



RTDI EVALUATION STANDARDS

Research, Technological Development and Innovation
Evaluation Standards

Project Acronym:	EVAL- INNO
Project full title:	Fostering Evaluation Competencies in Research, Technology and Innovation in the SEE Region
Project No:	SEE/B/0025/1.3/X
Funding Scheme:	SEE Transnational Cooperation Programme
Project start date:	1 May 2011
Project duration:	36 months
Project partners:	Centre for Social Innovation (ZSI), Austria (Coordinator) National and Kapodistrian University of Athens – Center of Financial Studies (NKUA-CFS), Greece Applied Research and Communications Fund (ARC Fund), Bulgaria Public Foundation for the Development of Industry (IFKA), Hungary University of Montenegro (International Relations Office) (UM-IRO), Montenegro Mihajlo Pupin Institute (MPI), Serbia
Contact:	Mr. Martin Felix Gajdusek (Project Coordinator): eval-inno-office@zsi.at

EVAL-INNO is a project funded under the South East Europe Transnational Cooperation Programme. The information contained in this publication does not necessarily reflect the position or opinion of the Joint Technical Secretariat (JTS) or European Commission (EC). The authors are solely responsible for the content. See Annex 2 for the full list of authors.

Publisher and editor

EVAL-INNO

c/o Centre for Social Innovation
Linke Wienzeile 246, A-1150 Wien
eval-inno-office@zsi.at

Technical editors

Ines Marinkovic, Tatjana Knezevic

Language editor

Maureen Patricia McNeill

Design

Krassimir Apostolov, Sofia, Bulgaria

Printing house

Saxoprint GmbH

Year of publication: 2012

Print run: 1500

ISBN 978-3-200-02844-9

TABLE OF CONTENTS

1. INTRODUCTION	5
2. FUNDAMENTALS OF RTDI EVALUATIONS	7
2.1. Rationale and Specifics of STI Policy	7
2.2. Terminology and Logic	9
2.2.1. Definitions	9
2.2.2. Functions and Purposes of Evaluations	11
2.2.3. Levels of RTDI Evaluation	12
2.2.4. Timing of Evaluations	14
2.2.5. Content of Evaluations	14
2.3. Evaluation Criteria	16
3. THE EFFECTIVE PLANNING OF EVALUATIONS	17
3.1. Evaluation in the Policy Cycle	17
3.2. Evaluation Framework	20
3.3. How to Ensure Evaluability	21
3.4. Institutional Aspects	24
3.5. Rules and Ethics for Evaluators and Commissioning Institutions	26
3.5.1. Competence of Evaluators	27
3.5.2. Potential Conflicts	28
3.6. Ensuring the Take up of Evaluation Results	30
3.7. Should Evaluation Results be Binding?	32
4. TENDERING EVALUATION	33
4.1. How to Proceed in Writing Terms of Reference (ToR)	33
4.2. Components of ToR at a Glance	33
5. RECOMMENDATIONS FOR AN EVALUATION ROAD MAP FOR SEE COUNTRIES	36
6. REFERENCES	40

ANNEX 1: EVALUATION METHODOLOGIES AND TECHNIQUES	42
ANNEX 2: NAMES OF AUTHORS.....	45
ABSTRACT	46

TABLES

TABLE 1: MAIN FUNCTIONS OF RTDI EVALUATIONS.....	11
TABLE 2: EVALUATION DESIGN MATRIX.....	21
TABLE 3: AREAS OF CONFLICT IN RTDI EVALUATIONS.....	28
TABLE 4: SELECTED EVALUATION METHODOLOGIES AND TECHNIQUES	42

FIGURES

FIGURE 1: PROGRAMME CYCLE	18
FIGURE 2: POLICY CYCLE.....	19
FIGURE 3: OBJECTIVES AND EFFECTS WITHIN THE PROGRAMME CYCLE	23

1. INTRODUCTION

The evaluation standards presented here aim to contribute to the improved implementation and exploitation of Research, Technological Development and Innovation (RTDI) measures by promoting meaningful evaluation procedures to foster strategic intelligence building and evidence-based decision-making in the field of science, technology and innovation (STI) policy. They address:

- a. authorities commissioning RTDI evaluations (often ministries in charge of research, technological development and innovation and their respective measures, programmes and policies);
- b. evaluators carrying out RTDI evaluation studies;
- c. organizations and stakeholders subjected to evaluations, such as funding agencies, public research organizations, universities or intermediary organizations (e.g. technology transfer offices, technology and science parks, and innovation centres etc.).

The evaluation standards provide information about the purposes and characteristics of evaluations in the field of STI. They introduce an internationally acknowledged terminology and evaluation theory framework, guide users in practical issues concerning governance, conduct and use of RTDI evaluations. Additionally, they provide many practical hints on how to plan and implement evaluations, including the writing of Terms of References (ToR) to procure external RTDI evaluations and the structuring of meaningful evaluation reports, to mention just a few.

The publishing of RTDI evaluation standards is motivated by the complexity and heterogeneity of research and innovation systems, which requires ERDF and IPA countries to possess strategic intelligence in order to design, implement and follow-up RTDI measures at different spatial levels (local, national, regional

and European) by addressing issues of relevance, efficiency, efficacy, impact and sustainability. Evaluations are one of the most essential tools for evidence-based decision-making. This is especially true in the South East European (SEE) region, which is characterised by an emergence of new RTDI policies, programmes, and (support) organizations and a funding transformation towards competitive schemes. At the same time, however, a lack of methodological and procedural know-how both on the side of evaluators and awarding authorities concerning purpose, design and use of evaluations is obvious¹ and the potential of evaluations to build strategic intelligence is not being fully exploited yet.

Especially under tight financial regimes, public spending for RTDI has to be justified through the identification of correct rationales and mechanisms for performance-based RTDI funding from the start. To secure the optimum use of taxpayers' money, the principles of good governance have to be respected. Adequately conducted RTDI evaluations are a proper tool for ensuring transparency and accountability and contribute to an efficient new public management. The evaluation standards guide presented here provides support to conduct proper and meaningful tenders to procure RTDI evaluations as well as to implement them in a way that secures strategic intelligence building and evidence-based decision-making.

The standards have been drawn up in an interactive process involving experts² from six countries through discussion of and reflection on existing RTDI experience and framework conditions in the South East European region. It is the first attempt of this kind at the SEE level. The authors have not been working from scratch and '*re-inventing the wheel*', but rather were using the existing practice of evaluation standards from EU countries (especially the Austrian Evaluation Standards in Research and Technology Policy³) and the USA⁴ as successful examples of good practice and trying to adapt them as comprehensibly concisely and usefully as possible to the particular situation and needs of the region.

Towards the end of the project, these RTDI Evaluation Standards will be revised and an updated version will be issued, enriched through the regionally focused experience that the consortium will gain through joint work under the umbrella of the EVAL-INNO project.

1 The EU INNO-Appraisal Project, which took stock of and assessed appraisal exercises such as evaluations in the area of innovation policy across Europe, concluded there is a significant difference in the application and use of evaluations between more advanced RTDI countries and new EU Member States, not to mention SEE.

2 The names of the experts are listed in Annex 2.

3 Evaluation Standards in Research and Technology Policy (full-length version), Platform – Fteval, Vienna

4 A detailed list of references can be found at the end of the document

2. FUNDAMENTALS OF RTDI EVALUATIONS

2.1. Rationale and Specifics of STI Policy

Science, Technology and Innovation (STI) policy:

- a. deals with the formulation, regulation, organization and administration of national research and innovation systems;
- b. is an umbrella policy for stimulating and driving knowledge generation, knowledge utilization and knowledge diffusion processes in an optimal mix to increase the economic competitiveness and societal well-being;
- c. is embodied in laws and regulations, strategies, policy initiatives and briefs, institutional and governance structures, public programmes and related human resources.

STI policy aims to establish a high level of competitive and efficient research and innovation activities within a distinct geographical area (e.g. regional, national, local). STI policy interventions can be multifaceted. They often materialize in measures (e.g. venture capital funds) and programmes with specified budgets (e.g. centres of excellence programmes or thematic programmes, for instance in the fields of nanotechnology, biotechnology or social sciences), but can be also embodied in dedicated organizations (e.g. agencies, universities, technology transfer facilities, incubators, innovation management service providers).

“Knowledge” has become a key factor for economic growth; understanding the dynamics of knowledge generation, knowledge diffusion and knowledge exploitation has become a determinant for successful STI policy. RTDI refers to the fact that it is not only research (and development) activities which are the focus of STI policy, but also the transfer and application of new knowledge in(to) new products, new process technologies and new marketing techniques, as well as organizational and social processes (i.e. innovation in a broader sense).

