technopolis |group|

14 January 2014

Evaluation Reference Model

For TAFTIE's Taskforce Benchmarking Impact, Effectiveness and Efficiency of Innovation Instruments

technopolis |group|

Evaluation Reference Model

In Search for a Benchmark of Impact, Effectiveness and Efficiency of Innovation Instruments

technopolis |group|, January 2014

Patries Boekholt (project leader)

Erik Arnold

Flora Giarracca

Matthias Ploeg

Table of Contents

Table of Figures

1. Introduction

1.1 The background for this Evaluation Reference Model

The TAFTIE Task Force on Benchmarking Impact, Effectiveness and Efficiency (TFBIEE) was set up to gain a combined insight in innovation agencies' impact, effectiveness and efficiency.[1] In order to fulfil TAFTIE's ambition to demonstrate the added value of innovation agencies to society, it is important to understand and demonstrate how the policies these agencies implement have an impact. The project has been conducted in close cooperation between the participating innovation agencies and Technopolis to emphasise the main aim of this project: learning at agency level. This study was a first experiment to assess to what extent it is possible to compare effectiveness between specific types of innovation instruments across different countries. At the outset it was clear that the design of the instruments and the dissimilar contexts in which agencies and instruments operate would make this an almost impossible task. In an ideal case we could compare the claims made on impacts as reported in evaluation studies.

The project's aim was to address three research questions:

- Is it possible to benchmark impact/effectiveness of policy instruments in an internationally comparable way?
- How do we estimate the impact/effectiveness of these instruments? Are we in line with (inter)national handbooks and guidelines in this matter?
- Is it possible to benchmark innovation agencies on key figures of the implementation process?

This Evaluation Reference Model is compiled as one building block of this exercise. It has been written to function as a stand-alone document for future use. The actual benchmarks can be found in a separate report with the title "*In Search for a Benchmark of Impact, Effectiveness and Efficiency of Innovation Instruments"* also a report for TAFTIE.[2]

In order to answer the first research question - whether it is possible to compare effectiveness of similar instruments - a necessary preceding step was to establish whether an objective and robust assessment of the impacts was made for these instruments. This is commonly done through evaluation studies. In order to reduce the influence of policy design and rationales on the comparisons, four categories of instruments were selected that were compared within their class (see below).

Thus the second research question was to review whether the impact assessments are conducted according to international 'good practice' and in line with international handbooks and guidelines. We accomplished this by synthesising the existing (generic) guidelines for good evaluations, combined this with other evaluation experience from the innovation field and translated this into a practical toolkit: the Evaluation Reference Model. We have adapted the generic guidelines to the four specific types of instruments. The Reference Model is to contribute to setting a TAFTIE reference model for ex-ante, midterm and ex-post evaluations for future policy learning.

The benchmark of 28 instruments showed that the key evaluation issues that would need to be addressed by the agencies are:

- Devote more efforts to defining the programme objectives and goals in a sufficiently concrete manner in order to render the evaluation of achievements and success more in

[1] TAFTIE is the European Association of leading national innovation agencies (see.taftie.org)

[2] Technopolis, 2013, In Search for a Benchmark of Impact, Effectiveness and Efficiency of Innovation Instruments, Amsterdam

line with the programme logic model. Develop an evaluation framework already at the start of the programme

- Define the data requirements for an evaluation at the start of the programme and develop baseline and monitoring data that would be needed for later evaluations
- Adapt the set of evaluation questions to the timing of the evaluation and be realistic about the time line at which one can expect to measure impacts at company level and as spill-overs to the rest of the economy and society
- Work on the access to appropriate micro-level data that can be linked to the instruments' population as well as any statistically constructed control group. If confidentiality is an issue, collaborations with national statistics offices have been successfully developed by a number of the TAFTIE agencies
- Insist from your evaluation team a methodology mix that is fit for purpose and allows for sufficient triangulation between the methods used. Insist also on where possible in view of the data availability and the instrument's target group on some form of counterfactual analysis for the assessment of impacts
- Insist from the evaluation that the conclusions can be easily derived from the evidence gathered and that all the methods used are transparent

1.2 The purpose of the Evaluation Reference Model

This Evaluation Reference Model intends to sketch state-of-the art methodologies and best use of evaluation tools for four types of innovation instruments:

- Business R&D grants
- Innovation vouchers
- Collaborative R&D grants
- Integrated cluster programmes (including competence centre programmes)

The choice of these innovation instruments on behalf of the innovation agencies slightly limits the general applicability of the reference model. The focus of the attention has been on innovation and grasping the socio-economic benefits from programmes that aim to improve the competitiveness of firms. Thus the reference model pays relatively little attention to the many evaluation approaches and techniques used to evaluate scientific and technological impacts.

Figure 1 Definition: Evaluation methodology and evaluation tool

DEFINITION
An **Evaluation methodology** is an ad hoc procedure specially constructed for a particular evaluation. It may include one or more tools.
An **Evaluation tool or a technique** is something that is used to carry out a standard treatment during an evaluation, for example: Interviews, Surveys, Case studies, Cost-effectiveness analysis, Benchmarking analysis.

Identifying a single state-of-the-art methodology for the evaluation R&D instruments is however challenging. Evaluation methodologies are always dependent upon factors that are external to methodological considerations, such as programme characteristics, scale, objectives and the reality of implementation.

Ten evaluation handbooks (presented in the table below) were used as a basis to develop this Reference Model. While some of the handbooks unpack generic evaluation guidelines, others are tailored for specific support measures such as enterprise support. Two of them focus of

impact assessments (e.g. Tekes and VINNOVA). While some focus exclusively on quantitative evaluation techniques (e.g. NL Dare to measure, DK CIM) and one only on qualitative techniques (UK Quality in qualitative evaluation), most handbooks address a combination of methods. The ERM draws on their diversity and complementary (instrument-wise and methodology-wise) in order to build a comprehensive framework. In addition we have used literature on specific evaluation issues.

Figure 2 Guidelines and handbooks used for this Reference Model

Country	Organisation	Name of the evaluation handbook
European Commission	JRC-IPTS and Joanneum Research	RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies (2002)
European Commission, DG Enterprise	Louis Legrand et associés	Smart innovation: A Practical Guide to Evaluating Innovation Programmes (2006)
European Commission, DG Regional Policy	Technopolis	Evaluation of Innovation Activities Guidance on methods and practices (2012)
Denmark	Danish Ministry of Science, Innovation and Higher Education	Central innovation manual on excellent econometric impact analyses of innovation policy (CIM) (2012)
Finland	Tekes	Better results, more value. A framework for analysing the societal impact of research and Innovation (2011)
Netherlands	Impact Evaluation Expert Working Group	Dare to measure: Evaluation designs for industrial policy in the Netherlands (2012)
Norway	NIFU (Ramberg I., Knell M.)	Challenges measuring effects of research and innovation policy interventions (2012)
Sweden	VINNOVA	VINNOVA's Focus on Impact, A Joint Approach for Logic Assessment, Monitoring, Evaluation and Impact Analysis (2008)
UK	HM Treasury	The Magenta Book. Guidance on evaluation (2011)
UK	Cabinet Office	Quality in qualitative evaluation: A framework for assessing research evidence (2003)
UK	Policy Research in Engineering Science and Technology PREST, University of Manchester	Assessing the Socio-economic Impacts of the Framework Programme (2002)

1.3 How to read and use the TAFTIE Reference Model

Methodology is an important but not sufficient criteria to assess the quality of evaluations. State-of-the-art evaluations are based on wider considerations that impact the methodology design and implementation of evaluation tools: utility of the evaluation for intended users, involvement of stakeholders in the evaluation, clarity and transparency in the analysis to name a few. This is why this Evaluation Reference Model presents a **process approach to evaluation** and is structured around six key steps:

- **Step 1**: Definition of the Programme Logic Model
- **Step 2**: Definition of the evaluation objectives and questions
- **Step 3**: Preparation of the evaluation
- **Step 4**: Identification of appropriate methodology for analysis (state-of-the-art methodology mix by type of instrument) and execution of the analysis
- **Step 5**: Conclusions and reporting of the evaluation
- **Step 6**: Apply the lessons from the evaluation

For each of these steps **key quality criteria** are identified that constitute the basis for best evaluation practices. A large part of **Step 1** considers actions that need to be performed before any evaluation takes place, preferably at the design phase of a programme. The activities described in this step are for programme owners and managers to perform rather than for evaluators. However if they are done well it will increase the quality and effectiveness of the evaluation. The same holds true for **Step 6**, which contains activities that need to be performed before and after the evaluation study is taking place and within the organisations that commission the evaluations. The activities in Step 6 will influence the effectiveness of the evaluations more than the quality of the studies as such. It is outside the scope of this exercise to elaborate in detail on organisational learning and political agenda setting related to evaluations.

Further details on the use of specific evaluation tools are available in 0. A **glossary** of the main evaluation terms used in this document is available in Appendix C.

For simplicity purpose, this Reference Model uses extensively the concept of 'programme evaluation', which is understood here in a broader sense and covers any type of programme, policy, instrument or intervention. We also refer mainly to national - and not European or regional/local – programmes, as most of the evaluations that were submitted to the benchmark cover national initiatives.

2. Evaluation reference model

2.1 Step 1: Definition of the Programme Logic Model

Criteria for good evaluation practice	Description
• The rationale of the programme is clearly outlined and detail which market failures and or failures in the national RDI system are addressed by the programme	This first step should be made before long before an evaluation takes place, already at the design phase of a programme. The step needs to be accomplished by the programme owners and managers. In the case of complex programmes this could be done in close consultation with the stakeholders to align the programme with the needs of the beneficiaries. For evaluators the task is mostly to reconstruct this Programme Logic Model on the basis of written material and interviews. Thus the more clearly this rationale is documented the more accurate the evaluators can reconstruct this in mid-term and ex post evaluations. This step includes explaining why the policy/programme was launched and what problems or needs it addresses in the national RDI system. Beyond social justice and equity, public support to RDI activities is justified in economic theory by the existence of market failures leading to under-investment in research and innovation compared to what would be socially optimal. The rationale for intervention is therefore to rectify these market failures: • Provide investment where science is a public or non-rival good • Reduce uncertainty and risk, caused by the high failure rate in RDI, e.g. collaborative projects are often considered as risky because they require increased caution in handling intellectual property and can create 'transaction costs' as well as disrupt 'normal' business • Address imperfect information leading to difficulty in assessing the likely success costs and benefits of an RDI venture, e.g. the rationale for innovation vouchers is that SMEs, contrary to big companies, do not have enough experience to make informed choice and invest in RDI • Address indivisibility, e.g. the large investments required might in some cases prevent enterprises of investing in R&D • Address positive externalities or spillovers, e.g. when a company that invests in innovation is unable to capture the full returns as it cannot stop other firms from copying or further developing the technology. • Problems of appropriability through knowledge, market and network externalities, e.g. with enterprises foregoing RDI because they will still benefit from RDI without having to invest in it. Another rationale for public intervention is the existence of systems failure in the national RDI system[3]: • Failures in the science and technology infrastructure (e.g. universities, research labs, national assets such as R&D staff). • Failures in formal or informal institutions that constrain innovation activity (e.g. legal systems, political culture) • Interaction failures between RDI stakeholders (e.g. lack of relations between RDI stakeholders) • Transition Failures, when firms get locked-in into technology paradigm and are unable to adapt new technology development • Capability and learning failures, when failures in competencies and resources (technological, organisational, restrict the firm's ability to learn and be innovative

