder surveys). This chapter first addresses the various approaches from the evaluator perspective and then moves on to more detailed methods of evaluation.

Approaches based on the type of evaluators

Expert evaluations

Expert evaluation of research funding usually means that an external expert is commissioned to do the evaluation. This, however, refers to two different types of evaluations: those carried out by consultants or other evaluation agencies using various methodologies and those carried out by expert panels consisting of scientific and/or other experts. Thus, peer panel reviews can be described as a subset of expert evaluations.

Evaluations using peer review and modified peer review panels

The term "expert panel evaluation" is used for evaluations of research conducted beyond the individual instance or piece of research or individual researcher, where the evaluator consists of a panel of scientific or other experts (peer panel evaluation and mixed panel evaluation) (Langfeldt, 2002). Expert panel evaluations are commissioned, often ad hoc, for evaluations at the programme, institutional, or disciplinary level, where the peer panel consists of researchers qualified in the area under review and the expert panel consists of both peers and other experts (e.g. experts on policy or commercialization of research) (Coryn, 2007). The expert panel approach can also be called a modified peer review. The academic peer review is often focused on evaluating the scientific outputs and outcomes of the programme, whereas a modified peer review may be better equipped to also evaluate the side-effects and other outcomes.

Evaluations using professional evaluators

The expert evaluations carried out by consultants or other evaluation agencies are actually fairly heterogeneous, having employed a variety of methods, including among others bibliometrics, surveys, case studies and network analyses. This means that an agency external to the funding organisation is carrying out an evaluation of the research programme and may consult various experts during the evaluation process. In this report, this type of evaluation is referred to as "professional evaluation". The professionalisation of evaluation implies that professional evaluators, such as consultants or evaluating institutions carry out the evaluations in addition to, or in place of, peer review/expert panels (Luukkonen, 1997). Professional evaluations may cover any of the evaluation focus areas, but may not be sufficient to evaluate the scientific quality. Professional evaluators are typically used in thematic evaluations of the outcomes of a group of programmes.

In evaluations commissioned of professional evaluators, a process of inviting tenders is usually used, because public organisations in some countries are required by law to do this. The calls for tenders for research programme evaluation usually include a description of the evaluation task based on the initial evaluation plan drafted by the funding or coordinating agency. The tenders are usually open to all organisations that have not taken part in the board, coordination, projects or steering groups of the programme to be evaluated. Table 8 shows an example of the kinds of questions that have been used to frame the tender for evaluation on the basis of which the professional, external evaluation would start the evaluation work.

Self-evaluation

Self-evaluation means that the researchers of the programme and the people participating in programme management evaluate themselves and the programme as a whole from the internal perspective. This can be an efficient method to assess the process and to some extent the effectiveness of the programme in relation to the immediate outcomes, but self-evaluation is insufficient for addressing for example the side-effects, final outcomes or cost-effectiveness of the programme, because an individual programme participant has a limited knowledge on the programme as a whole.

An example of self-evaluation is provided by the Flemish Impulse Programme for Nature Development, which was evaluated with the help of a questionnaire completed by project members and steering group members of the programme. The questions concentrated mainly on issues related to the management, resources and timescale of the projects, collaboration and the dissemination of results. This kind of evaluation may generate valuable information on the management procedures and other practical issues of the programme. Self-evaluation is also important from the participatory point of view.

Table 8. An example of evaluation questions in a call for tenders for evaluation

<p>Effectiveness of the programme</p> <ul style="list-style-type: none"> • What have been the main results of the programme? • How well has the programme succeeded in the achievement of the goals? • How well has the programme achieved the criteria of innovativeness, industrial relevance and good science? • Was the chosen portfolio appropriate to achieve the goals of the programme? • What kind of future impact the programme is to achieve?
<p>Operational performance</p> <ul style="list-style-type: none"> • How well have the programme board, advisory groups, programme coordinators and funding organisations co-operated and managed the programme? • How did the project partners in different countries cooperate compared to partners in the same countries?
<p>What is the value-added, possibilities and obstacles of the international cooperation in strengthening the network of expertise needed for implementation of projects?</p> <ul style="list-style-type: none"> • In creating a knowledge base enabling the development of innovative and eco-efficient products and processes? • In disseminating the results and using them for promoting the competitiveness and sustainability of the industrial sector?
<p>Recommendations for the future</p> <ul style="list-style-type: none"> • How to organise cooperation between countries? • How to organise industry cooperation with research and industry in other countries (already in the project preparation phase)? • How to organise international cooperation between financing organisations of different funding profiles (basic research, applied research, industrial research and development)?

In another research programme, the Environmental Cluster Programme of Finnish Ministry of Environment, a combined self-evaluation and internal evaluation was carried out by sending an e-mail questionnaire to project managers and supervisors. Six projects that received especially good remarks in the questionnaire were selected to provide more detailed information. Managers and supervisors of these projects were interviewed. The same was done for six projects that received the poorest remarks in the questionnaire. The interviews provided more specific information on the things that had worked well and that had caused problems and on the reasons behind successes and failures. An identical questionnaire and self-evaluation had been used in the previous years, thus enabling the long-term assessment of the programme.

The self evaluation method was fairly commonly used in the material received for this study. It was a part of programme evaluations and was also present in two of the evaluation guidelines out of five examined.

Internal evaluation

Internal evaluation is carried out by the funding agency staff or the organisation coordinating the research funding, for example, on behalf of the ministries. The internal staff may have more knowledge of the programme process and the results but may lack the outsider perspective. The question whether an evaluation should be internal or external depends largely on the main purpose of the evaluation and who will be the main user of the information. It has been argued in favour of internal evaluation that every organisation should be self-evaluating (Wildavsky, 1985) and that effective internal evaluation is able to provide indispensable support for managers of the programme (Love, 1991).

An interviewee stated that, according to a study of the important factors for the success of programmes, internally carried out mid-term evaluations were identified as a success factor, because internal evaluations better involve the programme management than other types of evaluations. Cooperation and interaction are generally important in internal evaluations.

Methods for carrying out evaluations

Bibliometric methods

In the evaluation of impacts or outcomes of basic research, bibliometric methods are prevalent in evaluating the scientific outcomes. These include publication counts, citation analyses, co-word analyses and patent analyses. (Kanninen and Lemola, 2006).

Bibliometric methods were used in at least three programmes examined for this study. For instance, citation analysis was used in one programme to evaluate research outputs. It was seen as an important method of quantifying whether the research is being used within the research community.

Science & technology indicators

Science and technology indicators emphasise inputs into the research process rather than outputs, and fail to capture the cultural and environmental impacts of science (Kanninen and Lemola, 2006). Thus, science and technology indicators are good for comparing the inputs into different research programmes, but do not tell us much if only a single programme is examined. Such indicators were used in at least two programmes examined for this study.

Economic evaluation

Economic evaluations are usually focused on the monetary costs and benefits of the programme. Cost-benefit analyses are most commonly used. The purpose of these methods, including both large-scale econometric studies and survey-based studies, is to arrive at a figure that indicates the value of research as a proportion of the benefits to the costs of the research effort (Kanninen and Lemola, 2006). The challenge of this type of evaluation is to find a monetary estimate for the benefits, i.e. how to measure the outcomes of the programme and especially the proportion of the outcome that is due specifically to the programme.

