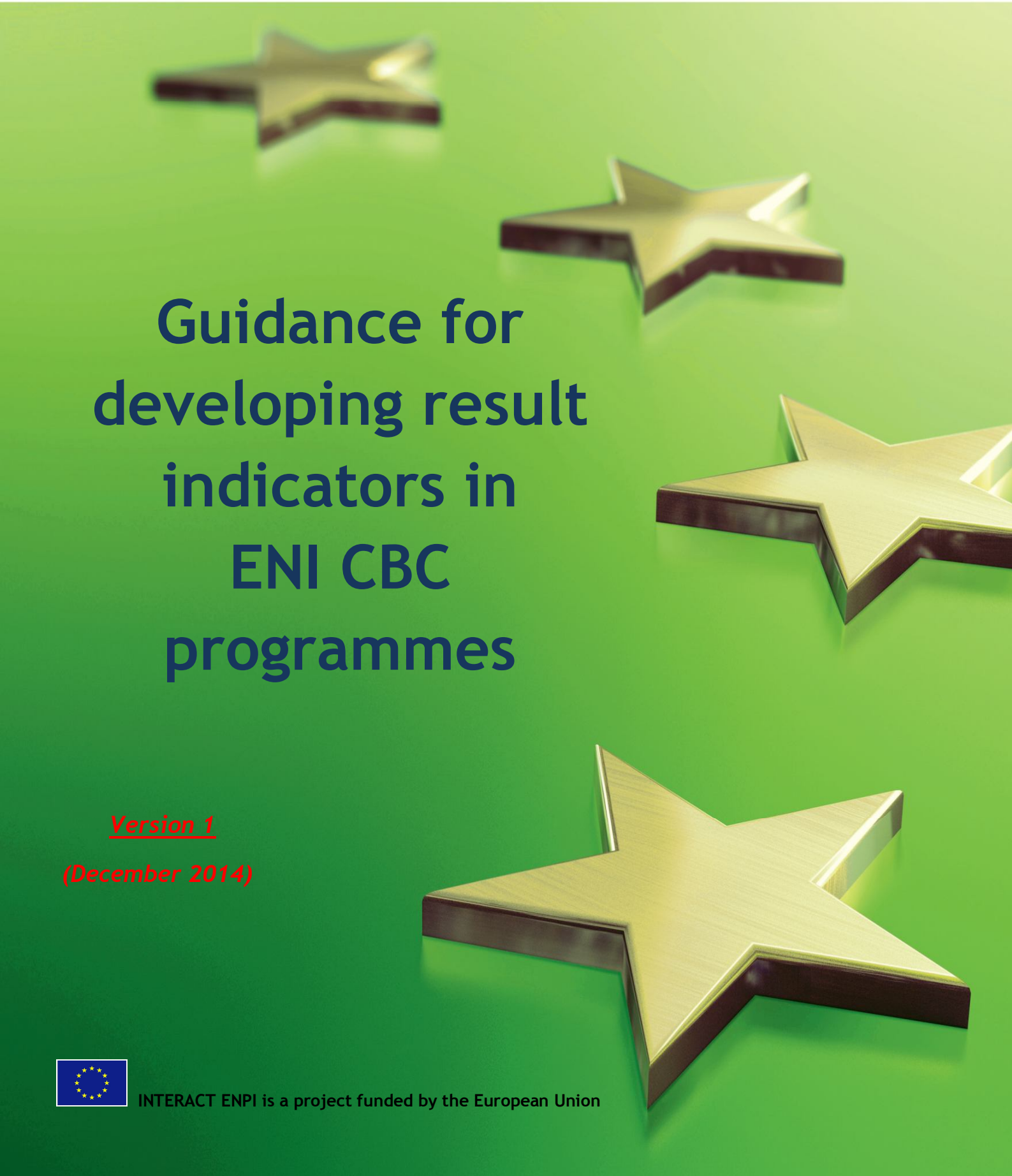
European Union

INTERact

Sharing Expertise ENPI CBC



# Guidance for developing result indicators in ENI CBC programmes

*Version 1*

*(December 2014)*

INTERACT ENPI is a project funded by the European Union

# Guidance for developing result indicators in ENI CBC 2014-2020 programmes

**Table of contents**


Preamble ..... 3
1. Introduction ..... 6
The objective of this paper ..... 6
The distinction between different types of indicators ..... 7
2. Generic prerequisites for result indicators ..... 10
The intervention logic ..... 10
General requirements for result indicators ..... 11
Additional prerequisites for ENI CBC programmes ..... 12
3. Developing result indicators ..... 13
How many indicators needed ..... 13
Quantitative and qualitative indicators ..... 13
Data sources ..... 15
Indicator data scales ..... 16
Data documentation ..... 17
Setting baseline and target values ..... 18
Disaggregation of national data ..... 20
4. Condensed checklist for result indicators ..... 21

Annex 1: Causality between output and result indicator ..... 22
Annex 2: Measuring only the effect of the intervention ..... 24
Annex 3: Case studies on the development of result indicators ..... 26