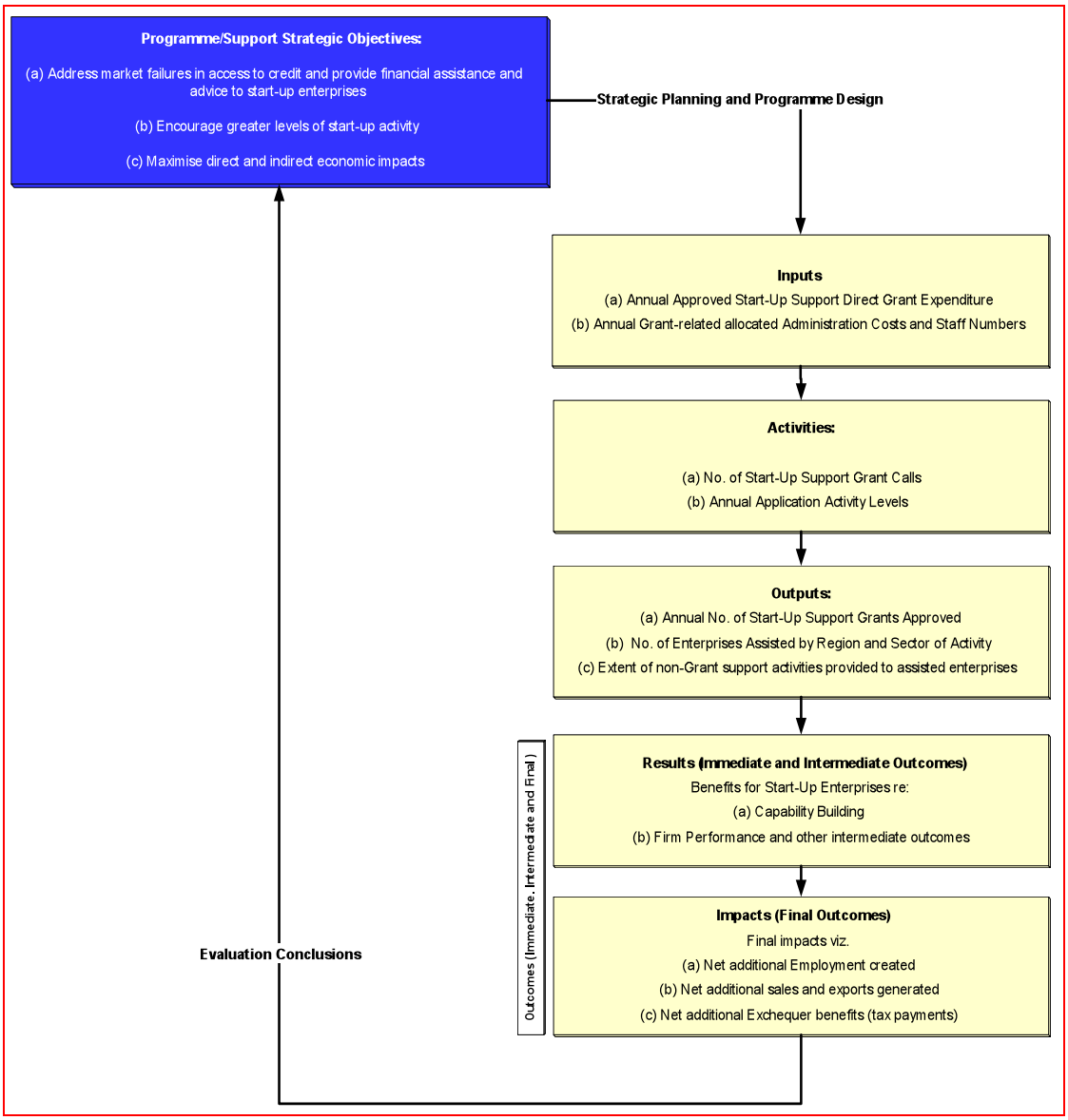
[3] UK Department of Business, Innovation and Skills, Occasional Paper no. 2 (August 2010), The economic rational for national design policy

Criteria for good evaluation practice	Description
• The objectives of the programme/ policy are clear, specific and clear targets are identified. • Where relevant, objectives are prioritised.	Ideally, this should be done at the programme design stage as part of an ongoing process of policy and programme design learning through evaluation. Best evaluation practices are built into programmes at the design stage, both for accountability and facilitation of future evaluation processes. Overarching policy objectives affect many aspects of the design and implementation of RDI programmes, and the evaluation questions by which their success may be judged. It is important for the evaluation that they are clear and specific. If not developed at the programme design stage, the list of objectives (and their prioritisation) should be agreed upon by key programme stakeholders, through interviews or focus groups, particularly for complex programmes such as integrated cluster programmes. In Step 3 key stakeholders can be involved again when preparing the evaluation study. Policy objectives correspond to how the specific measure intends to address market and system failures. For instance if an intervention aims to overcome the lagging investment in innovation by SMEs the objective of an instrument would be to encourage these SMEs to invest in innovation by means of a grant, in order for them to become more competitive and increase their future sales and profitability.
• The programme objectives are specified within the wider RDI context.	Programmes and policies are never standalone initiatives, but are part of a wider policy-mix of interventions aimed at supporting national RDI. Hence, evaluating a programme or policy requires to review the economic, social and political context in which he initiative is taken place: • Has the context of implementation changed since the start of the programme (or the last evaluation)? • What national/ regional/European strategies are the programme objectives aligned to? • What other policies are likely to impact the implementation and results of the programme? (e.g. other RDI support measures, or alternatively policies related to industry, higher education, skills). Refining the context can help refining the evaluation scope. In some case it might prove relevant to carry out joint evaluation of portfolio instruments that have a comparable policy objective and targets. Alternatively, past evaluations of alternative instruments can input the evaluation and the evaluator understanding of the situation. It is also more common to conduct ex ante impact assessments of planned instruments to assess the likelihood of impacts in different scenarios.[4]
• The evaluation framework is based on a Programme Logic Model (schematic representation of the programme logic).	A Programme Logic Model (PLM) lists the programme inputs, outputs, outcomes and impacts based on the programme objectives and rationale. It outlines how the resources deployed as part of the programme (inputs) are intended to produce the expected outputs, outcomes and impacts. • The inputs (i.e. financial, human and other resources employed in the delivery of the programme/support), outputs (i.e. direct results of the work enabled by the inputs), outcomes (the immediate benefits for beneficiaries of the support) and expected impacts (i.e. wider social and economic impacts of the programme). • Assumptions are made about how these elements link together which will enable the programme to successfully progress from inputs to outputs, to outcomes and impacts. • An assessment of important external effects that could have an impact on the programme objectives At this stage, assumptions about potential impacts over and above those directly intended can also be made, in order to inform the logic model. ***The framework to develop Programme Logic Models is presented in Figure 3.***
• Performance metrics are defined, based on the programme logic model, to measure inputs, outputs and outcomes.	The more precise the objectives and targets are defined the more precise the evaluation can report on the achievements of the programme. Programmes with fuzzy objectives usually result in fuzzy evaluations as it becomes difficult to conclude whether a programme is successful. Performance metrics should be SMART: • Specific, i.e. the chosen indicator is well defined and relates clearly to the specific aspect of the programme/support under which performance is being assessed; • Measurable, i.e. the chosen indicator can facilitate the measurement of progress towards achievement of programme/support objectives/goals; • Attainable, i.e. the data supporting the chosen indicator should be attainable or achievable in a cost-effective manner; (in some definitions the A stands for Acceptable i.e. the indicators are supported by programme owners and key stakeholders)

[4] See for instance for ex ante impact assessments http://ec.europa.eu/governance/impact/index_en.htm

Criteria for good evaluation practice	Description
	• Relevant, i.e. the chosen indicator should be relevant to the objectives of the support/programme being evaluated; and • Timely, i.e. measurement on the chosen indicator should be available in a timely manner. It is good practice to include Key Performance Indicators (KPIs) in the design of the programme, although their use is not without challenges (e.g. stimulating strategic behaviour). VINNOVA for example has developed an integrated approach: an impact logic assessment is designed and tested ex ante, together with indicators as a basis for determining whether the programme and its projects are progressing towards their impact goals. Impact logic is thus used to design initiatives and processes leading ultimately to different results and impacts, including monitoring of whether these results and impacts are being achieved.
• Baseline data necessary for future evaluation are defined and generated	In addition to programme monitoring data, the evaluation manager should identify all relevant data (e.g. databases, statistics, programme documents, past evaluation and studies) and key sources of data that will input the future evaluation evidence. If data are not available or accessible, effort should be made to generate those data or ask relevant organisations (such as national statistical offices) to prepare these datasets. A good practice is to systematically collect data on the baseline situation (e.g. R&D performance and investment in firms prior they receive a business grant) as part of the programme monitoring data, which can then be used to study any differences in performance and behaviour.

Figure 3 Example of Programme Logic model



Source: Indecon hypothetical example based on review of international best practice approaches.

2.2 Step 2: Definition of the evaluation objectives and questions

2.2.1 Criteria 1: Evaluation objectives

Criteria for good evaluation practice	**Description**
• The evaluation objectives are clearly stated, including expectations in terms of integration into policy-making • A discussion on the challenges, risks and limitations of the evaluation in meeting its objectives is included in the evaluation plan.	Define who the target end-users of the evaluation will be (programme managers, policy makers and analysts; other national/ regional/ local policy-making bodies; key stakeholders including industry bodies, HEIs, RIs, the public, local community groups and other interested parties) and what are the different expectations for how the results will be used, including any expectations on the timing of when the evaluation evidence might feed into decision making. Planners and policy makers in particular should be clear in relation to the purpose and objectives for the evaluation.
• The intended use of the evaluation (formative/ summative) is explained in further details	Evaluations can take place at any time – looking forward or looking back – and with different purposes – to guide an activity as it happens (known as formative evaluation) or to assess its effects in retrospect (known as summative evaluation). Summative and formative evaluations focus on different types of questions and, hence, different evaluation tools. It is therefore of prime importance to define the use of the evaluation from the start[5]: • Summative evaluation: 'summarises' the outcomes and impacts at a particular time in the programme life, with a view to judging and decide future resources allocation. Summative evaluation asks questions about the *impact* of a policy, programme or intervention on specific outcomes and for different groups of people. It looks back at achievements and is aimed at accountability. Hence, ex-post evaluations often tend to be summative. • Formative: takes place during the life of the programme, with a view to improving management and implementation (i.e. 'form' the programme). Formative evaluation asks *how*, *why*, and *under what conditions* something works, or fails to work, and is geared towards learning and programme or policy improvement. Formative evaluations are important for assisting the effective implementation and delivery of policies, programmes or projects. Hence, often mid-term evaluations are intended to be formative.
• Baseline data are used in the evaluation.	State-of-the-art evaluations start with a review what data are already available and what is already known about the programme. In terms of data, it is important to assess from the start what data are available and what needs to be collected. As already introduced in Step 1, at the start of a programme information needs should be to be defined and data collected, which are accessible to the evaluation team.
• The evaluation considers the implications of the nature of RDI policies for evaluation feasibility.	Whereas for most innovation instruments some effects may be expected within a five-year period, such as an increase in private R&D investment, this period is simply too short for wider impacts to emerge (new products, economic growth). Hence, an ex-post evaluation of RDI instruments needs to take this time lag into account. In the case of thorough impacts analysis, 15 to 20 years may be required to be able to grasp the full spectrum of impacts.