The target groups of STI policy (e.g. researchers, public and private non-profit RTDI organizations as well as innovative companies) are often directly or indirectly financially supported to reduce existing market and system failures. These failures, as well as externalities associated with knowledge dynamics, exist because of information asymmetries⁵, the necessary public good character of pre-competitive, especially basic research, and sometimes because of unwanted knowledge retention or spill-over impacts⁶, which may hinder the commercial appropriation of generated knowledge and may lead—if not counterbalanced—not only to insufficient utilisation of existing knowledge but also to insufficient RTDI investments in the private business–enterprise sector.

Evidently, STI policy aims to secure the highest possible *additionality* effects of its RTDI interventions employed to overcome existing market and system failures, which are measured within the framework of evaluations. The following additionality levels⁷ are typically distinguished:

- Input additionality is the extent to which R&D activity is increased as a result of government intervention.
- Output additionality is the extent to which additional outputs increase as a result of public intervention, e.g. the growth of new publications or product sales, export activity, patents, technology levels.
- Behavioural additionality is the extent to which beneficiaries and other stakeholders change their behaviour and become more competitive respectively goal-oriented.

Within the increasing importance of STI policy, RTDI evaluation practice evolved starting in the late 1960s in the United States of America and in the 1970s in Europe. In the 1980s, when STI policy theory became more system and mission oriented, RTDI evaluation gained additional momentum:

“The concern about the use of research results and the alignment of STI programs with societal needs resulted in a new social contract between science and society, which in turn had implications for the evaluation of science and technology policies (see, e.g., Dalpe & Anderson, 1993; Jaffe, 1998). Strategically targeted R&D called for evaluation practices able

5 ... which explains why investment in new knowledge development (typically R&D) is always suboptimal at the macro-level.

6 ... which explains why R&D funding cannot be directly connected to innovation output.

7 The additionalities are described in OECD (2006): Government R&D Funding and Company Behaviour – Measuring behavioural additionality. OECD, Paris

to assess whether the specific policy goals had been achieved and thus support the decision-making process.”⁸

Today, RTDI policy evaluation has become one of the most important strategic intelligence sources for policymakers in the STI field. One element generating this interest is the *“emergence of new public management approaches emphasizing the application to all public functions (including the funding of R&D activities) of management practices oriented to the control of outputs rather than simply monitoring processes and inputs”*. (ibid., p.71)

Examples of evaluation standards are the *Austrian Standards for Evaluation in Research and Technology Policy*⁹, which was jointly developed by all Austrian RTDI stakeholders, the *Standard Evaluation Protocol 2009-2015 (SEP)*, which is the fourth protocol for evaluation of scientific research in the Netherlands¹⁰, developed by several major Dutch stakeholders and the *British Research Evaluation Standards*, which were applied in the *Research Assessment Exercise*¹¹. In addition, wide-ranging literature proposes different standards with different levels of obligation and outreach. This includes the *White Paper in using bibliometrics in evaluating research published by Thomson Reuters*¹²; the standards of evaluation for qualitative research¹³; proposed specificities for the evaluation of transdisciplinary research projects¹⁴ and the proposed standards for research ethics evaluation procedures prepared by the *European Federation of Psychologists’ Associations*.¹⁵

2.2. Terminology and Logic

2.2.1. Definitions

Evaluation is a multi-actor multi-level process involving those who commission an evaluation (e.g. ministries), those who implement it and the recipients or customers of the evaluation. The term evaluation has various definitions, depending on the subject matter (e.g. RTDI evaluation, labour market evaluation,

8 Molas-Gallart, J. and Davies, A. (2006): Toward theory-led evaluation: the experience of European science, technology and innovation policies. In: *American Journal of Evaluation*, 27 (1), p. 71.

9 <http://www.fteval.at/>

10 http://www.knaw.nl/content/Internet_KNAW/publicaties/pdf/20091052.pdf

11 <http://www.rae.ac.uk/>

12 http://thomsonreuters.com/content/science/pdf/ssr/training/UsingBibliometricsinEval_WP.pdf

13 tackled by an NSF report: http://scholar.harvard.edu/sites/scholar.iq.harvard.edu/files/lamont/files/nsf_report_tackles_standards_of_evaluation_for_qualitative_research.pdf

14 http://www.transdisciplinarity.ch/documents/Evaluation_WorkingPaper.pdf

15 <http://science.efpa.eu/information-/proposed-standards-for-research-ethics-evaluation-procedures/>

development cooperation evaluation), applied methodology (e.g. process evaluation, ex ante or ex post impact assessment) and the application of results (e.g. project evaluation, programme evaluation, organization evaluation, policy evaluation). Below are few common definitions of evaluation:

“Evaluation can be defined as a systematic and objective process that assesses the relevance, efficiency and effectiveness of policies, programmes and projects in attaining their originally stated objectives. It is both a theory- and practice driven approach. Evaluation results feed back into the policy-making process, so that it is part of a continuous learning process. This brings transparency and accountability to the policy-making process and helps formulating and assessing policy rationales.”¹⁶

“[Evaluation is a ...] systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. Evaluation also refers to the process of determining the worth or significance of an activity, policy or programme. [Evaluation is an ...] assessment, as systematic and objective as possible, of a planned, ongoing, or completed development intervention.”¹⁷

“Evaluation is a process which attempts to determine as systematically and objectively as possible the relevance, effectiveness, efficiency and impact of activities in the light of specified objectives. It is a learning and action-oriented management tool and organizational process for improving both current activities and future planning, programming and decision-making.”¹⁸

Evaluations can be designed according to different functions and purposes, different aggregation levels, different timing points and content, as explained in the next few sub-sections.

16 European Commission (2002): RTD Evaluation Toolbox – Assessing the Socio-Economic Impact of RTD-Policies, IPTS. Technical Report Series, p. 15.

17 OECD (2010): DAC's Glossary of Key Terms in Evaluation and Results Based Management, p. 21.

18 A UNICEF Guide for Monitoring and Evaluation. Making a Difference? (1991) p. 2

2.2.2. Functions and Purposes of Evaluations

Since public funding in the field of RTDI (not exclusively) involves a variety of actors with differing interests and expectations —such as policymakers or the clients of their RTDI interventions (e.g. companies or researchers) —evaluations also differ in their function and purpose. Thus, depending on the basic conditions provided, evaluation can fulfil various functions, as the box below shows. Although evaluations may be retrospective, they are essentially forward-looking with regard to their purpose, which may include contributing to improving policy, procedures or techniques, or considering the continuation or discontinuation of a project, measure, organization or programme etc.¹⁹

TABLE 1: MAIN FUNCTIONS OF RTDI EVALUATIONS

	Evaluations...	Required complexity of analysis	Room for experimentation
Legitimizing function	justify the use of public goods	high	high
Information provision function	raise public awareness and inform policy	can be low or high	fairly high
Learning function	provide feedback to actors	high	high
Steering function	prepare for decisions and set future policies	high	low
Controlling function	check whether rules are abided by	it depends, but usually low	low
Mediating function	connect stakeholders	not relevant	not relevant

Source: The functions (first and second column) are listed originally in Evaluation Standards in Research and Technology Policy (full-length version). Plattform – Fteval, Vienna, additions (third and fourth column) are by BORSI, Balazs.

¹⁹ OECD (2010), Quality Standards for Development Evaluation, DAC Guidelines and Reference Series, p. 8.

A clear presentation of evaluation expectations (purpose, objectives and intended users) is relevant for all parties involved in the evaluation process and should be clarified at the very beginning. Broadly speaking, the purposes and expectations of evaluation processes are divided between two functional poles:

Formative evaluation: *“An evaluation concerned with examining ways of improving and enhancing the implementation and management of interventions. Formative evaluations tend to be conducted for the benefit of those managing the intervention with the intention of improving their work.”²⁰*

Summative evaluation: *“An evaluation concerned with determining the essential effectiveness of programmes. Summative evaluations tend to be conducted for the benefit of external actors (groups who are not directly involved in the management of a programme), for reasons of accountability or to assist in the allocation of budgetary resources.”²¹*

The specific objectives of an evaluation must clarify what the evaluation aims to uncover.²² For example, it should:

- ascertain results (output, outcome, impact) and assess the effectiveness, efficiency, relevance and sustainability of a specific intervention²³;
- provide findings, conclusions and recommendations with respect to a specific intervention in order to draw lessons for future design and implementation.

2.2.3. Levels of RTDI Evaluation

When it comes to STI policy evaluation, the following evaluation levels should be considered respectively differentiated:

- **Policy** evaluations (e.g. research and/or innovation policy on different spatial levels, such as national, regional or local): A policy²⁴ is typically described as a deliberate plan of action to guide decisions and achieve

20 EC (2002): RTD Evaluation Toolbox, p. 259.

21 EC (2002): RTD Evaluation Toolbox, p. 266.

22 OECD (2010), Quality Standards for Development Evaluation, DAC Guidelines and Reference Series, p. 22.

23 Not all evaluation criteria can be assessed for all programmes / projects —most notably effectiveness, impact and sustainability are assessed. If the intervention logic is flawed, i.e. activities are not appropriate to achieve certain objectives, then it does not make sense to evaluate their achievement.