Case study 1: Promotion of a low carbon economy -reduction of fossil energy consumption ..... 27
Case study 2: Improving the mobility of persons and goods – road upgrading ..... 29
Case study 3: Improving the mobility of persons and goods – development of strategies for connecting secondary rail networks ..... 31
Case study 4: Improvement of border management operations – improving border staff competence ..... 34
Case study 5: Supporting the integration of vulnerable population groups - fostering activities aiming at reduction of poverty and social marginalisation ..... 37
Case study 6: Promotion of entrepreneurship – support development of small enterprises ..... 40
Case study 7: Increasing capacity of local authorities –development of local governance ..... 43

## List of abbreviations

CBC – Cross-border cooperation

EEAS – European External Action Service

ENI – European Neighbourhood Instrument

EU – European Union

GDP – Gross domestic product

ICT – Information and communication technologies

ILO – International Labour Organisation

JMA – Joint Managing Authority

JPC – Joint Programming Committee

JTS – Joint Technical Secretariat

Km/h – kilometres per hour

Ktoe - kilo-tons oil equivalent

NGO – Non-governmental organization

NSI – National statistical institute

NUTS – Nomenclature of Territorial Units for Statistics

OECD – Organization for Economic Cooperation and Development

PPP – Purchasing power parity

SME – Small and medium-sized enterprise

TO – Thematic objective

## Preamble

On the basis of the regulatory framework, ENI CBC programmes have to develop result indicators, which should correspond to the expected results of each programme priority. Information for these indicators may be obtained from the programme itself, from the projects it supports or, more commonly, from external data sources.

### Which result indicators are covered by this guidance?

This guidance paper mainly focuses on indicators that utilise programme-external data, and where such indicators are expected to capture the short- or medium term societal impact that generally spans beyond the direct beneficiaries of a programme and that covers a wider group of society.

In terms of result measurement, a programme needs to strike a balance between an ideal framework, and what it is feasible to do. In CBC circumstances, a trade-off is often required.

### Is this guidance compulsory?

This paper is not intended as a compulsory list of requirements to be applied throughout the process of development of indicators. Rather, it should be viewed as good practice for innovative development of indicators in situations where, at first sight, data may not be easily accessible.

This guidance paper aims to put forward an understanding of the preconditions, options and choices that exist concerning the setting of result indicators.

CBC programmes often face methodological challenges and have to accept the fact that they are measuring things in a way that is not optimal. However, being aware of these challenges and spelling them out is already in itself a considerable step towards developing a better results framework.

In order to illustrate the application of the considerations raised in the document, seven case studies on the process of developing result indicators have been provided in Annex 3. Most of the cases depicted are fictional but a few of them are wholly or partially based on real-life examples from EU CBC programmes. The case studies cover a wide array of practical, day-to-day issues likely to be encountered by the programmes when developing their result indicators.

The case studies cover issues of relevance for land border, sea crossing and sea basin programmes. They generally depict the concrete difficulty in developing a coherent approach for measuring the results of a CBC programme. However, some of the cases also show that result monitoring – even in cases where dedicated surveys are conducted - does not need to be onerous or expensive.

Additional shorter examples provided in the body text are intended for illustrative purposes only.

Annexes 1 and 2 contain information regarding aspects of the development of indicators (causality conditions between output and result indicators, counterfactual impact measurement) likely to be less directly relevant to the practical reality in which ENI CBC programmes operate. These annexes are mainly for background information.

The tentative programme priorities and thematic objectives (TOs) covered in the case studies are based on the thematic objectives and indicative programme priorities included in Annex 1 of the Programming Document 2014-2020 ENI Cross-Border Cooperation.

It is generally recommended to combine the development of monitoring indicators with the actual process of programming. Leaving the indicator development as a separate exercise to be undertaken once the programming process is finished will most likely make the work much harder. The same applies for contracting out the development of indicators to persons not involved in the day-to-day programming work, although this may not always be possible to avoid completely.

The development of indicators in parallel with the programming process usually also sheds light upon issues that bear relevance for setting the objectives of the programme as such. When examining draft indicators, the originally perceived challenges may appear different in light of hard statistical facts, enabling refocusing of the programme priorities. Examining hard facts frequently also reveals additional pressing issues and risks that need to be accounted for in the design of the programme.

### For whom is this paper intended?

This paper is primarily intended for staff of the Joint Managing Authorities (JMA) and Joint Technical Secretariats (JTS) and external experts involved in the concrete development of monitoring indicators.

It may however also bear relevance for general practitioners involved in other parts of the programme administration. In particular, JMA and/or JTS staff in charge of the completion and submission of the operational programme to the European Commission may find useful the condensed checklist available under point 4.

Developing robust result indicators is a challenging task. Even though being far from exhaustive, it is our hope that this guidance paper acts as one additional aid for the ENI CBC programmes in developing their own result indicators for the 2014-2020 programming period.