[5] Technopolis (2011), Evaluation handbook for the International energy Agency. Accelerating Energy Innovation: Successful Strategies for Energy Technology RTD

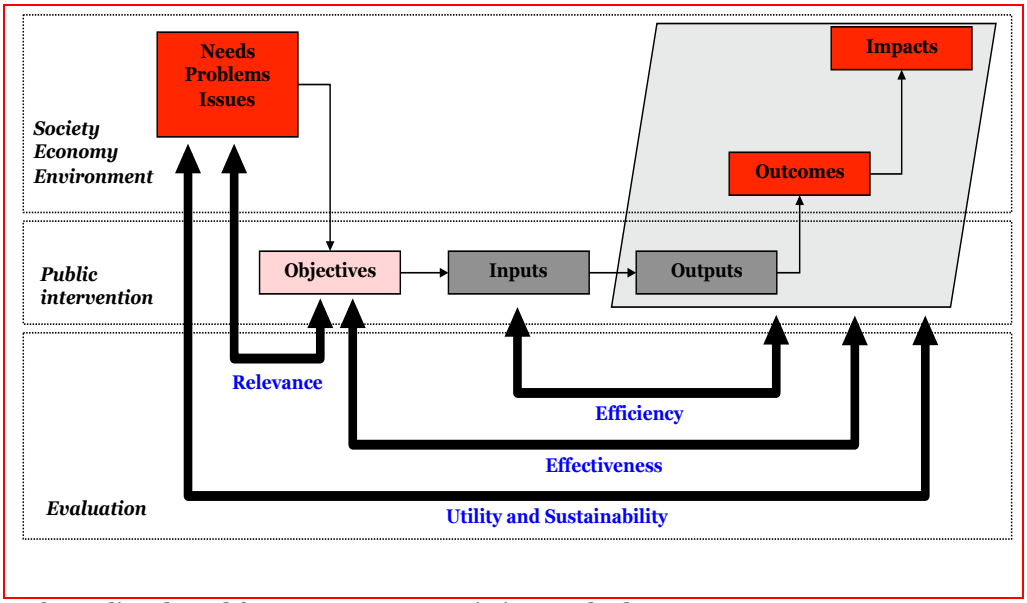
2.2.2 Criteria 1: Evaluation questions

Criteria for good evaluation practice	Description
• Completeness in the range of evaluation questions, which derive from the Inputs-Outputs-Outcomes-Impacts (I-O-O-I) model	Evaluation questions should offer an in-depth look into the programme. This includes testing the expected causal links between inputs, outputs, outcomes and impacts, and exploring rival assumptions and alternative hypothesis on the way the programme is delivering effects. To do so, evaluations can use the Inputs-Outputs-Outcomes-Impacts (I-O-O-I) model (Figure 4), which gives raise to a number of generic evaluation criteria: • Relevance. examining whether the objectives of an activity correspond with the needs, problems and issues it is intended to address (*Are we doing the right thing?*) • Effectiveness. asking whether results and impacts generated by the activities supported meet the objectives (*What happens as a result?),* and to which extent the impacts are additional and can be attributed to the activities. • Efficiency. examining the level of resource use (inputs) required to produce outputs and generate effects. In other words, optimisation of resource utilisation is concerned. An activity that is assessed as having an effect, may not necessarily be efficient: the same effect could have been reached with less resource. (*Are we doing it well and cost effective?*) • Utility. looking whether the intervention has contributed or solved the problem it set out to address. It also looks for expected and unexpected effects (i.e. those that were respectively identified and not identified at the design phase as objectives) and whether these, when they are positive, correspond with needs, problems and issues of different groups in society and the economy (*What shall we do next?*) • Sustainability. examining whether the positive impacts on critical clients and beyond would continue into the future, even after the ending of an activity (Will our intervention have a lasting effect ?) The I-O-O-I model is a straightforward, simple way to articulate knowledge about the programme and designing the evaluation questions and indicators. Its use is particularly recommended, where (for a reason or another) Programme Logic Models (PLM) are not part of the evaluation design.
• Multidimensional character of the evaluation, focusing (at least) on the criteria of relevance, effectiveness and efficiency of the programme.	The final list of evaluation questions is always based on the evaluation's objectives: an ex-post evaluation will in theory focus primarily on he effectiveness and sustainability criteria, while a mid-term evaluation will give more importance to the efficiency and relevance criteria. Best practice is however to include all the four evaluation questions in the evaluation. The OECD/DAC definition of evaluation stipulates that evaluation is the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results.[6] In other words, the strength of evaluation as a policy-making tool is that it questions not only the outcomes/ impacts of the programme but the whole chain from inputs to impacts (based on the I-O-O-I model above). Some studies are only focused on the effectiveness/impact criterion (impact assessment studies). These studies only question one dimension of the policy (its effects), leaving out the design and implementation dimension. As such, they cannot be considered as evaluations *stricto sensu*, even if their results can be used as part of evaluation studies. ***Examples of evaluation questions that are usually used in the evaluation of RDI and can form part of a balanced set of questions are presented in the frame below (*** Figure 5***).***
• Evaluation criteria and questions are proportionate to the scale of the programme/ the evaluation objectives and realistic given the resources and data	The evaluation exercise should be proportionate to scale and scope of programme it evaluates. For example in case of a small-scale pilot programme with limited participants a full economic impact assessment at national level would be unrealistic. Other factors that affect proportionality are the timing of the study in relation to the expected impacts and the availability of data to conduct expected quantitative analyses.

[6] OECD, DAC Working Party on Aid Evaluation (2002), Glossary of Key Terms in Evaluation and Results Based Management

Criteria for good evaluation practice	Description
available.	

Figure 4 The Inputs-Outputs-Outcomes-Impacts model



Technopolis, adapted from European Commission, Evalsed

Figure 5 Focus on evaluation questions

Evaluation Questions

A robust evaluation should, at least, address the criteria of relevance, effectiveness and efficiency. Are presented here a few examples of evaluations questions (by evaluation criteria) that can be included in state-of-the art evaluation methodology: A robust evaluation should, at least, address the criteria of relevance, effectiveness and efficiency. Are presented here a few examples of evaluations questions (by evaluation criteria) that can be included in state-of-the art evaluation methodology**:**

- **Relevance:**
  - Is the instrument well suited to address the needs of business R&D actors?
  - Is the instruments well suited for participation of particular target groups (e.g. SMEs and start ups, emerging sectors)?
  - How appropriate is the instrument to overcome barriers to private R&D investment?
  - Have the needs of the RDI system changed since the implementation of the instrument? Does the instrument fit with new needs of enterprises?
  - Is the programme aligned with its objectives and rationale?
  - What is the extent of synergies/ complementarities with other agency support programmes? How does the programme align with national policy? Does it duplicate any other support measure?

- **Effectiveness:**
  - What are the intended and unintended outcomes and impacts of the programme on participating business?
  - What are the additional RDI-related investments that the recipient companies make that that they would not have made if the programme had not existed? (Input additionality)
  - Additional RDI (jobs, innovation, new products, patents, market share, profitability) that would not have been achieved if the programme had not existed (Output additionality)?
  - To what extent has the support scheme induced the recipients to adjust their RDI processes/behaviour (production process, image, location of facilities, innovation process)? (Behavioural additionality)
  - Additionality questions can also be broken down as follows:
    1. What are the further economic activity stimulated by the direct benefits of the business grant programme (e.g. spillover/ multiplier effects on the national economy, including unexpected impacts)?
    2. What is the proportion of total outputs and outcomes that would have been secured without the

programme in question (deadweight effect)?

3. What is the proportion of private investment that has been reduced elsewhere in the target area for the intervention as a result of the intervention (crowd-out/ displacement effect)?
4. Has the instrument had any effect on non-target groups of enterprises (leakage effect)?
5. Are there any negative substitution effect (e.g. enterprise substitutes a jobless person to replace an existing worker to take advantage of the public sector assistance)

- **Efficiency (including programme implementation):**
  - How has the mode of delivery/ operations contributed to the achievement of programme's objectives?
  - What does it cost to run the programme and is that good value for money?
  - Has a communication/ raising awareness strategy been developed around the programme and how diffusion activities have contributed to the achievements of the programme's objectives?
  - What do the beneficiaries think of the responsiveness, timeliness, helpfulness, clarity of programme administration and operations?
  - Are there any barriers to business participation in the way the programme is administered?
  - Would an alternative delivery approach be more efficient?
- **Utility**
  - Did the programme contribute to solving the problem it was addressing?
  - Is the problem still worth addressing through state intervention?
  - Will the problem reoccur after the programme's termination?
  - Will continuation of the programme still have an effect on addressing the problem or does it need major changes?
  - Do unexpected and negative side effects of the programme outweigh the positive effects on the problem it was aimed to address?

- **Sustainability**
  - In how far is the programme expected to have lasting effects after the intervention?
  - Does the programme still have positive effects on the target group after the intervention has ceased?
  - Is the programme still contributing to solving the problem after the programme's intervention?

2.3 Step 3: Preparation of the evaluation

Criteria for good evaluation practice	Description
• Key stakeholders are involved in the evaluation process (inclusivity).	Stakeholders are those individuals or groups who have an interest in a project, project outcomes, or are part of the project's target population. Stakeholders are key to the evaluation process. Stakeholders involved in programme delivery should be communicated what the evaluation seeks to address, what input will be required from them, and how they might benefit from the findings. Evaluations also involve other stakeholders – including people and organisations directly or indirectly affected by the programme. The level of involvement and method of engagement will be specific to the policy and stakeholders in question, but may include inviting them onto a steering group, informing them about the evaluation, or including them as participants in the research. Stakeholder's analysis/ mapping, when they are conducted as part of the evaluation design process, are a good practice to make an inventory of the stakeholders in the programme being evaluated ('stakeholders analysis') and map their contribution/ role in the programme: • Funders: e.g. Ministries or international agency that has initiated and funded the programme • Programme managers: e.g. staff responsible for programme implementation and management at the Innovation Agency • Direct beneficiaries: e.g. cluster organizations that have been funded has part of the programme • Indirect beneficiaries (where relevant): cluster members' organisations (enterprises, research institutes, HEIs)
• Evaluation resources and duration are in line with the objective of the evaluation and the scale of the	The size of the evaluation is proportionate to the cost of the programme, but all R&D programmes, regardless of budget size, deserve a minimum standard of sophistication in terms of evaluation. The duration of the evaluation depends of its objectives. Robust impact

Criteria for good evaluation practice	Description
programme.	analysis can take up 18 to 24 months to implement. The required time however decreases with the experience in conducting evaluation and the adoption of good practice: completeness of monitoring system with direct extraction of data for evaluations, templates.
• The composition of the evaluation team and governance system ensures transparency, independence and quality of the evaluation process	It is good practice that evaluations are not directly (or solely) commissioned by policy officers responsible for the day-to-day management of the programme, as they might steer towards a positive outcome of the study. They obviously need to be involved in the evaluation as they have the most relevant background information and are those that want to learn from the evaluation. Equally programme managers should not be directly involved in research methods that could lead to socially acceptable answers from the beneficiaries. The evaluation team is comprised of people with different background experience, i.e. with sufficiently capable and competent evaluators with both subject or disciplinary expertise and evaluation expertise: • For R&D business grants and innovation vouchers: evaluation team should have expertise on the analysis of business and national/ regional economic statistics, survey techniques, in-house business innovation processes, financial engineering. • For collaborative projects and cluster programmes: international evaluators might be preferred if there is need for an external perspective and to compare with other systems and learn from them, especially in large, complex and ambitious programmes. An evaluation manager is selected and an evaluation steering group is formed with key programme's stakeholders and where relevant experts and beneficiaries.
• Define future use	When launching an evaluation consideration should be given to the future use of the evaluation study. This could have an emphasis on accountability (e.g. convincing programme funders that public money is well spent) an emphasis on formative use (the programme management wants to learn from the study how to make the programme more effective and efficient). An evaluation could also have the goal to involve certain stakeholder groups and to improve their commitment and participation in the programme. The main purpose of the evaluation can influence the choice of methodology mix, for instance an emphasis on evaluating programme processes rather than impacts or vice versa. It affects the role that stakeholder involvement plays in the design of the evaluation.