Economic evaluations were used on three programmes examined for this study. According to one evaluator, econometric evaluations are used to measure outcomes if economic indicator data is available. In the absence of this kind of data, outcomes are evaluated by interviews and surveys. Ex-post information gathering regularly after a programme has ended may provide important information for a statistical analysis of outcomes.

Stakeholder surveys

Surveys are useful for gathering programme-related information from a wide range of people, including researchers, policymakers, representatives of industry and the funding agencies to build up a comprehensive picture of the effectiveness and outcomes of a programme. With surveys, evaluators are able to collect both quantitative and qualitative information from the participants of the programme and the potential end-users of the research results (Kanninen and Lemola, 2006). Surveys are commonly used in external evaluations but are also a part of self-evaluations, where the group participating in the survey is smaller.

Stakeholder surveys were carried out on at least four programmes examined for this study. According to a funding agency representative, the surveys suffer from low response rates, and often interviews are more effective in eliciting the opinions of key stakeholders. In general the inclusion of different stakeholder groups in evaluations depends on the focus and purpose of the evaluation, dictated by the funding agency.

Intra-national & international comparison, benchmarking

To be able to ascertain the productivity or efficiency of a programme, some points of reference are needed. The programme or intervention that is being evaluated has to be compared against something. This can be done, for example, by comparing it to similar programmes within the same country or internationally. A benchmarking exercise can also be done, comparing the programme with the best programme in the field. (Dror, 1968; supplemented by Vedung, 1997)

Benchmarking will be part of the external evaluation exercise of the Irish ERTDI research programme currently being undertaken by a consultancy. In its invitation to tender for conducting the evaluation, the programme's implementing institution, EPA, defined the research programmes and funding agencies that need to be compared with the ERTDI programme. These included both national and international programmes and agencies. The benchmarking exercise will also focus on programme management and publicity/ dissemination.

Case studies and historical tracing of events

Case studies or historical tracing of events have been described as good examples of evaluating the ex-post impacts of policies in general or research in specific (e.g., Piric and Reeve, 1997; Georghiou and Roessner, 2000). A case study evaluation should typically be carried out by a professional evaluator with expertise in methods of social science research. Evaluators could put greater emphasis on identifying positive examples rather than 'averages', even if they are small in number; additionally, other lessons that might be learnt from 'failures' as much as from 'successes' could be emphasised (Perrin, 2002). It has been argued that in order to emerge, innovations need failures; we need to learn from past mistakes to achieve future successes. Thus, case studies can provide important examples of the significant individual impacts that a research funding programme has generated. Historical tracing studies may fail to account for the indirect effects of research, including dead ends (from which substantial learning takes place), spillovers, and synergistic effects (Georghiou and Roessner, 2000). This is not the case with the case study method.

Case studies in the form of a so-called bottom-up evaluation are sometimes, but rarely, used in programme evaluations. In one case where the climate change impacts were thematically evaluated in a group of programmes, case studies of technologies beneficial to climate change mitigation were used to complement the programme evaluation by carrying out an international comparison of how these technologies have been developed and commercialised (Hiltunen et al., 2003; interview). The bottom-up approach in general was identified by one evaluator as having the risk that this kind of evaluation, if it focuses on individual technologies or other specific outcomes, cannot capture the additional value of programme activities compared to independent project funding, namely the networking of actors.

2.5

Evaluating the environmental dimension

The societal outcomes and results of research cannot be outlined in a uniform way to different research fields because they vary (Oksanen et al., 2003, 111). Moreover, the societal outcomes or impacts of research must be distinguished in evaluation from the results of research, such as new information for citizens or bases of innovations, new research methods and tools, or solutions to complex technological problems (Oksanen et al., 2003, 110). For instance, a contribution of research programmes to environmentally friendlier technological innovations may indicate a potential societal outcome in terms of reduced environmental impacts, but the actual outcome

does not occur until the innovation is adopted and widely diffused (Kivimaa and Mickwitz, 2006; interview). Thus, an evaluation could focus on the impact potential of research funding programmes and the realisation of the impacts separately, where the programme has more influence over the impact potential than on the realisation of impacts (Hjelt et al., 2003; Box 3). In the workshop discussions it was also recognised that the complexity of environmental issues and the random use of the research results by policymakers may be reasons for evaluating the potential impacts of the programme, not merely the actual impacts.

Box 3. Evaluating the achievement of climate policy targets

*The evaluation of the climate change impacts of a group of research funding programmes of the Finnish Funding Agency for Technology and Innovation was carried out by professional evaluators. The framework for evaluating the climate change impacts focused on the **impact potential** of the programmes and the **realisation of the impacts**. It was assumed that the programme as such could only directly influence the impact potential, while the realisation was dependent on many different actors and factors. Since the programmes evaluated were funding technologically oriented research and development, the impact potential was built from two parts: **technological potential** and **actor potential**. The technological potential consisted of the CO₂ reduction capability of the technologies financed and the existing or anticipated market potential of these technologies. The actor potential included the role of the actors funded (technology developers or users) and the scope (national, international) of their actions. (Hjelt et al., 2003). This kind of method is best suited to evaluating technology funding programmes, because the intervention logic tracing the impacts of social science or natural science research is more complicated. Even in evaluating the climate change impacts of technology-funding programmes, the evaluation may only focus on the reduction potential, because the actual CO₂ reductions are dependent on many other factors and are thus immeasurable. (Interviews).*

According to one evaluator, there is no standard way suitable for evaluating the environmental impacts of research programmes; rather the approaches have to be case-specific. What could, however, be included in all evaluations of environmental impacts is an analysis of the operational context of the programme to identify the actors whom the programme influences and those actions of these actors that are relevant from the environmental perspective. This kind of analysis enables illustrating the intervention logic of the programme in a given operational context and helps to plan the evaluation.

The following questions are relevant when planning an evaluation, also considering the environmental impacts or outcomes of the research funding programme:

- Which environmental objectives and issues should be included in programme scope and/or evaluation?
- How will the achievement of environmental objectives be evaluated?
- How is uncertainty addressed in the evaluation?

Several challenges related to evaluating environmental research programmes were identified by the workshop participants. The environmental context is characterised by a multitude of different stakeholders. This sets requirements on the evaluation: stakeholders' views should be taken into account in the evaluation and their awareness of the programme issues should be measured. In environmental research there is typically a need for inter- and transdisciplinarity. However, there are not enough incentives for researchers to do this kind of research, as it is often more difficult to

publish the results of interdisciplinary research in recognised publications. Thus, evaluation should put emphasis on other issues than the number of publications in order to encourage inter- and transdisciplinary research in the environmental field.

In the workshop, it was mentioned that inter- and transdisciplinary research creates a need for integrating different types of knowledge: for example, scientific knowledge based on natural or social science and knowledge of various stakeholders. Traditional scientific evaluation criteria may not be enough when evaluating multidisciplinary environment-related research. Thus, an interesting aspect that evaluation could cover is the environmental outcomes of interdisciplinary research.