24 Definitions used according to EVALSED (2009) and fteval Evaluation Standards.

rational outcome(s), while policy cycle is the term used to describe the lifespan of a policy, from its formulation to its review. It comprises: needs assessment / agenda setting; planning / policy formulation; policy implementation; policy monitoring; and evaluation and feedback. While law can compel or prohibit, policy merely guides actions in a direction that is most likely to achieve a desired outcome. Policy is generally not limited by time or budget, in contrast to projects and programmes.

- **Systems** evaluations: e.g. RTDI funding systems, including direct and indirect funding components.
- **Sector** evaluations: e.g. main industrial branches or the service sector as a whole.
- **Organizational** evaluations: any organization can be evaluated, however in this context evaluation of universities, with their combination of research and teaching, research institutions and funding organizations or agencies as well as intermediary organizations, can be considered most relevant.
- **Portfolio** evaluations: e.g. comprehending a number of programmes, measures, organizations, etc. targeting similar objectives, e.g. venture capital support mechanisms, or certain fields of science, etc.
- **Programme** evaluations: a programme is a set of financial tools, organizational solutions and human resources mobilised to achieve a clearly stated objective or set of objectives within a given period. A programme is limited in terms of timescale and budget. Programme objectives are defined beforehand; efforts have to be made to systematically strive for coherence among these objectives. The three main steps in the life cycle of a programme are design, implementation and ex post evaluation. A programme is always under the responsibility of an authority or several authorities, which share in decision-making. Programmes are often broken down into axes, measures and projects.
- **Project** evaluations: a project is a single intervention directed towards the attainment of operational objectives, with a fixed time schedule, a dedicated budget and placed under the responsibility of an operator. Particularly, careful ex ante evaluations are made of major interventions (especially of RTDI infrastructure projects), using the cost–benefit analysis technique. Cost-benefit analyses need valuations and quantifications that are often lacking. In the case of RTDI they are often based on assumptions that may prove over (or under-) optimistic. Hence, if cost-benefit

analyses are performed they should be accompanied by complementary techniques, including sensitivity analyses and preferably also case-studies.

2.2.4. Timing of Evaluations

According to the time evaluations are carried out, and depending on the evaluation purpose, it is useful to differentiate between:

- **Ex ante evaluation** —the evaluation is conducted prior to the implementation of a strategy, piece of legislation, programme or project;
- **Interim evaluation** —the evaluation is conducted during the implementation of a strategy, programme, project or during the operations of an organization;
- **Terminal evaluation** —the evaluation is conducted immediately at the end of the implementation of a strategy, programme, project or piece of legislation;
- **Ex post evaluation** —the evaluation is conducted a short time after the end of the implementation of a strategy, programme, project or piece of legislation;
- **Periodical evaluation** —the evaluation is conducted regularly throughout the implementation of a strategy, programme or organization, for example biannually;
- **Ad hoc evaluation** —the evaluation was not foreseen during the development or implementation of a strategy, organization, programme, or project but is conducted to meet a need that emerged later.

2.2.5. Content of Evaluations

The content of evaluations²⁵ can be differentiated between:

- **Concept evaluations:** reviewing the mission, assumptions, fundamental hypotheses and basic conditions of programmes (institutions or projects). Concept evaluations assess the extent to which it is justified to use a particular policy, programme or project (including any methods of resolution it comes up with) to combat the problems identified earlier. Concept evaluations are typically component parts of ex ante evaluations.

²⁵ Evaluation Standards in Research and Technology Policy Platform —Fteval, Vienna, p.6.

- **Design evaluations:** deal with the effectiveness of the design of an intervention and its organizational structure. They assess the extent to which designated operational objectives are appropriate to respective problems, as well as whether the instruments used are suited to meet prescribed objectives and whether it is, or was, possible to reach them. Design evaluations are typically component parts of all stages of programme evaluations (**ex ante, interim and ex post**).
- **Process evaluations:** In the early stages of a new programme or new initiatives within a programme, evaluation questions often focus on programme processes, e.g. how well authorised activities are carried out and reach intended recipients. Evaluation studies designed to address the quality, efficacy and efficiency of programme operations are frequently called process or implementation evaluations. Process evaluations are typically component parts of interim evaluations and often occur in programmes demanding a great deal of programme management in terms of promotion, communication, timeliness, consultancy and control.
- **Impact evaluations:** seeks to answer cause-and-effect questions, and the changes in outcomes that are directly attributable to a policy, programme or project. Impact analyses assess the extent to which programme objectives have been attained and attempt to identify and to quantify, as far as possible, all of the effects brought about by the programme, directly or indirectly, intentionally or not. In doing so, a differentiation is made between the immediate '**output**' of a programme (e.g. the number of projects funded), the result or '**outcome**'²⁶, (e.g. the number of usable patents), and the effect or '**impact**' (e.g. the market profits or increases in turnover). In view of business RTDI, the effects of programmes are to be found where new inventions and developments interact with the market. They are usually expressed in economic terms. Impacts of (public) RTDI can also be measured in non-economic terms (e.g. publication citation rates; societal and environmental indicators). The focus is always on recording the return on a programme/project in social/environmental terms, going beyond mere private benefit. Impact analyses are important component parts of ex post evaluations.

26 Sometimes the outcome term used to refer to longer-term impact while the result term indicates the initial impact (European Commission (1997), Evaluating EU expenditure Programmes: A guide – Ex post and intermediate evaluation, 1st edition).

2.3. Evaluation Criteria

When evaluating programmes, projects and other interventions, it is useful to consider the following OECD-DAC²⁷ evaluation criteria:

Relevance means the extent to which an activity is suited to the priorities and policies of the target group, recipient and donor.

Efficiency measures outputs —qualitative and quantitative —in relation to inputs.

Effectiveness is the extent to which an intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.

Impact is understood as positive and negative, primary and secondary long-term effects produced by an intervention, directly or indirectly, intended or unintended.

Sustainability means the continuation of benefits from an intervention after major assistance/funding has been completed. This includes assessing the probability of long-term benefits. The resilience to risk of the net benefit flows over time. Projects need to be environmentally as well as financially sustainable.

²⁷ Glossary of Evaluation and Results Based Management (RBM) Terms, OECD (2010).

3. THE EFFECTIVE PLANNING OF EVALUATIONS

3.1. Evaluation in the Policy Cycle

Evaluations of policies, organizations, portfolios of programmes/measures or single programmes are extremely helpful in ensuring that they achieve their stated aims, and that —where they fall short —redirections can be made in order to put financial resources to optimal use. Evaluations provide information to policymakers in fields where economic indicators are often not readily available for use in impact assessments. This applies particularly to RTDI policy, where the scientific, social, environmental and economic impacts can take years to emerge, and when they do, are difficult to attribute to particular interventions within such complex systems.

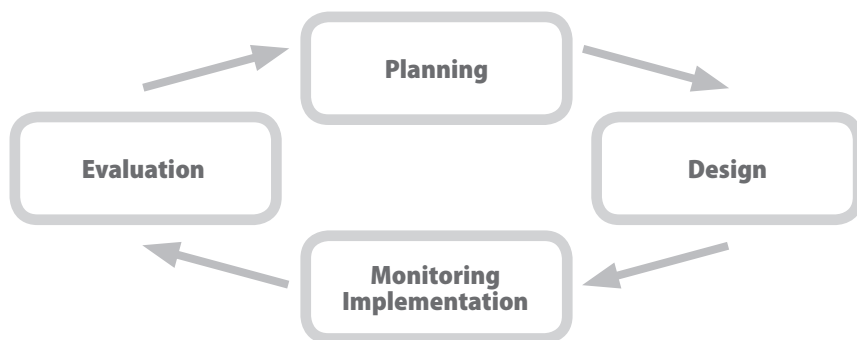
The timing of an evaluation is very important because it should reach decision-makers before major decisions are due to be made regarding a specific intervention. This requires data to be collected by the managing authority (or the evaluators) before an analytical assessment takes place. As a basis for this planning process, the timing of evaluations should be embedded into programme guidance documents to ensure that the managing authority (e.g. the agency managing programme implementation) has enough time, budget and fixed deadlines to collect proper data and commission the evaluation (often from an external source) in time. When embedding the timing of evaluations into programme documents at the strategic level, the length of time before a programmatic intervention is realistically expected to begin to achieve its objectives should be considered.

Figure 1 features four important steps of the ‘programme cycle’:

1. Planning of an intervention (e.g. deliberating the objectives of a programme and its main characteristics in terms of thematic orientation and budget);

2. Design (including decisions about the duration and substructure of a programme, its organizational implementation, flanking measures, assumptions and pre-conditions, evaluation requirements);
3. Implementation (e.g. via a dedicated number of calls for proposals with clear ex ante project evaluation procedures) and monitoring (including data collection);
4. Evaluation of the entire intervention.