2.4 Step 4: Identification of appropriate methodology for analysis (state-of-the-art methodology mix)

2.4.1 Criteria 1: Methodology mix and evaluation approach

A first good practice criterion in the evaluation of RDI support is the suitability of the methodology mix and quality of the overall evaluation approach.

Figure 6 Good practice criteria for the methodology mix

Criteria for good evaluation practice	Description
• The methodology mix is fit for purpose	• The design of the evaluation methodology is well suited to the objectives of the evaluation and challenges/ limits of evaluation tools • The methodology mix is well suited to the type of instruments, the programme objectives and rationale, the evaluation objectives and key characteristics of the programme design • The methodology mix combines several of the preferred evaluation tools for each instrument • The methodology mix is appropriate for the purpose and timing of evaluation

There is no standard 'golden rule' for the appropriate methodology mix as this depends on aspects such as the type of instrument being evaluated, the programme objectives, the timing of the evaluation, the availability of data and the amount of resource available for the evaluation. Some general rules of thumb can be provided for all types of instruments. Good practice would be to combine a mix of methods that can answer the full spectrum of evaluation questions and allows for **triangulation**: evidence from one method can be cross-checked with evidence from another method. Often a mix of methods is chosen that covers the **breadth of the programme** (evidence from as many participants as possible) as well as the **depth of the programme** (e.g. understanding underlying processes that define success and failure). Methods commonly used to cover the breadth of a programme are (electronic) surveys. The drawback of surveys is that while in theory they could go deeply into the evaluation questions, practical considerations as response rates and willingness to have a large group of beneficiaries undergo a time-consuming research exercise, limit this method. Similarly while in-depth interviews could raise more understanding of underlying processes, they are also very time consuming for the beneficiaries as well as for the evaluators.

Another important factor driving the methods mix is the approach to the **counterfactual analysis,** i.e. the combination of methods used to establish the net effect of an intervention, often by means of comparing the outcomes of beneficiaries with a comparable control group. Ideally, the net impact is established through a quantitative econometric counterfactual approach in an experimental setup, but usually evaluation setups have to 'fall back' to a quasi-experimental approach or more qualitative self-reporting of additionality in surveys or interviews with participants and non-participants. Figure 7 gives a good overview of factors determining the feasibility of a quantitative impact assessment, and Annex A gives more background information on the types of techniques available for counterfactual analysis.

Figure 7 Feasibility of quantitative impact assessment and counterfactual analysis.

Source: UK Magenta Handbook / HM Treasury

And finally there are **budgetary considerations** to be made in the choice in the method mix. Face-to-face interviews are relatively labour intensive and therefore expensive compared to surveys. Econometric analysis could also be very labour intensive if significant effort needs to be put in identifying and cleaning data.[7]

With regard to the timing of the evaluation there is a **difference between a mid-term evaluation** (say after 2 years in a 4 to 5 year programme) **and an ex-post evaluation** after the programme has had a number of years of operation and it is expected that effects and impacts are already visible.[8] A minority of programmes have been in existence for over a decade, which provides them with a large pool of data on participants and projects. In these cases longitudinal studies could be made with these data, particularly if historical micro-data are available for the participating companies. It is relatively rare for project participants to be questioned about impacts more than two years after the programme. An example where this does happen systematically is at the Austrian Agency FFG, where beneficiaries are surveyed two years after project completion.

In a mid-term review the objectives of the evaluations are most likely to focus on processes, outreach to the right target groups, the appropriateness of certain funding rules and organisational aspects of the programme. In terms of the methodology mix we would expect:

- An effort to understand the reach of the programme through analysis of programme data and the participants the programme is reaching. This can mostly be done by desk research and interviews with the programme management.
- An understanding of what is going well and what is not going well in the programme. This can usually be achieved by a set of interviews with a representative sample of the programme participants, with the programme owners (policy makers who have launched the programme and have the policy responsibility for the programme's result), with key stakeholders, with persons involved in the programme management processes (e.g. representatives of project review boards, management of boards of competence centres) and representatives of target groups that not have been reached sufficiently by the programme. At this stage it is important to use tools that help to understand issues in depth in order to make amendments that will make the programme more effective.
- Depending whether in the perception of key stakeholders aspects of the programme are need adjustment, more dedicated analysis and therefore methods might be applied. For instance if a certain modality of an instrument is not used or a large target group not reached, a survey could be launched to understand the breadth of the problem, alongside interviews that can be used to understand the issues in more depth.
- Large scale quantitative studies, surveys of the complete set of beneficiaries and control groups that are usually used to identify the socio-economic impacts seem premature in a mid-term evaluation as the numbers of finished projects will most likely not be significant and effects are too early to assess.

In a fully-fledged ex-post evaluation, where the aim is to analyse the full spectrum of evaluation questions, the expected methodology mix would be quite more sophisticated and extensive. The tool set applied depends for instance on the time frame when effects and impacts can be expected. When it is not likely that economic impacts occur in the short to medium term (e.g. with instruments that are geared to creating long term impacts such as competence centres for the life sciences) using econometric analysis after a four-year

[7] See for extensive discussion on social science methods for instance Bryman, A. (2012). Social Research Methods. 4th edition. Oxford University Press; Hoyle, R., Harris, M. & Judd, C. (2002). Research Methods in Social Relations, 7th edition, Thomson Press

[8] This study focuses on comparing impacts of instruments and thus the attention has been on ex-post evaluation studies. Ex-ante evaluations use a whole different set of forward looking methodologies that have not been included in this reference model.

programme does not make much sense. If on the other hand effects could be expected in the short to medium term, such as with voucher schemes or innovation grants, socio-economic impacts could be expected shortly after the programme. If a large number of participants have taken part in the programme and their characteristics are fairly homogeneous, quantitative analysis of beneficiaries compared to a similarly large control group could add value. To make any causal inferences we would expect a minimum pool of participants of 50 and equally a minimum pool of non-participants of that size. The following Figures 7 to 11 gives an overview of data requirements and methods used for mid-term reviews and ex-post evaluations. We have divided methods in four categories (desk research, qualitative methods, quantitative methods and other methods) realising that the distinction is not always clear cut.

Figure 8 Preferred methodology mix for mid-term and ex-post evaluations– Generic

	Mid term review	Ex-post evaluation
Data requirements	• **Minimum level** – Programme and project data to date • **Preferable option** – Baseline data on target group before intervention	• **Minimum level** – Full programme and project data across life cycle – Micro-level company data participants – Micro-level data participating R&D organisations • **Preferable option** – Baseline data on target group before intervention – Annual micro-level company data entire target group – Annual micro-level data non-successful applicants
Desk research	• **Minimum level** – Programme material on rationale and objectives – Review of all programme material and project data • **Preferable option** – Mapping of beneficiaries across target groups	• **Minimum level** – Review of all programme material and project data – Mapping of beneficiaries across target groups • **Preferable option** – Review of reports on broader RTDI context and policies – Review on thematic reports on markets, technologies, finance where relevant (thematic programmes)
Qualitative methods used	• **Minimum level** – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option** – Stakeholder mapping – Focus groups with stakeholders	• **Minimum level** – Reconstructing Programme Logic Model – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option** – Case studies – Interviews non-participants – Stakeholder mapping – Focus groups with stakeholders – Peer reviews (if relevant)
Quantitative methods used	• **Minimum level** – No quantitative methods used • **Preferable option** – Survey with focus on dedicated issues programme performance	• **Minimum level** – Survey of beneficiaries – Survey unsuccessful proposers – Counterfactual analysis – Analysis administrative cost • Preferable option – Cost-benefit analysis – Counterfactual through econometric

	Mid term review	Ex-post evaluation
		analysis with external micro-level data, depending on the context
Other methods applied		• **Minimum level** – Context analysis RTD-context • **Preferable option** – Context analysis markets & technologies – Benchmark similar initiatives (not all programmes relevant) – Social network analysis (not all programmes relevant)

The following sections discuss the specific method mix decisions for the four specific types of programme in our benchmark. Although in broad lines they follow the generic model the design of the four types of instruments and their different rationales ask for variations in the methodology mix.

2.4.1.1 Methodology mix - R&D Business grants

The aim of R&D business grants is to offer direct financial support to enterprises to undertake product development, enhancing product design, prototyping, process innovation, technology acquisition, organisational change or improvements to product marketing. Typical evaluations of such instruments are focused on assessing the extent to which they have encouraged firms to invest in R&D, increased innovation performance and subsequently improved their economic position.

An ex-post evaluation of R&D business grants should build on a review of individual grants at firm level to give an overall assessment of the programme's relevance, effectiveness and efficiency. It is important to evaluate not only the effectiveness/ impacts of the programme but also the design and operations of the instrument insofar as the programme is concerned. In R&D grants aspects such as selection procedures, communication about the programmes, and size of funding are of prime importance in the overall success of support instruments.

The methodology mix used for R&D business grants does not differ greatly from the generic set of methods presented in Figure 7. Annex C has the full list of methodologies that can be used for these instruments. The R&D business grant instrument is a relatively simple instrument with companies as the direct beneficiaries. What is particular to R&D business grants in terms of methodology is the following:

- The counterfactual analysis could be done using a number of methods such as interviews with beneficiaries (e.g. hypothetical additionality – asking what if?) surveys with beneficiaries and control groups. If we look at Figure 7, we can conclude that R&D grants in general could be considered as quite suitable for more quantitative econometric counterfactual analysis as well. While an experimental setup is unlikely, a quasi-experimental technique could be feasible if the intervention group and target population are both sufficiently large.
- In an ex-post evaluation, the use of social network analysis is not very relevant as the individual business grants do not aim to create linkages or collaborations. The same holds true for peer reviews which are mostly used for more complex instruments which have a public research component that needs to be assessed (e.g. on scientific quality)
- For generic R&D business grants, not targeting a particular sector or thematic area, extensive stakeholder mapping and focus groups with stakeholders are less relevant. If a business R&D grant scheme is targeted to a specific sector and there are concerns whether the programme is reaching the appropriate target groups, stakeholder mapping (positioning the beneficiaries against the potential target group) could be introduced for instance in a mid-term review, in order to reorient the programme for its next term.
- In an ex-post evaluation an analysis of the R&D context could be helpful to see whether the programme is still relevant – especially in the case of thematic rather than generic grant schemes (the market failure rationale remains).

2.4.1.2 Methodology mix - Innovation vouchers

Policymakers typically introduce voucher schemes in order to allow knowledge held by research and technology organisations to play a role in developing new products, processes and/or services. In general, they are aimed at broadening the basis for innovation beyond the traditionally strong large firms, involving SMEs in innovation and encouraging them to collaborate with research organisations. An SME that is awarded one or several vouchers can seek academic or technological expertise, usually from pre-approved universities, research institutes and colleges, to solve a specific problem or obtain new ideas for its business.

Evaluations of such instruments generally focus on assessing the extent to which they have encouraged SMEs/ start-ups to establish relations with public sector knowledge providers, invest in innovation activities and reduce barriers that hinder SMEs' capacity to innovate and successfully commercialise new products, services or processes. Owing to the widespread adoption of innovation voucher schemes across Europe over the past few years, benchmarking can be included in the wider methodology mix to complement control group approaches, interviews and surveys, wherever the allocation of evaluation resources allows for a more in-depth methodology.

