technopolis |group|

Version 6 February 2015

Measuring innovation policy across Europe

Reference model on indicator selection, operationalisation and sourcing

Part of the TAFTIE Task Force on Benchmarking Impact, Effectiveness and Efficiency

technopolis |group|

TAFTIE Task Force on Benchmarking Impact, Effectiveness and Efficiency

Task Force Composition: Pieter de Bruijn (RVO.nl, chair), Kirsten Voje (RCN), Paul Istvan Bencze (RCN), Ascension Barajas (CDTI), Malene Strømberg Rasmussen (DASTI), David Grønbaek (DASTI), Tanel Rebane (EAS), Kevin Flynn (Enterprise Ireland), Andreas Wildberger (FFG), Jeroen Fiers (IWT), Leo van de Loock (IWT), Pascal Fabin (LuxInnovation), Zuzanna Popis (PARP), Miroslav Janeček (TACR), Barbara Líšková (SIEA), Jussi Kajala (TEKES)

Report prepared by

technopolis |group|, February 2014

Matthias Ploeg

Geert van der Veen

Erik Arnold

Table of Contents

Table of Figures

1. Background and Context

1.1 Purpose of this reference model

This reference model, developed by Technopolis within the context of the TAFTIE Task Force on Benchmarking Impact, Effectiveness and Efficiency (TFBIEE) is an overview of ***criteria for best practise* in selecting, operationalizing and sourcing indicators for monitoring and evaluation in innovation policy**. This model (or guide) is aimed at monitoring and evaluation professionals in innovation agencies and other public organisations implementing innovation policies. Its goal is to provide a practical yet grounded set of guidelines that supports working with indicators innovation policy. Note that the model is mostly focused on the effects of innovation policy on firms.

This reference model can be seen as an specific extension of the TAFTIE TFBIEE Reference Model on Evaluation, which provides criteria for best practise for the entire monitoring and evaluation process in innovation policy. The original reference model is available from the TAFTIE website.

1.2 Background and context

TAFTIE (The European Network of Innovation Agencies) is an association with 29 European innovation agencies as its members that is aimed at promoting collaboration on the implementation of innovation policies. One of the main elements of TAFTIE's strategy is to learn from each other and exchange best practises on the design, implementation and evaluation of innovation instruments. A principal tool in sharing expertise and the joint development of knowledge are the Task Forces, which are focused initiatives of at least six members. In 2013, eleven members decided to initiate a task force on Benchmarking impact, effectiveness and efficiency.

The Task Force, assisted by Technopolis Group, pursued the benchmarking of 40+ innovation instruments across four types:

- Business R&D grants
- Collaborative R&D grants
- Competence centres and clusters
- Innovation Vouchers programmes

One of the main findings of the task force was, that while benchmarking has a large potential for giving valuable insight in an instrument's performance, the current evaluation practises are too divergent to already arrive at a meaningful (quantitative) benchmarking of indicators for effectiveness, efficiency and impacts. In order to address this, the Task Force developed a reference model for evaluations that is aimed at providing practical guidance in bringing evaluations to a comparable 'standard of good practice'

The benchmark study points out that policy makers should be careful in drawing conclusions from benchmarking quantitative indicators for effectiveness, efficiency and impacts, since instrument design aspects, as well as specific context can heavily influence outcomes. Benchmarking efficiency was found to be particularly challenging, since arriving at a comparable 'full-cost' figure for programmes is hampered by completely different internal accounting systems across different agencies:

> *"At best we can compare outcomes of similar questions in surveys across similar programmes, even though survey questions are also opinion based and have the usual bias issues. [...] This could be achieved if the agencies would agree to align some of their survey evaluation questions to the participants. [...] Where external data are used to compare the participants with a control group, an agreement*

could be made to use a set of common indicators. CIS data can be used as a common dataset across countries." [Technopolis: In search for a benchmark of impact, effectiveness and efficiency of innovation instruments, final report, 2014]

.

1.3 Objectives

Noting the current inconsistency between indicator selection and operationalization between agencies as well as the opportunities offered by potential benchmarking, the TAFTIE Taskforce on Benchmarking Impact, Effectiveness and Efficiency decided to initiate a follow-up project to build on the work carried out in 2013 and early 2014.

The objectives of this project were threefold:

1. To develop an addition to the reference model for practical guidance on indicator selection and operationalization, based on international best practises,
2. To develop a list of programme-level 'common indicators' based on logical framework analysis for measuring effectiveness and impacts which improves the opportunities of benchmarking as agencies are more likely to use identical indicators
3. To promote the use and relevance of these products by engaging the entire task force in an integral manner.

The reference model and list of common indicators have been developed for four specific innovation instruments, which were selected based on the fact that they encompass relatively similar and standardised interventions that lend themselves for benchmarking:

- Business R&D grants
- Collaborative R&D grants
- Competence centres
- Innovation Vouchers programmes

As can be seen, this selection has been aligned to the previous study, although cluster initiatives have been excluded due to the relatively complex and varying nature of intervention. The reference model and indicator list will be developed with an integral policy cycle in mind, including programme design (ex-ante), monitoring (continuous or mid-term) and ex-post evaluation.

This report presents the result of the reference model developed in the context of the task force described above. The results from the common indicator list are presented separately.

1.4 Approach and methodology reference model

The reference model provides key criteria for *good practise* in indicator use in innovation policy. It is designed to give both practical as well as theoretical guidance on how to select, operationalize and construct indicators for evaluating effectiveness and impact.

The reference model is based on three main sources:

- **Relevant literature on the use of indicators in innovation policy**
  - Recently published evaluation reports of innovation agencies or other public agents involved in public support for innovation
  - Relevant OECD / NESTA (e.g. the Frascati/Oslo manuals) and EU manuals
- **Best Practise case studies of monitoring and evaluation (see below)**
  - Case studies of specific evaluations of individual instrument of Task Force Members
  - Case studies of general monitoring and evaluation systems of Task Force Members

- Two external case studies
- **Collective knowledge and experience**
  - Of Task Force Members
  - Of Technopolis Group

1.4.1 Case studies

Practical case studies on good practises have been another key input for the reference model. Various Task Force Members have provided case studies based on a template developed by Technopolis. Case studies can either describe general monitoring and evaluation strategies or specific evaluations of instruments.

The case studies used for this report are (starred case studies are used explicitly in this report, others have provided background input)

- Task Force
  - SIEA Evaluation Strategy
  - RCN: Innovation Norway (On collaboration with statistical agencies)*
  - CDTI (On timed surveying and collaboration with statistical agencies)*
  - RCN: IPN (on long-term standardised monitoring systems)*
  - RVO.nl: Monitoring & Evaluation strategies:
  - TEKES: SHOK evaluation (specific)
  - DASTI: InnovationDenmark database (on integral firm-level database management) *
- External
  - NESTA (on integrating monitoring and evaluation into the policy cycle)
  - Czech Republic Research Centre Indicator Scheme / ERDF (on operationalizing indicators in and ERD setting)*

2. Introduction to indicator measurement in innovation policy

This section presents the key highlights of indicators for innovation policy performance measurements. The various existing methodologies and approaches mentioned here have fed into the development of the good practise criteria.

2.1.1 Policy performance measurement

The need for the measurement of policy performance through monitoring and evaluation is obvious, and has been discussed quite extensively in the *Reference Model on Evaluation*, of which this indicator reference model is an extension. In general, the two key goals of monitoring and evaluation are *learning* and *accountability*. The first is mostly aimed at supporting the improvement of policies internally (but others may learn from your lessons as well), the latter is aimed at ex-post justification of the use of public funds. There are various methodologies of assessing policy performance, and many national innovation agencies have their own manuals and procedures[1] related to policy performance measurement. As was identified in the Reference Model, all main methodologies use a logic model as the basis for an analysis of (among others) the effectiveness and efficiency of an intervention. A logical model or logical framework[2] establishes and makes explicit the (assumed) logical relationship between various levels of programme results. Within this reference model, we use a slightly adapted version of the OECD/DAC levels:

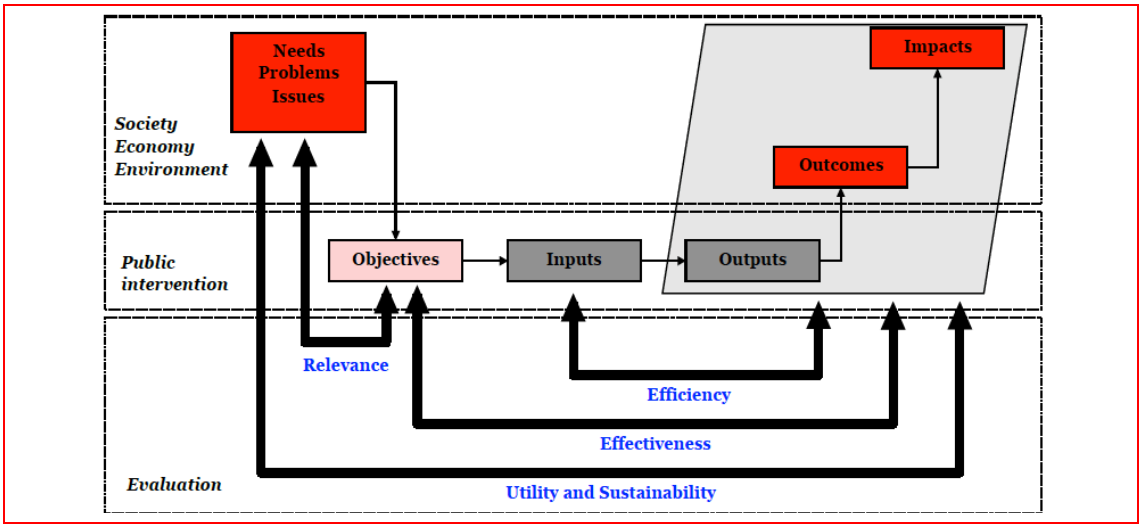
- **Inputs:** Resources used for instrument activities
- **Activities:** Scope and character of the actions taken by the agency
- **Output:** Service delivered by the intervention at the target group
- **Outcomes:** Short/Medium term effects at the level of participants
- **Impacts:** Long term effects at the participant / system level

These levels of programme objectives are directly related to the key aspects of evaluation (see figure below). Relevance is the relationship between the programme objectives and the societal needs; Effectiveness is the degree to which the activities result in the desired outcomes and impacts; Efficiency measures the relationship between the inputs and the effectiveness (value for money); Sustainability assesses whether an intervention has durably addressed the societal need and utility is the extent to which the underlying societal needs have actually been meet.