FIGURE 1: PROGRAMME CYCLE

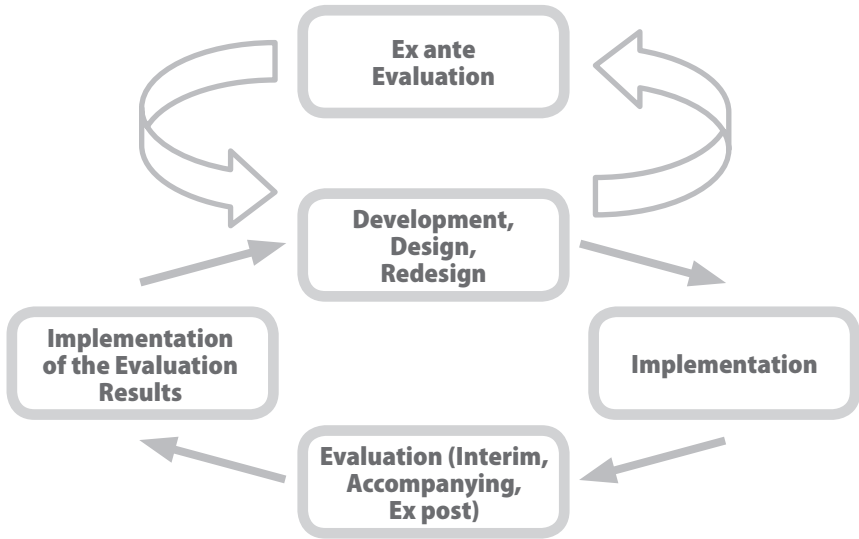


Source: ILO, Design, monitoring and evaluation of technical cooperation programmes and projects

Within this programme cycle, several evaluations can take place at different levels and at different times. Figure 2 shows a generalised ‘policy cycle’ for RTDI interventions which can be described as follows: Typically a programme is developed and designed to address a certain problem (e.g. societal or technological challenge) which should be overcome. Proposals are made and deliberated, stakeholders are consulted and some potential interventions with varying objectives and designs are drafted for deliberation at political level. An ex ante evaluation reviews the adequacy and design, including the intervention logic, of a new intervention. After eventual modifications, the planned intervention will be implemented, which often means that a dedicated agency is mandated to undertake this activity with programme management, often based upon a guidance document. After implementation has commenced, an interim evaluation may take stock of the progress made up to a certain point in time by assessing the progress made by funded projects and by assessing programme management with a view to providing

recommendations and to suggesting corrective actions for the programme's further implementation.

FIGURE 2: POLICY CYCLE



Source: Evaluation Standards in Research and Technology Policy; Platform — Fteval, Austria

The implementation of evaluation results —from both the interim and terminal evaluations—leads to consequences, which are either to prolong the programme under scrutiny, to terminate it or to modify and adapt it. The accompanying evaluation usually indicates ongoing possibilities to modify and optimize the design, and respectively to manage and implement an intervention.

An evaluation usually includes and addresses different levels: the strategic level, which is responsible for the development and design of an intervention, the implementation level (e.g. a mandated agency) and the target groups and beneficiaries of the intervention. It is important that a clear division of labour and clear communication structures exist between these three levels. It is, for instance, very important to transfer evaluation results 'upwards' to the strategic level, where they become part of the appraisal, learning and decision-making activities of those responsible for STI policy. If the information flow in the policy cycle is not closed, the danger of micro-actions, retention of unintended continuities, as well as ad hoc breaks and non-evidence based policy delivery increases. Evaluation is not a replacement for policy delivery!

3.2. Evaluation Framework

An evaluation framework lays out the overall logical structure and requirements of an evaluation process prior to the start of the evaluation itself. An evaluation framework should be fixed in the guidance documents of programmes or projects to ensure its sustainability. When designing evaluation activities, a number of questions should be addressed, including:

1. Who is performing the evaluation (i.e. external evaluators, managing authorities, or funding agencies) and how will transparency be ensured?
2. What are the objectives of the evaluation?
3. What is the overall evaluation methodology?
4. What tools will be used to perform the evaluation?
5. What data sources are needed to perform the evaluation and how are data obtained?
6. At what times must different evaluation outputs be finalized (reports, meetings, etc.)?
7. What will be the results/outputs of the evaluation?
8. How will the evaluation outputs be used, published and communicated to decision-makers?

While all of the above questions are essential to consider when establishing a logical framework in the development of an evaluation plan, questions 2–5 require particular attention. An evaluation design matrix is helpful in systematically considering these questions and developing an evaluation plan that best addresses these questions given the time and financial resources available. The following table provides an example of an evaluation design matrix²⁸:

²⁸ 'Designing Evaluations', Unites States Government Accountability Office, Applied Research and Methods, 2012 Revision.

TABLE 2: EVALUATION DESIGN MATRIX

Research Questions	Information Required and Source (s)	Scope and Methodology	Limitations	What the Evaluation will allow us to Conclude
What questions are the evaluation team trying to answer?	What information does the team need to address each evaluation question and how will it be obtained?	How will the team answer each evaluation question?	What are the design's limitations and how will it affect the outcome?	What are the expected results?
<p>These questions should be based on achieving the objectives of an evaluation as laid out in the programme or policy documents.</p> <p>The questions should be specific, objective, measurable and achievable.</p>	<p>The types of documents and information necessary</p> <p>Identify sources of required information such as databases, studies, subject area experts, programme officials, models and others</p>	<p>The methods that will be used to collect information/data (i.e. questionnaires, case studies, use of existing databases)</p> <p>A description of how each strategy will be carried out over time</p> <p>The techniques that will be used to analyze the data (regression analysis cost- benefit analysis, case study summaries etc.)</p>	<p>Questionable data quality</p> <p>Inability to access certain data types or data from certain times</p> <p>Unable to accurately generalize from the data</p> <p>Discuss how limitations will affect the final product</p>	<p>Given the available data, time and financial constraints, available methods and limitations, can the research questions be answered? If not, the evaluation needs to be redesigned.</p>

Source: Designing Evaluations', Unites States Government Accountability Office, Applied Research and Methods, 2012 Revision

3.3. How to Ensure Evaluability

Understanding the objectives of any intervention (e.g. policy, programme, measure, project or organization) in question is a fundamental issue for any evaluation exercise. Therefore, the authorities responsible for an intervention

have to make sure that objectives are clearly formulated and structured providing a clear idea to all actors (policymakers, programme managers, evaluators and beneficiaries) of what constitutes an intervention success or failure. Also the intervention logic of any policy/organization/programme/project etc. needs to be clear, namely how specific inputs, activities, components and measures are connected and how they contribute to achievement of the objectives. If the intervention logic is flawed, i.e. activities are not appropriate to achieve certain objectives, or if the input resources are simply too marginal to create any momentum leading to the achievement of objectives, then it does not make sense to evaluate their achievement.

To ensure evaluability, it is recommended to formulate SMART intervention objectives, i.e. objectives which are **Specific, Measurable, Attainable, Realistic** and **Time-bound**. In practice, notably the measurability of objectives often becomes quite challenging. A quantification of the changes which are (or should be) induced by an intervention makes an assessment of whether these objectives have been achieved (or not) easier. However, quantification is not always feasible and in a few cases it might not be useful either. If this is so, verification should be addressed in a qualitative manner.

Objectives in umbrella programmes with several sub-programmes such as the '*Operational Programmes of Structural Funds*' are usually multi-dimensional. In such cases, identifying successes or failures is not straightforward, as a programme cannot be automatically regarded as successful if its components are successful. The existing synergies, externalities and complementarities among the programme's components and their contribution to the overall objective should be clear. A useful publication, which provides several methodological approaches addressing the multi-dimensionality of programmes, is Volume 4 of the *MEANS* collection²⁹.

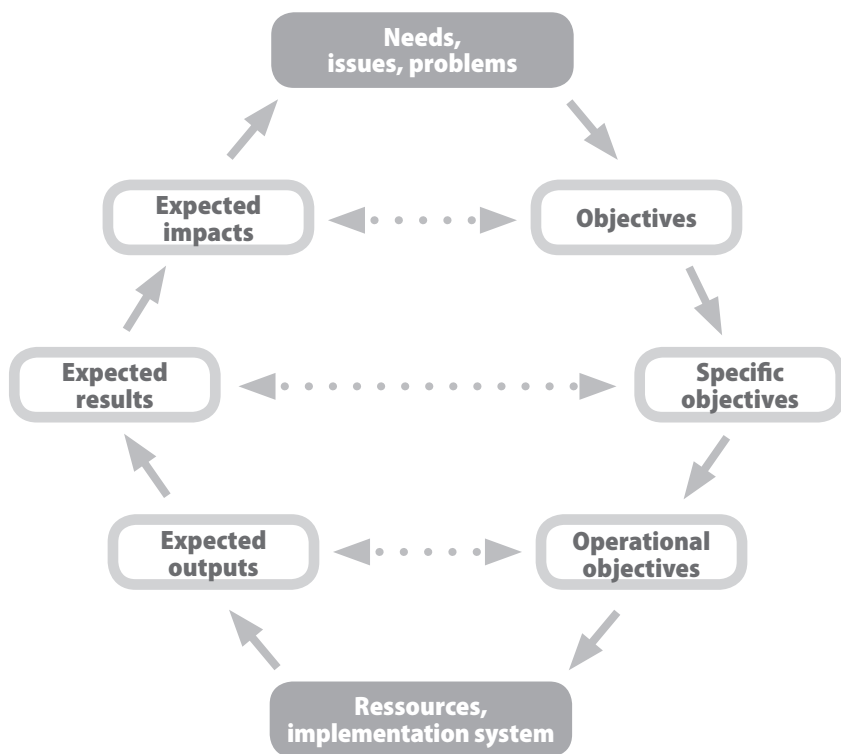
The following issues should be taken into account when formulating objectives (see also Figure 2):

- **Motivation:** What needs, issues and problems should be addressed? What do we intend to do? What is our mission?
- **Strategic objectives:** What are the main strategic objectives and how are they related with the desired impact?

