Short guide on P2P evaluation / impact assessment

Project no.	645782
Project acronym:	ERA-LEARN 2020
Project full title:	Strengthening joint programming in Europe
Funding scheme:	Coordination and support action
Start date of project:	1 January 2015
Duration:	36 months

Deliverable D 3.4a

Short guide on P2P evaluation / impact assessment

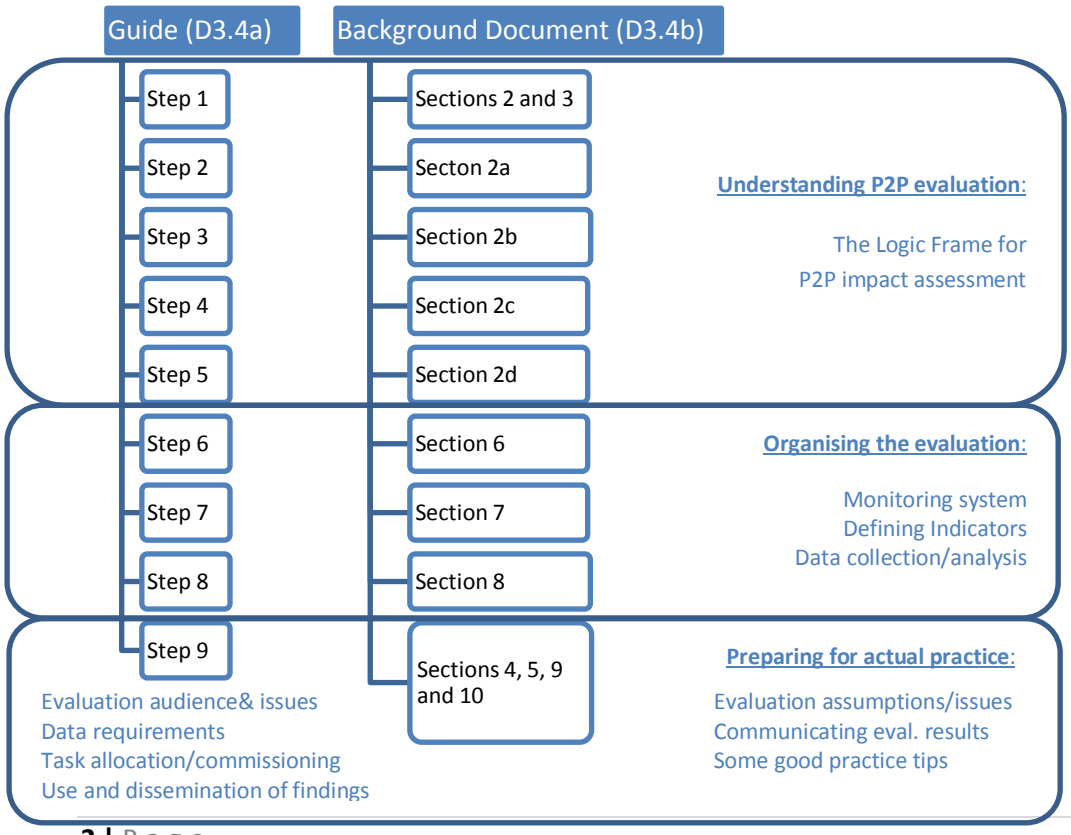
Due date of deliverable:	(M18)
Actual submission date:	(M18) Revised version (M24)
Dissemination Level:	PU
Organisation name of lead contractor for this deliverable:	The University of Manchester (UNIMAN), United Kingdom
Key authors:	Effie Amanatidou, Paul Cunningham, Deborah Cox
Contributors:	Angus Hunter, Optimat, United Kingdom
	Michael Dinges, Anja Köngeter, Susanne Meyer, AIT, Austria

Introduction

This Guide is targeted at Public-Public-Partnership (P2P) management actors responsible for evaluation tasks and aims to

- establish a common evaluation language / understanding of terms and concepts;
- improve understanding of the value and usefulness of impact assessment (IA) activities and results;
- address the needs of the different stakeholders, i.e. higher-level policy-makers and public officials at European, national and regional levels as well as P2P network coordinators and participants, in identifying and demonstrating P2P impacts;
- help the P2P community understand what it takes to carry out IA exercises alone or in collaboration with external evaluators;
- provide advice and recommendations on how to use impact assessment results at both the strategic/programme and operational levels.

Impact evaluation is a complex and constantly evolving topic, based on extensive past experience and evidence. Therefore, it is strongly advised that this short Guide be read in conjunction with Deliverable D3.4b "Background document on P2P evaluation / impact assessment", which provides additional information on the concepts used in the Guide as well as examples from P2P-relevant work. Both the Guide and the Background document are available at https://www.era-learn.eu/monitoring-and-assessment/Monitoring-and-impact-assessment-of-networks. The correspondence between the structures of the two documents 'Guide for P2P evaluation / impact assessment' (D3.4a) and the Background Document (D 3.4b) is shown below.



Contents

Setting the scene: Evaluation/impact assessment of P2Ps

Evaluation/impact assessment of policy measures is important as it generates new knowledge on the functioning of a policy intervention that can be used for accountability, learning and steering of an initiative.

Evaluations can be conducted at a number of stages of intervention planning and execution: before (ex-ante), during (interim) or after the intervention has been implemented (ex post). Ex ante evaluation brings together evidence and arguments concerning the likely consequences of the intervention's activities and tests the assumptions underlying the design of the intervention and its anticipated results and outcomes against the rationale for its creation.

In the interim mode, evaluation information is produced on how effectively the initial objectives set are still relevant and which progress in achieving these objectives has been made by the specific measure/programme. This also helps policy-makers to decide when and if to act on the programme (e.g. to continue, stop, extend, modify it, etc.) or to revisit the rationale of the programme and to assess its continued relevance in response to the specific problem or challenge set. Thus, this use is generating policy feedback. Ex post evaluations are conducted after the conclusion of a particular phase of the measure or after its lifetime.

At the same time, evaluations produce information on the effectiveness of the operationalisation of the programme, i.e. the design, management and implementation of the measure. By using this information programme owners and programme management can learn lessons as to how to improve policy design, management and procedures. This is the use of evaluation for operational learning purposes.