2.6

Utilisation of evaluation results

The use of evaluation results is crucial to facilitate learning and change. The interviewees perceived that the use of evaluation results has improved in recent times, while the evaluations themselves have also improved. According to one evaluator, the efficient use of results requires openness to critique on the part of the funding agency and programme management (Box 4, see also Box 5). According to this study, the utilisation of evaluation results concerns many different groups of people, including:

- Project managers – improving the research carried out and its dissemination
- Programme managers – improving the design of new programmes and, in the case of mid-term reviews, improving the management and focus of the on-going programme
- Researchers and others in the topic field of the evaluated programme – the development of the field nationally and internationally
- The funding agency – improve the coordination of research funding programmes and the focus of research
- Policymakers – learn about the latest priorities in research as well as the utilisation of research results
- Businesses – learn about the latest priorities in research as well as the utilisation of research results
- Media – hard to reach but could disseminate evaluation results to a wider audience

Box 4. Use of evaluation results in the development of the funding agency's activities

In the case of one funding agency studied, both the external evaluators and the funding agency representative saw that the programme design and management activities had considerably developed during the last decade. Many changes had been made partly due to results from previous programme evaluations:

- *Tightening the requirements for programme results*
- *Changes in the criteria for project funding decisions*
- *Changes in the programme board activities*
- *Increased focus on the commercialisation of innovations*

One of the most important results from evaluations has been that the programme board's role is crucial for the success of programmes. The programme board needs to consist of members who are committed and influential in their own field. They need to be able to get feedback to their decisions and develop their skills during the programme.

An evaluator takes the view that there is room to improve the use of evaluations by tackling some of the barriers to using the results. The interviewees identified the following barriers for the utilisation of evaluation results:

- Evaluations may be regarded by the programme owners (representatives of the funding agency) or programme management as audits done to criticise them rather than as opportunities to develop future activities.
- Often planning and running research funding programmes consumes ample resources, and people lack the time and energy to read evaluation reports, especially if they appear after the programme has ended (importance of mid-term reviews).
- Some research funding programmes emerge from the demand-side, i.e. they are influenced by companies and researchers. This limits the influence of programme funding agencies on programme design and thus to the utilisation of evaluation results.
- A special challenge related to thematic evaluations is that the managers and coordinators of programmes are mainly interested in their own programmes and often do not see the sense of thematic evaluations.
- A general challenge that was identified by one interviewee was creating and establishing a "culture of value for money" in the programme, so that self-evaluation would be naturally linked to all activities during the whole course of the programme.

Box 5. The utilisation of the results from a thematic evaluation

In the case of the thematic evaluation looking at the climate change impacts of a group of programmes the results were mainly used in the preparation of new climate change related research programmes. The success of utilising the evaluation results will be seen when the subsequent programmes are evaluated. The evaluation also generated knowledge about the status and development of the research field nationally.

Holmes (2007) and Furman et al. (2006) highlight in previous SKEP reports the various methods for communication and implementation of results. The list includes the use of seminars and workshops. Extending the programme evaluation to include stakeholder and user seminars and workshops could bring further knowledge into the evaluation and simultaneously increase the utilisation of the evaluation by including a wider audience in the evaluation and by raising awareness of it and its potential use. This is linked to a broader issue of collaboration: could the evaluators be included in the planning of programmes and to what extent should stakeholders be brought into programme management?

2.7

Synthesis

The planning of evaluations is an important part of programme management. The context of the programme, partly arising during the implementation of the programme, defines the focus and methods that would be beneficial in evaluation. However, certain areas, such as the general need for the evaluation, the budget and the main purpose of the evaluation can already be decided on at the programme initiation stage. More detailed planning can take place closer to the time of evaluation.

Peer panel evaluations, evaluations carried out by external professional evaluators, internal evaluations and self-evaluations are all useful methods for evaluating some elements of research funding programmes. Peer panels are excellent for evaluating the scientific quality and outcomes of the programmes, while professional evaluators are useful for evaluating the management process or other outcomes of the programmes. According to Vedung (1997), external evaluations are generally more credible as objective enterprises than internal evaluations. External evaluators are often more skilled in evaluation methods and theories, have experience from evaluating other programmes and can thus compare them (Vedung, 1997). However, if there is a need for rapid use of the results and immediate learning from the evaluation, internal evaluations have some advantages (Vedung 1997). According to one interviewee, internal evaluations were found especially suited for mid-term evaluations to secure the commitment of the funding agency staff. Self-evaluations are a useful complement to the other evaluations.

Different evaluation methods are in practice rarely used as "pure" methods. One method is often combined and supplemented with other methods. Mickwitz (2006b) discusses exhaustively and demonstrates the usefulness of triangulation in evaluations of policies which have an impact on the environment. According to interviewees, there is no standard procedure for methods to be used and a wide variety of methods were used in programme evaluations, including surveys, interviews and workshops. Peer reviews are used in basic research focused programmes, but for more applied research expert evaluations are used.

Evaluating the environmental dimension is challenging, but should be an important part of environmental research. The example provided analysed the impact potential of research separately from the realisation of the impacts of research, because realisation is largely dependent on other actors than researchers and programme coordinators. The diffusion and utilisation of research and evaluation results may influence among other things the realisation of environmental benefits. The potential users of evaluation include programme and project managers, researchers, the funding agency, policymakers, businesses and media.

Part 3 – Evaluations linked with supra-national collaboration

3.1

Evaluating at the EU-level

Since the end of the 1970s, the European Commission has developed and implemented the evaluation of research. The Evaluation Unit of the Commission is the body responsible for conducting evaluations. For mid-term and ex post evaluations, the Commission uses external evaluation, where a panel of independent experts acts as the evaluator. (Bobe and Viala, 1997) At the EU level it has often been felt that the use of panels better guarantees the independence of the evaluation and has a higher political impact than other types of evaluation (Massimo, 1997).

The choice of the panel members, especially its chairman, is one of the most important responsibilities of the Evaluation Unit. The panel members must be independent of the programme directors and the Commission and well known experts in the field. (Bobe and Viala, 1997; Massimo, 1997). The Commission maintains a database of potential evaluators and uses it when selecting evaluation panels. It is open for anyone to register. The database is used mainly for evaluating project proposals, but according to an interviewee, potential evaluators have also the option to register as ex-post reviewers.

The parameters for designating the panel members include (i) a good coverage of scientific knowledge relevant to the programme, (ii) an interest in evaluation methodology, and (iii) familiarity with socio-economic impact. This means that the panel, usually 5-8 members, should include scientists specialising in the field of the programme to be evaluated, experts in social sciences and people representing the users of results. In the panel there should also be a balance among the EU Member States, in that two experts from the same country are not chosen and large, medium-sized and small countries are equally represented. The evaluation panels may also include experts from outside the European Community. (Bobe and Viala, 1997).