As with the business R&D grants the methodology mix of voucher schemes is mostly in line with the generic methodology mix shown in Figure 8, but with some specific features:

- An additional dimension in voucher schemes is whether those with a demand (the companies) and those that supply (the providers of technological support) are well matched. Evaluations and impact studies of voucher schemes should thus look beyond solely the business users of the vouchers to understand the strengths and weaknesses of the programme fully
- Voucher schemes, more so than any other type of instrument, could be implemented in a first tier experimental design setting, especially if the voucher scheme is targeted at a large relatively homogenous sector (e.g. SMEs in the manufacturing sector). Since vouchers often aim to reach companies other than the 'usual suspects and resources may be limited to reach an entire population, a partly randomised distribution of vouchers could be justified.
- As with business grants, social network analysis, and peer reviews are less relevant. In terms of desk research review of market, technology and other trend reports are most likely not relevant, as vouchers schemes are generic and typically aim at a broad set of SMEs

2.4.1.3 Methodology mix - Collaborative R&D projects

Collaborative R&D projects are one-off support offered to a consortium of partners working on a joint R&D project. Typically, collaborative R&D supports consortium of public and private partners, with participation of two of several of the following stakeholders: business, research institutes, research technological organisations and HEIs. Policy-makers might also be included in cases where research involves legislative or regulatory aspects. The rationale for collaborative R&D grants is broader than single business R&D grants as, in addition to tackling the traditional market failure for R&D, such programmes aim to change behaviour, in particularly organisations' propensity to collaborate with public and private partners. Evaluations of collaborative R&D grants typically focus on assessing the extent to which they have fostered technology transfer, encouraged collaborative behaviour between public and private research and increased R&D outputs and eventually economic performance. Thus apart from measuring effectiveness through output additionality an evaluation would also want to establish behavioural additionality. If the objective is to create medium to long-term collaboration, the study will need to address evaluation questions on sustainability.

While a control group approach is used in some evaluations, a counterfactual analysis is difficult, due to the indirect effects of the collaborative aspects of these R&D activities. In most collaborative R&D instruments, the direct beneficiaries of public funding are research institutes, not the companies. So effects will most likely occur at a later stage and will be more difficult to attribute to the public intervention. The typical approach for this is through surveys of non-participants and comparing the results with those of participants. An

alternative is to carry out interviews with non-participants – i.e. rejected applicants or groups of enterprises, HEIs or research institutes that match some of the parameter of consortium partners but did not apply for funding. In addition, a baseline analysis of firm-level data (from the applicant records or from official statistics) can be conducted to review the economic impacts on participating enterprises.

In comparison to the generic methodology mix presented in Figure 8 for collaborative R&D programmes:

- In terms of the mix, more or less the same set of methods can be used with the exception of peer reviews. Benchmarking with similar initiatives abroad has limited value due to the large influence of national contexts, unless it is done on very specific topics such as for instance the involvement of specific target groups
- A counterfactual analysis aiming to measure economic benefits will, more so than in 'company only instruments' be distorted by the contextual (controlling for external factors) and attribution (linking the effect to the intervention) factors. Thus this ask for a larger diversification in the methodology mix
- Ask for a larger emphasis on qualitative methods to better understand the behavioural additionality and organisational learning that has occurred as a result of the programme, possibly leading to more substantial economic impacts, but at a later time. This would require more emphasis on methods such as interviews and case studies
- Context analyses such as market and technology trend studies could be particularly useful if the programme is thematic and focuses on a specific technology domain and/or sector, in order to understand the impact of external factors better

2.4.1.4 Methodology mix – Integrated cluster and competence centre programmes

For this study, we took a broad approach to integrated cluster policies, including support for clusters as well as public-private partnership programmes such as competence centres. This affects comparability as the designs of the programmes benchmarked show significant variations. Cluster/ competence centre policies aim to stimulate innovation by addressing coordination and information barriers that prevent knowledge and technologies being diffused, transferred and used in the economy. Cluster and competence centre measures can take the following forms: i) funding a cluster organisation (with an office/cluster manager) to undertake activities to strengthen co-operation between businesses, intermediaries, ii) funding projects from a cluster/domain (or a set of clusters) to boost business innovation by influencing the intensity of co-operation, iii) setting up physical and virtual centres involving to enhance medium to long term strategic alliances in a cluster/domain iv) strengthening the framework conditions for cluster development such as support to human resource upgrading, improving the business environment, research infrastructure and support to internationalisation.

Depending on the timing and objectives of the evaluation, the focus can be on the efficiency and effectiveness of the cluster management in implementing actions to strengthen the cluster, and/ or output in terms of new forms of co-operation within the cluster or between the cluster participants and other regional or inter-regional clusters in complementary sectors or technologies, and/ or impact of the cluster measure on the innovation activity and resulting economic performance of the firms. In the case of competence centres, there are additional objectives such as a good governance structure for the centre, the development of a robust and ambitious common business plan, strategy and research agenda, the development of transparent processes to involve the stakeholder community and the stimulation of excellent science and technology. The question of sustainability is important for this type of instrument, as the objective is to create medium- to long-term relationships.

A control group analysis is extremely challenging in the evaluation of clusters and network policies, as collaboration structures are often unique to a region or sector and allegedly the majority of the stakeholders that are eligible to participate are part for the cluster or network organisation. In addition it is more difficult to distinguish between participants and non-participants as these policies are not merely providing funding to a particular company and more about involving companies in a suite of activities ranging from taking part in seminars

to innovation projects. The facilitating character of the intervention makes it more difficult to identify the beneficiaries. At best an economic impact analysis of part of the programme (for instance a business grant scheme as one component of a wider cluster package) can be assessed.

The literature on cluster evaluation mentions the difficulties of data availability, the complexity of indirect intervention models and consequences for attribution of effects to cluster policy, the time lag and the danger of missing the core aspects of cluster policy.[9] An alternative might be to compare active cluster participants and less active participants but such an analysis might be sensitive and is liable to a high degree of distortion from self-selection.

There are a small number of examples of econometric analyses of cluster policies, using data sets that go back in time, which have managed to identify control groups using for instance the difference-in-difference methods.[10] Such a study would require a robust baseline study at the start of the programme identifying the potential target group in a particular domain and comparing that with the situation a considerable time after the start of cluster policies. Usually the availability of the data and the resources needed to retrieve and clean these data impede the use of econometric techniques. The decision whether or not to do an econometric analysis would need to reflect these issues of data availability, the feasibility of identifying a representative control group outside the cluster/ competence centre and the relative efforts needed to conduct these studies. As an econometric study alone would not be sufficient to understand the factors for success and failure it would need to be accompanied by other methods as well.

The literature and guides favour instead the use of international peer review and benchmarking to provide a comparative basis for policy learning. In addition to focusing on the evaluation at the programme level, there are suites of methods developed to assess individual competence centres (using peer reviews and reviews of the CC as organisation as additional tools) or the performance of individual clusters. [11] A cluster / competence centre programme evaluation can make use of these approaches and aggregate the results at the programme level.

Thus in terms of the methodology mix used for cluster programmes and competence centres, compared to the generic set provided in Figure 8 the differences are:

- Participation in Cluster and CC programmes are often quite fluid and have open access for firms to join or not join and to take part at the core of the activities (e.g. large R&D projects) or have a relatively modest participation (e.g. strategic workshops, matchmaking events). In many cases this is not regulated through formal application processes. The distinction between participants and non-successful applicants is therefore not always as straightforward as with for instance collaborative R&D programmes. Therefore studies comparing the target group and a control group will be more difficult to construct in a statistically sound manner. Also, cluster programmes may already involve the majority of the active R&D-intensive companies in a specific sector, and a suitable control group is difficult to identify. However, this is very dependent on the design of these programmes and the rules of participation. Using quantitative counterfactual techniques should be done with great care in view of these considerations, but could be appropriate for specific sub-parts of a cluster or CC programme, such as a business grant. Naturally, more qualitative aspects of counterfactual analysis through interviews or surveys (hypothetical additionality) can and should be used.
- In CC programmes which aim to establish sustainable centres with state-of-the-art research for industry, peer reviews with preferable international experts are a good method to assess the quality and relevance of the research performed in the centre. More

[9] See for instance Schmiedeberg, C. Evaluation of Cluster Policy: A methodological overview, Evaluation, 2010, 16:389

[10] Uyarra, E. and R. Ramlogan, The Effects of Cluster Policy on Innovation, NESTA Working Paper Series, No 12/05.

[11] See for instance Lämmer-Gamp, T. G. Meier zu Köcker, T Alslev Christensen, Clusters Are Individuals, The Danish Ministry of Research, Innovation and Higher Education, Copenhagen/Berlin, 2011.

so than with classic academic peer review, the panel of experts could be a mix of academic researchers, industrial R&D experts and research users.

- Even more so than with collaborative R&D, in order to understand the behavioural additionality and organisational learning that the programme is aiming to achieve, evaluations should emphasise qualitative methods such as interviews, case studies, stakeholder focus groups and organisational reviews. For cluster programmes and CCs the governance, organisational and management aspects of the instruments form a crucial success factor, which ask for qualitative methods.
- Benchmarking these cluster and instruments could be useful in terms of the organisational set up of the programmes, more as a formative evaluation approach (learning from good practices) rather than a summative evaluation approach as impacts are too much dependent on national and sector contexts

2.4.2 Criteria 2: Use of evaluation methodologies

The table below gathers together a few criteria that can be used to assess good practice in the use of evaluation methodology in the R&D field.

Criteria for good evaluation practice	**Description**
• Transparency: in the way data were collected and interpreted is transparent and clearly outlined in the report	• The report includes a description of the overall methodology and details how the research strategy was designed to meet the goal of the evaluation • The data used during the evaluation and their limitations are described • A description of strengths and weaknesses of evaluation framework and methodology mix is included in the report • The processes involved in relation to evaluation research design, sampling and data collection is fully described • The background of analytical constructions, such as categories and codes, is clearly explained • All data manipulations undertaken for methodological purposes are clearly documented and retraceable
• Robust sampling: the sampling strategy is robust and allows for generalisation of the results	• The size of the sample allows for generalisation of the results and the sample coverage is balanced and representative of the wider population. Hence, the data sources are credible and representative of the wider population • The sample profile is clearly outlined and consideration is given to implications of the sample size and coverage.
• Triangulation and sensitivity checks	• Triangulation of data: Data are triangulated and results are corroborated • Internal and external validity of quantitative data is tested (e.g. Test for statistical robustness, Sensitivity checks, treatment of outliers) • Robustness in reporting: proper use of statistical terms, inclusion of other statistics besides means, such as medians, standard deviations. • A systematic and thorough analysis is conducted, including atypical cases and emerging issues and alternative assumption • Additionally, if comparing findings with previous evaluations or studies, similar methods and approaches are used to make comparison credible

The quality criteria for the use of evaluation methodologies are explained further in 0, detailing good practices for each of the main evaluation method identified in this reference model: i) (Quasi-) Experimental Counterfactual analysis, ii) Surveys, iii) Interviews, iv) Case studies, v) Cost-efficiency/ effectiveness analysis, vi) Benchmarking, vi) International peer review, vii) Social Network analysis and viii) Context analysis. The purpose of counterfactual analysis is to establish the *net* effect of an intervention taking account of effects that would have taken place without the intervention. Various methodological approaches and tools can be used for this purpose which are described in short in Appendix A with further reading suggestions. Good sampling, robust analysis and triangulation of data are common requirements in interviews, surveys and case studies.