[1] For a full overview, see the introductory chapter of the Evaluation Reference Model

[2] The term Logical Framework is officially referring to a specific matrix-based method that had its origins in international development planning, however in this report it will also be colloquially used for describing the whole set of similar methods. For more information on the original method, see also: The Logical Framework Approach, Handbook for objectives-oriented planning, Fourth edition, NORAD, 1999

Figure 1 Relationship between logical model and evaluation



Technopolis, adapted from European Commision, Evalsed

Key further reading:

- OECD DAC Evaluation Standards, 2007
- World Bank, Khandker, Koolwal & Samad (2010). Handbook on impact evaluation

2.1.2 Performance indicators

Using this general policy performance framework, all five of its levels can be translated into programme objectives, the progress of which can subsequently be measured through (key) performance indicators. Indicators are:

> *variables, based on measurements, representing as accurately as possible and necessary a phenomenon of interest to human beings*[3].

Using the adapted OECD/DAC structure, five key types of indicators for policy performance assessment emerge: input, activity, output, outcome and impact indicators. Indicators are measurable proxies of objectives (or other concepts) and are as such always an imperfect representation of both abstract programme logic as well as implementation reality. This is the essence of the evaluator practitioner's challenge: how to develop indicators that excellently represent the logic behind public interventions which are actually feasible to implement. This reference model aims to provide practical good practise criteria that are specifically relevant for innovation policy instruments, but there are various general frameworks that have been a key inspiration and source for the good practise criteria. An important and accessible framework that helps in assessing the quality of indicator is the RACER-framework, developed for the European Commission[4]:

- **R**elevant: the indicators measure the concepts with validity and reliability
- **A**ccepted: the indicators are accepted by the main stakeholders, i.e. the TAFTIE TFBIEE members

[3] Joumard, R & Gudmundson, H. (2007) Functionalities of indicators and role of context.

[4] Giljum, S. & Lutter, S. (2009). Development of the RACER Evaluation Framework, SERI.

- **C**redible: the indicators are unambiguous and easy to expert (also for non-experts)
- **E**asy: the indicators are feasible in terms of data requirements
- **R**obust: the indicators are robust against manipulation / strategic behaviour.

Key further reading:

- Pidd, M. (2012) Measuring the Performance of Public Services: Principles and Practise. Cambridge University Press
- Handbook on monitoring and evaluating for results. UNDP, Evaluation Office, New York, 2002.
- Giljum, S. & Lutter, S. (2009). Development of the RACER Evaluation Framework, SERI.

2.1.3 Measuring innovation

There is a specific body of literature relating to the measurement of innovation using various types of indicators. The key starting point for the measurement of innovation has been the process driven by the OECD and Eurostat to enhance the standardised measurement of innovation *activities* and innovation *outcomes*. After a number of experimental surveys, the foremost standardised work became the Oslo manual (1992), including guidelines on measuring R&D (Frascati Manual). The most recent edition is the third edition (2005), and it the standard departing point for any evaluation practitioner when looking for guidance in the area of indicators for innovation policy. A revision process towards a fourth edition is currently taking place.

The Oslo manual has also been the key driver behind the *Community Innovation Survey*, which is a European enterprise survey in which Member States[5] have carried out 5 waves of surveys on innovation activities and outcomes. It should be noted that these are based on varying sample sizes and cover only a small part of the total enterprise population[6]. The indicators used in the CIS-survey are among the most standardised indicators within the context of measuring innovation[7].

Since supporting innovation is also one of the key pillars (or 'axes') of the European Regional Development Fund, the European Commission has undertaken various efforts to develop and standardise indicators for public (regional) interventions in the area of innovation policy. The EC uses a comparable system as the I-A-O-O-I model described before, although it only uses output and 'result indicators', the latter closely related to the outcome-level in the approach used in this report. A list of common indicators was developed for output indicators[8].

It is noteworthy to mention the Compendium of Evidence on Innovation Policy, developed by Manchester Business School and NESTA[9]. It provides an overview of almost all available evidence in both academic and policy literature (including evaluations) on the evidence of effectiveness and efficiency of various policy instruments.

---

[5] And Norway & Iceland

[6] It also excludes all companies smaller than 10 employees.

[7] With the possible exception of patent statistics, which are also well documented and standardised

[8] See: EC (2014). The Programming period 2014-2020. Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund.

[9] See also http://www.innovation-policy.org.uk/compendium/

Further reading & information:

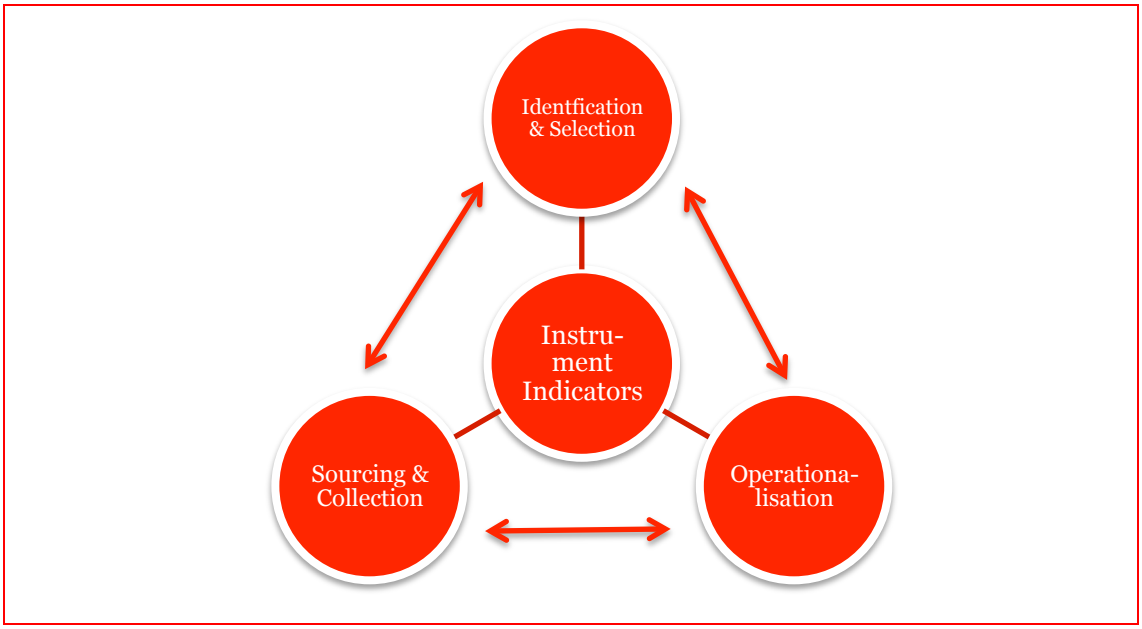
- OECD / Eurostat Oslo Manual (3rd edition)
- EC (2014). The Programming period 2014-2020. Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund.
- Gault, F. (2012) Handbook Of Innovation Indicators And Measurement. Elgar Publishing.
- Compendium on Evidence on Innovation Policy, Manchester Institute of Innovation Research (http://www.innovation-policy.org.uk/compendium/)

3. Reference Model on Indicators

3.1 Overview

Based on an analysis of good practises in using indicators for measuring effectiveness and efficiency of innovation policy instruments[10] using literature review, case studies and 'collective learning', a total of 15 good practise criteria were identified. These criteria can roughly be divided across three phases. Indicator identification and selection is generally the first step, when the relevant indicators for a specific programme are identified and selected. In the operationalisation phase, indicators are made concrete, measurable and objective. The sourcing & collection refers to the strategies and challenges around data collection for indicators. These three 'phases' are not linear, but rather are an iterative thinking cycle that is used when indicators are designed and implemented. The arrows between the phases in the figure highlight this interdependence, and each 'phase' has one specific good practise criterion on its relationship to the other phases. However, the use of these phases help to guide the thinking and strategy process around indicators in a useful way as it follows the main logical dependencies in indicator design. Specifically, it is designed to provide a good balance between the quest for the perfect indicator and the day-to-day pragmatism of working with indicators.

Figure 2 Good practise framework



The overview of 15 good practise criteria is given in the Table below. These criteria will be discussed in the next section of this report.

[10] In particular the four focus instruments: R&D grants, collaborative R&D grants, innovation vouchers and competence centres.