### IMPORTANT!

Even though not explicitly exhibited in the indicator examples that are provided, this guidance paper implicitly presupposes that the actions foreseen by a programme do aim for cross-border impact and cross-border value-added. In other words, the entire set-up of the document presupposes that any CBC intervention should be more than just a collection of local or regional interventions. In reality there may however exist also cases where a particular project may include specific actions not having *per se* a cross-border dimension, even if the project as a whole does. This could for example include pilot projects where a particular issue is tackled in different ways in different parts of the programme area, or tested in just one or a few regions alone. In such cases, the result measurement would include aspects with only regional value.

The guidance has been elaborated by Mr Tomas Hanell from Eurofutures Finland in collaboration with the INTERACT ENPI team. In the process of elaboration, valuable contributions have been received from both the European External Action Service and Directorate General Development Cooperation of the European Commission. Comments from the ENI CBC Support to Programme Preparation project have been incorporated as well. Our acknowledgement goes to all involved parties for their engagement in making this guidance possible.

# 1. Introduction

## The objective of this paper

Result indicators have a significant communicative value as well as present an important tool for the monitoring and evaluation of a programme. Adequately defined, they are able to partly answer the question of how well a programme is performing in terms of concrete results.

In order to achieve better accountability on the use of the public financing and to be able to demonstrate what has been achieved with it, the ENI CBC 2014-2020 programmes are required to pay sufficient attention to the development of programme-level indicators that can be used for this purpose. This is a challenging task for the programmes, especially taking into account their geographical coverage and the wide scope of possible actions.

In view of the above, the objective of this paper is to present brief and concise guidance acting as an aid in the development of result indicators for the ENI CBC 2014-2020 programmes.

This paper is strongly influenced by the European Commission's Guidance Document on Monitoring and Evaluation[1] that has been produced for use in the European Regional Development Fund, Cohesion Fund and European Territorial Cooperation programmes.

According to the draft regulatory framework, result indicators are mandatory in the ENI CBC 2014-2020 programmes. Article 4.3(c) of the ENI CBC Implementing Rules (Regulation (EU) 897/2014) stipulates that the joint operational programmes shall include:

"a description of objectively verifiable indicators, in particular:

a. The **expected results for each priority**, and the corresponding result indicators, with a baseline value and a target value;

b. The **output indicators for each priority**, including the quantified target value, which are expected to contribute to the results."

This paper is primarily concerned with result indicators alone (point *a* above). However, a number of references are made also to output indicators[2] (point *b* above), particularly concerning the conceptual boundaries between output and result indicators and the logic link that should exist between them.

Due to the regional character and wide thematic spread of ENI CBC, result indicators in these programmes are by nature programme-specific. Aggregated result measurement at the level of the instrument tends to be challenging.[3] This in practice implies that each programme needs to develop its own result indicators targeted specifically at the issues it aims to address.

Therefore this paper aims to aid the process of identification of appropriate result indicators by highlighting a number of attention points and some of the general standards that programmes should adhere to in order to guide their work. The focus lies on quantitative result indicators albeit also qualitative ones are touched upon. The approach is illustrative rather than exhaustive.

1 COM (2014): Guidance Document on Monitoring and Evaluation – European Regional Development Fund and Cohesion Fund – Concepts and Recommendations, European Commission, January 2014.

2 Concerns both programme-specific output indicators as well as output indicators common for all ENI CBC programmes (for further information refer to the "List of Common Output Indicators" distributed to the programmes by EEAS on 24 September 2014).

3 For example, the World Bank, regional development banks or DFID results Framework all include result indicators aggregated across countries.

## The distinction between different types of indicators

Although this guidance paper is primarily concerned with result indicators only, experience from the programming period 2007-2013 has highlighted the difficulty in setting precise boundaries in particular between output and result indicators[4].These boundaries are not always evident, which calls for logical reasoning regarding their set-up. The different types of indicators are briefly presented below.

**Info box: The different types of monitoring indicators**

| Indicator | Scope | Indicative examples |
|---|---|---|
| **Input indicators** | Measure the financial, administrative and regulatory resources. | Indicators connected to financial, human, material, organisation or regulatory resources mobilised during the implementation of the intervention (e.g., allocation or actual payment in euro, or as percentage of total allocation or payment). |
| **Output indicators** | Measure the direct products of the chosen activities. They concern the direct beneficiaries[5] of the projects and are only affected by what the project actions lead to, being insensitive to any external factors | Indicators connected to the specific deliveries of projects, such as kilometres of new road upgraded, number of additional households connected to wastewater treatment facilities, number of participating NGOs, etc. |
| **Result indicators** (the term outcome indicator also used) | Measure the broader societal impact of a particular objective or priority **in the short-or medium term**. They generally span beyond the direct beneficiaries of the support and cover a wider group of society (such as final beneficiaries or the entire target group).Result indicators should to a certain extent be affected by the outputs of the programme, but in general they are also affected by external factors that lay beyond the activities of the programme. | Indicators connected to the wider societal short- or medium term impact of the actions, and logically connected to the corresponding outputs, such as increase in average speed on upgraded road stretches or decrease in road accidents on them, decrease in untreated wastewater discharge, etc. |
| **Impact indicators** | Measure the **longer term consequences** of the results (or outcomes) | Indicators connected to the wider societal **long-term impact**, such as a decrease in mortality (as a result of less road accidents) or better health status of population (as a result of less contaminated agricultural produce), etc. |

*Source: Adapted from DG DEVELOPMENT (2002): Guidelines for the use of Indicators in country performance assessment*

4 INTERACT ENPI Project (2014): "Utilisation of monitoring indicators in ENPI CBC 2007-2013 programmes, INTERACT ENPI: 2014", submitted to programmes on 30 April 2014.