²⁹ European Commission (1999), Evaluating socio-economic programmes: Technical solutions for evaluation within a partnership framework, *MEANS* collection Volume 4.

- **Specific and operational objectives:** What are the specific and operational objectives that serve the strategic objective? What outputs and results have to be attained to fulfil these objectives? Is it feasible and useful to quantify these objectives? If it is not feasible, can a procedure be set-up to review the attainment of these objectives in a qualitative manner?
- **Intervention logic:** How is the programme (or any other intervention) designed to achieve its objectives? What are the mechanisms for doing so? What are the main assumptions? When are the different types of effects expected to be visible? What pre-conditions have to be realized before starting the programme/intervention?

FIGURE 3: OBJECTIVES AND EFFECTS WITHIN THE PROGRAMME CYCLE



Source: European Commission (1999), Evaluation design and management, MEANS Collection Volume 1

3.4. Institutional Aspects

Institutional aspects are important in creating a favourable environment for procuring, conducting and using evaluation results. Some of these aspects are formulated in questions and deliberated below.

■ **Is it necessary to have institutions specialized in evaluations?**

...and on what level and to what purpose?

Examples from different countries show that there are different forms of institutionalized evaluation competencies. These could be specialized departments at the ministerial (or corresponding public administration bodies) level, or specialized more or less independent public evaluation agencies focusing on programmes and institutions; it could be the court of auditors, but it could also be competent private evaluation providers (profit or non-profit). There are pros and cons in relation to any institutional set-up. For instance, one advantage of a centralized institutional set-up (e.g. a specialized department within a ministry or public agency) is that relevant knowledge can be accumulated in-house (provided that a good human resources and knowledge development system is in place). Moreover, there is usually a mission behind such an internal institutional setup which drives the evaluation agenda and which also increases the binding feature. Contrasts for such centralized in-house solutions could include increased bureaucracy, the emergence of senseless evaluation automatisms, or the danger of political influence and lack of other independence. For smaller countries, in-house solutions are rather unlikely, unless enough critical mass for evaluations becomes available over a long period of time. The advantage of private providers is that they can be procured case-by-case on the market without the necessity of establishing a permanent and probably costly in-house unit, which exists even when no evaluations are implemented. The market approach requires, however, that enough local competencies are available which can be addressed. Otherwise, external evaluators from abroad have to be approached who might bring in valuable views from outside, but who may also lack insight into the RTDI system of a country or who are not capable of understanding the language. Moreover, the hiring of external experts might fall under public procurement regulations, the implementation of which requires effort on behalf of the procuring authority. On the other hand, independent and fresh views from experts who are operating outside the core system of STI policy might increase the objectivity and creditability of evaluations.

- **Is there a need to create a legal basis that would make evaluation compulsory?**

The provision of a legal basis enhances the regularity of evaluations, legal enforceability (to varying degrees) and the establishment of adequate structures, as well as contributing in general to the development of an evaluation culture. However, there is a danger that evaluation will be perceived as something imposed and overwhelming. It can also happen that form prevails over substance. In order to overcome the fragmented evaluation practice existing in many South East European Countries and to mainstream RTDI evaluations, it is suggested that legal precautions be introduced at least for interventions over a certain budget level.

- **Should there be strict pre-defined evaluation requirements or ad hoc evaluations for specific purposes?**

In the USA, all government offices have to legitimize their activities —and consequently also their budgets —through rigid evaluation procedures, while in Austria, for instance, evaluations vary, depending on their purpose. This does not, however, mean that in Austria an ad hoc system prevails. In Austrian RTDI funding regulations, evaluations are called for if specific thresholds are passed. Thus, programmes are typically more often evaluated since they have both a time-binding factor and a higher budget allocation.

- **Is evaluation predominantly a steering instrument or an instrument for learning?**

A balance between self-evaluations/voluntarily commissioned external evaluations and truly external evaluations which are commissioned by third parties (e.g. a ministry as programme owner commissions an external evaluation which also assesses the programme management of the mandated agency) should be found. Voluntarily implemented evaluations (either self-evaluations or external evaluations) better foster learning, although the objectivity of self-evaluations might be questioned. External evaluations are more objective, but usually put emphasis on findings more suitable for controlling and steering a programme. In any case, the policy circle should be closed by communicating evaluation findings to all concerned levels (ministries, agencies, other stakeholders) to foster reflection, understanding and learning.

- **Cost–benefit ratio in evaluation:**

Evaluation is a costly process, and therefore funds have to be reserved for this purpose. International practice shows that approximately 1–2% of the budget of

a larger programme (i.e. more than €1.5m) should be spent on an evaluation. If smaller programmes are evaluated, the percentage reserved for the evaluation increases because of critical minimum size effects.

Large scale programmes should be subjected to ex ante, interim and ex post evaluations by external evaluators. While the definition of what constitutes a 'large-scale' programme may vary in time and budget by the country and region where the programme is to take place, a programme which lasts longer than 5 years and whose volume is at least €1,500,000 would be considered 'large scale' in developed countries. Programmes with a smaller budget and of a shorter duration would be subject to fewer evaluations, such as a small ex ante assessment by external experts and a small midterm or terminal evaluation. While formal evaluations are very important in ensuring that programmes are conducted in such a way that they achieve their strategic objectives, light accompanying methods such as workshops with project leaders can also be helpful in establishing a reflective discourse about the progress of programme implementation.

The managing authority mandated by a government should also put a system in place to measure its own internal functioning and efficiency, as well as to collect data about funded projects on a regular basis by employing a monitoring system; this also incurs additional costs.

3.5. Rules and Ethics for Evaluators and Commissioning Institutions

Evaluation has to be based on relevant professional and ethical guidelines and codes of conduct for individual evaluators. Evaluation should be undertaken with integrity and honesty. Commissioners of evaluations (i.e. those who procure evaluations), evaluation managers and evaluators must respect human rights and differences in culture, customs, religious beliefs and practices of all stakeholders. Evaluators should be mindful of gender roles, ethnicity, ability, age, sexual orientation, language and other differences when designing and carrying out an evaluation³⁰.

The purpose of ethical principles and guidelines is to promote³¹:

- responsible behaviour toward all stakeholders, particularly those affected by interventions under evaluation;

³⁰ OECD, 2010: Quality standards for development evaluation, p.6.

³¹ Definitions used according to GEF Evaluation Office Ethical Guidelines (2007),p.2.

- the credibility of evaluation results, through establishing impartiality, inclusiveness and comprehensiveness of the work undertaken by applying appropriate approaches and methods;
- the responsible use of resources.

In planning and carrying out evaluations, evaluators and those commissioning an evaluation must comply with certain rules in order to ensure that the evaluation process can and will lead to a transparent and fair assessment of the RTDI intervention in question. These rules, which can collectively be seen as a kind of code of conduct, primarily refer to³²:

- the competence of evaluators;
- the systematic planning and implementation of evaluations;
- the correctness and credibility of evaluators.

3.5.1. Competence of Evaluators

Evaluators have to be methodologically and professionally competent for particular evaluations and should have a range of expertise derived from technical sciences, social sciences, or evaluation disciplines. In case some fundamental competences for the implementation of an evaluation cannot be covered, evaluators should make clear which content based and/or methodological limits are to be expected in the evaluation and ensure coverage of the missing competences by bringing in third parties³³.

In order to complete evaluations successfully, evaluators should include not only knowledge and skills, but also strategies and routines for applying their knowledge and skills in specific contexts. They should also dispose of inappropriate emotions and attitudes and, above all, possess an effective self-regulation of these competencies³⁴.

An evaluation team's impartiality and independence is essential. It is necessary that evaluators have no conflicts of interest. They have to retain independence of judgment in order to avoid pressure from any party to modify evaluation findings. Evaluators should, however, establish good relationships with clients as well as

32 Definitions used according to Evaluation Standards in Research and Technology Policy Platform – Fteval, Vienna, p.24.

33 Evaluation Standards in Research and Technology Policy Platform – Fteval, Vienna, p. 23

34 Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation*, 26, p. 43–59.

other evaluation stakeholders and protect their rights³⁵. They have to continuously strive for fair judgement, which has to be based on the transparent disclosure of their findings as well as the respective approaches and methods used.