A third major use of evaluation is to inform a system impact approach. This is when the information produced is about the broader impacts of the measure, i.e. insights as to how the programme in question complements a range of other programmes; how it fits into the broader research and innovation system at national or international level; how it affects actors other than its beneficiaries and how it creates wider social, environmental, economic and technological impacts.[1]

All three of these uses of evaluation are relevant for P2P networks and the respective use and focus of evaluation activities should be tailored towards the needs of programme owners and the main stakeholders of the initiatives. Evaluations can produce information about the effectiveness of mobilisation of policy and research and innovation communities around their respective network activities (e.g. mobilisation of trans-national R&I funding around a commonly agreed topic, alignment of national R&I programmes, research support in the form of research grants, summer schools, training, etc.). This can support policy-decisions in relation to the continuation of specific activities. Evaluations can also produce information on the efficiency of management of certain activities (e.g. the operationalisation joint calls, exchange schemes, etc.). This information is relevant for improving and even harmonising the operational procedures of participating agencies/ministries across Europe in the P2P examined.

[1] The three uses of evaluation were adjusted from Miles, I. and Cunningham, P. (2006): A Practical Guide to Evaluating Innovation Programmes. Brussels: ECSC-EC-EAEC.

A key objective of P2Ps at European level is to create a European innovation system by promoting the coordination of national activities related to three key functions of R&I systems[2]:

Orientation: definition of the objectives to be reached by the research system (areas in which research projects should be conducted or challenges that should be addressed) and the adoption of the appropriate budgets.

Programming: translation of these high-level objectives into research programmes including the definition of scientific priorities and of the financial modalities (

Execution: conduct of the research projects by the researchers in the research performing organisations (research centres, universities, private companies).

Evaluations and impact assessments of P2Ps therefore may consider **impacts** at various actors and levels:

- on researchers and organisations that are the direct beneficiaries of the P2P networks' activities (execution function)
- on the funding / management agencies participating in the P2P (national ministries and research funding organisations / programming function),
- on the national/regional research and innovation systems policy level in the P2Ps and at the European level (national governments, Council, GPC).

Together, they produce systemic impacts and thus it is highly relevant to examine how the P2P network fits into the broader research and innovation system at national or international level and how it affects actors other than its direct R&I beneficiaries and creates wider types of impacts.

The clearer the intended purposes of an evaluation are outlined in advance, the more useful an evaluation can be. An evaluation framework should be fixed in the guidance documents of programmes or projects to ensure its sustainability. When designing evaluation activities, a number of questions should be addressed, including[3]:

1. Who is performing the evaluation (i.e. external evaluators, managing authorities, or funding agencies) and how will transparency be ensured?

2. What are the objectives of the evaluation?

3. What is the overall evaluation methodology?

4. What tools will be used to perform the evaluation?

5. What data sources are needed to perform the evaluation and how are data obtained?

[2] European Parliament (2016), Public-public partnerships in research: The joint programming process, October 2016.
[3] Eval-Inno (2012), RTDI – Evaluation Standards, fostering evaluation competencies in research, technology and innovation in the see region, 2012.

6. At what times must different evaluation outputs be finalized (reports, meetings, etc.)?

7. What will be the results/outputs of the evaluation?

8. How will the evaluation outputs be used, published and communicated to decision-makers?

The following principles enhance usability of evaluations:

- the more the policy-makers are informed, the more useful the evaluation;
- when policy-makers and/or programme managers are more engaged during the course of an evaluation, the more useful evaluations become;
- the greater extent to which wider stakeholders are involved in the evaluation, the more useful evaluations are and vice versa;
- dissemination of the results widens the usefulness of an evaluation - utilisation and dissemination plans should be part of the evaluation design;
- we need to understand how decisions and activities occur in a diffused decision-making model such as a P2P network;
- we need to recognise that networks evolve through stages of development and that their shape and structure are important influences on their development;
- we need to acknowledge that it takes time to organize networks effectively and show results - however, it is possible to see what progress is being made.

Step I: Defining the intervention logic of the P2P

Understanding the objectives and expected results of an intervention is a fundamental issue for any evaluation exercise and the authorities responsible for an intervention have to make sure that objectives are clearly formulated and structured, providing a clear idea to all actors (policymakers, programme managers, evaluators and beneficiaries) of what constitutes an intervention success or failure. [4]

A useful tool to ensure "evaluability" of P2Ps and to manage expectations is to construct a 'logical framework' or 'log-frame' of a P2P intervention. A logical framework outlines the connection between ends and means of an intervention. It comprises the underlying rationales of an intervention (a specific challenge to be addressed), formulates specific objectives that should be achieved by the intervention, and provides an overall roadmap on how specific activities of the intervention are intended to produce immediate outputs, intermediate outcomes that can be attributed to the activities of the intervention, and ultimate impacts.

By formulating this logic chain of objectives, activities, and expected results, key performance measures that should result from the P2P can be clearly identified and mechanisms designed by which they can be captured either by routine monitoring or by dedicated evaluation data collection approaches.

It is strongly recommended that the construction of a logical framework is undertaken prior to the design and implementation of a P2P and that all relevant stakeholders are involved in the exercise. In practice, it is often the case that the rationale and design of a programme is formulated in such a way that a logical framework is implicitly conceived rather than being explicitly constructed. Thus, the creation of a logical framework may have to take place post hoc, after the implementation of the P2P. However, it is important that all relevant stakeholders are engaged at whatever stage the log-frame is constructed.

For setting up an intervention logic it is important to first revisit the rationale for the P2P – i.e. to consider the reasons for which the P2P was established, and the problems, situation or challenges it aims to tackle. Next, the primary objectives that the P2P was intended to achieve in order to address these problems and challenges can be identified. In turn, this leads to the main activities or actions that are undertaken in order to best address these objectives. These will generally entail the input of resources, typically time and human or financial resources, although political and infrastructural preconditions may also have to be met. Each action or activity will be expected to generate outputs and results that are relevant to the objectives of the P2P. Some of these will have immediate outcomes and effects, whilst others may lead to impacts that will take longer to emerge. Several outcomes and effects may combine to eventually realise such longer term impacts. As a final consistency check, an examination of the eventual outcomes and impacts should find that they map, in some way, onto the original problems and challenges which the P2P was originally set up to address.

The basic elements of a Logical Framework are illustrated in Figure 1 below.

[4] cf. Eval-Inno (2012), RTDI – Evaluation Standards, fostering evaluation competencies in research, technology and innovation in the see region, 2012.

Figure 1: Basic elements of Logical Framework for impact assessment framework

An illustrative Logical Framework for P2P evaluation is provided in Figure 4 (section 3 'A suggested Logical Framework for P2P evaluation') in the Background Document (Deliverable 3.4b) available at https://www.era-learn.eu/monitoring-and-assessment/Monitoring-and-impact-assessment-of-networks. Examples of Logic Frames of specific P2Ps that have embarked in building an evaluation framework are also provided in the same section of the Background Document.