2.5 Step 5: Conclusions and reporting of the evaluation

Criteria for good evaluation practice	Description
• The conclusions are robust and enables a better understanding of the programme's performance (summative evaluation)	• Conclusions/ findings address the original set of evaluation questions and objectives of the study • Conclusions are evidence-based and stem from the triangulation of data • The main assumptions and theoretical background on which the evaluation is based are discussed • Consideration is given to rival assumptions/ explanations/ theories, in order to refine the original Programme Logic Model and take into account unexpected effects • Findings/ conclusions have a coherent logic • Key findings are credible and the validity of results have been checked (if needed through the use of a peer review group of experts or alternatively through discussions within the steering group) • Findings/ conclusions are coherent with other knowledge and research evidence • A discussion on the limitations of evidence and what remains unknown or unclear with regards to the programme is included
• The results of the evaluation can be used to inform further policy-making (formative dimension of the evaluation)	• Clear action (improvement/ changes) are identified as to what modifications are necessary in the programme, including timetable and effects of any proposed change • The evaluation clearly highlights the key success factors and barriers in programme design and implementation • The evaluation clearly highlights what are the impacts of evaluation findings in terms of policy-learning (e.g. review of programme delivery, reflections on the quality of design and implementation, review of the wider context for delivery of the programme, highlight of wider good practices at national/ international level) • The evaluation includes a reflection on how to run similar programmes better in future
• The evaluation report is clear and accessible to end-users	• The report can be easily accessed and read by all target groups • Key messages are summarised and highlighted • There is clarity in structure and text • Main study documents are reproduced in the report (e.g. survey questionnaire, interview guides, letters of approach)
• The evaluation includes further reflections on how to improve future evaluations	• This can include reflections on how to improve internal monitoring system, recommendations for future evaluation methodologies or identification of further needs in terms of evaluation/ studies • Where possible, the source data should be archived to allow subsequent secondary analysis.

2.6 Step 6 Apply the lessons from the evaluations

As aforementioned the Evaluation Reference Model presents a **process approach to evaluation**. In simplified conceptual framework of the policy cycle, an important additional step after the evaluation has been formally completed is the sixth step: apply the lessons of the evaluation. It is considered important to add this into the process model as a separate and 'last' Step 6. Nevertheless Step 6 is closely linked to Step 1, where programme managers design the programme including its objectives and the programme's framework for future evaluation. Most likely in Step 1 some arrangements have been made how and when the programme will be evaluated and who are target group for this report. Sometimes the studies are for internal use only, sometimes the reports receive ample public attention and need to inform specific target groups (e.g. Finance Ministers, Parliaments, National Court of Auditors, the European Commission in case of Structural Funds' Operational Programmes)

How evaluation studies are used is dependent on many context variables related to the national governance structures and cultures. In the scope of this TAFTIE benchmark we have not been able to include how the evaluations studies are used, internally within the agencies

and externally with programme owners, political actors (e.g. members of Parliament), programme stakeholders, media, user groups and any other possible stakeholders. Particularly in integrated cluster and competence centre programmes considerations should be made how the stakeholders will use the results of the evaluation for the next phase of their initiative. The topic of the 'use of evaluation' contains a large body of academic research going back decades with different schools of thought.[12] A recent study that has covered the use of evaluation in innovation policy is the InnoAppraisal study commissioned by the European Commission, DG Enterprise. [13] The following grid provides three sub-criteria that can be used as reference points to assess whether agencies are situated in a governance context that supports policy learning from evaluations.

Criteria for good evaluation practice	**Description**
• The timing of the evaluation fits well in the policy cycle of the instrument and related policy instruments	• Evaluations are conducted in a timely manner allowing the lessons from that evaluation to be taken on board for the continuation of that programme or the redesign of follow-up programmes • The main conclusions from previous evaluations even from other similar programmes are taken into account in the design of new programmes
• A culture of policy learning is developed within the agencies	• Expertise on evaluation designs and methods and the outsourcing of evaluation studies is codified or coordinated within the organisation, avoiding that each new programme manager has to reinvent the best possible approach • Evaluations considered as an investment in policy learning and not only as a cost to the programme officer. Appropriate levels of resources (internally and externally) should be foreseen at the programme design stage adapted to the timing and complexity of the study.
• Ensuring political support for conducting and using evaluations	• The lessons learned from evaluations have visibility and support at highest political levels to ensure that this knowledge is taken on board when proposing new measures or amending existing instruments • Consistent transparency on the outcomes of evaluations will increase the trust in the validity of evaluations Lessons on how to ensure political support and policy learning from evaluations are very context specific and depend on the Agencies' role in the national innovation system, the legal and regulatory environment and many cultural factors such as instance the use of good governance and new public management principles in government bodies. Considerations on how the results of evaluations are better used are rarely made within the evaluation studies benchmarked. Good practices can be found for instance in Poland where an extensive legal framework and system for evaluations is made as part of the Evaluation of the Operational Programmes for 2007 -2013. This specifies clear roles for Agencies and Ministries to respond to and implement the evaluation results.

[12] See for instance Patton, MQ, Utilization-Focused Evaluation, 4th Edition, Sage, 2013; Mark, M., G. Henry (2004), The mechanisms and Outcomes of Evaluation Influence, Evaluation, vol. 10: pp 35-57; Caracelli, V., H.Preskill (editors) (2000), The expanding use scope of evaluation use, New Directions for evaluation, A publication of the American Evaluation Association, San Francisco, Jossey-Bas

[13] See Inno-Appraisal, Understanding evaluation of Innovation Policy in Europe, 2010, by Manchester Institute of Innovation Research and partners, Study for European Commission, DG Enterprise.

Appendix A Counterfactual Methods

A.1 Overview

One of the most essential yet also most challenging aspects of measuring effectiveness is the concept of counterfactual evaluation. The purpose of counterfactual is to establish the *net* effect of an intervention, which essentially means to account for effects that would have taken place without the intervention as well (called 'deadweight'). All counterfactual methods rely on the principle of comparing the development on specific indicators of the intervention group with the (hypothetical) development these indicators of a similar control group. Counterfactual studies are generally quantitative, but may also include semi-quantitative indicators such as those based on Likert scales.

In measuring causal effects of intervention, one can generally identify a ranking of methods preferred in terms of robustness and validity. An evaluation of high quality will follow the best available model. Note that model selection mainly depends on data availability, but specific programme (selection) characteristics may influence methodology choice as well. The choice for a specific (combination of) methodologies should be well explained, and should generally follow the 'best-available-model'. Data for these approaches is usually based on a combination of monitoring data (for inputs), external data (e.g. statistical agencies) and often surveys, especially for behavioural additionality and other qualitative indicators

A.2 (First-tier): experimental design

The best way to establish the causal effects of interventions is to use an experimental design, where 'treatment' is delivered randomly to the target group. By doing so, al other simultaneous factors influences the measured outcomes are controlled for given a large enough sample. Although this setup is possible for some voucher schemes, an experimental design is usually not feasible or desirable, the problem being compounded by the fact that the intervention design cannot be changed ex-post. Most evaluations use a quasi-experimental design instead, in which the treatment group is compared with a statistically constructed control group. *Example study: Voucher Scheme Evaluation by NESTA*[14]

A.3 (Second-tier): quasi-experimental design using control groups

The control group comprises of non-participants who have the similar characteristics as participants. In practice, it is however very difficult to assemble a robust control group and different techniques can be used to control the selection bias. Three main approaches[15] can be distinguished, in (general) order of preference:

A.3.1 Regression discontinuity design (RDD)

RDD is based on the principle of comparing those firms that just fell within the acceptance criteria and those which just fell outside. Note that this design needs a scoring mechanism with a relatively continuous and linear scoring range. Needs a relatively large sample size and does not measure the average effect, but rather the local average treatment effect. However, when the data and programme design meets these design criteria, RDD is generally among the most preferred second-tier designs as it leads to a relatively unbiased estimate

A.3.2 Difference-in-difference analysis.

In its simplest form, difference in difference analysis compares the intervention group with a control group, both before and after. Preferably the model is extended via a fixed-effect

[14] http://www.nesta.org.uk/publications/reports/assets/features/creative_credits_report

[15] Another main tool is *instrumental variables analysis* using a 2-staged approach. Since this method is rarely used in practice in this application domain due to the lack of instruments and data, it is not discussed in further detail here.

analysis when panel data is available. *Fixed-effects analysis* [16]takes advantage of the additional possibilities offered by panel data (i.e. observations across years) as it yields a more robust estimation of effects. The advantage is that difference in difference can account for differences between the control group and intervention group as long as both groups share a common trend. This assumption should be explicitly tested and discussed, and relevant control variables (which are not fixed) should be included.. [17] *Example study: Evaluation InnovatieKredieten/Uitdagerskredieten (NL Agency)*[18]

A.3.3 Propensity Score Matching (PSM)

PSM is a method where each participant is coupled to a mirror firm that roughly shares the same characteristics. Matching can be based on various statistical or qualitative criteria, such as number of employees, annual turnover, export region, sector or location. This technique should be applied when there are mirror firms available. A good multivariate matching design is to use propensity score matching, in which a prior probit-regression analysis is done to determine the likelihood of companies entering the programme. This technique can be combined with other model designs.

A.4 (Third-tier) Alternative methods:

When no panel data (i.e. only ex-post data) is available, the minimum requirement for a regression model is too include as many as possible control variables and/or use matching techniques (such as described above). Still, the evidence remains quite weak.

When no control group at all is available, there are options to analyse variations within the intervention (for instance grant size, see discussion on binary vs. linear interventions above). Another fall-back option is to use self-reporting of additionality in surveys and ask recipients what would have happened if they would not have received support. *Example study: Evaluation of the Innovation Partnerships (Enterprise Ireland)*

A.5 Further reading:

- Handbook on impact evaluation: quantitative methods and practises. (2009). Khandker S. et al. World Bank training series
- Experimental and quasi-experimental designs for generalised causal inference (2012), Cook, T. Thomas Cook Publishing.

[16] And, by extension, dynamic panel data models with endogeneous terms.

[17] An alternative/extended approach is to use the *Heckman 2-step adjustment procedure:* this technique involves the formulation of a single equation to explain the selection procedure and then a second equation that explains performance change based on the factors included in the selection equation.

[18] http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2013/08/23/evaluatie-uitdagerskrediet-en-innovatiekrediet.html

Appendix B Good practice criteria in the use of main methodology tools for RDI evaluation

Some of the key aspects of quality criteria for the use of individual evaluation methodologies are explained in this Appendix detailing good practices for each of the main evaluation method identified in this reference model: i) Counterfactual analysis, ii) Surveys, iii) Interviews, iv) Case studies, v) Cost-efficiency/ effectiveness analysis, vi) Benchmarking, vi) International peer review, vii) Social Network analysis and viii) Context analysis.

Figure 9 Good practice in the use of main methodologies

Tool	Good practice criteria in the use of key tools
(Quasi-) Experimental Counterfactual analysis or alternatives	Any evaluation that has the goal to assess the effectiveness of a policy instruments needs to take into account counterfactual aspects. Evaluations designs should implement the **best available model**, as discussed in Appendix A. For some evaluations, this will account to a full-fledged econometric (quasi-) experimental approach, whereas other evaluations will use more qualitative methods such as interviews with unsuccessful applicants. Of course, counterfactual analysis is carried out with various other methods (surveys, data analysis, interviews), and all good practice criteria for these methods apply. However, there are also some general good practice criteria that can be identified: • The **composition of control group** and possible selection bias needs to be discussed thoroughly. As a minimum, background characteristics need to be presented and discussed. • **Matching criteria:** When selecting enterprise control groups, the minimum requirements are to take as many different factors as possible into account, but this also depends on the instrument being analysed. Control groups are to be comprised of enterprises approximately equally likely to use or participate in the instrument, but yet have not. • **Binary vs. Linear interventions:** The methodology should consider to not only measuring the effect of participation vs non-participation, but also measuring the effects of different intervention sizes • **Internal validation through triangulation of sources:** By linking self-reported survey data with external data, the validity and reliability of the analysis can be improved substantially.
Surveys	Survey are often used in the evaluation of RDI support programmes to gather quantitative data on a large sample of the population and collect baseline data as well as outputs, outcomes and impacts information that are not captured by the programme monitoring system and external data systems. • **Good Sampling:** if the total population of recipient firms (and possibly control group) is too large, it is possible to use a sampling strategy. Sampling needs to take account of a (as broad as possible) set of background characteristics (see matching criteria before), but also intervention parameters (such as size of grant, role in projects). Note that the sampling strategy should take account of the research questions in terms of sub-groups: if the evaluation wants to establish differences between for instance different types of actors (SME, non-SME), enough observations should be available for statistical analysis for each specific subgroup. • **Respondent analysis:** When the final survey response is available, a respondent analysis should be carried out in order to verify whether the respondent group is representative of the total target group based on background characteristics. • **Triangulation:** A survey should employ multiple type of questions for important indicators, such as open questions, multiple choice, likert scales, numerical and question modes such as statements (positive and negative) or questions. This will allow for a more robust construction of indicators as responses can be verified and triangulated.