The work of evaluation panels has fixed terms of reference set by the EU. Some general guidelines were drafted as early as in 1986. The scientific and technological achievements of the programme (taking into account the original objectives), the effectiveness of programme management and the use of resources, and the programme's contribution to the development of EU policies and to the social and economic development of the EU are the three areas that the evaluation should cover. The panel evaluation is sometimes complemented with evaluation data obtained through questionnaires, interviews and bibliometric studies. External consultants can be employed for this purpose due to the limited time available for the evaluation panel. (Bobe and Viala, 1997; Massimo, 1997).

While the evaluation of project proposals that seek funding from the framework programmes is fairly strictly defined, some of the interviewees representing the

Finnish secretariat for EU R&D did not recognise systematic requirements set for ex-post evaluation of research projects and programmes funded under the framework programmes. According to one interviewee, the evaluation activities of the framework programmes are strongly focused on the evaluation of project proposals, while ex-post evaluations are not as common among the framework programmes. Mid-term evaluations focusing on the available results and the management structure of programmes are more common, because evaluation-based decisions regarding further research funding need to be made before sufficient time has elapsed to carry out ex post evaluation (Massimo, 1997). A typical evaluation budget has been between 0.5% and 1% of the total cost of the research programme (Massimo, 1997).

3.2

Examples of evaluation in other ERA-Nets

ERA-Nets are collaboration instruments for European research funding organisations and SKEP is one of the ERA-Nets funded by the Commission. The ERA-NET scheme aims at increasing the cooperation and improving the coordination between national and regional research activities in the member and associated states.

Many ERA-Nets have established or plan to establish a joint research programme on a particular theme that in most cases reflects the topic of the ERA-NET in question. There are many ERA-Nets that are already at the stage in which the management and evaluation of the programme have been agreed on.

Currently, the member countries of ERA-Nets may have different rules and practices for funding, managing and evaluating research programmes. Knowledge of these practices is important when planning and setting up a joint research programme. Furthermore, the ERA-NET scheme encompasses research funding organisations of many kinds, ranging from scientific institutions to policy departments. This poses a challenge to supra-national research programmes and their evaluation, as different partners may have different research interests and objectives.

Following from this, many ERA-Nets have work packages that aim to identify best practices in programme management and evaluation among their partner organisations, and as a result, several ERA-Nets have produced studies concerning good practices in programme management and evaluation. An underlying aim of the studies has been to identify barriers to jointly funded and managed research programmes. Studying the current evaluation practices of the ERA-NET partners also helps to plan the evaluations of joint programmes effectively and to find ways to overcome challenges encountered. Some ERA-Nets have also taken steps towards developing common guidelines to be applied in the management or evaluation of the joint research programmes.

Studies undertaken in the ERA-Nets indicate that there is a rather wide variation in evaluation procedures within different countries and research funding organisations. At the same time there is considerable interest in finding common approaches in programme management and evaluation and to overcome barriers to jointly funded and managed research programmes. The following sections present approaches to joint programme evaluation in two ERA-Nets in more detail (BONUS and VISION-ERA-NET).

Case: Bonus

BONUS is an example of an ERA-NET that has resulted in a jointly funded environmental research programme. BONUS brought together key funders of research around the Baltic Sea. One of the tasks of the partnership was to develop a common

evaluation scheme for the joint research programme. This was done with the help of a questionnaire and by organising a workshop for the partners. (BONUS, 2006).

The evaluation guidelines of BONUS pay special attention to common procedures for evaluating project proposals within the joint programme. In addition, recommendations are made concerning the mid-term evaluation and final evaluation of the programme as a whole. The programme evaluation will be carried out by an international panel appointed by the programme steering committee. It is recognised that carrying out the evaluation requires different kinds of expertise in addition to scientific expertise. Thus, researchers, policymakers, people with previous experience in programme management and other stakeholders are needed. It is stated in the guidelines that clear and measurable goals should be set in the planning phase of the programme, and that the goals should be measurable during the programme, at the end of it and some time thereafter. According to the guidelines, it is important that the various partners are unanimous about the goals and how to measure them. (BONUS, 2006).

Both a mid-term evaluation and final evaluation are planned to be carried out. The final evaluation is planned to be carried out in two phases. The first phase will concentrate on the scientific quality and management processes, while the focus of the second phase will be on the impacts of the programme. It is recommended in the guidelines that the mid-term evaluation and the first part of the final evaluation should be undertaken by an evaluation panel appointed by the programme steering committee, while the second part of the evaluation could be done by the Baltic Marine Environment Protection Commission (HELCOM) or by the representatives of the European Commission. These bodies are seen to have the expertise needed for evaluating the impacts of the programme. The importance of the timing of the evaluation is also stressed in the guidelines. The social impacts of a research programme may not be evaluated too soon after the programme has ended. However, an evaluation tends to lose some of its value if it is done too long a time after the programme has ended. (BONUS, 2006).

Case: Vision ERA-NET

VISION ERA-NET is a collaborative network of nationally and regionally leading innovation policy agencies. The network includes twelve partners from ten countries, and develops shared knowledge bases for innovation policy. The project is part of the 6th Framework Programme of EU and runs between 2005 and 2008. (www.visioneranet.org) Vision ERA-NET seeks to

- Coordinate European research on innovation and technology
- Improve the utilization of research and evaluation intelligence in policy making
- Expand and advance European knowledge base on innovation environment
- Identify common knowledge and development needs (www.visioneranet.org)

The partners are committed to explore and prepare jointly cross national research mechanisms that address issues beyond national scope, including

- Joint research programmes
- Joint research management structures
- Jointly executed evaluation and monitoring exercises of innovation policies and instrument
- Collaboration and benchmarking for policy research managers and officers. (www.visioneranet.org)

Vision ERA-NET launched a pilot call for research proposals “Collaborative and open innovation: Future challenges for national innovation policies in the emerging European Research Area” in March 2007. According to a representative of Vision ERA-NET, the budget of this research programme is so small that no plans to evaluate the programme mid-term or ex-post have been made.

In general, Vision ERA-NET has capabilities to develop evaluation with many top organisations in the member countries. The challenges relate to creating cooperation among European research funders when the funding mechanisms in use and research funding structures differ greatly. The idea of Vision is to promote policy learning and exchange of information between the member organisations. Thus, learning based evaluation exercises in the science-policy interface might be useful.

The work plan of Vision ERA-NET includes a specific component of evaluation. The operation of the network is evaluated through self-evaluation by the members as well as an expert evaluation conducted by a consultancy. The expert evaluation will focus on identifying areas which the member organisations are willing to fund, on determining a feasible funding model for jointly funded programmes, and on assessing the operation of the ERA-NET itself. According to a representative of Vision ERA-NET, traditional evaluation tools may not be feasible in the context of ERA-Nets. While evaluation often looks back, a new instrument like ERA-NET may need other analytical approaches.

3.3

Outcomes from the workshop – How to evaluate a jointly funded research programme

SKEP partners have expertise in nationally and jointly funded project and programme management and evaluation. Synthesising this expertise and the knowledge held by researchers is of value for developing programme evaluation. Bringing the expertise together in the form of a workshop, where group discussions were held (see the Chapter on methods), led to further insights on evaluation of jointly funded research programmes.