There are certain factors that affect the degree to which an evaluation/impact assessment study is useful. When evaluating a (P2P) network and in order to manage expectations, a number of issues need to be taken into account:[5]

- the clearer and more verifiable the objectives of a programme, the more useful its evaluation;
- we need to consider that networks have a "chain of impact" that includes the network's impact on its members, the members' impacts on their local environments, and the members' combined impact on their broader environment. Evaluations designed to examine impact must understand the relationship between these three levels and be clear about where their focus lies;
- we need to recognise that network impacts evolve through stages of development and that their shape and structure are important influences on their development.

The following sections (Step 2-5) provide explanations and guidance on how to 'build' each of the framework elements.

Step 2: Linking the challenges and objectives

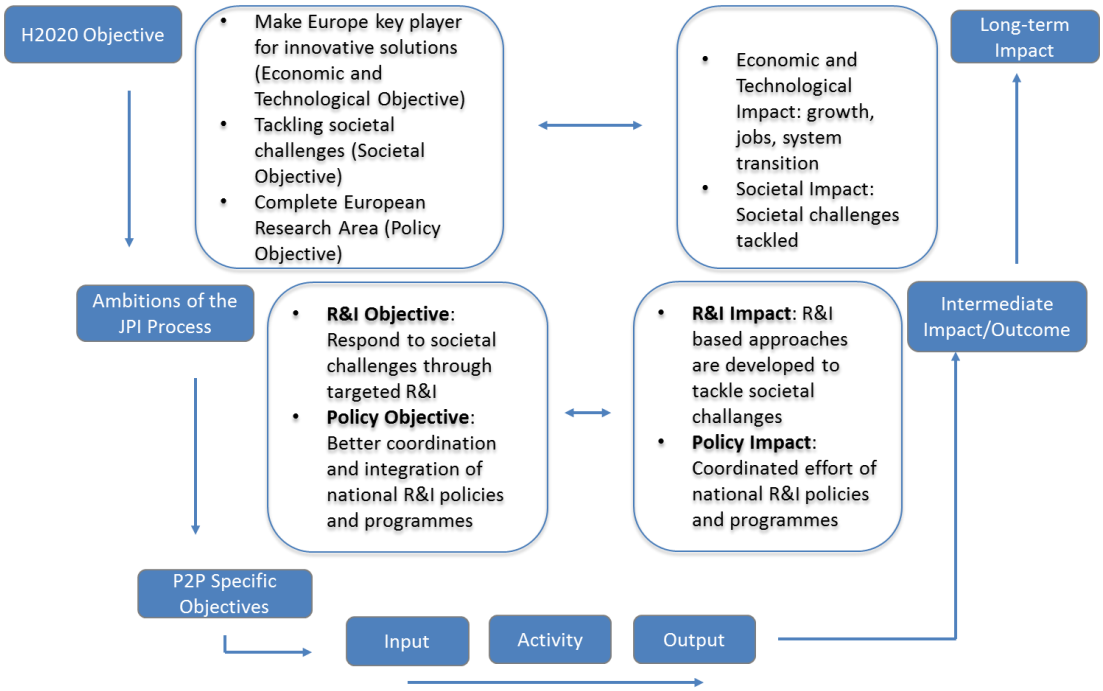
For devising an intervention logic of P2Ps it is important to understand that the P2P intervention does not take place in isolation from its wider environment, but the P2P objectives are embedded into the long term strategic objectives of the EU 2020 priority of enabling smart and inclusive growth, the Innovation Union objectives, H2020 objectives and R&I activities, and the national policy objectives and activities. In this regard Figure 2 illustrates how overarching European Union policy objectives, and H2020 objectives in particular, relate to the overarching ambition of P2Ps that can be structured along two main analytical dimensions:

1. The ambition to respond to societal challenges through the provision of targeted research and innovation activities, and

[5] Adjusted from Network Impact and Center for Evaluation Innovation. 2014. Part 1 of a Guide to Network Evaluation Framing Paper: The State of Network Evaluation, July 2014.

2. To enhance the likelihood of success to solving the challenges through better co-ordination and integration of the national R&I policies and programmes.

Figure 2: Basic elements of Logical Framework for impact assessment framework



Source: Illustration by AIT - Austrian Institute of Technology

By way of illustration, some examples of challenges from the Joint Programming Initiative JPND shows how societal challenges are linked to specific R&I objectives and related R&I policy objectives:

- The ageing of the European population.
- The number of European citizens suffering from neurodegenerative diseases.
- The rising costs of healthcare.
- The debilitating and largely untreatable character of disorders that are linked with age.
- A need to improve understanding of neurodegenerative diseases and to provide new approaches for prevention, diagnoses and treatment.
- A need to effectively provide healthcare, social care and support to optimise quality of life at all stages of the illness.

Then there is the assumption that an effective response to these challenges calls for certain improvements in relation to policy:

- A more coordinated and harmonised approach for research efforts.
- Reduction of unnecessary duplication and fragmentation of research activities.
- A more holistic, multi-factoral and multi-disciplinary research approach.

In relation to the specific scientific area addressed are the following requirements:

- Strengthening the linkages between basic, healthcare and clinical research.
- Stimulating the interaction between different disease factors and research disciplines.
- Development of a longitudinal approach in research.

This line of argumentation forms the rationale for the existence and design of JPND.

Source: Abida Durrani 2016 JPND Framework for JPND monitoring, evaluation and impact assessment. Presentation at the ERA-LEARN Workshop 18 May 2016, ZonMw.[6]

Following the identification of its rationale, i.e. the major challenges and problems that led to the establishment of the P2P, as with any policy intervention, the next step is to consider the specific objectives that must be achieved.

Taken together, P2P objectives can be combined in what is known as an **objectives hierarchy**, i.e. an interlinked yet often temporally distinct set of results, effects, outcomes and impacts. These types of objectives are listed below.

Operational objectives provide a basis for assessing an intervention in relation to its outputs. The latter can be defined as what is directly produced / supplied through the activities and actions carried out during the implementation process. Indicators at this level are called output indicators.

Specific objectives provide a basis for assessing an intervention in relation to the short-term results that occur at the level of direct beneficiaries/recipients of assistance. Indicators at this level are called outcome indicators.

Intermediate objectives provide a basis for assessing an intervention in relation to its short to medium-term effects (or intermediate impacts) on both the direct and indirect beneficiaries/recipients of assistance. Indicators at this level are called impact indicators.

Global objectives provide a basis for assessing an intervention in relation to longer term and more diffuse effects (or global impacts). Indicators at this level are also called impact indicators.