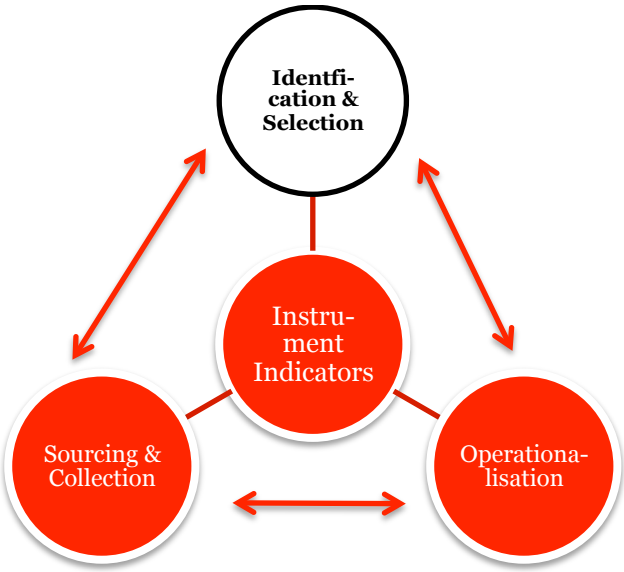
Figure 3 Overview of good practise criteria

Identification and selection	Operationalization	Sourcing & Collection
1. Derivation of relevant indicators through a logical framework analysis 2. Integration of indicator identification and selection in the policy cycle 3. Integration of indicator selection and identification into an integral monitoring 4. Indicator selection supports triangulation but avoids duplication 5. Indicators selection takes into account the operationalization & sourcing requirements	6. Indicators should be designed in a robust way 7. Indicators have to be well-communicated and acceptable to stakeholders while maintaining relevance and independence 8. Indicators should align with international standards and take into account opportunities for benchmarking. 9. Indicators operationalization should have attention for counterfactual aspects 10. The relevance of the selected indicator should be guarded throughout operationalisation	11. Indicators collection should be sourced from a credible, reliable and independent data source 12. Data collection is carried out in context of a long-term data strategy (together with key partners) 13. The data collection/sourcing should be carried out in a cost-effective manner 14. The data collection process does not cause unnecessary or disproportional burden on beneficiaries 15. Data collection provides feedback to indicator operationalisation and selection

Merit and risks of good practise principles

The good practise scheme summarised above and presented in further detail below provides very useful principles to guide indicator identification, operationalization and sourcing. Whereas the principles above, based on a logical framework approach, provide for structured guidelines helping monitoring & evaluation professionals, it should be stressed that any set of principles should not deter them from independent and holistic analysis. Particularly, the logical framework approach has the risk of focusing on *intended* consequences, whereas overlooking potential *unintended* consequences (either good or bad). If there is a risk of major unintended consequences, it is worthwhile to also include indicators for these unintended effects. Similarly, external effects may have profound impacts on both the objectives as well as the causality links between the objectives, warranting inclusion of external indicators.

3.2 Identification and selection of indicators

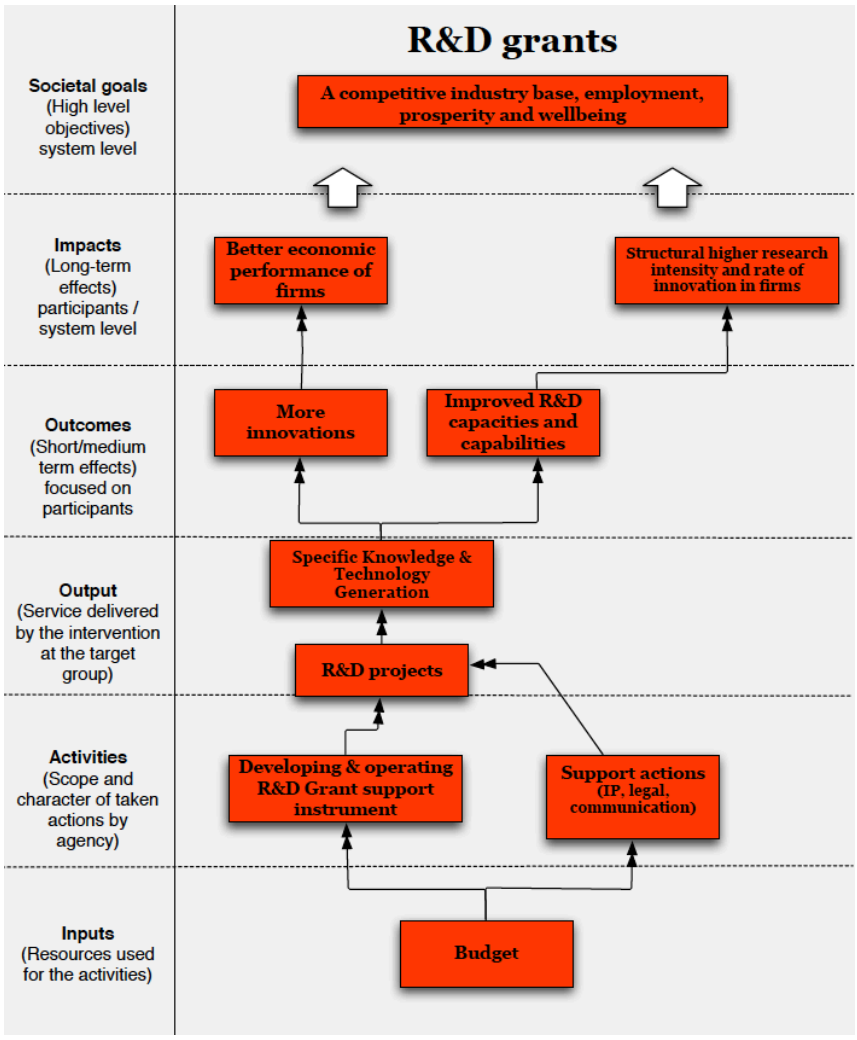


3.2.1 Key good practise criteria

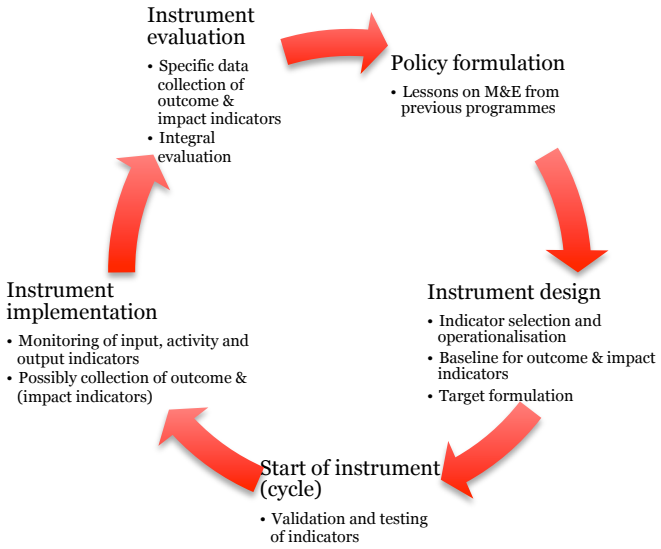
The first 'phase' in indicator design is the identification and selection of indicators. Of course, it is rather obvious that an M&E (monitoring & evaluation) practitioner should first identify the required indicators at a general level before operationalisation and sourcing. However, it is key in good practise criteria that indicators are derived using a logical and coherent approach, that their use is integrated in the policy cycle, is part of a broader monitoring strategy and uses the right level of prioritisation.

Figure 4 Identification and Selection

Good practise criteria	Description
1. Derivation of relevant indicators through a logical framework analysis	Indicators are a tool in monitoring and evaluating the effectiveness, efficiency and impact of a public intervention, such as an innovation policy instrument. In order to measure the effectiveness, efficiency and impact of programmes, it is good practise to identify the key outputs, outcomes and impacts an instrument intended to achieve, as well as the logical relationship between these various levels of results. These programme logic models, or logical frameworks, are the key input for identifying suitable indicators. Programme logic model elements (such as specific outputs or outcomes) can subsequently be translated to indicators. Using this method (instead of just listing various results or available indicators) has the following advantages: • The chosen indicators are almost by definition *relevant*, i.e. in principle they measure aspects directly related to a programme's effectiveness, efficiency or impacts • The logical framework provides a coherent and holistic overview of the various expected effects and the relationship between these effects. This ensures that important indicators are not left out, but also it promotes to leave out redundant information

Good practise criteria	Description
	• The approach to start from a programme logic reduces the risk of perverse indicator effects at the side of the policy makers. Of course, policy goals should inform the design of an instrument, and not the easiness of measuring. Starting with policy goals first and only bringing in feasibility later helps to ensure that the priorities stay clear. • The use of various structured levels of results (outputs, outcomes and impacts), as well as the logic between these results significantly expands the possibilities for advanced analysis of drivers and barriers to effectiveness, efficiency and impacts. It is particularly helpful to inform counterfactual method designs. For the reasons above, it is considered good practise to use the logical framework (or equivalent methods) integrally when selecting indicators for programme monitoring and evaluation[11]. An example of a logical framework is given below: 

[11] It should be stressed that logical frameworks present a 'model version' of actual reality While very useful in bringing a coherent and logical approach to monitoring and evaluation, the logical framework model is at times criticised for being overly linear and simplistic. A good evaluation takes into account the complexities of reality, such as feedback loops, heterogeneity of beneficiaries, dynamic causality patterns etc.

Good practise criteria	Description
2. Integration of indicator identification and selection in the policy cycle	Appropriate timing is key for indicator selection. It is therefore good practise to integrate indicator selection directly into the policy cycle (see figure below):  • **Policy formulation** should generally be informed by evidence on a societal need, a justification of government intervention and the performance of earlier programmes. The latter (generally provided by instrument evaluations) naturally provides input into the type and scope of intervention required, but should also provide reflection on the monitoring and evaluation strategy of the previous programmes. This feedback should be used to provide lessons for the formulation of an M&E strategy (including indicator selection) of the upcoming programme. • **Instrument design** is the phase when the policy priorities are operationalized into an actual instrument. During this phase, when the key modalities of an instrument are determined, monitoring and evaluation should be an integral element. Since instrument objectives should be completely clear at this stage, this is the phase when indicator selection and operationalization should take place. It is good practise to have a complete M&E framework in place during this phase, including data collection strategies for each indicator. In principle, this phase also has a baseline carried out for outcome and impact indicators in order to be able to set realistic yet ambitious targets • At the **start of the instrument implementation**, there is a good opportunity to test and validate all the indicators in the actual implementation settings. Should there be indicators found not to be acceptable or feasible, it is still possible to make adaptations with relatively modest implication. • During the **instrument implementation phase**, the chosen input, activity and output indicators are collected directly through programme monitoring, both internally at the agency as well at the level of beneficiaries (generally at specific moments such as contract signing, progress reviews and contract closure). When data collection at the outcome and impact level also requires annual monitoring, there will already be data collection at this level as well. However, in general this is only used for non-survey data collection (e.g. information from statistical agencies). For certain instruments it may be worthwhile to carry out a mid-term evaluation, where a survey of outcome indicators or expectations on output indicators (generally impact indicators are to far away at this stage) • During **instrument evaluation phase,** all indicator collection will be finalised and used for evaluation purposes. This generally means that data collection (surveys) on outcome and impact indicators is carried out (either by the agency