Several recommendations concerning the evaluation of jointly funded research programmes were made in the workshop discussions. Concerning the planning of evaluation, it was deemed important that there should be continuity between ex-ante and ex-post evaluation, so that the objectives and the evaluation criteria would be coherent during the whole programme cycle. Criteria should be made public at an early stage of the programme, which would add to the transparency of the evaluation process.

To carry out international comparisons, it is necessary to have a common evaluation protocol for all funding partners. It is also important to pay additional attention to creating good and trustful relationships between the partners at an early stage and to ensure that the evaluation is also a joint activity with shared responsibility.

It is important that the scientific quality does not dominate as the focus of evaluation and that room is made for other focuses as well. Policy relevance especially was recognised as an important focus of evaluation. In the case of a joint research programme it would also be essential to evaluate the added value of a jointly funded international programme compared to a national research programme. This could include evaluating different funding systems, e.g. comparing a common pot model with a model in which each country has an individual pot.

When communicating the results of programme evaluation, emphasis should be put on the positive experiences, highlighting success stories rather than failures. Bad experiences should be communicated on a more general level, not mentioning

individual projects by name. This would support the idea of evaluation being an activity that promotes learning from both positive and negative experiences and is not aimed at criticising individual projects or people.

An ideal evaluation of a jointly funded research programme would include both self-evaluation and external evaluation: self-evaluation would provide programme managers with information about the strengths and weaknesses of the programme, while the external evaluation would give an objective view of the impacts of the programme. A suitable external evaluator could be a panel consisting of representatives from different funding countries. The evaluations should be planned realistically in accordance with the resources available.

The importance of mid-term evaluations for a jointly funded programme was recognised in the group discussions, as its results can be used for improving the programme during its course. Thus, sufficient capacity in the programme should be reserved for undertaking a mid-term evaluation. Some challenges were identified in the evaluation of international jointly funded research programmes. Regional and cultural differences between partners and possibly conflicting national interests set extra requirements for evaluation. Different national needs, interests and requirements should already be given enough space in the planning phase of the evaluation.

3.4

Synthesis and recommendations on evaluations linked with supra-national collaboration

Evaluations are important from the point of view of justifying the use of public money for research. This can be regarded as crucial in jointly funded supranational research programmes, where money flows beyond national borders. The report sought to discover previous experiences in evaluating research programmes at the EU-level, both the Commission's Framework programmes and ERA-NET activities.

While the European Commission has carried out mid-term and ex-post evaluations of its programmes, it is still somewhat a neglected area. EU publications on such evaluations are hard to find and ex-post evaluations are rare. ERA-Nets are too a recent instrument to have generated many experiences in evaluation.

The previous experiences of the evaluations of supra-national programmes and the opinions in the workshop discussions do not present very different elements from mid-term and ex-post evaluation more generally. Yet some important elements to consider in evaluating jointly funded research programmes can be identified:

- Evaluations need to take into account the settings in the various funding countries. At the same time, a common evaluation protocol or uniform terms are needed to ensure fair and consistent evaluations.
- The composition of the evaluation panels, if they are used, is important. There needs to be a balance in panel members, so that each funding country or organisation is sufficiently represented.
- Modified peer review panels, which combine scientific and other experts, are often regarded as important, because policy and business level outcomes often highlighted in EU-level research also need to be evaluated. They should be used if possible.
- The evaluations should focus especially on the added value of joint funding. Regional and cultural differences and conflicting national interests are also important evaluation focuses.
- The size and budget of the programme affects the methods available for the evaluation. An evaluation component that corresponds to the size of the programme should therefore be included in the programme management.

As ERA-Nets are a new kind of networking instrument, new evaluation approaches may also be needed. Some innovative tools for mid-term or continuous evaluation have been developed, such as PROTEE described in Chapter 1.4, but their use for evaluating jointly funded programmes has not been tested. The pilot research funding programmes of ERA-Nets could be important players in developing and testing evaluation processes for supra-national research programmes.

Acknowledgements

The authors would like to thank all the interviewees in Finland and Ireland, the workshop participants and the experts who kindly provided information and material for the study. In addition, the authors acknowledge Mr. Pekka Harju-Autti and Mr. Sauli Rouhinen from the Ministry of the Environment of Finland for their collaboration. Professor Mikael Hildén and Dr. Helena Valve from SYKE and Ms. Vera Rabelt from UBA provided the authors with valuable insights on the report. Finally, we thank the SKEP ERA-NET coordination team for all their support and the EU for financial support via the SKEP project.