5 The term direct beneficiaries refers to entities directly benefiting from the projects. This includes project partners, as well as entities, persons and organisations that are actually involved in the activities of the projects.

The distinction between output and result indicators is clear. Where output indicators reflect the direct actions of a project (and when aggregated, also of a priority), result indicators in turn are quantitative or qualitative expressions of the cumulative change brought about by the project actions.

The distinction between result and impact indicators is less obvious. As a rule of thumb, one may observe the time span: a programme's activities are likely to affect an impact indicator in the long term, whereas the effect on a result indicator would be more immediate. Both impact and result indicators can be affected by other factors than the programme activities alone. In the box below, a hypothetical example illustrates these differences in practice.

**Example: defining output, result and impact indicators**

A hypothetical programme priority aims for "stimulating employment for youth" with the specific objective to enhance language skills for young persons in order for them to be able to benefit from a larger cross-border labour market. The types of actions supported primarily relate to language training of such youth on both sides of the border.

An example of a programme specific **output indicator** for this hypothetical case is "nr of additional young people taking part in language education – as a direct consequence of the support – lasting for at least two semesters and resulting in a diploma", with the baseline consequently set at zero, and the target set for example at 400 participants. The data for this indicator will be collected from actual project reports with the output indicator hence being directly related to the concrete actions of the priority.

The corresponding **result indicator** in this case is connected to the specific objective ("enhance language skills for young people") whereas the **impact indicator** concerns the larger priority ("stimulate employment for youth") of the entire thematic objective.

A hypothetical result indicator comprises "share of persons 20-24 years that are able to function in the language spoken across the border at least at CEFR Level B2[6]" with the baseline set at the rate at the starting of the programme, and the target set at "a three percentage unit increase during the programme period". Both the baseline and the target data will be acquired through a survey. Such an indicator is likely to change more or less immediately as a consequence of the activities undertaken.

A hypothetical impact indicator is tied to the larger priority and constitute "Employment rate for persons aged 20-24 years" with the baseline set at the rate at the starting of the programme, and the target set at "a three percentage unit increase during the programme period". This data will be acquired from administrative registers.

In the hypothetical case above, the output indicator reflects the concrete actions of the programme (enabling language courses) whereas the result indicator reflects a wider societal impact (better language skills) as a consequence of these actions.

As required, the output indicator is concerned with the direct beneficiaries of the proposed action (i.e., only the participants in training and as a direct consequence of the support are measured). The proposed result indicator in turn spans beyond the direct beneficiaries of the support and covers a wider group of society as the conducted survey[7]includes direct beneficiaries and non-beneficiaries alike.

6 Exemplifying classification based on "Common European Framework of Reference" for language skills.

7 Provided that the survey sampling is performed satisfactorily.

In this illustrative case there also exists a logic chain between the output and the result indicator, as more persons taking part in language training should by definition[8] also result in better language skills in the age group.

The long-term effect of better language skills could finally be plausibly connected to wider employment opportunities, under the assumption that better language skills increase the employability of a person specifically in an integrated border area.

8 Provided that they do not move away after course completion.

# 2. Generic prerequisites for result indicators

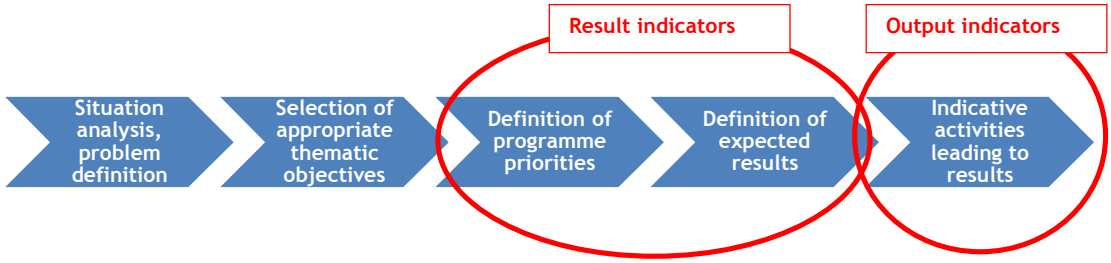
There are a number of general prerequisites for result indicators that are applicable regardless of which programme is concerned. These relate both to the inherent methodological quality of the indicators themselves (i.e., how they relate to the actual programme logic), as well as to the specific transnational setting in which CBC programmes operate.