Good practise criteria	Description
	itself or an external actor, see also good practise criteria 12). The indicators are used to assess effectiveness and efficiency of the programme (and thereby also help to assess utility and sustainability). The evaluation should explicitly include an assessment of monitoring and evaluation practises in order to provide lessons for the next programme cycle.
3. Integration of indicator selection and identification into an integral monitoring strategy	Ideally, indicator selection and collection for a specific instrument does not take place in isolation of other instruments in the portfolio of an innovation agency. The advantage is seeking synergy between monitoring and evaluation strategies are especially large when there is a significant overlap between target groups of instruments, which often is the case in typical innovation instruments such as direct R&D grants, collaborative grants and competence centres. Best practise in this regard is to have an integral, client-based monitoring system as the key linking pin between evaluations (see Innovation Denmark case study below for a concrete example). The key idea here is that there is a central database of beneficiaries (usually identified by their companies' registration numbers) that keeps track of participation in individual instruments and can contain key background characteristics (e.g. company size, sector etc.). Monitoring data of instruments is actively linked to this central database, which can subsequently be linked to the databases of statistical agencies (e.g. to collect data about economic performance). There are a number of key advantages: • The ability to carry out portfolio evaluations allows for much more robust impact assessment methods. Additionally, it gives agencies a much better insight in their client base and their medium to long-term evolution of needs for innovation support • The burden on individual programme monitoring and evaluation can be substantially reduced, as indicators collect for a particular programme can be used for other instruments as well. This can reduce programme overhead and administrative burden on beneficiaries. A second-best alternative, when such a system is not available, is of course to consult with staff responsible for the monitoring and evaluation in other programmes with similar target groups, and to discuss indicator selection and collection strategies and seek for opportunities for co-ordination.
4. Indicator selection supports triangulation but avoids duplication	Indicators rarely measure an instrument outcome or impact with fully validity and reliability. In general, indicators are proxies that measure only part of the effects studied in an evaluation. For this reason, triangulation, the use of multiple methods/indicators to measure the same concept, is a useful strategy when it is particularly hard to measure a complex concept (as many outcomes and impacts of innovation policies are). In general, triangulation is particularly important at the outcome level, as it is outside the direct monitoring of an instrument programme, but still less general than impacts. • For instance, 'improved R&D capacities and capabilities' is a relatively broad concept. It might be useful to choose two indicators for evaluation: o FTE R&D personnel o R&D expenditure It is clear that these measures together provide a more reliable overview of the concept R&D capacities and capabilities. However, refraining from increasing the number of indicators unnecessarily is important as it increases the burden on respondents and evaluators (the parsimony principle). For instance, adding an indicator 'capital expenditure on R&D' to these two indicators would add very little information not already captured by these two existing indicators.
5. Indicators selection takes into account the operationalization	As discussed in good practise criteria 1, indicators are tools to measure progress of instrument objectives and should therefore follow programme logic. In an ideal setting, the indicators set are at all times completely tailored to the programme objectives. However, there are often significant constraints in terms of the feasibility

Good practise criteria	Description
& sourcing requirements	of certain indicators, both in terms of operationalization (such as acceptability to stakeholders) and data collection possibilities (certain indicators might not be available). Indicator selection therefore has to take into account these constraints in a balancing act between the ideal situation and a pragmatic approach where close proxies are taken instead of an 'ideal' indicator. This means in practise that there is a series of feedback loops between indicator selection, operationalization and collection, and it may be necessary to going through this indicator 'cycle' several times before arriving at a final set of indicators. It is good practise to make **clear reasoned and transparent decisions** taking both programme logic and feasibility into account, and document these in the monitoring strategy. In this way any mid-term or final evaluations can take into account these indicator challenges and adapt responsively to the considerations made during indicator selection.

CASE STUDY SUMMARY: Integrating Indicator design and selection into the Programme cycle (NESTA Creative Credits Evaluation)

Creative-Credits is a business-to-business voucher mechanism to support Small and Medium Enterprises (SMEs) in accessing creative services in the Manchester City Region.

NESTA, the UK Innovation Foundation, carried out a randomized control trial impact assessment of the scheme and published its report in 2013. This case highlights a – one could say almost extreme – example of integration between programme implementation and monitoring and evaluation.

The evaluation methodology started with identifying the logic model, based on earlier evidence and policy strategy.

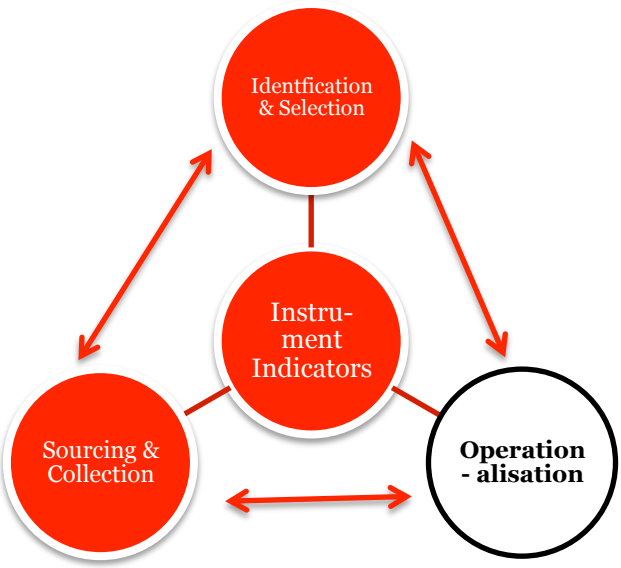
A key interesting point of this impact assessment has been that the monitoring and evaluation strategy has played a central role already during design stage, due to the choice of a randomized control setup

This had also a big impact on careful indicator design, especially given the repeated monitoring of non-participants, where attrition is always a big risk. Indicators were chosen to align with OECD definitions where possible with tailored additions, but generally measured on simple scales to enhance the ease of survey completion. In general, the number of indicators chosen was rather limited in order to further improve survey response rates. Key indicators were future innovation intentions (Likert Scale), sales growth (scale) and expected duration of sales benefit (scale).

Interesting indicators used: Future innovation intentions of firms, duration of sales benefits

Full version of case study is available in the Appendix.

3.3 Indicator operationalisation



3.3.1 Key good practise criteria

Operationalisation is the phase when indicators are specified and made more concrete, until they are measurable in a relevant and reliable way. Other key criteria are the acceptability of the indicators to the various stakeholders, alignment with international standards and a thorough assessment of counterfactual analysis opportunities.

Figure 5 Indicator Operationalisation

Good practise criteria	
6. Indicators should be designed in a robust way	It is good practise to take the **robustness criterion for indicators** into account both during the programme design phase (when robustness should be assessed ex-ante) as well as the programme evaluation phase (when the robustness of indicators should be validated ex-post and adjusted for future programmes). Robustness can be separated into two main issues: • **The indicator is not susceptible to manipulation.** It is obvious that potential for manipulation by stakeholders. An effective way to screen for potential manipulation is to identify stakeholders that both have an **interest** in and (partial) **control** of indicator collection. Interests varies by type of stakeholder, for instance companies may be interested in future grants of the agency. Interests do not only have to be those vis-a-vis the grant provider, but can also be on a personal level. For instance, scientists may have an incentive to overreport publication outcomes for personal career reasons. Note that instrument management itself is almost by definition an agent with high interest and control over indicator collection. It is good practise to improve robustness by working towards reduced control of stakeholders in data collection through more independent data collection (see criteria 11), but careful operationalization can also reduce the level of interest of stakeholders substantially. This is mostly achieved by working on objective operationalizations and clear definitions, thereby

Good practise criteria	
	decreasing the room for interpretation: o Acquisition of R&D-related equipment with a total value of more than 100k EUR in the last two years *instead of* 'substantial new investments in R&D capacity' o 'A new or substantially improved product or service that has been introduced on the market' *instead of* 'an innovation' • **The indicator does not lead to perverse effects** Perhaps even more worrying than manipulation of indicators (which leads to wrong evaluation results) is the risk of perverse side effects that certain indicators may cause. This has not just effect on measurability but may actually decrease effectiveness and efficiency. In research, an example of perverse effect has been the introduction of citations scores, which have increased the number of citations for each paper enormously, thereby obfuscating the actual purpose of citations. Indicators especially at risks are those that aggregate multiple different concepts (e.g. an innovation 'index'). Another set of perverse effects are those on the policy level, where indicators are used to oversimplify information on the programme for political purposes.
7. Indicators have to be well-communicated and acceptable to stakeholders, while maintaining relevance and independence	Acceptability (the A in RACER) is a key criterion during operationalization, and reflects the need of indicators to be responsive to the needs of key stakeholders in evaluation. Indicator operationalization (and selection) is not just a technical matter, but is based on interaction with a policy environment where stakeholders have interests and agendas. Evaluation practitioners should be aware of this and are in charge of maintaining support from stakeholders while not compromising an evaluation's relevance and independence. Examples of acceptability issues with stakeholders are given below: • **Various departments** within an agency may have different priorities regarding monitoring and evaluation. Good internal communication is the first basis of an acceptable indicator set. • **Beneficiaries** can be reluctant to divulge sensitive company information, such as profitability, margins on new products or detailed information on upcoming market introductions. • Evaluation practitioners might be put under pressure to ex-post facto align indicators with a **new policy agenda,** even if indicator operationalization and collection would not be feasible or relevant. An example is a strong focus at the moment to measure employment gains, even if a policy only very indirectly contributes to new job creation. • **Public co-funders** have a clear interest in indicator operationalization and might exert influence to include or exclude specific indicators. For example, a regional authority might want to see economic effects measured by region instead of at a national level. Dealing with stakeholder needs and interests can be challenging. It is good practise to **inform and consult key stakeholders** during indicator selection and operationalization and to take valid needs and concerns seriously. There is a clear need for a strong communication and co-operation between monitoring and evaluation departments and the 'core implementation' business units in all phases of indicator selection, operationalization and sourcing. Clear internal and external communication and dissemination are key strategies to ensuring a long-term commitment from stakeholders However, any major change away from a first-best indicator list (from a technical relevance perspective) for reasons of acceptability should be clearly explained in a report's methodology section (e.g. explain how sensitivity of business information has led to the choice of a second-best indicator).
8. Indicators should align with international	Alignment with international standards where possible is in general a good strategy during indicator operationalization. Key indicator frameworks to