Source: European Commission 2004

When defining objectives the following five (SMART) criteria can be used as far as possible:

- **S**pecific: should be precise and concrete enough to avoid being open to varying interpretation.
- **M**easurable: Objectives should refer to a desired future state (as compared to the baseline situation) so that it is possible at some later date to see whether the objective has been achieved or not.

[6] https://www.era-learn.eu/events/era-learn-2020-workshop-on-evaluation-and-impact-assessment-of-p2ps

- **A**ccepted: If objectives and target levels are intended to influence behaviour, they must be accepted, understood and interpreted similarly by all of those who are expected to take responsibility for achieving them.
- **R**ealistic: Objectives and target levels should be ambitious but they should also be potentially achievable so that those responsible see them as meaningful.
- **T**ime-dependent: Objectives and target levels remain vague if they are not related to a fixed date or time period.

More examples of P2P objectives and Objectives' Hierarchy are provided in the Background Document (section 3.b 'The Objectives of P2Ps').

Guiding questions for defining the Rationale and developing an Objectives' Hierarchy

- Why was the P2P established? Which challenge, problem, or situation does it aim to address?
- What are the main assumptions about the way the specific challenge, problem, or situation should be addressed?
- What are the short-term, operational objectives, the medium-term, intermediate objectives and longer-term, global objectives of the specific P2P?

Following the basic elements of a P2P impacts assessment framework (Figure 1) the next section deals with the inputs and activities.

Step 3: What are the inputs and activities that will achieve the objectives

Having identified the key objectives of the P2P, the intervention logic of the P2P now needs to be clarified, i.e. the assumptions that are made on how (by what inputs, activities, outcomes and impacts and their interplays and by which interplays among them) the desired objectives are expected to be achieved.

Firstly, we will focus on inputs and activities. *The Background Document provides a list of possible P2P activities (section 3.c 'The inputs and activities of P2Ps') that are further described in https://www.era-learn.eu/joint-activities.*

Activities of P2Ps comprise all programme specific actions that are being taken by the various actors of the P2P (i.e. programme owners, programme management, direct beneficiaries (R&I and various user communities).

Each activity will rely on one or more forms of input. These are the means used to support activities and action and to produce outputs. Inputs include budgetary costs (financial, administrative and human resources), costs for the beneficiaries or target population (co-financing, compliance costs stemming from administrative burden), costs for third parties (Member States, intermediary organisations, etc.). At the same time inputs also comprise the management and governance structures, and processes governing the operation of the P2P.

Guiding questions for deciding on the key activities and inputs of a P2P

- Which activities are expected to help achieve the objectives set (research collaboration, dissemination of results, training, mobility, sharing of infrastructure, network expansion, strategy building, monitoring and evaluation of the P2P, etc.)?
- What are the key inputs that these activities require (in terms of money, time, human resources, and capacities)?
- What are the governance and management structures and processes that will govern the implementation of these activities (decision making processes, management processes, internal/external reporting processes, etc.)?

Following impact assessment framework (Figure 1) the next section deals with the outputs of the activities.

Step 4: What are the outputs of the programme and project activities?

Naturally, all activities supported by a P2P should result in one or more outputs that will eventually contribute, at least in part, to the objectives of that P2P. An output is defined as a direct, typically tangible product of an activity. When specifying outputs we need to ask i) what should be delivered, and ii) at what time?

Outputs are items directly produced by certain activities (e.g. workshop reports, Strategic Research and Innovation Agendas, databases of programmes, etc.) and are typically produced within the short-term.

Based on which activities a P2P decides to implement, outputs may vary among projects supported under joint calls, joint strategy documents stemming from strategy building activities, training modules of students/researchers trained, databases with mapped national/regional programmes, new partners from different Member States brought together, dissemination events, etc.

Step 5: What are intermediate and global impacts?

The next underlying assumption in the logic framework is that every output will lead to some form of effect. Since this guide is concerned primarily with impact assessment, we devote some attention to this topic. Broadly speaking, impacts are the effects stemming from the activities and outputs supported under the policy intervention (i.e. the P2P). They may fall into a number of categories, generally defined by the time they take to manifest or the scope of their extent.

Outcomes can be defined as the likely or achieved short-term and medium-term effects of an intervention's outputs[7].

Impacts can be defined as positive and negative, primary and secondary long-term effects produced by an intervention, directly or indirectly, intended or unintended[7].

[7] (OECD) 2010, Glossary of Evaluation and Results Based Management (RBM) Terms, OECD (2010).

While outcomes rather relate to those directly addressed by an intervention impacts rather refer to the wider environment surrounding the policy intervention.

It should also be noted that short-term or intermediate impacts may in turn engender longer term effects, or may broaden their scope of impact. This characteristic can have important implications for the precise timing of the evaluation. Thus, impact assessment exercises oriented towards intermediate/long-terms impacts are not usually done earlier than 3-5 years after the end of the measure/programme. It should also be noted, however, that the longer the time has passed since the end of the intervention, the less the ability for direct attribution of the impacts identified to the specific intervention. Comparing the three types, outputs are items directly produced by certain activities (e.g. workshop reports, SRAs, databases of programmes, etc.) and they are produced within the short-term. Intermediate impacts are rather medium-term and may refer to both direct and indirect beneficiaries while global impacts are longer-term and refer to the wider environment surrounding the policy intervention.

Impacts may also vary according to their content. *Some major types are explained in the Background Document (section 3.d 'Outputs, outcomes and impacts of P2Ps') covering both the project and the network levels.* Naturally, not all impact types are to be expected by a P2P activity or a network as a whole. The types of impacts to be anticipated depend on the nature of the activity/ies examined, the timing of the evaluation and the stage of development of the P2P. The suggested list in the Background Document shows the diversity of impacts to look for when relevant. Moreover, certain impact types will occur before others. *You can read more about the periodicity of impacts in the Background Document (section 3.e).*

Guiding questions for identifying on the outputs, intermediate and global impacts of a P2P

- What are the key outputs of the activity examined? How are the outputs likely to influence direct beneficiaries in the short-term time? How are they to affect indirect beneficiaries?
- How are they to affect a broader set of stakeholders in the medium to long-term?
- How are the various outcomes of the P2P likely to affect the wider environment in which the P2P is situated, particularly in the long-term? (consider all levels: national, cross-national and European/international).
- What is the correspondence between outputs, intermediate and global impacts with the Objectives' Hierarchy?

Step 6: Setting up a monitoring system of inputs, outputs and outcomes

Any evaluation task can be significantly facilitated when an effective monitoring system is in place even before any evaluation activity starts.