Tool	Good practice criteria in the use of key tools
Interviews	Interviews with beneficiaries and/ or key programme stakeholders and/ or wider experts and stakeholders are typically included in most RDI evaluations. Interviews can be exploratory (intended to test hypotheses at the start of the evaluation on a small number of stakeholders/ beneficiaries) or can be conducted at a later stage of the evaluation to collect data for analysis. They often address the full range of evaluation questions from relevance to coherence, effectiveness and efficiency. • **Good Sampling:** in most cases, R&D support programme (in particular business grants, vouchers and collaborative projects) are too large to enable the evaluation of the effects on all participants. Sample should be as much as possible representative of the larger beneficiary population and allow for generalisation. The size of the sample for beneficiaries interviews takes into account the size and scale of the instrument, within the given evaluation budget and time. In the case of R&D Business grants, the sampling strategy is often a mix of different sampling methods. Representativeness and balance of the sample can be based, among others, on the following criteria: – Range and types of beneficiary organisations (e.g. firms, SMEs, start-ups) or range/ type of consortium (collaborative project)/ clusters – National/ international/ regional distribution of support – Thematic distribution of support (e.g. top-industry sector, emerging sectors) – Type of RDI supported (product/process innovation, new product, new features added to a product, improvement in process) or type of activities supported (technology transfer, networking, information dissemination) – Number of participations (where relevant) and period of participation (past/ ongoing/ future participation) – Other criteria that are relevant to the objectives of the programme: e.g. women participation. – Where they are conducted prior to interviews, survey might also be a good opportunity to select potential candidates for interviews, by mixing participants who reported important effects of their participation and those for whom the impact was more restricted. • **Transferability:** interviews' analysis includes a discussion on what can be generalised to the wider population from which the sample interviews is drawn • **Analysis of opinions:** The evaluation includes a discussion of the nature and source of any divergence in opinions from business beneficiaries Rival explanations and new assumptions on the effects of the programme on business participants are considered in the interview guide • **Triangulation of data:** The main results of interviews are corroborated/ enables to confirm the main finding of the control group and survey

Tool	Good practice criteria in the use of key tools
Case studies	Case studies are often carried out after towards the end of the data collection periods and information gathered through interviews and surveys provides useful input for the selection business sample to case study. In the case of integrated cluster programmes, case studies are typically focused on the cluster/ network in itself and investigate the outputs, outcomes and impacts of the cluster/ network on participants and on the wider economy. In the case of business grants or innovation vouchers, the unit of analysis is typically the business having received the grant, while collaborative R&D might focus either on individual consortia participants (and, for example, their successive participations in the programme) or the project in itself . • **Good Sampling:** Case studies are often carried out after towards the end of the data collection periods and information gathered through interviews and surveys provides useful input for the selection business sample to case study. In addition to representativeness and balance criteria (elaborated above as part of the description of interviews), the best sampling methods are usually based on a combination of the following techniques: – Perception of successful stories (e.g. business/ consortia/ clusters that have reported very high level of benefits) – Inclusion of 'typical' (e.g. organisations reporting the expected benefits) and 'atypical' case (e.g. organisations reporting unexpected benefits) – Inclusion of deviant cases (e.g. organisations which have declared that the support they received had none or only negative impacts on their activities) – Balance between organisations for which different types of impacts have been reported in the literature/ surveys/ interviews (e.g. economic, social, environmental, or policy impacts) • **Protocol for common data collection approach**: Case studies are based on a common protocol that enables cross-analysis of data collected through case studies and coding of information • **Combination of data:** Case studies are based on a combination of quantitative (e.g. programme monitoring statistics of business inputs and outputs, business data on changes in R&D activities and performance before and after participation) and qualitative data. Interviews are conducted with project manager in charge of the business grant, staff working on the project supported by the grant, and business executives/ programme managers in charge of the R&D strategy or the given field area. This allows for an in-depth analysis of the outcomes and impacts of the grant on participating business, including unexpected impacts and benefits. • **Transferability:** the analysis of case studies includes a discussion on what can be generalised to the wider population from which the sample case studies are drawn • **Analysis of opinions**: The evaluation includes a discussion of the nature and source of any divergence in results from business beneficiaries. Rival explanations and new assumptions on the effects of the programme on business participants are considered in the case study protocol. • **Triangulation of data:** The main results of case studies are corroborated/ enables to confirm the main finding of the interviews, survey and control group analysis. • **Robustness in analysis:** The conclusions of the case study are evidence-based and clearly linked to the analysis of the data collected during the case study field work.
Cost-efficiency/effectiveness analysis	Cost effectiveness analysis is method in which programme benefits and costs are consistently monetised in order to determine the net rate of return. This method usually builds on the effectiveness measured in a counterfactual design, but extents this towards second-order and third-order effects as well. On the cost side, all internal and external costs of the programme should be included, such as handling costs but also invested time of (unsuccessful) proposal writers. If a cost-effectiveness is included (which is preferred), the following requirements can be identified. • The analysis should be **inclusive** of all important internal and external costs and benefits, including second and third order effects • Obvious exclusions for methodological/data availability reasons need to be properly motivated. • The analysis should include a **time perspective** (i.e. discounting and using Net Present Value) • The analysis should use **uncertainty bounds** and posit these clearly in the results

Tool	Good practice criteria in the use of key tools
Benchmarking	Several of the R&D support instruments described in this reference model have become a main feature in many European countries, allowing for benchmarking between different types of instrument at national or international level. • **Good sampling:** A first good practice is to include more than one comparator country/ instrument in the analysis in order to enhance the robustness of the analysis. The selection of comparator collaborative R&D instruments can be based on a wide range of criteria, depending of the programme design and nature. The following criteria are typically considered: – Country-related criteria: Size of the countries, Economic/ R&D specialisation, Similarities in terms of R&D performance and barriers within the RDI innovation system – Instrument-related criteria: mode of delivery (e.g. competitive design, random selection, calls), targets and eligibility criteria, mode of operations (e.g. selection and attribution procedures, reporting procedures), budget, average funding allocated to R&D consortia, any specific target in terms of participants and field of R&D. • **Combination of questionings:** A minimum requirement is that benchmarking provides insight not only into the effectiveness of the programme but also into the processes and design behind the programmes. A whole array of indicators (including performance indicators but also indicators related to programme rationale, objectives and design) is used to compare performance between programmes, in order to capture the specificities and good practice of each single instrument and to outline areas where programmes can be compared and areas that differ between programmes. • **Combination of data:** A benchmarking analysis can build on a combination of quantitative (e.g. monitoring data on programmes' inputs, outputs and outcomes, country-level R&D) and qualitative data (e.g. interviews with programme managers, review of existing documents, studies and evaluations).
International peer review	As part of the benchmarking exercise or as an alternative to it, an international peer review can be organised to review specific features of collaboration and clusters programmes, most importantly the requirements on the selection of partners and interdisciplinarity, intellectual property and legal and governance issues of collaborative entities (if the instrument design include a formal partnership). • **Selection of panel members:** The panel should comprise a mix of domain (technology, sectoral), R&D co-operation and governance expertise. • **Organisation of peer review process:** all relevant data (e.g. cluster monitoring data, self-assessment reports from cluster or competence centre management) should be made available on time to peer reviewers

Tool	Good practice criteria in the use of key tools
Stakeholder Analysis/ Mapping	Stakeholder analysis is a tool used to systematically analyse the actions, interests, roles and behaviours of key individuals or groups who have an interest in a project, project outcomes, or the project's target population, and to assess the influence and resources they bring in the implementation and decision-making process. Stakeholder mapping is a type of stakeholder analysis that focuses on the assessment of a large number of actors linked together by various forms of relationship. [19] Stakeholder analysis/ mapping is most often used to assist policy definition, rather than evaluate. It is however a relevant framework for state-of-the-art evaluation, especially in the case of complex, multi-actors policies such as cluster instruments. It is used to assess the relevance/ consistency between goals, institutions and actions within a wider RDI system. It can also be used at the start of an evaluation to input the evaluation design: identify target population, identify key resources for data collection, draft/ update the Programme Logic Model. In practice there exists a wide range of methods and approach depending on the use of the analysis. Beyond the diversity in techniques, here are a few good practices: • The analysis should include **all programme funders, managers and beneficiaries** but state-of-the-art stakeholders analysis. Depending on the evaluation questions and type of instruments, it can however be that the analysis is limited to mapping the programme beneficiaries within the wider context in which they operate (within their innovation system, within their sector). • **Systemic view:** Stakeholders analysis/ mapping is based on the Programme Logic Model (whenever available), however the unit of analysis is the RDI system and the policy sector as a whole rather than the programme itself. Conducting a stakeholders analysis/ mapping means 'thinking outside the box' and considering all relevant aspects of the policy/ programme (e.g. what relevant actors/ institutions have been left out of the programme? is there a need for new institutions/ actions?). • **Prerequisites:** a stakeholder analysis requires certain knowledge of the RDI system and familiarity with the policy field. • **Combination of data**: state-of-the-art models of stakeholders analysis apply a variety of tools to understand stakeholders, their positions, influence with other groups, and their interest in a particular reform. It is a good practice to complement desk research and documentary review with interviews with key stakeholders or with experts. • **Participative use**: stakeholders analysis provides a means to engage different stakeholders in the discussion of their position and contributions to a programme or instrument and they can involve other techniques such as focus groups.
Social Network Analysis (SNA)	A (social) network analysis is a useful tool to investigate the relationships developed in a R&D-cluster or collaborative programme. In a SNA, links between organisations are represented graphically through nodes and lines, where relative positions in the resulting web chart are based on the centrality of an organisation. For an intervention promoting collaboration, it may be useful to present a before-and-after network analysis in which graphically and numerically the level of integration of the network can be compared. • As a minimum, the analysis should **include all beneficiaries**, which are clearly distinguishable by type of organisation (e.g. SME, research organisation, semi-public organisation). • Preferably, the **links between actors are not just binary** (yes or no), but represent a quantity such the amount of licensing/contracting taking place. • The graphics should be backed up with **clear numbers and statistics** that support the graphical representation. • A network analysis is especially **useful for a 'medium' number of beneficiaries**, too little and the analysis provides little additional information, too many and the figure/analysis is too complex to easily understand.