References

- Bobé B., Viala H. 1997. Panel of experts of the European Commission. In: Callon, M., Laréro, P., Mustar, P. (eds). *The strategic management of research and technology: Evaluation of programmes*. Paris: Economica International.
- BONUS 2006. Guidelines for a Common Evaluation Scheme for a Joint Baltic Sea Research Programme. BONUS Publication Nr. 4. <http://www.bonusportal.org/uploads/4gybgd.pdf> (Last visited 25 May 2007).
- Calás, M.B., Smircich, L. 1999. Past postmodernism? Reflections and tentative directions. *Academy of Management Review* 24(4): 649–671.
- Callon, M., Laréro, P., Mustar, P. 1997 (eds). *The strategic management of research and technology: Evaluation of programmes*. Paris: Economica International.
- Capron, H., van Pottelsberghe, B., 1999. Public support to R&D programmes: An integrated assessment scheme, in: *Policy Evaluation in Innovation and Technology: Towards Best Practices*. Paris: OECD, 35–47.
- Chen, H.-T., 1990. *Theory-driven Evaluations*. Newbury Park: Sage.
- Coryn, C. L. S. 2007. *Evaluation of researchers and their research: Toward making the implicit explicit*. Unpublished doctoral dissertation. Kalamazoo: Western Michigan University.
- Dror, Y. 1968. *Public Policymaking Re-examined*. Scranton: Chandler.
- Furman, E., Kivimaa, P., Kuuppo, P., Nykänen, M., Väänänen, P., Mela, H., Korpinen, P. 2006. *Experiences in the management of research funding programmes for environmental protection*. Finnish Environment 43. Helsinki: Finnish Environment Institute.
- Gallopin, G., Vessuri, H. 2006. Science for sustainable development: articulating knowledges. In: Guimarães Pereira, Â., Guedez Vaz, S., Tognetti, S. (eds). *Interfaces Between Science and Society*. Sheffield: Greenfield Publishing.
- Georghiou, L., 1999. Issues in the evaluation of innovation and technology policy. In: *Policy Evaluation in Innovation and Technology: Towards Best Practices*. Paris: OECD, 19–33.
- Georghiou, L., 2006. Innovation, Learning, and Macro-Institutional Change: The Limits of the Market Model as an Organizing Principle for Research Systems. In: J. Hage, M. Meeus (eds). *Innovation, Science and Institutional Change: A Research Handbook*. New York: Oxford University Press.
- Georghiou, L., Roessner, D. 2000. Evaluating technology programmes: tools and methods. *Research Policy* 29: 657–678.
- Guimarães Pereira, Â., Guedez Vaz, S., Tognetti, S. 2006 (eds). *Interfaces Between Science and Society*. Sheffield: Greenfield Publishing.
- Halme, K., Kanninen, S., Lemola, T., Autio, E., Arnold, E., Deuten, J. 2004. *Competitiveness through internationalisation: Evaluation of the means and mechanisms for promoting internationalisation in technology programmes*. Technology programme report 10/2004. Helsinki: National Technology Agency of Finland Tekes.
- Häkkinen, L., Pyykkö, P. 2006. SMP2006 Science meets Policy: 8th workshop, Helsinki, Finland, 19–20 October 2006 During Finland's EU presidency. Report of the workshop. November 2006. <http://www.sciencemeetspolicy.eu>
- Hildén M., Lepola J., Mickwitz P., Mulders A., Palosaari M., Similä J., Sjöblom S., Vedung E. 2002. Evaluation of environmental policy instruments – a case study of the Finnish pulp & paper and chemical industries. *Monographs of the Boreal Environment Research* 21. Helsinki: Finnish Environment Institute.
- Hiltunen J., Ahvenharju S., Eskelinen S., Vanhanen J. 2003. *Ilmastomyönteisten teknologioiden kehityksen ja käyttöönoton edistäminen: Kansainvälinen vertailu case-esimerkkien valossa*. [Promoting the development and implementation of climate-beneficial technologies: An international comparison in the light of case examples]. Helsinki: The Finnish Funding Agency for Technology and Innovation (Tekes).
- Hjelt, M., Luoma, P., Hiltunen, J., Vanhanen, J. 2003. *Teknologiiahjelmat ja ilmastonmuutos*. [Technology programmes and climate change]. Helsinki: National Technology Agency of Finland Tekes.
- Holmes, J. 2007. Study of dissemination and implementation of research in SKEP Member organisations. Vol 1. Main report. http://www.skep-era.net/site/files/WP4_reort_main.doc
- Jordan, G. B. 2006. Factors Influencing Advances in Basic and Applied Research: Variation due to Diversity in Research Profiles. In: J. Hage, M. Meeus (eds). *Innovation, Science and Institutional Change: A Research Handbook*. New York: Oxford University Press.
- Jordan, G. B. 2007. Evaluation and Performance Measurement for Research and Technology Development Programs. Presentation at the SKEP-ERA-NET workshop 18th April 2007. http://www.skep-era.net/site/files/WP4_Aland_Jordan.ppt
- Kanninen, S., Lemola, T. 2006. *Methods for Evaluating the Impact of Basic Research Funding: An Analysis of Recent International Evaluation Activity*. Helsinki: Publications of the Academy of Finland 9/06.
- Kivimaa P, Kautto P, Hildén M, Oksa J. 2008 (in press). *Green Markets and Cleaner Technologies (GMCT) What drives environmental innovations in the Nordic pulp and paper industry?* Copenhagen: Nordic Council of Ministers.
- Kivimaa P., Mickwitz P. 2004. Driving Forces for Environmentally Sounder Innovations: The Case of Finnish Pulp and Paper Industry. In: K. Jacob, M. Binder, A. Wiczorek (eds). *Governance for Industrial Transformation Proceedings of the 2003 Berlin Conference on the Human Dimensions of Global Environmental Change*. Berlin: Environmental Policy Research Centre, 356–372.
- Kivimaa P., Mickwitz P. 2006. The Challenge of Greening Technologies – Environmental Policy Integration in Finnish Technology Policies. *Research Policy* 35(5): 729–744.

- Knaap, G.J., Kim, T.J. 1998. Environmental Program Evaluation: Framing the Subject, Identifying Issues. In: G.J. Knaap, T.J. John (eds). *Environmental Program Evaluation: A Primer*. Urbana and Chicago: University of Illinois Press, 1–20.
- Lafferty, W., Meadowcroft, J. 1996. Democracy and the environment: congruence and conflict – preliminary reflections. In: W. Lafferty, J. Meadowcroft (eds.) *Democracy and the Environment: Problems and Prospects*. Cheltenham: Edward Elgar, 1–17.
- Langfeldt, L. 2002. *Decision-making in expert panels evaluating research: Constraints, processes and bias*. Norsk institutt for studier av forskning og utdanning, Oslo. <http://www.statsvitenskap.uio.no/fag/polit/disputas/fulltxt/langfeldt.pdf> (Last visited 20 March 2006).
- Love, A. 1991. Internal Evaluation: Building Organizations from Within. Newbury Park: Sage.
- Luukkonen, T. 1997. The increasing professionalisation of the evaluation of mission-oriented research in Finland: implications for the evaluation process. In: *Policy Evaluation in Innovation and Technology: Towards Best Practices*. OECD Proceedings. Paris: OECD, 347–356.
- Lycett, M., Rassau, A., Danson, J. 2004. Programme management: a critical review. *International Journal of Project Management* 22: 289–299.
- Massimo, L. 1997. Lessons from the experience of European Commission. In: Callon M., Laréro P., Mustar P. (eds). *The strategic management of research and technology: Evaluation of programmes*. Paris: Economica International.
- McNally, R., Valve, H. 2007. *Evaluation of innovation projects: why "PROTEE"?* Abstract 10 April 2007.
- Mickwitz, P. 2006a. *Evaluations for the EU strategy for better regulation*. Presentation at the conference: "Evaluation in Society: Critical Connections", 4–6 October 2006, London.
- Mickwitz, P. 2006b. *Environmental policy evaluation: Concepts and practice*. Commentationes Scientiarum Socialium 66, The Finnish Society for Sciences and Letters.
- Mickwitz, P., Kivimaa, P. 2007. Evaluating Policy Integration: The Case of Policies for Environmentally Friendlier Technological Innovations. *Evaluation* 13(1): 68–86.
- NERC 2006. *Economic benefits of environmental science: A study of the economics impacts of research funded by the Natural Environment Research Council*. November 2006. http://www.nerc.ac.uk/publications/corporate/documents/economic_benefits_report.pdf (last visited 9 March 2007)
- Oksanen, T., Lehvo, A., Nuutinen, A. (eds.) 2003. *Suomen tieteen tila ja taso: Katsaus tutkimus-toimintaan 2000-luvun alussa* [Scientific research in Finland: a review of its quality and impact in the early 2000]. Helsinki: Publications of the Academy of Finland 9/03.
- Owen, J.M., Rogers, P.J. 1999. *Program Evaluation: Forms and Approaches*. London: Sage.
- Pawson, R., Tilley, N. 1997. *Realistic Evaluation*. London: Sage.
- Perrin, B. 2002. How to – and How Not to – Evaluate Innovation. *Evaluation* 8(1): 13–28.
- Pinch, T.J., Bijker W.E. 1987. The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In: W.E. Bijker, T.P. Hughes, T.J. Pinch (eds). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, Massachusetts: MIT Press.
- Piric, A., Reeve, N. 1997. Evaluation of Public Investment in R&D – Towards a Contingency Analysis. In: *Policy Evaluation in Innovation and Technology: Towards Best Practices*. OECD Proceedings. Paris: OECD, 49–64.
- PROTEE Final Report 1999. Project funded by the European Commission under the Transport RTD Programme of the 4th Framework Programme.
- Sandström, U. 2006. *Förslag: explorative evaluering av COPE-programmet*. Utcast 2, april 2006.
- Scriven, M. 1991. *Evaluation Thesaurus*. 4th edition. Newbury Park: Sage.
- Scott, A. 2005. *Science meets Policy 2005. Next steps for an effective science-policy interface*. Report of London conference held as part of the UK's presidency of the European Union, 23–25 November 2005. NERC, EA, DEFRA.
- Shadish, W.R., Cook, T.D., Leviton, L.C. 1991. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park: Sage.
- Similä, J. 2007. *Regulating Industrial Pollution. The Case of Finland*. Publications of the Faculty of Law at the University of Helsinki: Forum Iuris.
- Solomon, L.S. 1998. Evaluation of Environmental Programs: Limitations and Innovative Applications. In: G.J. Knaap, T.J. John (eds). *Environmental Program Evaluation: A Primer*. Urbana and Chicago: University of Illinois Press, 45–60.
- Sørensen, K.H. 2002. Social Shaping on the Move? On the Policy Relevance of the Social Shaping of Technology Perspective. In K.H. Sørensen & R. Williams (eds). *Shaping Technology, Guiding Policy*. Cheltenham: Edward Elgar, 19–36.
- Stein, J. 2004. Is there a European knowledge system? *Science and Public Policy* 31 (6): 435–447.
- Valtakari, M., Pulkkinen, M., Leminen, S., Pelkonen, T., Rajahonka, M., Kiuru, P., Laurila, H. 2004. *Kaupallistaminen ja innovaatiotavoitteet teknologiaohjelmassa* [Commercialisation and innovation targets in technology programmes], Technology programme report 11/2004. Helsinki: National Technology Agency of Finland Tekes.
- Valve, H., McNally, R., Pappinen, A. 2007. *Research and the realities of the public world: a "PROTEE"-mediated evaluation of the field trial as a risk assessment tool for gm-trees*. Manuscript 4 May 2007.
- Vedung, E. 1997. *Public policy and programme evaluation*. New Brunswick: Transaction publishers.
- Wildavsky, A. 1985. The Self-Evaluating Organization. In: E. Chelmsky (ed). *Program Evaluation: Patterns and Directions*. Washington D.C.: American Society for Public Administration, 246–265.