Monitoring is a continuous and systematic process of data collection about an intervention. It generates factual information for future evaluation and impact assessments and helps identify actual implementation problems.[8]

In the first instance a monitoring system should collect contact details of individuals, groups, organisations or agencies that are participating in an activity or are subject to the intervention and where possible, the contact details of a control or comparison group to enable counterfactual analysis of impacts. It should also capture and retrieve evidence on the inputs required by an activity/measure as well as the outputs and possibly some short and medium-term outcomes. **Input indicators** should appropriately reflect the type of input (financial data relating to policy/programme/project expenditure, human resources, capacities, time, management structures, etc.) that activities require and the quantitative and/or qualitative nature of the activities. Often, quite simple indicators can be selected to monitor activities: for example, the number of workshops or meetings held, or the number of MoUs signed.

At the same time, the monitoring system should collect evidence about the anticipated outputs of the activities undertaken. **Output indicators** monitor and assess the qualitative and quantitative extent of outputs and should be chosen with care. In some cases, the choice of indicator will be relatively straightforward: for example, publication counts form a tried and tested means of measuring scientific output in the academic environment (although they are now subject to policy pressures which may strongly influence institutional and individual behaviours when choosing how and where to publish the results of scientific research). On the other hand, patent counts are often selected as an indicator of innovation output/outcome despite large variations in the motivation underlying their production and in the sometimes doubtful strength of their linkage to the full range of innovation-oriented activities.

A monitoring system should also be able to capture certain outcomes with appropriate **outcome indicators**. For instance, the identification of common areas of interest could be captured as an outcome of a P2P activity by monitoring relevant workshop reports. The level and the trend of engagement could be captured when monitoring the number of stakeholders participating in relevant P2P activities over time. Changes in priority areas of agencies could be monitored when P2P members are asked to report on such changes at regular intervals.

Annex 1 contains a number of suggested output, outcome and impact indicators at both the network (Table 1) and project level

[8] http://ec.europa.eu/smart-regulation/guidelines/docs/br_toolbox_en.pdf

Table 2*) including also the source of information for each of these indicators.* As can be seen most of the indicators can be addressed during the monitoring of the P2P's operation and its supported projects. *More about the monitoring system and the kind of information to gather for P2P networks and projects at various times is available in Section 6 of the Background Document.*

Some key points about setting up a monitoring system and defining monitoring indicators[9]

- The monitoring system should be established right from the outset of designing an activity/measure, i.e. before the actual implementation phase begins. During the lifetime of an activity/measure you can measure inputs to an activity as well as the immediate outputs and some medium-term outcomes.
- Ensure that the monitoring system works from the outset and that data will be collected reliably and smoothly from the appropriate sources.
- Many monitoring indicators can only be created and developed when the activity is implemented, because you need the cooperation and agreement of stakeholders in developing them and in collecting the relevant information.
- Make adequate use of the collected data by ensuring the soundness and reliability of the proposed methods and instruments for collecting, storing and processing follow-up data.
- Design indicators that will allow collection of data relevant for improving the implementation and later evaluation of the policy intervention.
- Before proposing new data requirements, you should carefully assess to what extent the existing data reflect the objectives set and whether the missing key data can be collected via existing monitoring structures. It is not unusual that the reporting requirements for P2P-supported projects have to comply with both national and P2P rules, thus causing extra administrative burden for both P2P members as well as beneficiaries. This should be avoided.
- Consider carefully and provide for the cost of setting up and maintaining a monitoring system over the time life of an intervention. This cost may not be negligible; either to the administrators who need to collect the required information and data or to those expected to deliver it.

Guiding questions for setting up a P2P monitoring system

- How much will the setting up of the monitoring system cost and what kind of resources will be needed to run it? Are these resources (human, financial, time) in place and can they be ensured?
- What data needs to be collected and how is it to be used (for inputs, outputs as well as outcomes)?
- When should the relevant data be collected (during the monitoring phase, ex-post, how often, etc.)?
- By whom should the required data be collected (e.g. P2P management team, project team, a centralised P2P systems)? Is the required capacity both in time and skills resources available?
- Can the process be aligned with the monitoring/reviewing process of the P2P partners? Can the process be aligned with the reporting schedule for the evaluation/impact assessment?

[9] Amended from http://ec.europa.eu/smart-regulation/guidelines/docs/br_toolbox_en.pdf.

- How will required data be gathered and stored? Where will the data be stored? Can this be aligned with other monitoring systems?
- What are the necessary data protection protocols to ensure the system will meet security and data sharing requirements?
- How and by whom will the data be verified to ensure it is accurate and consistent with the relevant requirements? Are the required skills and resources available for this task?
- What are suitable methods and instruments for collecting, storing and processing follow-up data?

Step 7: Defining output, outcome and impact indicators

An indicator is a quantitative or qualitative measure of how close we are to achieving a set objective. Indicators need to apply to different levels of effects (short-term outputs, intermediate impacts, long-term impacts) in a similar way that the logic model establishes a hierarchy of linked objectives at different levels (operational, immediate, intermediate, and global). Since the intervention logic is highly likely to be subject to some evolution during the lifetime of the policy intervention due for instance to change of assumptions over time, it is important that indicators are also allowed to be revisited.

There are several ways of capturing outcomes and impacts although there tends to be a preference, especially among policy-makers, for the use of (generally quantitative) indicators, which often attain higher visibility in policy debates than do qualitative impact statements.

Overall, indicators must be meaningful and contain information that is relevant to the evaluation of the P2P and the extent to which it is achieving its objectives. For example, a head count of the participants at a workshop will convey little information whilst additional data on the affiliation of those participants could provide information on the scope of the networking activity and level of interest across various stakeholders.

Some key points on indicators

There is no clear-cut rule on the appropriate level of detail for indicators – this will depend on the type of activity, the complexity of the intervention logic and the hierarchy of objectives constructed for a particular intervention. In principle, however, the "smarter" the policy objective, the easier it is to define a corresponding indicator.

It is essential to understand that indicators are subject to a number of limitations. They cannot measure all aspects of the reality and are often just a proxy of an intended outcome, while indicators that are defined ex-ante can only capture intended impacts. Societal impacts appear especially difficult to measure but, in certain cases, the cost of developing a good information system may be as costly as the intervention itself.

It may be the case that the most accurate indicators are extremely resource intensive to collect; thus a balance will have to be struck between indicator suitability and ease of collection. However, care should be taken to establish the quality of the monitoring data being collected as poor quality or partial data will affect the scope and scale of monitoring data's contribution to an evaluation.

Qualitative indicators can be highly illustrative of the outputs and impacts of activities but are more difficult to aggregate and to subject to quantitative analyses.