3.5.2. Potential Conflicts

Typical conflicts arising in the practical implementation of an RTDI evaluation are often referred to as the ‘delivery gap’, the ‘customer gap’ and the ‘management gap’³⁶ as summarised in Table 3 below:

TABLE 3: AREAS OF CONFLICT IN RTDI EVALUATIONS

The ‘Delivery Gap’		The ‘Customer Gap’		The ‘Management Gap’	
<i>What policymakers want ...</i>	<i>What evaluators respond ...</i>	<i>What evaluators want ...</i>	<i>What policymakers say ...</i>	<i>What programme managers say ...</i>	<i>What evaluators answer ...</i>
information in time for decision-making	research may take years to have effects	clearly defined hierarchical objectives for each intervention	programmes are a compromise involving multiple and also conflicting interests	I have tons of work and a lot of customer relationships	Fine! Where is the documentation?
clear attribution of effects to investments	linear models are rare and additionality is complex to assess	guaranteed independence	recommendations must be within realistic policy constraints	Look at this nice development!	Where are the social returns?
independent evidence of research excellence	peers usually defend both their subject field and their colleagues	time and resources to do the job	they need results in 3 months!	We collected lots of facts about our projects!	Yes, but not a single number is useful for additionality measuring.
key indicators to monitor and benchmark	too crude regimes distort performance measurement and can be manipulated	full access to information and stakeholders	everyone is overworked and busy		

Source: Boden, M. and Stern, E. (2002): User Perspectives. In: RTD Evaluation Tool Box

35 GEF Evaluation Office Ethical Guidelines (2007),p.5–6.

36 Boden, M. and Stern, E. (2002): User Perspectives. In: RTD Evaluation Tool Box

Besides and sometimes also because of the 'delivery', 'customer' and 'management' gaps, further potential areas of conflict can emerge:

- Evaluators should strive to fulfil the request of the organization which commissions the evaluators, but their work may contradict other (public) interests.
- In practice evaluation means working on research hypotheses, and carrying out evaluations in a systematic manner, which may conflict with the interests of some stakeholders (e.g. policymakers, programme owners, beneficiaries).
- In an evaluation situation it may happen that the competence and credibility of the evaluators comes into question in a conflicting situation.
- Third parties can have legitimate interests, which may conflict with the interest of the organization commissioning the evaluation.
- As evaluation results emerge and solidify during its progress, the potential detrimental consequences to certain parties also become more and more apparent, which can be a source of conflict.
- Lack of appropriate data is common, especially in countries with less developed evaluation regimes, which increases the significance of reliance on qualitative social science research methods; these are sometimes subject to judgement, discourse and criticism.
- Evaluators may have interests of their own in connection with the object of the evaluation and, thus, interpretation of evaluation results by third parties may be different from theirs.

Cases in which the credibility of evaluation results could be questioned should be avoided through objective tendering, transparent communication, constructive conflict resolution or the inclusion of further independent expert(s). Evaluators should not only look for failures (as auditors are supposed to do), but should also appraise good results.

It is strongly recommended that evaluators always work in teams, because group reflection is an important corrective measure. Evaluators are accountable for their evaluation findings and recommendations. Thus, evaluations have to be methodologically sound and should be accurately implemented. Sometimes evaluators are also asked to accompany the implementation of their recommendations. Evaluators should be aware that this is no longer part of an evaluation, but rather organizational development support, etc. They must

decide whether or not they are ready and capable of doing this. If they accept, they change roles and are not anymore free of conflict of interest regarding any subsequent evaluation of the concerned subject.

3.6. Ensuring the Take up of Evaluation Results

Utilization of evaluation results in policy making is a considerable challenge. Two aspects are of high importance. On the one hand, it is important to create supporting structures and conditions to facilitate the utilization of evaluation results. On the other hand, results should be prepared and presented in a way to make them 'easily digestible' for those involved.

Demand for evaluation results could be encouraged by:

- ensuring that decision-makers at higher levels of the hierarchy support the evaluation;
- integrating stakeholders in the evaluation process in order to motivate them to actively acknowledge the evaluation and utilize its results;
- setting realistic expectations.

An appropriate environment for evaluations could be created by:

- including evaluation in the policymaking routine and thus performing evaluations regularly;
- following up on the implementation of results and especially linking evaluation to budget planning;
- taking the needs of the stakeholders into consideration in evaluation objectives;
- ensuring the relevance of results for policymaking and for stakeholders;
- integrating those involved and affected;
- ensuring that the quality of evaluations meet international standards and take into consideration good practices;
- communicating the idea of an evaluation and its findings in an appropriate way;
- making the evaluation process transparent and the evaluation results available to all interested parties.

Besides creating a favourable environment for meaningful commissioning and take up of evaluations, the reporting and communication of evaluation results is of utmost importance.³⁷ This can be improved by taking the following issues into account:

- Findings of evaluations are communicated in due time internally and externally and in a way that meets the specific needs of those affected. In addition, interested parties in the wider community are identified and targeted to maximize use of the findings.
- Conclusions, recommendations and lessons are clear, to the point and well-focused. In addition, recommendations should be realistic within the specific country context.
- If agreed upon by the commissioning authority, results should be communicated by combining a written report with other forms of communication (e.g. stand alone executive summaries, leaflets with an illustrative presentation of the main points, articles in the press or specialized publications, references in the programme or at the managing or funding agency's website).

Furthermore, to increase the take-up of an evaluation's findings, an evaluation report should:

- contain an executive summary, which provides an overview of the report, highlighting the main findings, conclusions, recommendations and any overall lessons;
- describe the context of the intervention which was evaluated, providing policy context, objectives and strategy, socioeconomic context, institutional context and stakeholder involvement;
- describe the intervention logic (e.g. RTDI programme), including underlying assumptions and factors affecting the success of the intervention;
- describe and explain the evaluation methodology and sources of information used and explain any limitations in the methodology, processes, sampling or data used; discuss validity and reliability;
- make explicit any assumption underlying analyses;

³⁷ Some of the recommendations in this section have been based on the OECD (2010), Quality Standards for Development Evaluation, DAC Guidelines and Reference Series.

- make sure that conclusions are substantiated by findings and analyses;
- present findings, conclusions, recommendations and lessons to be learned separately in a distinctive way which allows the reader to follow the logic of the arguments;
- acknowledge any unresolved differences of opinion within the team;
- reflect the comments of stakeholders on the draft report and acknowledge any substantive disagreement.

3.7. Should Evaluation Results be Binding?

Naturally, the results of project evaluations should be as binding as possible, because stop or go decisions are usually attributed to project evaluations. It depends largely on the specific context in the case of programme evaluations, organization evaluations or instrument evaluations and there may be different degrees of binding. This depends on the political culture of a country, the mission of policy interventions, the interests which are assigned to evaluations, the portfolio of available policy instruments and —last but not least —on the robustness of evaluation findings. Regarding the latter, the level of binding is directly connected to the precision of findings, which means that consequences should be more binding if evaluation findings are more definite and well substantiated. Quantifying, clearly defining objectives in advance and reducing unpredictable factors to a minimum increase the value of an evaluation and its suitability as a basis for action and consequences.

4. TENDERING EVALUATION

4.1. How to Proceed in Writing Terms of Reference (ToR)

The terms of reference (ToR) document defines a more or less detailed framework of how evaluators will have to conduct an evaluation. Basically, the stakeholders who initiate an evaluation are the first party responsible for writing the ToR. Writing a ToR for an evaluation is an important and responsible step in the evaluation process, as it defines the objectives and scope of the planned evaluation, outlines the responsibilities of the evaluators, and provides a clear description of the resources available to conduct the evaluation. *“Developing an accurate and well specified ToR is a critical step in managing a high-quality evaluation. The evaluation ToR document serves as the basis for a contractual arrangement with one or more evaluators and sets the parameters against which the success of the assignment can be measured.”*³⁸

4.2. Components of ToR at a Glance

A few basic principles and guidelines are presented here for ToR development, through specific content and format will vary to some degree based on organizational requirements or, for instance, the type of assignment. The components which should be included in the Terms of Reference³⁹ are the following:

1. **Background and rationale:** *The opening section of the ToR typically provides an orientation about the overall programme, project, or another intervention to be evaluated.*

38 The World Bank (2011): Writing Terms of Reference for an evaluation: a how to guide. p. 2.

39 The components of the ToR are following definitions and structure provided in: The World Bank (2011): Writing Terms of Reference for an evaluation: a how to guide. p. 2.