APPENDIX I

Organisations that provided material for the study

Federal Ministry of Agriculture, Forestry, Environment and Water Management, Austria
Federal Ministry of Education, Science and Culture, Austria
Federal Ministry of Transport, Innovation and Technology (BMVIT), Austria
Federal Office for Scientific, Technical and Cultural Affairs (OSTC), Belgium
The Flemish Environment, Nature and Energy Department (DENE), Belgium
Academy of Finland, Finland
Finnish Funding Agency for Technology and Innovation (Tekes), Finland
Ministry of the Environment (FiMoE), Finland
Agency of the Environment and Energy Control (ADEME), France
French Ministry of Environmental and Sustainable Development (MEDD), France
Environmental Protection Agency (EPA), Ireland
Committee for the Evaluation of Research (CIVR), Italy
Research Council of Norway (RCN), Norway
Institute of Environmental Protection (IEP), Poland
Swedish Environmental Protection Agency, Sweden
Environment Agency for England and Wales (EA), United Kingdom
Natural Environment Research Council (NERC), United Kingdom
National Institute for Public Health and the Environment (RIVM), The Netherlands

APPENDIX 2

List of participants: SKEP WP3 & WP4 Workshop 18-20th April 2007

Ariana Nastaseanu	National Authority for Scientific Research – Romania
Sauli Rouhinen	FiMoE – Finland
John Holmes	University of Oxford – UK
Paula Kivimaa	SYKE – Finland
Erik Fellenius	SwEPA – Sweden
Nils Raynaud	MEDD – France
Eeva Furman	SYKE – Finland
Simon Gardner	EA – UK
Pekka Harju-Autti	FiMoE – Finland
Natasha Martineau	EA – UK
Manuela Kienegger	BMLFUW – Austria
Mathieu Jahnich	MEDD – France
Martine Vanderstraeten	BELSPO – Belgium
Gretchen Jordan	Sandia National Laboratories – USA
Hanna Mela	SYKE – Finland
Susan Senecah	State University New York College of Environmental Science and Forestry – USA
Bozena Kozera-Sucharda	IEP – Poland
Jennie Savgård	SwEPA – Sweden
Pierre Demoitie	BELSPO – Belgium
Jean Marc Merillot	ADEME – France
Jeroen Cockx	FG-ENED – Belgium
Oliver Hochadel	University of Klagenfurt – Austria
Philippe Van Haver	FG-ENED – Belgium
Ingunn Borlaug Lid	RCN – Norway
Vera Rabelt	UBA – Germany
Bettina Wanschura	Ministry of Science and Research – Austria
Jon Greaves	IEPA – Ireland
H.J. van Veen	SKB – Netherlands
Catarina Johansson	SwEPA – Sweden
Karin Totland	RCN – Norway
Alister Scott	The Knowledge Bridge – UK
Dick van Lith	Ministry of the Environment – the Netherlands
Gaetano Battistella	APAT – Italy
Shane Colgan	IEPA – Ireland
Hana Soszka	IEP – Poland

APPENDIX 3

SKEP WP3 & WP4 Workshop Programme Åland islands 18–20th April 2007

Wednesday 18 April

- 9.00–11.00 Arrival and coffee
- 11.00 Welcoming words
- 11.15 Setting the scene: SKEP activities in the field of programme evaluation
- 11.30 “Programme Evaluation as a Path Towards Improved R&D”
– Gretchen Jordan
- 12.15 Discussion
- 12.30 Lunch
- 13.30 Experiences of programme evaluation of the SKEP partners:
draft report
- 14.00 Group work I
- 15.00 Discussion
- 15.15 Coffee
- 15.45 Group work II
- 16.45–18.00 Discussion
- 19.00 Dinner