The appropriateness of indicators is case and context dependent.

All indicators should be 'RACER', i.e.:[10]

- **R**elevant, i.e. closely linked to the objectives to be reached. They should not be overambitious and should measure the right thing;
- **A**ccepted (e.g. by staff, stakeholders). The role and responsibilities for the indicator need to be well defined (e.g. if the indicator is the handling time for a grant application and the administrative process is partly controlled by Member States and partly by the EU then both sides would assume only partial responsibility).
- **C**redible for non-experts, unambiguous and easy to interpret. Indicators should be simple and robust as possible. If necessary, composite indicators might need to be used instead – such as country ratings, well-being indicators, but also ratings of financial institutions and instruments. As they may be difficult to interpret, they should be used to assess broad context only.
- **E**asy to monitor (e.g. data collection should be possible at low cost).
- **R**obust against manipulation (e.g. administrative burden: If the target is to reduce administrative burdens to businesses, the burdens might not be reduced, but just shifted from businesses to public administration).

Annex 1 contains a number of suggested output, outcome and impact indicators at both the network (Table 1) and project level (

Table 2*).* Naturally, these need to be adjusted to the objectives and activities of each specific P2P. As can be seen from Annex 1, the measurement of activities does not typically require any complex or sophisticated collection techniques and simple routine monitoring procedures should be able to adequately capture them.

Sources of examples of indicators that have been developed by P2P networks are included in Section 7 of the Background Document.

Guiding questions for defining P2P output, outcome and impact indicators

- Which are the key outputs, outcomes and impacts of an activity?
- For each one of these, how can they best be measured/captured?
- Is there a quantitative or qualitative way (or both) that they can be measured / captured?
- What is the added value of applying a quantifiable or a qualitative measurement or a combined approach in measurement?
- Is the cost of gathering data for the indicators set affordable? Are the necessary skills available?

[10] http://ec.europa.eu/smart-regulation/guidelines/docs/br_toolbox_en.pdf

- Are any relevant indicators already being estimated at Member State or EC level? Can access and compatibility be ensured?

Step 8: Data collection and analysis methods

There are a variety of data collection and analysis methods available to those conducting evaluations and impact assessment. However, their suitability for use is determined by a number of factors, such as the timing of the evaluation, the nature of the specific activities and actions being evaluated. For the suitability of various methods for the different timing of evaluations you may refer to Figure 11 under Section 8 'Data collection and analysis methods' of the Background Document.

After analysing more than 200 evaluation reports, the INNO-APPRAISAL study[11] found some interesting links between certain data collection / analysis methods with the evaluation issues addressed. Descriptive statistics are the most common approach applied, while case studies – to understand contexts and developments over time – are performed less often possibly due to the increased resources they require, although they can provide detailed explanations of how and why policy measures work (or not). More sophisticated, quantitative approaches such as econometric analysis and network analysis are used more selectively. Interestingly, the majority of the cases examined claim to use monitoring data and existing surveys and databases as a basis for the analysis. However, it appears that this kind of data is insufficient to be used for specific evaluation questions such as networking or behavioural additionality. The most important pro-active data collection is done through interviews and participant surveys.

Overall, the methodological approaches applied are tailored to the evaluation issues addressed and impacts to be covered. For example, evaluations interested in strategy development and policy issues more generally also look at consistency as the evaluation issue and vastly use interviews and generally qualitative methods such as focus groups or document search. Evaluations more concerned with effectiveness rely on (often simple) statistical analysis and data. Those evaluations more concerned with efficiency and project level issues, in turn, tend to look for different kinds of additionality and rely on surveys, interviews and, less broadly, though, on case studies.

A short description of the various data collection and analysis methods for evaluation and impact assessment and their pros and cons can be found in Annexes A.2, and A.3 of Technopolis and MIoIR (2012), http://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/eval2007/innovation_activities/inno_activities_guidance_en.pdf. For more on the relations between the evaluation issues addressed and the methods most often applied you may visit Section 8 'Data collection and analysis methods' of the Background Document. The same section also refers to examples of templates for surveys and interviews as well as reports on impact case studies.

The timing of the data collection also needs to be considered carefully – the eventual impacts of a policy may take many years to materialise, and are likely to be too distant to be collected as part of an evaluation project. In such cases, it may be important to build in the collection of data related to intermediate or

[11] http://www.isi.fraunhofer.de/isi-wAssets/docs/p/de/publikationen/INNO-Appraisal_Final_Report_100228.pdf.

proxy[12] outcomes which can be used to measure impact in a shorter timeframe. These outcomes might then be "translated" into final outcome measures using the logic model framework.

For P2Ps it should be remembered that, in each country, several organisations may take part in P2Ps. Thus, the same national agency/ministry may take part in a significant number of P2Ps. Retrieving data and information for each P2P separately from the same organisation using different templates and methods may lead to so-called 'evaluation fatigue', which can quickly diminish the participation rate in an evaluation and reduce the quality (and quantity) of the data supplied.

Data collection and analysis - Key points

- The collection of data required for an evaluation should be planned before policy activity commences: in the absence of a clear data collection strategy an evaluation may not be possible or be limited.
- Ethical and data protection requirements need to be taken into account and planned for prior to collecting data, particularly for monitoring data collected from individual participants.
- Administrative, long-term structural survey and monitoring data are important sources of evaluation data but where they are not available, or inappropriate, alternative data collection methods can be used.
- Monitoring and evaluation are complementary activities, and ideally the design and requirements for each should be considered together, so that the comprehensive data needs of the policy can be considered in the round. This will facilitate the collection of relevant and high quality data and avoid duplication or missed opportunities for the collection of key data. For example, routine ongoing collection of publication data arising from project activities can be far more cost efficient than retrospective requests for participants to supply such data and reduces the problem of wrongly attributing certain publications to the project activities. Early identification of any existing data, or other ongoing data collection processes, that can be utilised for the evaluation will ensure best use of resources and effort.
- It is important to design data collection tools so that they are consistent with relevant existing, or previous, data monitoring and collection tools to enable comparison and ensure data consistency.

Source: The Magenta Book, https://www.gov.uk/government/publications/the-magenta-book

Summary: Suggested guiding questions for selecting methods:

- What are the evaluation issues and associated evaluation questions that we need to examine?
- What are the key activities and associated outputs, outcomes and impacts that can contribute to answering the respective evaluation issues and questions?
- Which of the relevant output, outcome and impact indicators we shall examine?

[12] A proxy indicator is an indirect measure or sign that approximates or represents a phenomenon in the absence of a direct measure or sign. Number of female members of a chamber of commerce, for example, is a proxy indicator of the percentage of female business owners or executives. http://www.businessdictionary.com/definition/proxy-indicator.html.