## The intervention logic

Any programme wishing to intervene in society must establish the *raison d'être* for this intervention, as well as the suitable means for this policy intervention. In brief, for the ENI CBC programmes this – in principle – logical chain of decisions includes:

1. identifying the need for something to be changed (through situation analysis);
2. choosing a suitable frame of addressing this challenge (by selecting the appropriate TO);
3. prioritising within these objectives (by selecting suitable programme priorities);
4. defining specific expected results for each chosen priority (what specifically needs to be altered/affected); and finally
5. selecting the practical way in which these goals could be achieved (which types of activities should be supported).

The "positioning" of the different monitoring indicators in this intervention logic chain calls for specific attention. While result indicators in principle should address the effects of points nr 3 and 4 above (i.e., at the level of the priority), output indicators in turn, being connected with concrete actions, largely relate to the fifth point on the list.[9]




**A basic precondition is naturally that the entire programme adheres to a logic chain of reasoning.** Without that, any indicator, regardless of how carefully it is constructed, will most likely not meet its intended purpose.

The intervention logic implies that a causal chain of reasoning between the output and the result indicator should be present in a similar way as the causal link between the identified problem and the chosen intervention. A change in an output indicator should hence be expected to induce a change in its corresponding result indicator. In terms of a logical framework approach, this relationship is expressed as "if … and … then". **If** outputs are delivered **and** certain assumptions hold good / expected risks do not materialise, **then** the expected results will occur. For a more detailed discussion on causality issues between indicators, please refer to Annex 1.

9 Albeit they may eventually be aggregated upwards.

## General requirements for result indicators

When designing result indicators, there are a number of general aspects that should be taken into consideration.

First, and most important, the indicator should be an OVI (**objectively verifiable indicator**), implying that the information collected should be the same if collected by different people (i.e., not open to the subjective opinion/bias of one person). This aspect is emphasised in situations where the result indicator stems from data sources connected to the programme itself (such as project reports, programme database, etc.).

A result indicator should further have a **high responsiveness to action** so that it as much as possible reflects the effects of the indicative actions.

The result indicator should also be such that it can be **interpreted unequivocally**: any change in a particular direction (e.g., more, less) must have a clear and accepted interpretation (i.e., favourable, unfavourable).

The definition of the indicator, as well as its unit of measurement, should be **as precise as possible** and the **quality and the integrity of the indicator** should be possible to assess also externally. Transparency of practices and procedures when assembling the data should be made clear. For a CBC programme, the international comparability of the indicator also needs to be assessed.

The indicator should also be **available when needed**, which needs to be taken into account when designing the monitoring and evaluation plan of a programme.

Finally the indicator should preferably be of a type that **does not place an unnecessarily heavy burden to report**. This includes choosing indicators that do not require substantial calculus and/or estimation to be performed in order to report them, albeit avoiding this might not always be possible.

Additionally, another requirement is **reliability and statistical validation** (i.e., being resistant to errors caused by abnormal deviations). This requirement is not always necessary; statistical indicators based on own surveys, as well as expert panel assessments, are by nature excluded from this requirement.

Finally, in a situation where no data exist for the whole or parts of the programme area, **indirect measurement** (as opposed to direct measurement) is one option that could be explored. Whereas "direct measurement" refers to indicators measuring precisely the concept that one aims to measure, "indirect measurement" in turn utilises either 1) a commonly accepted or 2) a scientifically verified substitute indicator for the concept one aims to measure.[10] The same applies in situations dealing with so called latent variables, which are variables that by definition do not lend themselves to measurement at all.[11] In these cases, indirect measurement remains the sole option.

[10] For example, in the absence of statistical information on the skill level of 20-24 years old population in an area, school enrolment rates for persons 16-19 and 20-24 years could constitute one out of many indirect measurements of that concept. In this example, the measurement is not perfect (as skills can also be obtained outside the formal education system), but most likely still a sufficiently satisfactory proxy.

[11] Quality of life or wellbeing constitute prime examples of latent variables. There exist no single statistic able to capture such concepts and proxy indicators remain the only viable way to address these issues via numerical data.

## Additional prerequisites for ENI CBC programmes

When designing result indicators for ENI CBC programmes, there are a number of specificities that need to be considered. These concern primarily the specific territorial setting in which these programmes operate; this is, the multi-national set up of the programmes and the programme areas covering both entire countries as well as parts of countries. These factors lead to at least the following additional prerequisites:

- **Availability of regional data**. The fact that all ENI CBC programmes concern parts of some countries implies a necessity for availability of data at a lower level than the nation state. This is a restricting factor as the amount of data at the regional level in general is much less than that available at the corresponding country level.

- **Aggregation possibility across the programme area**. The indicators must be defined so that there exists the possibility for aggregating them across the entire programme area into one indicator. In practice this implies that the indicators in the first instance should only concern ***absolute numbers*** and not include ratios, percentages, categorical data[12], or other numerical types that do not accommodate straightforward addition.