DOCUMENTATION PAGE

<i>Publisher</i>	Finnish Environment Institute (SYKE)			<i>Date</i> April 2008
<i>Author(s)</i>	Paula Kivimaa, Hanna Mela & Eeva Furman			
<i>Title of publication</i>	SKEP ERA-NET Work Package 3 – Approaches and practices in the evaluation of environmental research programmes			
<i>Publication series and number</i>	The Finnish Environment 13/2008			
<i>Theme of publication</i>	Environmental protection			
<i>Parts of publication/ other project publications</i>				
<i>Abstract</i>	<p>SKEP ERA-NET is a partnership that aims to improve the coordination of environmental research in Europe. This report is a result of Work Package 3 of the SKEP ERA-NET, led by the Finnish Environment Institute (SYKE) and the Finnish Ministry of the Environment (FiMoE). SKEP ERA-NET is part of the ERA-NET scheme, which is one of the tools of the European Union to create an integrated European Research Area for innovative knowledge production.</p> <p>The report concentrates on the mid-term and ex-post evaluation of environment-related research programmes. It synthesises experiences of programme evaluations by reviewing evaluation literature and analysing material that has been collected from various research funding agencies. The material includes both written documents and interviews. The report also presents outcomes of an evaluation-related workshop held in spring 2007. Different approaches are provided concerning the planning of programme evaluation and choosing the type, focus and methods of the evaluation. Special attention is paid to the challenges of evaluating the societal and environmental outcomes of a research programme. The utilisation of the evaluation results is also covered.</p> <p>As the ERA-NET scheme aims at increasing supranational research collaboration in the form of e.g. joint research programmes, one part of the report is devoted to the evaluation of jointly coordinated and funded research programmes.</p>			
<i>Keywords</i>	Evaluation, environmental research programmes, research funding			
<i>Financier/ commissioner</i>				
	ISBN (pbk.)	ISBN 978-952-11-3078-6 (PDF)	ISSN (print)	ISSN 1796-1637 (online)
	No. of pages 56	Language English	Restrictions Public	Price (incl. tax 8 %)
<i>For sale at/ distributor</i>				
<i>Financier of publication</i>	Finnish Environment Institute (SYKE), P.O.Box 140, FI-00251 Helsinki, Finland Phone +358 20 490 123, Fax +358 20 490 2190 Email: neuvonta.syke@ymparisto.fi , www.environment.fi/syke			
<i>Printing place and year</i>				

KUVAILULEHTI

Julkaisija	Suomen ympäristökeskus (SYKE)			Julkaisu-aika	Huhtikuu 2008
Tekijä(t)	Paula Kivimaa, Hanna Mela & Eeva Furman				
Julkaisun nimi	SKEP ERA-NET Työpaketti 3 – Näkökulmia ja käytäntöjä ympäristötutkimusohjelmien arviointiin				
Julkaisusarjan nimi ja numero	Suomen ympäristö 13/2008				
Julkaisun teema	Ympäristönsuojelu				
Julkaisun osat/ muut saman projektin tuottamat julkaisut					
Tiivistelmä	<p>SKEP ERA-NET on yhteistyöhanke, jonka tavoitteena on parantaa ympäristötutkimuksen koordinoitua Euroopassa. Tämä raportti on tuotettu SKEP ERA-NET:in työpaketissa 3, jota vetävät Suomen ympäristökeskus (SYKE) ja ympäristöministeriö (FiMoE). SKEP ERA-NET kuuluu ERA-NET –ohjelmaan, joka on yksi EU:n työkaluista Eurooppalaisen Tutkimusalueen (ERA) luomiseksi ja innovaatioiden edistämiseksi.</p> <p>Raportti keskittyy ympäristöaiheisten tutkimusohjelmien puoliväli- ja jälkikäteisarviointiin. Raportissa käydään läpi arviointikirjallisuutta, analysoidaan tutkimusta rahoittavista organisaatioista kerättyä kirjallista ja haastatteluaineistoa sekä esitellään keväällä 2007 järjestetyn arviointityöpajan tuloksia. Raportti tarjoaa useita eri näkökulmia tutkimusohjelmien arvioinnin suunnitteluun sekä arvioinnin tyyppiin, fokuksen ja menetelmien valintaan. Erityistä huomiota kiinnitetään tutkimusohjelmien yhteiskunnallisten ja ympäristöön liittyvien vaikutusten arviointiin. Raportissa käsitellään myös ohjelma-arvioinnin tulosten hyödyntämistä.</p> <p>ERA-NET –ohjelma pyrkii lisäämään kansainvälistä tutkimusyhteistyötä muun muassa monenkeskisten tutkimusohjelmien kautta. Siten yksi osa raportista keskittyy käsittelemään yhteisrahoitteisten ja yhteisesti koordinoitujen tutkimusohjelmien arviointia.</p>				
Asiasanat	Arviointi, ympäristötutkimusohjelmat, tutkimusrahoitus				
Rahoittaja/ toimeksiantaja					
	ISBN (nid.)	ISBN 978-952-11-3078-6 (PDF)	ISSN (pain.)	ISSN 1796-1637 (verkköj.)	
	Sivuja 56	Kieli Englanti	Luottamuksellisuus Julkinen	Hinta (sis.alv 8 %)	
Julkaisun myynti/ jakaja					
Julkaisun kustantaja	Suomen ympäristökeskus (SYKE), PL 140, 00251 HELSINKI Puh. 020 490 123 Sähköposti: neuvonta.syke@ymparisto.fi , www.ymparisto.fi/syke				
Painopaikka ja -aika					

PRESENTATIONSBLAD

Utgivare	Finlands miljöcentral (SYKE)			Datum April 2008
Författare	Paula Kivimaa, Hanna Mela & Eeva Furman			
Publikationens titel	SKEP ERA-NET Arbetspaket 3 – Synpunkter på utvärdering av miljöforskningsprogram			
Publikationsserie och nummer	Miljön i Finland 13/2008			
Publikationens tema	Miljövård			
Publikationens delar/ andra publikationer inom samma projekt				
Sammandrag	<p>SKEP ERA-NET är ett samarbetsprojekt som syftar till att förbättra samordningen av miljöforskning i Europa. Denna rapport är ett resultat av SKEP ERA-NETs arbetspaket 3 som drivs av Finlands Miljöcentral (SYKE) och Finlands Miljöministerium (FiMoE). SKEP ERA-NET är en del av ERA-NET programmet, som är ett av EU:s arbetsverktyg i skapandet av Europeiska Forskningsområdet (ERA) och befrämjandet av innovationer.</p> <p>Rapportens fokus ligger på utvärdering halvvägs i program och efter programslut. Utvärderingslitteratur samt material från diverse finansieringsanstalter används i analysen av evaluering av projekt. Både skriftligt och intervju material ingår i analysen. Rapporten redovisar även resultat av en utvärderingsworkshop som hölls under våren 2007. Rapporten tillhandahåller olika lösningar till förberedning och implementering av utvärderingsaktiviteter. Särskild uppmärksamhet fåsts vid utvärderingen av vilken social och miljömässig inverkan ett projekt har haft, samt på hur resultaten från programutvärderingen används.</p> <p>Eftersom ERA-NET programmet har som syfte att öka internationellt forskningssamarbete genom t.ex. multilaterala forskningsprogram, behandlar rapporten utvärdering av samfinansierade och samkoordinerade program.</p>			
Nyckelord	Utvärdering, miljöforskningsprogram, forskningsfinansiering			
Finansiär/ uppdragsgivare				
	ISBN (hft.)	ISBN 978-952-11-3078-6 (PDF)	ISSN (print)	ISSN 1796-1637 (online)
	Sidantal 56	Språk Engelska	Offentlighet Offentlig	Pris (inneh. moms 8 %)
Beställningar/ distribution				
Förläggare	Finlands miljöcentral (SYKE), PB 140, 00251 Helsingfors Tel. 020 490 123 Epost: neuvonta.syke@ymparisto.fi, www.miljo.fi/syke			
Tryckeri/tryckningsort och -år				



S Y K E

ISBN 978-952-11-3078-6 (PDF)

ISSN 1796-1637 (verkkoj.)