2. **Specific objectives of the evaluation and evaluation questions:** *The framing and presentation of evaluation objectives is usually a brief but important section of any ToR. Common understanding of, and consensus around, stated objectives and evaluation questions will be important throughout the negotiation and implementation of the assigned tasks. Evaluation objectives might relate to predetermined criteria, such as relevance, effectiveness, efficiency, impact or sustainability or other —often more horizontal —features, such as gender issues.*
3. **Scope of the evaluation:** *This section presents the scope and limits of the evaluation. The scope should be realistic given the time and resources available for implementing the evaluation study.*
4. **Approach and methodology:** *Specifying the approach for an evaluation can be the most challenging part of developing a ToR. This section should outline how the evaluation will be conducted. However, many ToRs leave room for the evaluator(s) to define a more detailed methodology in line with the prescribed scope and objectives, which is recommended.*
5. **Governance and accountability:** *This section of the ToR specifies the governance and management arrangements for carrying out the evaluation. Any decision-making arrangements (such as a steering committee or an advisory group) should be described here in terms of their organization and function(s). Participation of other stakeholders (for example, beneficiary representatives in validating results) and the lines of accountability should also be noted with, at minimum, clear guidance on who will review and approve the evaluation plan and subsequent products of the evaluation (e.g. inception report, draft report and final report).*
6. **Guiding principles and values:** *ToR specifies research ethics or procedures that evaluators should follow. These might include fundamental principles of the organization commissioning the task(s), basic tenets that should guide the study (for example, transparency, cost-effectiveness, collaboration with beneficiaries, hiring of local consultants, involvement of local agencies, etc.) or practices expected to be taken into account by the evaluators (for example, confidentiality of data, anonymity of responses, making data publicly available in a usable format, and so on).*
7. **Professional competencies:** *The mix of requested knowledge, skills, and experience will depend on the scope and methodology of the evaluation. The*

ToR should specify as clearly as possible what the profile of the evaluator or team should be to attract the strongest candidates for conducting the study.

8. **Deliverables and schedule:** *The outputs and reporting requirements expected for an evaluation should be specified, along with the required or proposed timeline for a study. Clear guidance in this section will help ensuring that the outputs from an evaluation meet expectations.*
9. **Budget and payment:** *The commissioner of an evaluation should consider what funds are available to support the tasks envisioned for the evaluators. In cases where a limited budget will likely constrain the scope and methodology of a study, an effective practice is to state the available budget and ask proposers to describe what they expect to achieve. This allows for value-for-money assessments.*
10. **Structure of the proposal and submission guidelines:** *ToRs can be used to request proposals from potential evaluators as part of a competitive bidding process. In this case, a ToR should provide instructions regarding the proposal format, content, and submission process.*
11. **Additional references or resources:** *A high-quality evaluation will draw on existing knowledge regarding relevant previous and ongoing studies and programme experiences. To the extent possible, a ToR should identify useful information sources for the evaluator to better ensure that this body of knowledge is taken up in planning and conducting the evaluation.*

5. RECOMMENDATIONS FOR AN EVALUATION ROAD MAP FOR SEE COUNTRIES

Research, Technological Development and Innovation (RTDI) evaluations emerged in the late 1960s in the Anglo-Saxon policy environment and proliferated rapidly; RTDI became the cornerstone of competitiveness policies. At the same time, New Public Management rules called for improved public policy efficiency. The creation and adoption of evaluation standards is a policy tool aimed at raising awareness and developing a good balance between processes, timing, number, quality and types of evaluations. In countries where the evaluation culture is still emerging, standards can contribute in helping stakeholders agree on priorities and establish a road map leading to a high quality national RTDI evaluation policy. *In that sense, each South East European (SEE) country should adopt RTDI evaluation standards to guide this process in an aligned manner.*

Evaluations may address organizations, programmes, instruments/measures or projects, but regardless of their type they are conducted in the spirit of continuously improving RTDI policies; not only their individual components but also their combined portfolios and entire systems. *For this reason, it is important that all stakeholders agree on certain basic standards regarding the transparency, timeliness and frequency of evaluations.*

The objective of public research and development (R&D) policies is to increase the quality and quantity of R&D, leverage business-financed R&D, create conditions for R&D cooperation and create research-efficient public research infrastructures. Innovation policies go beyond this into the launching of new (R&D based or non-R&D based) products and services into the market. Clearly stated objectives for each competitive or institutional funding regime, programmes or measures are essential cornerstones for designing and implementing good evaluations. In many countries with a limited demand for RTDI policies from the business sector, top-down policy design tries to compensate for a lack of experience by copying best practices from abroad. While this is a practical way to facilitate policy

development, there is a risk that such ideas are copied but not adapted. *In order to avoid inappropriate imitation, programme owners and policymakers have to be committed and get used to very clearly stating the objectives of each measure and/or portfolio of measures they adopt or create and ideally integrate performance indicators into a programme or measure from the very beginning, so that evaluations can focus on assessing the achievement of specific objectives.*

Evaluations have many functions and when an evaluation is launched it is important to know the primary function it should serve (legitimizing, information provision, learning, steering, controlling or mediating). Evaluations can be conducted for individual measures or programmes, portfolios of programmes, entire policies or a whole system of innovation (at regional, national or sectoral levels). They can also address individual organizations (research institutes, universities, agencies, intermediary organizations). In terms of timing, evaluations can take place prior to implementation (ex ante), during implementation (interim or real time), at the end of implementation (terminal), a shorter or longer time after the end of implementation (ex post), or regularly throughout implementation. Ad hoc evaluations are not foreseen during the development or implementation of a strategy, programme or piece of legislation, but are conducted to meet a need that emerged later. *By and large, as evaluations are costly, countries that need to invest in RTDI evaluations are expected to start by systematically evaluating all their programmes either in real time or ex post. However, these evaluations need to be timely in order to feed into the next policy cycle.* At the same time, it is expected that all universities will be evaluated in the context of the Bologna process.

Once the first evaluations have become routine (in some cases even mandatory, although this is not recommended for SEE countries at their current stage) and all stakeholders participate and feel comfortable with them, the next step is to organize portfolio and system evaluations in regular time intervals (e.g. 5 years). Ad hoc evaluations will emerge automatically, once policies increase ambitions.

The standards at hand include a series of relevant topics, discussing the content of evaluations and criteria for evaluations in terms of relevance, efficiency, effectiveness, impact and sustainability. Stakeholders need to be aware of them and what they can offer.

These standards offer a guideline for programme owners on how to plan, tender and conduct evaluations. Good planning starts with a good understanding of the objectives and their likely achievement over time. The next step is to agree on potential evaluators. In general, for larger programmes external evaluations are more reliable, as they are more objective and are not associated with agency problems and individual interests. Smaller interventions can be evaluated

internally. *It is important to have experienced evaluators in a country (public, private non-profit and profit-oriented service providers), but as some SEE countries do not possess a sufficient base of professional evaluators to ensure competitive tendering, external evaluations can be tendered internationally or domestic evaluations can be required to include experienced evaluators from abroad. In this case, a by-product of an evaluation would be a learning process for national professionals. By obtaining the necessary skills they can form the basis of a future national supply.*

As the market develops, it is important to adopt clear rules and ethics for commissioning institutions and evaluators to ensure responsible behaviour, credibility of results and cost-effectiveness. Evaluators need to be independent, credible and impartial. These rules are well defined in existing standards and it is important to adopt them immediately in each country so that an evaluation culture can be created on solid ground.

Standards help commissioning institutions in particular to prepare evaluations adequately and make the best use of them. Present standards include recommendations on how to design an evaluation framework before launching a tender and describe the basic elements of the Terms of Reference. *When programme owners first launch calls to procure external evaluation services they can comply strictly with standardized recommendations, but may create a more complex Terms of Reference for composite evaluations once they gain experience.* A classical trap to avoid when tendering evaluations is over-specification: While the tendering authority needs to know exactly what result is expected, it is important to leave certain degrees of freedom to the evaluator concerning the methodologies to be applied. Over-specification facilitates the comparison of proposals but diminishes the variety of approaches that may enrich the information content of an evaluation.

In short, the idea of this document on standards for the SEE countries includes good practices extracted from countries which already have a developed evaluation culture and proposes a road map as follows:

1. Adopt standards (those suggested verbatim or an adapted variation) agreed-upon by all relevant national stakeholders.
2. Start with a commitment to regularly evaluate larger programmes and public organizations (incl. universities) by external evaluators. Three to four years might be needed for this first stage, in which programmes will set out clear objectives and a budget earmarked for evaluation ranging from 1–2% of their total funds (depending on the size of the programme).

3. During this process, commissioning organizations will gain experience, evaluators will be trained and a market for evaluations will be created.
4. Programme owners will, based on their own experience, increase their ambitions for RTDI policies by tendering more complex evaluations (portfolio and systemic), whereas national public, private non-profit and profit-oriented units will emerge to cover market demand.

6. REFERENCES

European Commission (1999): Evaluation design and management, MEANS Collection Volume 1. Luxembourg.

European Commission (1999): Evaluating socio-economic programmes: Technical solutions for evaluation within a partnership framework, MEANS collection Volume 4. Luxembourg.