Good practise criteria	
standards and take into account opportunities for benchmarking.	align with are: • OECD / NESTA indicators for innovation (Frascati and Oslo manuals) • Eurostat Community Innovation Survey (mostly based on the OECD work) • Eurostat Business Data • TAFTIE indicator lists There are two main reasons to align indicators with those from international frameworks : • **To benefit from the embodied knowledge and experience** that led to this indicators. The CIS survey has arrived at indicators after countless hours of discussion and research by top scholars in the field of innovation policy measurement. Their design has been assessed along the equivalence of all the good practise criteria in this report. **To greatly improve the strategic value of monitoring and evaluation by allowing for more and better benchmarking.** By aligning with international standards, comparison (or benchmarking) with similar programmes in other countries becomes possible. This has enormous opportunities for evaluation, as it allows for a better assessment of performance, but also to key insights that lead to very relevant follow-up questions. While international alignment is a very valuable strategy, as described above, there is a real risk that too much focus on harmonisation may lead to a decrease in willingness to pilot new, innovative indicators in innovation policy. As theoretical insights and data availability evolve, so should indicator operationalization, and harmonisation should not be a barrier to progress. In such situations, it may be wise to keep using the 'old indicator' in parallel while a new indicator is gaining international acceptation.
9. Indicators operationalization should have attention for counterfactual aspects	With the increasing scope and availability of quantitative data relating to programme effects, there is an increasing need to use counterfactual methodologies to assess the *net effects* of government intervention (or additionality). Various counterfactual methods (for more information see the TFBIEE Reference Model on Evaluation) have different data requirements for indicators. Since there is a clear order of preference between various methods, it is crucial to take into account the data requirements necessary during the evaluation phase (which is often years away). In practise, many counterfactual techniques make use of the following data charactersistics: • **Various data points over time.** Regular monitoring of outcome and impact indicator levels at the start of the programme (baseline), specific intervals during the implementation phase and possibly extending afterwards greatly enhances the options for counterfactual assessment • **Collect information on background characteristics of beneficiaries.** Counterfactual methodologies often rely on matching strategies or other identification methods for control groups that require a base set of background characteristics • **Keep track of rejected applicants or other potential control groups.** Although an ideal control group also includes complete programme outsiders, having a list of rejected applicants with their contact details and background characteristics can contribute to a better counterfactual methodology. Also, an indicator on the reason for rejection can be extremely useful in constructing a control group.
10. The relevance of the selected indicator should be guarded throughout operationalisation	While relevance - the fact that an indicator measures the result in a valid way – is in principle ensured by deriving indicators from a logical framework, it is important to ensure validity throughout the entire operationalization phase as well. In general, indicator operationalization (and later sourcing and collection) requires concessions and pragmatism to find indicators that are acceptable to stakeholders and feasible in terms of implementation. However,

Good practise criteria	
	this pragmatism should not go as far as to undermine the validity of the indicator. It should in principle be avoided to select indicators based only on what is already available or easy to collect. In particular, its should be avoided to select internal programme indicators which provide no meaningful measure for measuring impact, effectiveness and efficiency, such as: • The number of successful projects (to the outside world, both the terms 'project' and 'successful' are ambiguous and their scale is difficult to interpret) • The number of supported companies (what does 'support' entail? Does sending a newsletter count as well?) Often these problems can be resolved by being more precise and measured from the beneficiaries' perspective, for instance: • The share of granted R&D projects that met or exceeded the technical objectives • The number of companies receiving R&D grants

CASE STUDY SUMMARY: Indicator operationalization in ERDF setting (Czech Research Centre Scheme)

In the Czech OP RD&I research centres outside Prague are supported, under ERDF, to acquire state-of-the art equipment and set up research programs. The aims are to create research centres of international excellence (Priority Axis 1 (PA1), 6 centres supported) and strengthen technology transfer to industry (PA2, 42 centres). All the centres have to report regularly on monitoring indicators, in relation to ex-ante defined targets (inspired by the requirements of ERDF, implemented by the Management Authority for the OP that is part of the Czech Ministry for Education, Youth and Sports).

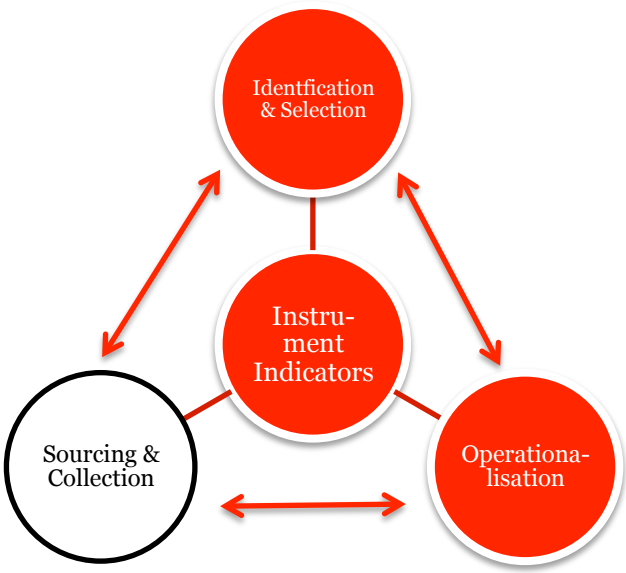
The OP RD&I has the dual aim to develop research as well as to promote (knowledge transfer to) industry. The choice of the monitoring parameters reflects this dual aim. The set of parameters (Figure 1 in the Appendix) is focused on activities, outcomes and outputs and not so much on impacts. This is in line with the goals of the programme: impacts (research impacts at universities and economic impacts for industry) will only be realised in the long to medium term and therefore after the present financing period (until 2015).

The centres themselves collect all the present indicators. In this way they are responsible for their own monitoring, which helps them focus on the management actions to achieve the targets.

Most important challenge was to develop a simple, uniform monitoring methodology for a large set of centres, in a country where there is little experience with the management of professional research institutes focused on international excellence and valorisation of research results in industry

Full version is available in Appendix A

3.4 Indicator Sourcing and collection



3.4.1 Key good practise criteria

Indicators are in the end all based on data, either sourced from third parties or collected through programme monitoring or participant surveys. There are various key aspects that constitute good practise, including cost-effective data collection and sourcing, building long-term partnerships with data partners, the use of a credible and reliable source and the reduction of administrative burden on participants.

Figure 6 Indicator sourcing and collection

Good practise criteria	
11. Indicators collection should be sourced from a credible, reliable and independent data source	A good indicator has high external **credibility**, both to external experts and the public at large. Credibility is closely linked to objectivity and independent data collection. In general, indicators collected by independent agencies (especially statistical agencies) is considered highly credible, whereas data dependent collected or provided by stakeholders with a direct interest can be less credible. Especially when indicators are self-reported (e.g. through participant survey), credibility and independence of data collection is an essential good practise criterion. Besides the fact that a close relationship between programme management and indicator data collection can lead to public *perceptions* of bias, there is also a real threat of *actual* measuring bias due to the fact that such data collection may elicit strategic behaviour on the side of the respondents. This is especially an issue when beneficiaries are dependent on the agency to receive financing in the future. This is one of the reasons why evaluation and impact assessment is often outsourced to either relatively **independent** evaluation units or to external evaluators. This specifically applies to *outcome* and *impact indicators*. Output indicators fall under the programme implementation, are often easily verifiable and can be directly monitored by the programme management without too much risk of bias. **Reliability** is especially important when monitoring and evaluation relies on regular (e.g. annual) updates of indicator values. Data providers should have

Good practise criteria	
	clear, consistent and standardised data collection protocols to ensure reliability throughout the years
12. Data collection is carried out in context of a long-term data strategy (together with key partners)	While many indicators can be sourced directly from programme monitoring data, or will be obtained using dedicated surveys, agencies will be dependent on co-operation with statistical agencies for certain specific indicators of a confidential nature. These indicators include economic performance of firms (turnover, export, employment etc.) as well as indicators on innovation performance (e.g. data from the Community Innovation Survey). As opposed to publicly available external data sources (e.g. patent data, webometrics), these indicators are not available in-house but require a special procedure in working with the statistical agencies. In a best practise situation, an innovation agency will therefore have a close, long-term co-operation with their national statistical agency. There are three main models: • Innovation agencies contract statistical agencies to match their own internal beneficiary database and to provide regular update on key indicators at an aggregate level • Innovation agencies have dedicated staff with a good knowledge of the statistical data available, which work (either on distance or on location) regularly with the statistical data in a confidential setting (see CDTI case study) • Innovation agencies contract a third party to manage the data collection process, either ad-hoc for an evaluation or with regular intervals for a monitoring strategy. Since the matching process is quite technical and requires good preparation and agreements in terms of confidentiality, it is considered best practise to have a clear working relationship with the statistical agency (e.g. through a MOU). These can be particularly productive if the relationship is beneficial both ways. For instance, in Denmark (see case study Innovation Denmark), the Statistical agency using the innovation agency's integral monitoring system to validate the results from the R&D and Innovation survey.
13. The data collection/ sourcing should be carried out in a cost-effective manner	Innovation instruments are spending taxpayers' money and are often subject to tight budgets. It is therefore essential that not just an instrument itself is cost-effective (i.e. delivering value for money), but that the same applies to monitoring and evaluation. Good, appropriate data collection for good indicator use can put pressure on M&E budgets. There are various strategies to improve cost-effectiveness: • A clear view on the **priorities for monitoring and evaluation** in the context of the instrument. M&E is a tool for learning and accountability, not a goal in itself. Focusing on measuring the most important programme objectives (e.g. by using the 80-20 rule of thumb) makes sense for both applications. It may for instance not be necessary to measure in detail all the specific outcomes various support activities (e.g. coaching, networking) if this is not the key part of the intervention and could be covered using more qualitative methods. Of course, an instrument which primary focus is on such activities cannot discard such outcomes. • Apply the **principle of proportionality** in a sensible way. Large, complex instruments require larger data collection efforts than small focused interventions. The relationship is not linear, there is a certain minimum and there are certain scale benefits. • Use **long-term integral data collection strategies** (see also criteria #12) to save on costs. • The use of clear, focused terms of references when outsourcing the collection of indicators.