- **Avoiding monetary units**. The transnational nature of the ENI CBC programmes entails that the scale of measurement of an indicator should preferably not be connected with a specific currency unit. Fluctuating exchange rates, differences in purchasing power, and inflation/deflation, all render comparison of monetary units across a programme area technically demanding, and if not properly performed, likely to lead to misleading conclusions.

- **International comparability**. The transnational nature of the ENI CBC programmes further entails that the entities that are measured by the indicator should as far as possible be roughly similar in definition across the programme area.[13] If not adequately considered, such a comparison of "apples and pears" runs the risk of leading to faulty conclusions.

12 See section on indicator data scales and data re-scaling.

13 For instance, the definition of 'public, private or third sector entities' may vary substantially across a programme area, or an 'immigrant' may be defined differently, or a 'full-time equivalent' may be calculated in a different manner, and so on.

# 3. Developing result indicators

## How many indicators needed

As result indicators should cover all priorities of a programme, the minimum number of result indicators equals that of the number of priorities of the programme.[14]

Depending on the scope and width of a particular priority, there may emerge a need to have more than one result indicator for a particular priority. Such is generally the case where a priority is expected to lead to several parallel, but thematically disconnected, results. For example, a priority might aim for "enhancing the access to health" by means of both (1) upgrading of existing (physical) health care infrastructure as well as by (2) improving the telemedicine network. In such a case result indicators should need to be able to reflect both of these two strands of results.

However, utilising the hypothetical priority above, the programme may alternatively wish to measure the results of its actions by an approach which is able to encompass both strands in a single indicator. That could for example be a survey among the residents where changes in the "quality of their access to health" would be charted. Conceptually both parallel strands ("upgrading of health care infrastructure" and "improved telemedicine") are covered by this single indicator.

## Quantitative and qualitative indicators

Result indicators can, depending on the information needed, be either quantitative or qualitative, neither of which is superior to the other.

In contrast to common comprehension, the difference between quantitative and qualitative indicators is not very sharp and the two types often intervene. Many quantitative indicators have qualitative aspects embedded in them, and vice versa. In aiding this division, an indicator can be characterised by its *substance* and its *count*, both of which can be either objective or subjective.

***Count***

The term *count* denotes the mode in which the data has been obtained.

- The count is **objective** if the number is obtained, for example, from a register or other comparable source that is not affected by someone's subjective assessment. Number of unemployed persons stemming from an unemployment register is a typical example.
- If the number of unemployed persons is obtained through a labour force survey instead (i.e., where the respondents of the survey themselves assess whether or not they are unemployed), then the count is **subjective**.

In both of the examples above, the substance (unemployed as opposed to employed, or completely outside the labour force) is objective.

***Substance***

As stated above, the *substance* of measurement (i.e., the phenomenon which is being measured) can also be either objective or subjective:

[14] Please refer further to the Communication letter from DG DEVCO accompanying this guidance.

- An example of an **objective substance** would be for instance the length of a motorway in kilometres, which is an undisputable fact. In this case also the count would be objective (kilometres measured from a map).
- The quality of that same motorway in turn (e.g., obtained through a survey on a five-item scale from "1. Very bad" to "5. Very good") would be an example of a **subjective substance**, where also the count is subjective. Asking a sample of the population within a programme area to state the number of kilometres that they drive daily on a motorway would be an example of a measurement where the substance (kilometres) is objective but where the count is subjective.

In light of the issues above, both quantitative as well as qualitative indicators can take many different forms. "Purely" qualitative indicators would rather take the form of verbal expert assessments indicating either a status (such as good, acceptable, negotiable, etc.) or a trajectory (such as better, worse, etc.). An example of qualitative target setting is provided in Annex 3 (case study 7).

Quantitative indicators are generally considered more objective, having a higher validity, and being more verifiable than qualitative ones, which is not necessarily true. Quantitative indicators are only as "good" as is their weakest link: if what is being measured is measured incorrectly, then they fail to deliver a correct message. For programme monitoring purposes, qualitative indicators can be equally relevant provided that they meet required quality criteria (see chapter 2. Generic prerequisites for result indicators on page 10).

Since many CBC programmes are dealing with "soft" priorities, qualitative data should also be considered since they are often able to capture the effects of such actions better than purely quantitative indicators. Quantitative and qualitative indicators may also be used side by side in supplement of each other. The box below indicates some (hypothetical) examples of quantitative and qualitative result indicators used in parallel.