European Commission (2002): RTD Evaluation Toolbox – Assessing the Socio-Economic Impact of RTD-Policies, IPTS. Technical Report Series, EUR 20382 EN.

European Commission (2009): EVALSED Guide: The Resource for the Evaluation of Socio-Economic Development.

FTEVAL (2003): Evaluation Standards in Research and Technology Policy. Vienna.

Global Environment Facility (2008): GEF Evaluation Office Ethical Guidelines. Washington, D.C.

Independent Evaluation Group, World Bank (2011): Writing Terms of Reference for an evaluation: a how to guide. Washington, D.C.

INNO-Appraisal (2010): Understanding Evaluation of Innovation Policy in Europe. Final Report. Pro Inno Europe.

Kuhlmann, S. (2001): Future Governance of Innovation Policy in Europe – Three Scenarios. In: Research Policy, Special Issue “Innovation Policy in Europe and the

US: New Policies in New Institutions”, edited by Hans K. Klein, Stefan Kuhlmann, and Philip Shapira, vol. 30, 6/2001, 953-976.

Molas-Gallart, J. and Davies, A. (2006): Toward theory-led evaluation: the experience of European science, technology and innovation policies. In: American Journal of Evaluation, 27 (1), 64-82.

OECD (2006): Government R&D Funding and Company Behaviour – Measuring behavioural additionality. OECD, Paris.

OECD (2010): DAC’s Glossary of Key Terms in Evaluation and Results Based Management. Paris.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. In: American Journal of Evaluation, 26, p. 43-59.

UNICEF (1991): A UNICEF Guide for Monitoring and Evaluation: Making a Difference? Evaluation Office, New York.

US Government Accountability Office (GAO) (2012): Designing Evaluations’: Applied Research and Methods, Washington, D.C.

ANNEX 1: EVALUATION METHODOLOGIES AND TECHNIQUES

TABLE 4: SELECTED EVALUATION METHODOLOGIES AND TECHNIQUES

Methodology	Data requirements	Strengths	Limitations
Foresight/ technology assessment	Qualitative data scenario	<p>Consensus building to reduce uncertainty under different scenarios</p> <p>Combination of public domain and private domain data</p> <p>Articulation and road mapping of new technology development</p>	Impossible to detect major RTD breakthroughs
Cost-efficiency	Micro data profit & cost estimates	<p>Provides an estimate of socioeconomic effects of intervention</p> <p>Good approach to assess the efficiency of an intervention</p> <p>Outputs are usually in non-monetary terms</p>	<p>Requires high technical capacity</p> <p>Some degree of judgement and subjectivity, depends largely on assumptions made</p>

Methodology	Data requirements	Strengths	Limitations
Cost-benefit analysis	Micro data profit & cost estimates	Provides an estimate of socioeconomic effects of intervention Good approach to assess the efficiency of an intervention Addresses by making explicit all the economic assumptions of intervention impact	Requires high technical capacity Some degree of judgement and subjectivity, depends largely on assumptions made Not easily comparable across cases Careful interpretation of results when benefits are not easily quantifiable in monetary terms
Innovation surveys	Micro data expenditures profits patents, innovation	Detect innovation trends and insights on the soft side of innovation Findings from interviewed sample can be generalized to the population Permits identification of size and distribution of impacts Provides group comparisons and changes over time	High cost and time consuming Processing and analysis of data requires large amount of human resources Some types of information are difficult to obtain Long time series generally not available
Benchmarking	Science and technology indicators	Comparison method across different sectors Support to systemic evaluation of institutions and systems	Data detail requirements Non transferable
Macroeconomic modelling and simulation	R&D expenditures R&D output macroeconomic data	Social rate of return to R&D Capture R&D spillovers Estimate long term policy intervention impact Scenario simulations for policy supported geographical areas	Average returns Robustness of results Time lags for observation of effects

Methodology	Data requirements	Strengths	Limitations
Productivity analysis	Micro data expenditures profits R&D, patents	Estimation of effect of R&D on productivity Estimate the rate of return to R&D	Quality of data Deflation of series Required assumptions for measurement of stock variables
Microeconomic evaluation studies	Micro data expenditures profits patents	Results based on explicit formulation of theory based causal relationships R&D additionality Control for different effects: firm size, expenditures, innovation capacity	Quality of data Persuade participant and nonparticipant entities to disclose information Only private rate of return to R&D
Control group approaches	Micro data expenditures profits patents	Capture the impact of policy intervention on the programme participant entity	Requires high technical capacity High implementation cost High data demand
Interviews and field/case studies	Project programme data	Observation of the socio-economic impacts of intervention under naturalistic conditions Good as exploratory and descriptive means of investigation Good for understanding how contexts affect and shape impacts	Results not generalizable
Expert panels/peer review	Project programme data	Evaluation of scientific merits flexibility Wide scope of application Fairness	Peer independence Economic benefits not captured
Network analysis	Project programme data	Comprehensive empirical material Compilation for policy purposes Cooperation linkages	Time involved in collecting survey information Persuasion requirements

ANNEX 2: NAMES OF AUTHORS

Balázs BORSI, policy advisor, professor, external expert for the Public Foundation for the Development of Industry (IFKA)

Zoya DAMIANOVA, Applied Research and Communication Fund (ARC Fund)

Katharina DEMEL, Consultant and Evaluation Expert, Austria

Martin Felix GAJDUSEK, Centre for Social Innovation (ZSI)

Robert HICKEY, Applied Research and Communication Fund (ARC Fund)

Tatjana KNEŽEVIĆ, University of Montenegro – International Relations Office (UM-IRO)

Đuro KUTLAČA, Mihajlo Pupin Institute (MPI)

Daniela MINEVA, Applied Research and Communication Fund (ARC Fund)

Ines MARINKOVIĆ, Centre for Social Innovation (ZSI)

Nikos MAROULIS, Consultant and Evaluation Expert, Greece

Marija MOŠUROVIĆ – RUŽIČIĆ; Mihajlo Pupin Institute (MPI)

Klaus SCHUCH, Centre for Social Innovation (ZSI)

Dušica SEMENČENKO, Mihajlo Pupin Institute (MPI)

Nikos SIDIROPOULOS, National and Kapodistrian University of Athens – Center of Financial Studies (NKUA-CFS)

Lena TSIPOURI, National and Kapodistrian University of Athens – Center of Financial Studies (NKUA-CFS)

Mira VUKČEVIĆ, University of Montenegro – International Relations Office (UM-IRO)

ABSTRACT

The evaluation standards presented here have been developed within the project “*Fostering Evaluation Competencies in Research, Technology and Innovation in the SEE Region (EVAL-INNO)*”. The major objective of EVAL-INNO is to promote the role of RTDI evaluations as an important element of strategic policy intelligence for the improvement of national research and innovation systems.

In particular, the project aims to:

- promote the role of RTDI evaluation as a crucial precondition for a reflexive learning research and innovation system;
- develop needed capacities and competencies for comprehensive RTDI evaluations;
- provide procedural and methodological know-how and tool kits both on the side of evaluators and on the side of awarding authorities to facilitate the use of RTDI evaluations.

The activities aimed at building up RTDI evaluation intelligence in the South East Europe region include: training activities, workshops and conferences, the establishment of exchange platforms and databases, the implementation of pilot evaluations and —last but not least —the publication of evaluation standards.

Professional evaluation standards are crucial to guide upcoming RTDI evaluations in South East Europe at an international state-of-the-art level.

They address:

- authorities commissioning RTDI evaluations;
- evaluators carrying out RTDI evaluation studies;
- organizations and stakeholders (their programmes and measures respectively) subjected to evaluations.

The standards inform about the purposes and characteristics of evaluations in the field of science and technology, provide users with an internationally acknowledged terminology and evaluation theory framework, guide users in practical issues concerning governance, conduct and use of RTDI evaluations and offer practical hints on how to plan and implement evaluations. The present version of evaluation standards was developed by a team of experts from Austria, Bulgaria, Greece, Hungary, Montenegro and Serbia.

You are cordially invited to use, review and adopt these standards as a policy tool for assessing and improving research, technological development and innovation in your country. We invite distinguished readers and stakeholders to provide feedback and comments in order to publish a new and improved issue of these evaluation standards, based on a larger consultative process, at the end of the project.



RTDI Evaluation Standards

The evaluation standards presented here have been developed within the project *“Fostering Evaluation Competencies in Research, Technology and Innovation in the SEE Region (EVAL-INNO)”*. The major objective of EVAL-INNO is to promote the role of RTDI evaluations as an important element of strategic policy intelligence for the improvement of national research and innovation systems. The present version of evaluation standards was developed by a team of experts from Austria, Bulgaria, Greece, Hungary, Montenegro and Serbia. The standards are published in English, Bulgarian, Greek, Hungarian, Montenegrin and Serbian language.

We invite distinguished readers and stakeholders to provide feedback and comments in order to publish a new and improved issue of these evaluation standards, based on a larger consultative process, at the end of the project.