[19] Mohammad Hosein Rezazade Mehrizi, Fereidoun Ghasemzadeh, Jordi Molas-Gallart, Stakeholder Mapping as an Assessment Framework for Policy Implementation Evaluation October 2009 15: 427-444

Tool	Good practice criteria in the use of key tools
Context analysis	To assess the relevance and effectiveness of an instrument there are a number of context analyses that can be made particularly in an ex-ante or ex-post evaluation. A context analysis can be used to understand what problems the programme is addressing and whether the intervention is (still) relevant to address the appropriate issues. A context analysis can also look at the wider policy portfolio to assess whether a particular programme has added value in relation to other existing programmes, aiming to establish similar effects or perhaps even the opposite effect. We can distinguish a number of commonly used forms of context analyses: • An analysis of the **socio-economic environment** in which the beneficiaries operate (looking at markets, technologies, financial markets, human resources, value chains, regulations, depending on their relevance to the programme) to understand how the intervention could help to improve the competitiveness of the targeted beneficiaries. An example is the Finnish evaluation of the thematic programmes in the telecoms sector, where fast changes in technologies and markets asked for a re-check of the relevance of the programme's objectives • An analysis of the **research and innovation system** to understand how an instrument or series of instruments fit with the key challenges and opportunities in the wider (national or regional) innovation system, in order to understand the relevance and effectiveness of the evaluated instrument better. • An analysis of the **policy mix** and particularly to identify overlaps, unnecessary duplication, gaps, complementarity and synergies between a portfolio of instruments

Appendix C Glossary of main evaluation terms

Additionality	Positive results that would not occur if the activity in question did not take place.
Baseline	Conditions present when activity begins. Changes can be measured against the baseline.
Beneficiaries	Members of the target groups who are the clients or beneficiaries of an action or policy.
Control group	A sample which can be directly compared with the target group, but which does not take part in the activity in question. A control group is used to indicate changes which have taken place in baseline conditions, whatever the activity in question.
Criteria	Principles, standards or values against which a thing is judged.
Deadweight	Expenditure to promote a desired activity that would in fact have occurred without the expenditure. Within the additionality framework these are the outputs that would arise under the basecase/counterfactual.
Displacement	The degree to which an increase in productive capacity promoted by government policy is offset by reductions in productive capacity elsewhere. Within the additionality framework it is the proportion of the project outputs accounted for by reduced outputs elsewhere.
Effectiveness	One of the criteria used in evaluation: The issue of effectiveness consists of asking whether results and impacts generated by the activities supported meet the objectives
Efficiency	One of the criteria used in evaluation: The issue of efficiency consists of examining the level of resource use (inputs) required to produce outputs and generate effects.
Evaluation	The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results (OECD).[20]
Formative Evaluation	Evaluation that takes place during the life of the programme, with a view to improving management and implementation (i.e. 'form' the programme). Formative evaluation asks how, why, and under what conditions something works, or fails to work, and is geared towards learning and programme or policy improvement.
Indicator	Condition that can be measured and that is present when the objective of the activity is being met.
Impact	Impact is achieved if the opinions of policy-makers or the day-to-day activities of practitioners are changed when outcomes of activities or partnerships are applied.
I-O-O-I model	The I-O-O-I (Inputs-Outputs-Outcomes-Impacts) model is a way of representing the intervention logic of the programme, based on the list of inputs,outputs, outcomes and impacts.
Monitoring	The process of collecting and recording information systematically to check progress against objectives during the life of a programme.
Objective	What a policy or programme sets out to achieve
Outcomes	The effects or end results of the programme's activities for its beneficiaries, clients, staff and other people and organisations – for example, the number of jobs created, or qualifications gained, or the number securing employment or entering further training.
Outputs	Conditions generated as a result of the activity – usually quantifiable (for example, training hours).
PLM	The Programme Logic Model lists the programme inputs, outputs, outcomes and impacts based on the programme objectives and rationale and outlines how the resources deployed as part of the programme (inputs) are intended to produce the expected outputs, outcomes and impacts (assumptions).
Relevance	One of the criteria used in evaluation: the issue of relevance consists of examining whether whether the objectives of an activity correspond with the needs, problems and issues it is intended to address.

[20] OECD, DAC Working Party on Aid Evaluation (2002), Glossary of Key Terms in Evaluation and Results Based Management

Reliability	The extent to which results would be the same, regardless of whoever conducts the research.
Stakeholders	The people and organisations either directly involved in a programme or interested in its work. They include direct beneficiaries; delivery staff (trainers, counsellors, advisers); representatives of local, regional or national public or semi-public agencies; enterprises; trade unions; voluntary or professional organisations; and funders.
Stakeholder analysis	Approach to systematically analyse the interests and roles of key individuals or groups who have an interest in a project, project outcomes, or the project's target population.
Stakeholder mapping	A type of stakeholder analysis that focuses on the assessment of a large number of actors linked together by various forms of relationship.[21]
Summative evaluation	Evaluation that 'summarises' the outcomes and impacts at a particular time in the programme life, with a view to judging and decide future resources allocation. Summative evaluation asks questions about the *impact* of a policy, programme or intervention on specific outcomes and for different groups of people.
Sustainability	One of the criteria used in evaluation: The issue of sustainability consists of examining whether the positive impacts on critical clients and beyond would continue into the future, even after the ending of an activity
Targets	Quantifiable expression of what an activity should achieve.
Utility	One of the criteria used in evaluation: The issue of utility consists of looking for expected and unexpected effects (i.e. those that were respectively identified and not identified at the design phase as objectives) and whether these, when they are positive, correspond with needs, problems and issues of different groups in society and the economy
Validity	How close evaluation methods get to providing a measure of the condition they claim to quantify.

Appendix D Overview of preferred methodology mix per type of instrument

Figure 10 Preferred methodology mix – R&D Business grants

	Mid term review	Ex-post evaluation
Data requirements	• **Minimum level** – Programme and project data to date • **Preferable option** – Baseline data on target group before intervention	• **Minimum level** – Full programme and project data across life cycle – Micro-level company data participants • **Preferable option** – Baseline data on target group before intervention – Annual micro-level company data entire target group – Micro-level data non-successful applicants
Desk research	• **Minimum level** – Programme material on rationale and objectives – Review of all programme material and project data • **Preferable option**	• **Minimum level** – Review of all programme material and project data • **Preferable option** – Review of reports on broader RTDI context and policies

[21] Mohammad Hosein Rezazade Mehrizi, Fereidoun Ghasemzadeh, Jordi Molas-Gallart, Stakeholder Mapping as an Assessment Framework for Policy ImplementationEvaluation October 2009 15: 427-444

	– Mapping of beneficiaries across target groups	– Review on thematic reports on markets, technologies, finance where relevant
Qualitative methods used	• **Minimum level** – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option**	• **Minimum level** – Reconstructing Programme Logic Model – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option** – Case studies – Interviews non-participants
Quantitative methods used	• **Minimum level** – No quantitative methods used • **Preferable option** – Survey with focus on dedicated issues programme performance	• **Minimum level** – Survey of beneficiaries – Survey unsuccessful proposers – Counterfactual analysis – Analysis administrative cost • **Preferable option** – Cost-benefit analysis – Counterfactual through econometric analysis with external micro-level data
Other methods applied		• **Minimum level** – Context analysis RTD-context • **Preferable option** – Context analysis markets & technologies – Benchmark similar initiatives

Figure 11 Preferred methodology mix – Innovation Vouchers

	Mid term review	Ex-post evaluation
Data requirements	• **Minimum level** – Programme and project data to date • **Preferable option** – Baseline data on target group before intervention	• **Minimum level** – Full programme and across life cycle – Numbers of vouchers issued and returned – Micro-level company data participants • **Preferable option** – Baseline data on target group before intervention – Micro-level company data entire target group – Micro-level data non-successful applicants
Desk research	• **Minimum level** – Programme material on rationale and objectives – Mapping of beneficiaries across target groups	• **Minimum level** – Review of all programme material and project data – Mapping of beneficiaries across target groups • **Preferable option** – Review of reports on broader RTDI context and policies
Qualitative methods	• **Minimum level** – Interviews sub-set participants, programme management, – Interviews research and technology centres where	• **Minimum level** – Reconstructing Programme Logic Model – Interviews sub-set participants, programme management, key

	Mid term review	Ex-post evaluation
	vouchers are used – Analysis of programme processes and governance • Preferable option – Mapping of beneficiaries	stakeholders – Interviews research and technology centres where vouchers are used – Analysis of programme processes and governance • Preferable option – Case studies – Interviews non-participants
Quantitative methods	• **Minimum level** – No quantitative methods used • **Preferable option** – Survey with focus on dedicated issues programme performance	• **Minimum level** – Survey of beneficiaries – Survey unsuccessful proposers – Counterfactual analysis • **Preferable option** – Cost-benefit analysis – Counterfactual through econometric analysis with external micro-level data
Other methods applied		• **Minimum level** – Context analysis RTD-context of target group – Policy portfolio analysis • **Preferable option** – Benchmark similar initiatives

Figure 12 Preferred methodology mix– Collaborative R&D grant instruments

	Mid term review	Ex-post evaluation
Data requirements	• **Minimum level** – Programme and project data to date • **Preferable option** – Baseline data on target group before intervention	• **Minimum level** – Full programme and project data across life cycle – Micro-level company data participants – Micro-level data participating R&D organisations • **Preferable option** – Baseline data on target group before intervention – Micro-level company data entire target group – Micro-level data non-successful applicants
Desk research	• **Minimum level** – Programme material on rationale and objectives – Review of all programme material and project data • **Preferable option** – Mapping of beneficiaries across target groups	• **Minimum level** – Review of all programme material and project data – Mapping of beneficiaries across target groups • **Preferable option** – Review of reports on broader RTDI context and policies – Review of technology trends in thematic field of programme
Qualitative methods used	• **Minimum level** – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option** – Stakeholder mapping – Focus groups with stakeholders	• **Minimum level** – Reconstructing Programme Logic Model – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option** – Case studies – Interviews with non-participants – Stakeholder mapping – Focus groups with stakeholders
Quantitative methods used	• **Minimum level** – No quantitative methods used • **Preferable option** – Survey with focus on dedicated issues programme performance	• **Minimum level** – Survey of beneficiaries – Survey unsuccessful proposers – Counterfactual analysis – Analysis administrative costs • **Preferable option** – Cost-benefit analysis – Counterfactual through econometric analysis with external micro-level data
Other methods applied	• Minimum level • Preferable option	• **Minimum level** – Context analysis RTD-context • **Preferable option** – Context analysis markets & technologies – Social network analysis

Figure 13 Preferred methodology mix– Cluster programmes

	Mid term review	Ex-post evaluation
Data requirements	• **Minimum level** – Programme and project data to date • **Preferable option** – Baseline data on target group before intervention	• **Minimum level** – Full programme and project data across life cycle – Micro-level company data core participants – Micro-level data participating R&D organisations • **Preferable option** – Baseline data on target group before intervention – Micro-level company data entire cluster / thematic domain
Desk research	• **Minimum level** – Programme material on rationale and objectives • **Preferable option** – Mapping of beneficiaries across target groups	• **Minimum level** – Review of all programme material and project data – Mapping of beneficiaries across target groups • **Preferable option** – Review of reports on broader RTDI context and policies – Review of reports on socio-economic context of cluster/ CC
Qualitative methods used	• **Minimum level** – Interviews sub-set participants, programme management, key stakeholders – Analysis of programme processes and governance • **Preferable option** – Stakeholder mapping – Focus groups with stakeholders	• **Minimum level** – Reconstructing Programme Logic Model – Interviews sub-set participants, programme management, key stakeholders, research organisations – Analysis of programme processes and governance – Analysis of governance and organisation cluster /CC • **Preferable option** – Case studies – Interviews non-participants – Stakeholder mapping – Focus groups with stakeholders – International peer reviews
Quantitative methods used	• **Minimum level** – No quantitative methods used • **Preferable option** – Survey with focus on dedicated issues programme performance	• **Minimum level** – Survey of beneficiaries – Analysis administrative cost • **Preferable option** – Cost-benefit analysis – Counterfactual analysis
Other methods applied		• **Minimum level** – Context analysis RTD-context • **Preferable option** – Context analysis markets & technologies – Benchmark similar initiatives – Social network analysis

technopolis |group| The Netherlands
Herengracht 141
1015 BH Amsterdam
The Netherlands
T +31 20 535 2244
F +31 20 428 9656
E info.nl@technopolis-group.com
www.technopolis-group.com