**Example: Hypothetical quantitative and qualitative result indicators used in parallel**

| Priority | Quantitative indicator | At least partially qualitative indicator |
|---|---|---|
| **a) Promotion of a low carbon economy by means of improved rail transport capacity** | Ratio of annual road to rail person-km, obtained from a transport modality survey | Respondent assessment of rail transport quality, obtained from a survey |
| **b) Developing ICT infrastructure by means of construction of optical fibre networks** | Nr of household equivalents connected to optical fibre networks | Business survey regarding quality of local ICT infrastructure, conducted among a sample of SMEs within the programme area |
| **c) Improvement of land border-crossing efficiency by raising staff competence** | Volume of cross-border cargo flows, estimations based on daily traffic and average cargo/unit | Qualitative assessment on changes in border-crossing efficiency, conducted among a sample of traffic operators |
| **d) Support accessibility to basic health services by means of investment in joint cross-border technical equipment** | Share of population living within less than 25 km from a basic health care unit, estimation based on accessibility modelling | Regionalised data from the European Social Survey[15], question B29: "State of health services in country nowadays?" |

[15] Not only Member States, but also countries such as Israel, Russia, Turkey and Ukraine are included in the European Social Survey.

## Data sources

Data sources can be divided into two general types. Secondary data sources are sources of data that someone else has collected previously; primary data in turn are data that one has assembled oneself. A programme may opt for either of these, or a combination of both.

Secondary data sources (such as administrative registers, existing surveys, etc.) have the advantage that they usually are cost-effective to utilise, which also explains their frequent usage in development programmes. However, the ease alone cannot be used as a justification for their utilisation, since second hand sources are also from a CBC programme monitoring point of view connected with a wide array of other issues that limit their feasibility.

A first bottleneck is finding data that allows for precise measurement of the programme's expected results. A programme hence needs to take a stance on whether or not an imprecise fit between the selected indicator of the programme's results and the corresponding data source can be acceptable from a programme monitoring perspective.

A second obvious limitation is the very probable lack of data altogether for some parts of the programme area, especially for the sea basin programmes that may contain a large number of different countries. The programme hence needs to take a position whether or not a result indicator should cover all countries of the programme area, or if some data gaps for certain countries are acceptable. Provided that national level data is available, the scarce availability of regional data is the next major constraint since all ENI CBC programmes include areas below the level of the nation state. This issue is discussed further in subchapter "Disaggregation of national data" on page 20.

Second hand data sources are also often connected with differences in transnational data comparability, as classifications differ from one country to the next. The programme needs to identify such differences and assess whether or not they affect the outcome of the indicator and how it is – or should be – interpreted.

When several countries or parts thereof are concerned, a longer data time lag in only one of them also automatically affects the usability of available fresher data in the remaining countries, pushing the last common data year further back in time. The programme hence needs to decide whether or not a flexible approach concerning data years could be utilised, both for setting the baseline as well as for future target monitoring.

A final challenge with secondary data sources is the future access to the same data. A programme needs to secure that there will be access to this data throughout the entire programming period, or to the least in a foreseeable future. This is an issue to consider especially in cases connected with data gathered "spontaneously": specific academic studies, data stemming from a particular project, etc.[16]

Considering the constraints above, in many cases programmes are left with the option of conducting own data gathering or surveys specifically for the purpose of monitoring their performance.

At face value it may appear ambitious to conduct a separate survey merely to be able to extract a few result indicators from it.

Surveys conducted specifically for a programme have however a general advantage over other existing data sources in that they can be designed to fit exactly the information needs of the programme, therefore considerably improving the measurement precision. Surveys can further be designed so that - as far as possible - they are able to identify the precise impact of an intervention and disregard other external factors (see further Annex 2). Such luxury is generally not available when utilising secondary data sources.

[16] Data for example from the ESPON Programme (www.espon.eu) is generally of this type.

The obvious drawbacks in utilising primary data are the monetary costs as well as resources spent on designing surveys or programme-specific data gathering, which need to be assessed.

If properly designed, specifically conducted surveys can also ease the evaluating exercises. Information not used in the result monitoring as such, but still valuable for the evaluators, can generally be added to a survey at a negligible added cost. It may thus well be the case that resources spent in conducting a survey are resources saved when conducting the evaluations. Connecting the evaluators to the indicator development and the actual survey design process is in that case important.

Specifically designed surveys are also a good guarantee for that the information will exist during the whole course of the programme eliminating the risk that no monitoring data will be available in the future.

Finally, specifically designed surveys can also be of aid in the development of programmes beyond 2020 due to the fact that the information obtained from them is tailored to the precise needs of the programme area.

It is worth to notice that surveys - or other such own data collection – do not necessarily need to be extremely onerous or expensive. For instance, if well chosen, a sample of a few regions of a programme area may well be sufficient and able to adequately reflect the activities and results across the entire programme area. An example of such a coverage-wise limited approach is provided in Annex 3 (Case study 3).

A final consideration regarding data sources concerns information obtained from projects, programme beneficiaries, or other parties directly affiliated to the programme. Although such information primarily concerns output indicators, at times such data may also be utilised in connection with result indicators (see for example Annex 3, case study 6). In principle, the same rules for objective data verification should be applied as is the case with programme-external data.