Good practise criteria	
14. The data collection process does not cause unnecessary or disproportional burden on beneficiaries	Administrative burden is one of the most common complaints of instrument beneficiaries. There is a concrete risk that indicator collection through surveys at the level of companies can risk increased administrative burden, especially when companies receive multiple surveys on their participation within a short timeframe. In general, data collection burden should be **proportional** to the scope of intervention received. There are a number of key good practise strategies that help to reduce administrative burden in data collection: • **Co-ordination of monitoring and evaluation** at the agency (and preferably even higher) level. Joint data collection, for instance across several instruments, with a flexible survey design prevents companies from being asked the same question twice • **Keep it short and simple**. Company surveys often suffer from being overly long and technical. In general, it is important to focus on the need-to-know indicators only. Furthermore, it is good practise to keep jargon at bay and use language to which instrument participants can relate. • **Use tailored surveys for each group of beneficiaries**. Do not send out a general survey to all beneficiaries, but ensure that each survey is tailored to their specific scope of involvement. • **Use user-friendly data collection methods**. Use electronic surveys that allow companies to start with the data provision at one point and continue later • **Use up-front, clear and transparent communication about expectations for contribution to M&E.** All communication, including those during the application phase should indicate that beneficiaries can expect to be asked to contribute to monitoring and evaluation of the programme. The perceived burden is generally decreased when participants are made aware in advance. The other advantage is that response rates may increase (thereby leading to an increase in validity and usefulness of the evaluation) when participants know this is 'part of the deal'.
15. Data collection provides feedback to indicator operationalisation and selection	Just as good practise criteria 5 and 10, the key message of good practise is to have a continuous cycle of feedback and indicator improvement during the programme design phase, but also afterwards (see good practise criteria 2). Specifically, certain reliability, acceptability or robustness issues may only arise during actual data collection. Although indicators frameworks should in general be consistent over time, there may be due cause to adapt the framework if certain indicators are simply not compliant to key good practise criteria anymore. Any change should be transparently and clearly document.

CASE STUDY SUMMARY: Collecting data efficiently using modular surveys (Innovation Projects, Norway)

Innovation Project in the Industrial Sector (IPN) is a policy instrument, administered by The Research Council of Norway (RCN), which objective is to stimulate research and development (R&D) activity in trade and industry. Møreforsking Molde, a research company, has supported RCN for more than 20 years with empirical studies of companies that have received support from RCN to user-driven innovation projects in industry. Data are available for a variety of user-driven projects from the portfolio from 1995 to 2012. The survey data are supplemented with project and user-specific data from the RCN application database. The surveys are conducted in three steps: the year after project start-up (baseline), the year after close-out, and a long-term post project survey four years after close-out.

The survey is designed using a modular structure, which' main idea is to collect empirical data in order to gain insight into the complex issues of economic impacts and rate of return of investment in innovation, as well as the external effects. This modular economic evaluation system for user-oriented research schemes, comprise:

- an *ex-post* module focusing on traditional impact indicators, implemented as soon as the programme has terminated;
- a long-term module which is used to evaluate the economic impacts about ten years after project start-up;
- an infrastructure module which links up with information from R&D institutions to enable identification of external effects;
- an econometric module which combines project information with company time series to allow estimation of the rate of return; and
- an *ex-ante* module describing the economic and external effects anticipated at programme start-up.

Over the years both the methodology and the surveys have been developed and improved, to continuously fit new needs for knowledge. Since the data from the studies can be broken down at the program level, they have also contributed to raise the awareness internally in RCN – among the program managers/teams – of benchmarking the results between the different innovation programs. Some programs order yearly a special analysis of their own activities, using the results to monitor their program development.

One of the advantages of having an independent research institute carrying out these studies, and not carrying them out in-house, is that the institute guarantees the respondents full confidentiality/anonymity. The Research Council cannot identify the companies and check up on their answers. The disadvantage of this is, of course, that the Møreforsk-data cannot be connected to other data sources (e.g. the R&D&I-statistics). The Council is, however, trying to find solutions to the latter problem – hoping to solve it in the near future.

Interesting indicators used: Socio-economic impact (including upgrading of skills and expertise, knowledge dissemination, market spillovers)

Full version available in Appendix A

CASE STUDY SUMMARY: Merging survey data with external data: (combination of CDTI, DASTI and RCN)

There are various opportunities for linking internal data sets and surveys with external sources. A key source is often the national statistical agency, which has access to company data. The examples below illustrate various methods in collaborating and working with external data providers. CDTI staff uses both statistical agencies and commercial data providers to provide input for impact assessments. DASTI is working on an integral connection between their own rich data set (Innovation Denmark Database) and the statistical agencies, where the eventual idea is a co-operation for mutual benefit. Innovation Norway has followed another strategy and partially outsourced economic impact assessment to the statistical agency itself.

CDTI (R&D Business Support)

The impact survey system of CDTI is based on two electronic surveys that supported firms must complete at two time points: 1st) after finishing the technological development of the R&D project (**results survey**) and 2nd) two years after the market launch of the innovations (**ex-post survey**). The first one is mandatory; meanwhile the second one is a voluntary questionnaire.

Data are stored and merged with the corporative database at the CDTI, so that every authorized CDTI employee will be able to see the individual questionnaire for any project. Quality control (coherence of answers) is carried out by a CDTI employee. Moreover, using statistic and econometric software, an exhaustive review of coherence is carried out.

This corporative data set is merged with external commercial sources (SABI database), in order to gain access to objective economic indicators. The output is a complete dataset containing information about firms, projects, R&D results and economic performance. Yearly the CDTI elaborates and publishes an exhaustive, descriptive report on the web site. This report contains a project-level analysis.

Since many indicators are similar to those used by the CIS, it is possible to have a first impression of the results obtained by the supported firms compared to the general innovation performance of the Spanish industry. In a second step, under a collaboration agreement with the Spanish National Statistic Agency (INE), a complete impact assessment is be carried out using control samples.

Interesting indicators used: Simultaneous innovations (product + process) , impact on exports, technological leadership achieved.

DASTI (InnovationDenmark database)

The InnovationDenmark database is an internal agency database which includes information on approximately 11,400 projects and 13,700 Danish and international participants. There are approximately 11,000 unique Danish companies in the database.

The data is collected as a part of the regular administration of the programmes, and reported to the InnovationDenmark database once or twice yearly.

The database is based on the principle of collecting the best data while focusing on minimising the administrative burden of all parties involved. This is addressed through focusing on key variables, which enable retrieval of detailed data through other registries. It is for instance possible to find a company's NACE-code, location and company form through the company's registration number.

The database is public and everyone can get a retrieval of the database. Researchers and other persons with access to Statistics Denmark's anonymised micro data can link the database with

business data, educational data or other relevant registries. This is done by Statistics Denmark who will merge the database with the relevant registries following confidentially procedures.

Furthermore DASTI have a plan to share the data with Statistics Denmark, whom will validate the data and use it to validate the Danish R&D and Innovation statistics as well as give researchers access to a unique database. At present it is not possible to say when this will be implemented.

Innovation Norway (Statistics Norway)

Innovation Norway and Statistics Norway have in 2013 started to collaborate on developing a method to calculate the effect of the support Innovation Norway provides.

Innovation Norway has outsourced the econometric analysis to the Research department of Statistics Norway.

Statistics Norway has defined control groups of comparable enterprises that have not received support (selected by propensity score matching), and then compared them with Innovation Norway's customers.

The matching is based first region and NACE Level 2 and then within this group on propensity score matching based on firm data (assets, concentration of ownership, NACE level 3) from the initial year.

Appendix A Full Case Studies

A.1 DASTI InnovationDenmark database

Author: Karina Lisberg, Malene Strømberg Rasmussen and David Grønbaek (DASTI)

A.1.1 Background and context of the case

The Danish Agency for Science, Technology and Innovation (DASTI) has increasing focus on collecting data from the research and innovation schemes managed by DASTI and the Innovation Fund Denmark. Data from 16 national and international schemes are combined and harmonised in one joint database called the InnovationDenmark database.

The InnovationDenmark database includes approximately 11,400 projects and 13,700 Danish and international participants. There are approximately 11,000 unique Danish companies in the database.

A.1.2 Indicator selection and operationalization process

The data are collected as part of the regular administration of the schemes, and reported to the InnovationDenmark database once a year. As a standard the following data are collected for all research and innovation schemes in DASTI:

- *Variables for each project:* Name of schemes, project title, grant status (rejection or approval), application year, start date of the project, end date of the project, total budget and total grant
- *Variables for the participating partners in each project:* Company registration number (CVR number), budget per partner and grant per partner

The database is based on the principle of collecting the best data while minimising the administrative burden of all parties involved. This is addressed through focusing on key variables, which enable retrieval of detailed data through other registries. It is for instance possible to find a company's industrial classification code (NACE code), location and company type through the company's registration number.

The next step is widening the scope of the database, either by including new schemes or by joining the InnovationDenmark database with databases of other business support schemes.

Furthermore DASTI is working towards sharing the database with Statistics Denmark. Currently everyone can get a retrieval of the InnovationDenmark database. Researchers and other persons with access to Statistics Denmark's anonymised micro data can link the InnovationDenmark database with business data, educational data or other relevant registries. This is done by Statistics Denmark staff who will merge the database with the relevant registries following confidential procedures. In the longer term, the plan is for Statistics Denmark to host the InnovationDenmark database so that the database will automatically become part of their anonymised micro data. It is not possible to say when this will be implemented.