- How can the data for the selected indicators best be captured (through monitoring data already available, through new data; quantitatively, qualitatively, both)?
- Based on the above what are the suitable methods to apply? (consider also time, and resources limitations).

Step 9: From planning to actually conducting a P2P evaluation

The previous steps are basic guidelines on how to plan an evaluation addressing all the key elements and concepts involved in the Logic Frame. This section[13] presents the basic decision tasks and poses important questions that need to be answered when preparing to carry out an evaluation/impact assessment exercise. By the time these questions are answered then all the required decisions will have been made to start the execution of the evaluation exercise.

Decision Task 1. Defining the audience for the evaluation

Important questions that need to be answered are, 'who will be the main users of the findings and how will they be engaged?' It is important to consider who the anticipated users of the findings are, and the requirements of policy makers and other stakeholders. Evaluations and impact assessments are important when they are acted upon and they are acted upon when they address the needs of their target-audiences and consider their expectations, and when they can ensure the desired level of robustness of results and scrutiny.

Decision Task 2. Selecting the evaluation issues and associated questions

Evaluation issues are essentially addressing the relations across the different elements of the Logic Frame (see Step 1 until Step 5 and Figure 1 above) and usually come from specific questions that we plan to answer through an evaluation task.

For instance the question 'to what extent is an intervention relevant with respect to the needs, problems and issues identified in target groups?' refers to the issue of Relevance. The question 'to what extent do the effects induced by an intervention correspond with its objectives as they are outlined in the intervention strategy?' refers to the issue of Effectiveness. The question 'how economically have the resources used been converted into effects?' refers to the issue of Efficiency. The question 'how do the effects of an intervention compare with the wider needs of the target populations?' refers to the issue of Utility. There are also other evaluation issues that can apply to P2Ps.

A list of suggested issues to examine in a P2P evaluation is presented in the Background Document (Section 5). Naturally, not all evaluation issues should be addressed simultaneously. It is important at this stage to carefully consider what is feasible to be evaluated depending on the level of development of a P2P manage expectations from the evaluation. Certain evaluation issues are also more relevant for specific timing of evaluations (see Figure 9 in the Background Document). The selection of the issues to address generally depends on:

[13] It is partly based on the Magenta Book, https://www.gov.uk/government/publications/the-magenta-book.

- the level of development of a P2P,
- the needs/interests of the direct audience of the evaluation,
- the availability and reliability of existing evidence,
- the feasibility of collecting/analysing new data,
- the availability of data sources for both existing and new data,
- the measurability of outcomes/impacts sought, as well as
- time and resource availability for the task of evaluation

Decision Task 3. Identifying the data requirements

At this stage the following questions need to be answered:

- What data are required? What is already being collected / available? What additional data need to be collected?
- At what point in time should the impact be measured?
- Who will be responsible for data collection and what processes need to be set up?

The specific data required for an evaluation will relate to the inputs, outputs, outcomes and impacts. It is here that an effective and robust monitoring system established long before the evaluation takes place will be most appreciated. **Thus, consideration of data requirements for monitoring and evaluation purposes should take place and be acted upon at the same time with designing the envisaged P2P activities.** Step 6 above provides guidance for setting up a monitoring system.

Decision Task 4. Conducting the evaluation

In relation to conducting the evaluation itself it is important to reflect on the following:

- When will the evaluation start and end?
- Will the evaluation be externally commissioned or conducted in-house?
- Who will be responsible for specification development, tendering, project management and quality assurance? What will the quality assurance processes be?
- When does any new data collection need to take place?
- Who will be the project manager, provide analytical support, be on the steering group?
- What budget is to be used for the evaluation and is this compatible with the evaluation requirements?

At this stage it is relevant to decide about the timing and duration of the evaluation including whether it should be commissioned externally or conducted (either partly or wholly) in-house and the implications of this. Having a clear idea about the internal available resources and capacities for the evaluation will influence selection of the most appropriate evaluation approach.

Some P2P networks have commissioned the evaluation task to external evaluators while others are considering to do it in-house. It may be useful to undertake a scoping or feasibility study to support this decision making process. This can foster greater understanding of what can and cannot be evaluated at the specific point in time, and therefore what level of investment as well as of capacities are required internally to determine whether the task should be assigned with an external evaluator or not. Commissioning the

evaluation exercise partly or wholly to external evaluators might also serve the purpose of safeguarding objectivity.

Evaluations, whether conducted internally or externally, however, will often require significant input to ensure they are designed and delivered successfully. For larger evaluations involving dedicated data collection, this will generally require an appropriate internal project manager with the relevant skills to oversee the evaluation, a supporting team, and a steering group to govern the evaluation. The level of input required of different members of the project team will be greatest at key points (in particular, the design, commissioning and reporting stage), but there will be an ongoing resource requirement even if the project is externally commissioned and this should not be underestimated.

Decision Task 5. Using and disseminating the evaluation findings

Evaluations are useful when their results are acted upon. As pointed out in Task 1 the primary audience needs to be identified early on and engaged in the whole process in order to ensure that their needs and interests and met to the degree possible. This will increase the possibility that the evaluation results will be considered for the purposes they were set out to be considered. Thus, at this stage it is important to answer questions like 'what will the findings be used for, and what decisions will they feed into?' and 'how will the findings be shared and disseminated?'

Along with the publication of the evaluation results it is important to consider how these will be presented and to whom. For example whether there will be one long report, an executive summary, a technical report, and/ or presentations, and what communication means will be used (on-line, off-line means, social media, etc.). Different audiences (policy makers, academics, researchers, businesses, society, youth, etc.) are receptive to different communication means and approaches. A differentiated communication strategy needs to be designed and applied to increase the impact of the evaluation findings. If you are externally commissioning the evaluation you will need to specify the format of the report and any presentations at the time of commissioning.

There is more on 'Communicating evaluation/impact assessment results' in the Background Document (section 9 'Communicating evaluation / impact assessment results') which also includes some good examples of communicating impacts in a short, but comprehensive and attractive way.
Section 10 of the Background Document also includes some good practice tips for preparing, organising, commissioning and carrying out an evaluation.

Annex 1: Suggested indicators

Table 1: Examples of indicators at P2P network level

Activity	Sub-activity	Output Indicators / nature	Outcome Indicators / nature	Impact indicators / nature	Source of information	Timing
Mapping national/trans-national activities	Mapping workshops/ meetings	• No of attendants (quant.) • Quality of report/ deliverable (qual.) • Programme clustering (qual.)	• Identification of common areas of interest (qualitative)	• Critical mass of research in certain areas (both quant. qual.)	Monitoring/ questionnaire	Interim/ ex-post
Foresight and common vision	Foresight exercise Vision building ws	• No of attendants (quant.) • Quality of report/deliverable (qualitative)	• Identification of common areas of interest (qualitative)	• Inform national and European policies (qualitative)	Monitoring/ questionnaire	Interim/ ex-post
Strategic Research Agenda / Implementation Plan	Interaction with AB, stakeholders Specific surveys	• No of attendants (quant.) • Quality of discussions (qual.) • Quality / level of approval of SRA (qual.)	• Identification of themes for calls (qual.) • changes in research priorities of agencies (qual.) • alignment of research strategies (qual.)	• Specific strategies for certain areas (qual.) • Influence national strategies/policies/ programmes (qual.) • Changes in national budgets (quant.)	Monitoring/ questionnaire	Interim/ ex-post
Joint calls	Building a portal Call management Evaluation of prop.	• User-friendliness of portal (quant. qual.) • No of proposals submitted/ approved (quant.) • Time to contract (quant.)	• Promotion of research area at national levels (quant.) • Change of national rules, timings (qual.) • Multinational evaluation schemes (qual.)	• Common rules, procedures, timing, and evaluation panels (qualitative) • Changes in legislation to allow payments to foreign researchers (qual.)	Monitoring/ questionnaire	Interim/ ex-post
Research Infrastructures	Shared use of infra. Building new infra.	• Hours of shared usage (quant.) • Quality of joint use (qual.) • New rules and procedures for new infra. (qual.)	• Increased efficiency in use of infra (quant.) • Increase use of infra (quant.)	• Harmonised rules and timing in participation (quant. qual.) • Harmonisation in research processes/ data (qual.)	Monitoring/ questionnaire	Interim/ ex-post
Stakeholder involvement	AB membership SRA involvement	• Composition of governance structures (quant.)	• More relevant research to stakeholders' needs	• Increased stakeholder participation in national /	Monitoring/ questionnaire	Interim/ ex-post

Activity	Sub-activity	Output Indicators / nature	Outcome Indicators / nature	Impact indicators / nature	Source of information	Timing
	Proposal evaluation	• Share of industry /society in AB, other structures (quant.)	(qual.)	European decision-making (quant.)		
Dissemination/ Up-take of research results	Website Newsletter Academic papers Brokerage events	• No. of hits (quant.) • No. of recipients (quant.) • No. of conferences (quant.) • No of highly-ranked papers • No. of attendants (quant.)	• No of papers with new counterparts • New proposals in funding programmes (quant. qual.)	• Spin-out companies for commercialisation (quant. qual.) • New collaborations (quant. qual.)	Monitoring/ questionnaire	Interim/ ex-post
Widening participation	Conferences / ws Signing of MoUs	• New partners (quant.) • Increased budgets (quant.)	• Better success rates of widening countries in H2020 (quant.)	• Increased research performance of widening countries (quant. qual.)	Monitoring/ questionnaire	Interim/ ex-post
Internationalisation	Conferences / ws Signing of MoUs	• New partners (quant.) • Increased budgets (quant.)	• New international proposals (quant. qual.)	• New collaborations (quant. qual.)	Monitoring/ questionnaire	Interim/ ex-post
Monitoring and evaluation (M&E)	Set up M&E system	• Quality of M&E reports • Budget allocated (quant.)	• Informed decision-making (qual.) • Improved network operation (quant. qual.)	• Improved engagement of P2P partners • Engagement of new partners	Monitoring/ questionnaire	Interim/ ex-post
Capacity building in national agencies	Internal training, exchange visits	• Number of events / schemes (quant.) • Participants / attendants at events (quant.)	• Increase of expertise in coordination / collaboration (qual.)	• Improved rules and procedures at national level (qual.)	Monitoring/ questionnaire	Interim/ ex post
Network health and connectivity	Guidelines for collaboration	• Formal/informal guidelines produced (qualitative)	• Increased engagement in activities over time	• Long-standing collaboration of partners	monitoring/ questionnaire	Interim/ ex post
	Governance structures	• Memberships (quant. qual.)	• Improved governance over time	• Self-sustained P2P	monitoring/ questionnaire	Interim/ ex post
	Network sustainability	• Resources availability (quant. qual.)	• Future intentions re collaboration (qual.)	• Changes in national budgets re international activities (quant.)	questionnaire	Ex post

Table 2: Examples of indicators at project level

Project activity	Output Indicators / nature	Outcome Indicators / nature	Impact indicators / nature	Source of information	Timing
Research collaboration	Publications (quant.) New staff, students, employees linked to project/theme (quant.) New methods, services, products (quant/qual) Co-authorships (quant.) New joint proposals/projects (quant.) New theories, practices, trans-disciplinary activities (qual.)	Changes to research programmes of organisations (qual.) Increased collaborations (quant.) Higher-research ranking (quant.) Increased reputation (qual.) Access to additional R&I funding (quant.)	New research trajectories / new areas of research (quant./qual.) Solutions to challenges (qual) international profile (quant./qual) Increased long-standing collabs (quant./qual)	Monitoring/ questionnaire	Interim Ex post
Research collaboration Academia – industry	Industry/HE co-publications (quant.) Prototypes of new methods/products/services (quant.qual) Patents, licenses, leasing, etc. (quant)	New methods/products/services (quant.qual) Spin-offs (quant./ qual) Market share figures (quant./qual) Commercial returns – turnover – employment (quant.) Reduced operating costs (quant.)	Solutions to challenges (qual.) Increased industry competitiveness (quant/qual) Improved business models (qual.)	monitoring/ questionnaire	Interim Ex post
Results dissemination Science-society	Raising awareness in society (quant./qual)	Change consumers behaviour (quant./qual)	More informed / concerned citizens (quant./qual)	Monitoring/ questionnaire	Interim Ex post
Research collaboration Dissemination to policy	Inputs to standards (qual.)	White papers, draft regulations (quant./qual) Changes in policies / regulations (quant/qual.)	Solutions to challenges (qual./quant) Improved policy-making (qual.) Improved service quality (qual) Reduced environmental impacts (quant.qual)	questionnaire	Ex post
Capacity building Training / knowledge transfer	Training schemes/activities (quant./qual.); Masters/PhD students (quant.); Conferences, workshops, seminars (quant./qual.)	Improved capacities at organisational level (quant./qual.) Changes to human resources Organisational changes (quant./qual)	Improved national capacity / performance in specific area (quant./qual) New practices for research organisation (qual)	Monitoring/ questionnaire	Interim Ex post