At present, DASTI provides a retrieval of the InnovationDenmark database to Statistics Denmark, who validates the data and use it to validate the Danish R&D and Innovation statistics.

A.1.3 Lessons learnt

- The InnovationDenmark database offers new opportunities for analysing the Danish research and innovation system. The following are examples of analyses based on the InnovationDenmark database.

- *Mapping users and linkages of the Danish research and innovation system:* The participants' company registration number is a unique key which makes it possible to identify participation in other schemes. In this way it was possible to examine whether the companies participate in one, two or more schemes.
- *Measuring the impact on companies' productivity of participating in a scheme:* The InnovationDenmark database was used to ensure that the companies have not participated in other research and innovation schemes two years before and two years after participating in the examined scheme. In this way the effect of each scheme was isolated. The next step is to analyse the impact on companies' productivity of participating in more than one research and innovation scheme.

- The administrative burden must be manageable. Hence, focus is on collecting key variables. On the other hand, it is important not to omit important variables. Including a variable retrospectively is demanding.
- The definition of data must be clear.
- The database is continually validated to increase the quality of data. It is important not to underestimate how time-consuming this task is.
- It is important to gather the data at the right level. DASTI did only gather company registration data, meaning that it is not possible to distinguish between different production units – for instance different departments of a large company.

A.2 CDTI Impact Assessment Methodology

Author: Ascension Barajas Inigo (CDTI)

A.2.1 Background and context of the case

How could an innovation agency put into practise a result monitoring system integrated in the operational management of the public instrument? This case aims to be an example.

The selected instrument is low-interest credits with partial non-reimbursable funding for individual R&D business projects.

The system is based on two electronic surveys that the supported firms must complete at two time points: 1st) after finishing the technological development of the R&D project (**results survey**) and 2nd) two years after the market launch of the innovations (**ex-post survey**). The first one is mandatory; meanwhile the second one is a voluntary questionnaire.

The results survey questionnaire is available to firms at the official CDTI software application, as soon as they declare that the last milestone of the project has been finished. They must complete one questionnaire per finished project. The supported companies are asked for the market launch date (year) of the innovations. Two years after this date they will receive the ex-post questionnaire, in order to contrast previous data and increase knowledge on the economic impact of the project.

A.2.2 Indicator selection and operationalization process

The results survey is based mainly on the Community Innovation Survey, although it includes other questions relevant to our specific case. The ex-post survey is shorter and more focused on economic impact. In general, there are six information blocks (for relevant indicators see annex):

1. Firm activity: general indicators
2. Technological results
3. Economic additionality
4. Behavioural additionality
5. Intellectual property instruments
6. Tax credits (only for results survey)

The list of indicators included in questionnaires is review once a year. We maintain all the existing questions and, if necessary, we add new ones. It is very important to maintain the internal coherence of the questionnaires over the time in order to build long time series. New questions could be proposed by CDTI personnel. The Studies Department assesses the convenience of the modification and decides the best way to include it.

Data are stored and merged with the corporative data base at the CDTI, so that every authorized CDTI employee will be able to see the individual questionnaire for any project. Quality control (coherence of answers) is carried out by a CDTI employee. Moreover, using statistic and econometric software, an exhaustive review of coherence is carried out.

This corporative data set is merged with external commercial sources (SABI database), in order to gain access to objective economic indicators. The output is a complete dataset containing information about firms, projects, R&D results and economic performance. Yearly the CDTI elaborates and publishes an exhaustive, descriptive report on the web site. This report contains a project-level analysis.

Until now, the results database contains information about more than 3,000 R&D projects finished in 2011, 2012 and 2013, with a total budget of 2,500 million € and a public co-funding of 1,800 million €.

Information extracted from the ex-post survey is also available for 300 projects with commercialized results. These data are very sensitive and have a great value for both; internal management and external diffusion of R&D results. Until now, one edition of the ex-post survey has been launched, with an answer rate of 60%. Data are managed exactly in the same way as the results survey information. Both data sets have been merged in order to analyse if expected economic and commercial results have been finally achieved.

Since many indicators are similar to those used by the CIS, it is possible to have a first impression of the results obtained by the supported firms compared to the general innovation performance of the Spanish industry. In a second step, under a collaboration agreement with the Spanish National Statistic Agency (INE), a complete impact assessment will be carried out using control samples.

A.2.3 Lessons learnt

- Surveys Design: with the CIS as reference point, the final questionnaires have been elaborated taking into account experience of CDTI technical staff and academic experts.
- Coordination between surveys: both, results and ex-post surveys are part of the same task; they should be designed from a complementary point of view.
- Timing: the term in which R&D results are achieved could be quite different between industry branches or even between firms in the same branch. Special care must be taken selecting the survey timing to capture reliable information.
- Data management: software should allow managing a large amount of data coming from different sources.
- Open text questions: Although this information is quite difficult to manage from a statistical point of view, it extends the knowledge about R&D results. Many firms use open questions to explain which market or financial obstacles they must face.
- Internal communication: Interaction with other corporative units. Relevant to get feedback from technical experts who are working with firms on a daily basis. Their indications could be very useful to understand results.
- External diffusion: Relevant to encourage firms to participate in the survey and to communicate to the public the achieved R&D results.

A.3 RCN Innovation Norway

Author: Paul Istvan Bencze (RCN)

A.3.1 Background and context of the case

Innovation Norway is one of the government's most important instruments for innovation and profitable business development throughout Norway.

The innovation agency is owned by The Ministry of Trade, Industry and Fisheries (51%) and county municipalities (49%)

Innovation Norway and Statistics Norway have in 2013 started to collaborate on developing a method to calculate the effect of the support Innovation Norway provides.

Three main types of instruments have been investigated: innovation grants ("innovasjonsoppdraget"), regional grants ("distriktsoppdraget"), and loans and guaranties ("bankoppdraget"). High-risk loans ("risikolån") are treated as grants, not (ordinary) loans.

A.3.2 Indicator selection and operationalization process

Innovation Norway has outsourced the econometric analysis to the Research department of Statistics Norway.

Statistics Norway has defined control groups of comparable enterprises that have not received support (selected by propensity score matching), and then compared them with Innovation Norway's customers.

The matching is based first region and NACE Level 2 and then within this group on propensity score matching based on firm data (assets, concentration of ownership, NACE level 3) from the initial year.

"The indicators which have been selected are differences in growth in turnover, labor productivity and rentability. In addition are survival rate for entrepreneurs an indicator. Supportive analysis and information has been produced regarding growth in the number of employees and value added.

A.3.3 Lessons learnt

The econometric approach, which was employed for the first time in 2014 as part of a new the Management By Objectives (MBO) system for Innovation Norway, has been well received. In 20014 the Ministry of Trade, Industry and Fisheries was awarded the 'Better state prize' for the development of new MBO system. The focus on clear goals for the value Innovation Norway is contributing to firms and society was highlighted by the jury.

Out of the total of 15000 firms having received support 7200 have been possible to match with control groups.

A.3.4 References

Identifying effects on firm performance of support from Innovation Norway: A methodological note. Arvid Raknerud, Statistics Norway. April 23, 2014

Innovation Norway report for 2013.

http://www.innovasjonnorge.no/PageFiles/68761/Annual%20report-2013-shortversion-in-english.pdf

A.4 RCN Innovation Project in the Industrial Sector (IPN)

Authors: Kirsten Voje and Paul Istvan Bencze, Research Council of Norway

A.4.1 Background and context of the case

Innovation Project in the Industrial Sector (IPN) is a policy instrument, administered by The Research Council of Norway (RCN), which objective is to stimulate research and development (R&D) activity in trade and industry, particularly activities that promote innovation and sustainable value creation. The Project Owner (the formal applicant that must be a company) and any partners will generally fund at least 50 per cent of the project costs. In general the partners will include research institutes and/or universities that contribute expertise and R&D services.

Møreforsking Molde has for more than 20 years conducted empirical studies of companies that have received support from RCN to user-driven innovation projects in industry. Data are available for a variety of user-driven projects from the portfolio from 1995 to 2012. The results are related to indicators for the measure of socio-economic impacts for the portfolio of projects, and include indicators on the development of expertise, knowledge dissemination, commercialization and input additionality.

All of the indicators are based on surveys among firms who are formally the project owners of the innovations projects. The survey data are supplemented with project and user-specific data from the RCN application database. The surveys are conducted in three steps: the year after project start-up (baseline), the year after close-out, and a long-term post project survey four years after close-out. Questionnaires are answered by project managers or executives in the firms responsible for the projects. Other participating firms and organisations are not surveyed, which leaves some impacts and effects unaccounted for.

A.4.2 Indicator selection and operationalization process

The indicators are used to illuminate competence building and corporate financial performance in firms, as well as external effects in terms of knowledge dissemination and other effects that form the basis for economic gains outside the supported firms. Innsatsaddisjonalitet, the extent to which projects would have been implemented without public support, is also an important element in performance measurement.

The main idea with this modular structure is to collect empirical data in order to gain insight into the complex issues of economic impacts and rate of return of investment in innovation, as well as the external effects. This modular economic evaluation system for user-oriented research schemes, comprise:

- an *ex-post* module focusing on traditional impact indicators, implemented as soon as the programme has terminated;
- a long-term module which is used to evaluate the economic impacts about ten years after project start-up;
- an infrastructure module which links up with information from R&D institutions to enable identification of external effects;
- an econometric module which combines project information with company time series to allow estimation of the rate of return; and
- an *ex-ante* module describing the economic and external effects anticipated at programme start-up.

The modular economic evaluation system		
Evaluation type	Assessments	Time of evaluation
Ex-post evaluation	Short-run economic and external effects	Just after programme termination
Long-term evaluation	Economic impacts and external effects	About ten years after project start-up
Infrastructure evaluation	Network and external effects	Just after, then ten years later
Econometric evaluation	Time series analysis combining company and project information to estimate rate of return	Time series
Ex-ante evaluation	Start-up expectations of economic and external effects	The year after start-up

Main conclusions from the studies:

- although there is great uncertainty regarding financial yield and profit, the result is probably satisfactory;
- there are considerable financial effects in the long run due to the companies' investments in the development of knowledge and the creation of networks;
- there are considerable external effects in other companies through the development of knowledge-based capital and R&D co-operation and networks;
- there is too low additionality and risk profile in the portfolio;
- Innovation Project in the Industrial Sector is probably a satisfactory instrument in cost-benefit terms, although these calculations suffer from great uncertainty.
-

A.4.3 Lessons learnt

These studies have been well received, especially in the Ministry of Trade, Industry and Fisheries. This ministry in particular has a strong focus on documenting the value creation and socio-economic impact of public investment in R&D&I. The Møreforsk-studies are (perhaps) the only studies that have been carried out over such a long period of time. They have led to better evaluations. Over the years both the methodology and the surveys have been developed and improved, to continuously fit new needs for knowledge. Since the data from the studies can be broken down at the program level, they have also contributed to raise the awareness internally in RCN – among the program managers/teams – of benchmarking the results between the different innovation programs. Some programs order yearly a special analysis of their own activities, using the results to monitor their program development.

One of the advantages of having an independent research institute carrying out these studies, and not carrying them out in-house, is that the institute guarantees the respondents full confidentiality/anonymity. The Research Council cannot identify the companies and check up on their answers. The disadvantage of this is, of course, that the Møreforsk-data cannot be connected to other data sources (e.g. the R&D&I-statistics). The Council is, however, trying to find solutions to the latter problem – hoping to solve it in the near future.

A.5 Research Centre Evaluation Model Czech Republic / ERDF

Author: Geert van der Veen (Technopolis Group), with input from Miroslav Janecek (TACR)

A.5.1 Background and context of the case

In the Czech OP RD&I research centres outside Prague are supported, under ERDF, to acquire state-of-the art equipment and set up research programs. The aims are to create research centres

of international excellence (Priority Axis 1 (PA1), 6 centres supported) and strengthen technology transfer to industry (PA2, 42 centres). All the centres have to report regularly on monitoring indicators, in relation to ex-ante defined targets (inspired by the requirements of ERDF, implemented by the Management Authority for the OP that is part of the Czech Ministry for Education, Youth and Sports). The monitoring parameters (commented on by each centre in a self assessment) are also an important input for an international peer review of every centre that is carried out after 3-4 years of the start of the centre. These evaluations are mainly formative, focused on improving the management of the centres. However the results of all the evaluations are also used for an assessment at programme level.

A.5.2 Indicator selection and operationalization process

The OP RD&I has the dual aim to develop research as well as to promote (knowledge transfer to) industry. The choice of the monitoring parameters reflects this dual aim. The set of parameters (Figure 1 in the Appendix) is focused on activities, outcomes and outputs and not so much on impacts. This is in line with the goals of the programme: impacts (research impacts at universities and economic impacts for industry) will only be realised in the long to medium term and therefore after the present financing period (until 2015).

The centres themselves collect all the present indicators. In this way they are responsible for their own monitoring, which helps them focus on the management actions to achieve the targets.

Key indicators that could be transferred to other 'competence centre programmes' are the volume of contract research (as indicator for interaction with industry) and the volume of funds for R&D from international resources (as indicator for international excellence). The number of publications in peer-reviewed journals is a useful indicator for the quantity of publications of a minimum quality standard, however it would be useful to include more qualitative information (e.g. in which impact factor quartile the publications are). 'Applied research results' is not considered useful as an indicator: it does not really show application potential and, in the case of patents, has a large time-lag.

The main strength of this methodology is however not in the indicators but in the exposure of centre management to foreign experts in a formal formative evaluation.

A.5.3 Lessons learnt

- Most important challenge was to develop a simple, uniform monitoring methodology for a large set of centres, in a country where there is little experience with the management of professional research institutes focused on international excellence and valorisation of research results in industry
- A robust set of monitoring parameters was developed, in line with the intervention rationale of the programme.

This set of parameters is used for monitoring, but also as the foundation for **formative** peer review of each centre, focused on bringing centre management at international level, with adequate attention for scientific excellence and application of research results by industry.

A.6 NESTA Creative Credits: Randomized control experiment

Author: Matthias Ploeg (Technopolis Group)

A.6.1 Background and context of the case

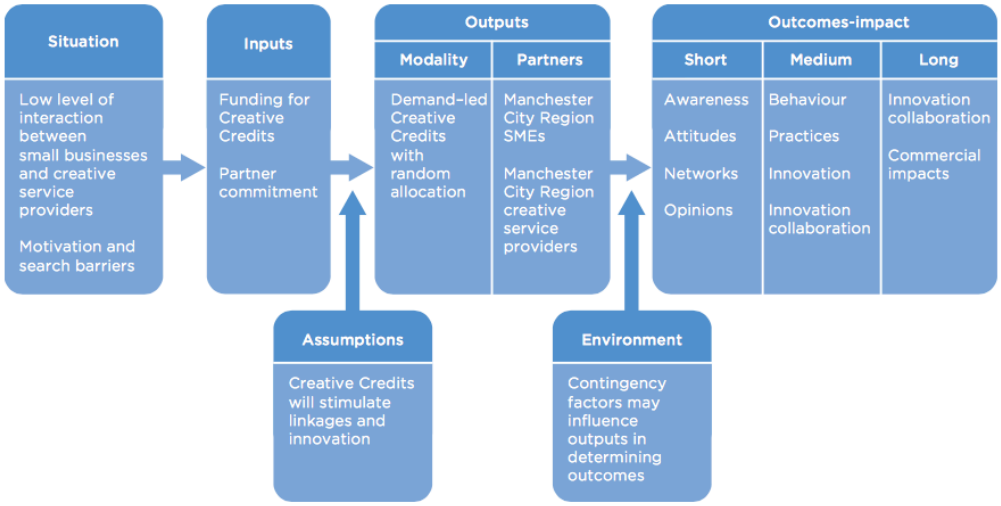
Creative-Credits is a business-to-business voucher mechanism to support Small and Medium Enterprises (SMEs) in accessing creative services in the Manchester City Region. The intervention logic behind this programme is that research has shown that creative companies can spur innovation in other businesses. During a pilot in 2009 and 2010, companies received 4000 GBP to spend on creative services, with a requirement to contribute 1000 GBP themselves.

NESTA, the UK Innovation Foundation, carried out a randomized control trial impact assessment of the scheme and published its report in 2013. This case highlights a – one could say almost extreme – example of integration between programme implementation and monitoring and evaluation.

For more information see the full report: Bakshi et al. (2013) Creative Credits: A Randomized Controlled Industrial Policy Experiment. NESTA

A.6.2 Indicator selection and operationalization process

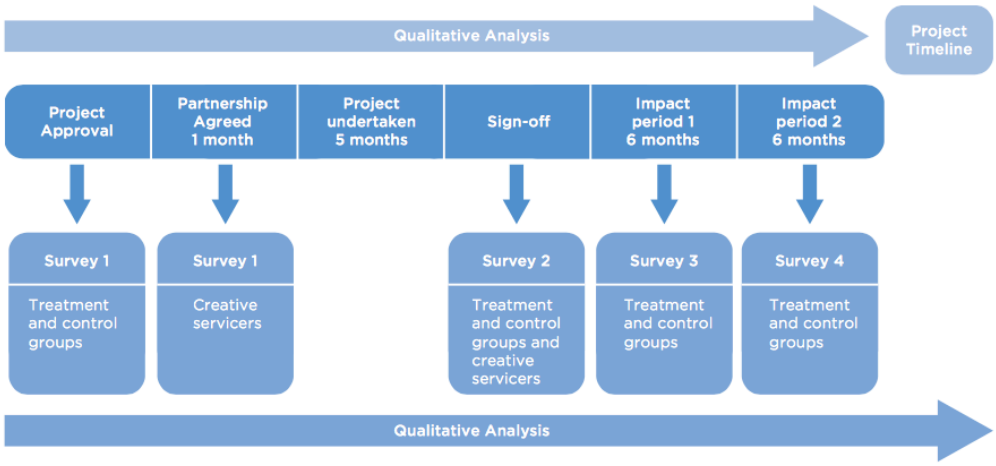
The evaluation methodology started with identifying the logic model, based on earlier evidence and policy strategy. The figure below (from the report) shows that there has been a clear process of logic model analysis.



A key interesting point of this impact assessment has been that the monitoring and evaluation strategy has played a central role already during design stage, due to the choice of a randomized

control setup[12]. An overview of the project timeline below (from the report) shows that various surveys have been organised around programme implementation.

Figure A1.1: Timeline for Creative Credits projects and their evaluation



This had also a big impact on careful indicator design, especially given the repeated monitoring of non-participants, where attrition is always a big risk. Indicators were chosen to align with OECD definitions where possible with tailored additions, but generally measured on simple scales to enhance the ease of survey completion. In general, the number of indicators chosen was rather limited in order to further improve survey response rates. Key indicators were future innovation intentions (Likert Scale), sales growth (scale) and expected duration of sales benefit (scale). Note that this evaluation used the system of input, output and behavioural additionality.

A.6.3 Lessons learnt

- It can be very productive and worthwhile to include monitoring and evaluation aspects right from the very start of the programme design phase.
- Randomized control experiments can be a feasible method in the case of vouchers, where there is a large group of potential participants.
- Keeping participant surveys short and to the point with acceptable and easy-to-understand indicators will help to improve response rates and thereby indicator reliability and validity.

[12] A RCT is a method where participants are randomly selected from a total population. As such, there is no (self-) selection bias which generally distorts the measurement of economic impact when a standard counterfactual design (non-random control group) is used.

technopolis |group| The Netherlands
Spuistraat 283
1012 VR Amsterdam
The Netherlands
T +31 20 535 2244
F +31 20 428 9656
E info.nl@technopolis-group.com
www.technopolis-group.com