## Indicator data scales

For quantitative as well as qualitative indicators alike, the scale of measurement affects how the data can be used. The four basic[17] measurement scales are:

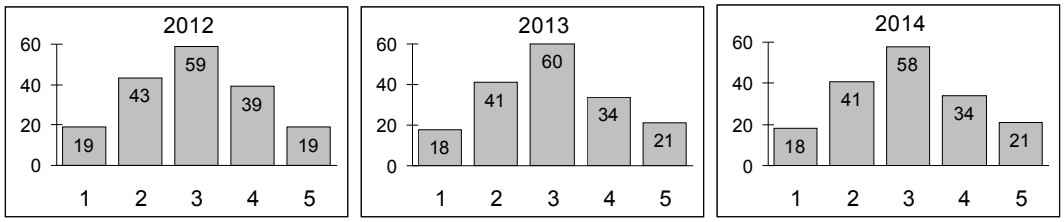
- **Nominal scale**(also referred to as categorical data), mutually exclusive data (e.g., “male/female”, “yes/no”, “green/red/blue”), cannot be ranked;
- **Ordinal scale**, categorical data (e.g., “agree much / agree some / disagree some / disagree much” or “0,1,2, …,10”), categories can be ranked, but their difference (the distance between them) remains unknown;
- **Interval scale**, the difference between two values is known *and* meaningful (e.g., difference between 100ºC and 90ºC is the same as between 90ºC and 80ºC), no clear definition of zero;
- **Ratio scale**, as all above, but with a clear definition of zero: when the value is zero, there exists none of that attribute (e.g., zero number of persons employed).

For programme monitoring purposes, data could in principle be on any of these data scales. However, in many cases it is practical to transform nominal, ordinal or interval scale data to a ratio scale.

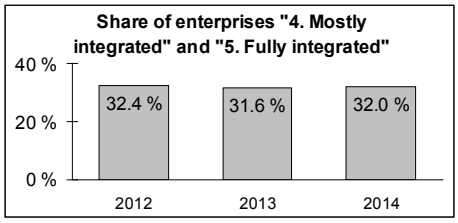
For instance, in a hypothetical survey conducted among the SMEs of a programme area, where the respondents are asked to choose their company’s level of integration across the border on a five-

[17] Additional measurement scales that bear less relevance for programme monitoring (such as the cyclic scale, graded membership scale, etc.) are not listed here.

item ordinal scale between 1 ("Not at all integrated") and 5 ("Fully integrated"), one would for three consecutive years obtain the following information:




It is difficult to assess the overall trend of the data set, to state any precise baseline, or to set up a clear target value with data at such a scale. The data can however easily be modified to a ratio scale for instance by counting the "number of enterprises stating either 4 or 5 as a percentage of all surveyed enterprises", thus obtaining the following information:




The new indicator is now on a ratio scale, which makes observing the overall trend clearer and the setting of baseline and target values easier.

## Data documentation

Meta data implies documenting all relevant technical aspects of an indicator. From a programming point of view, the aim of meta data is twofold.

On one hand a clear and unequivocal documenting of the method in which the indicator is constructed is a prerequisite for the transparency requirement. It provides stakeholders, evaluators, and the public, the possibility to objectively verify and assess the quality of the indicator.

On the other hand, meta data also aids a programme in the repetitive updating of an indicator during the course of it; an aspect of particular relevance in multiannual programmes, where the same indicator needs to be updated with frequent intervals. A well-documented method is hence also a thorough safeguard against changes in the programme personnel responsible for indicators.

Depending on the circumstances, some the following aspects should be reflected in the meta data:

- **Indicator data source(s)**, including all subcomponents (variables) of the indicator. For example, for an indicator like GDP/capita, if GDP in million euro stems from Eurostat, and the population in million people stems from different national statistical institutes (NSIs), all separate sources should be indicated. It is preferable for example to indicate not only "Eurostat" as the data source, but also the more detailed data base "Eurostat/Regional Statistics/Regional economic accounts/Gross domestic product indicators/Gross domestic product (GDP) at current market prices by NUTS 2 regions" and even, if available, an identifier of the data set ("nama_r_e2gdp" in this case). Alternatively, for printed sources, the name of the publication and tables utilised need to be indicated. This will aid future identification of the same data. Links to online data bases are also helpful, but they tend to change rather frequently. In case specific data extractions have been ordered (e.g., from a NSI), also the contact person responsible for the data extraction could be included;
- **Data source type and data population coverage**, whether the data stems from a register, a survey or other type of source, and whether that source covers the entire data population

(e.g., "all legally registered enterprises") or whether it lacks certain data populations (e.g., "all legally registered enterprises apart from enterprises in the primary production branches A.1.1 – A.5.7") that may affect the interpretation of it;

- **Measurement unit(s) of the indicator,** containing the unit of the actual indicator as well as the units of its possible subcomponents;
- **Data year(s) utilised**, containing the reference period of the actual indicator as well as the corresponding information of its possible subcomponents. If, for target value setting, time series have been used, it is preferable to indicate the availability of these as well;
- **Indicator construction method,** containing the possible method(s) of statistical processing of data: data transformation and/or amalgamation, data adjustment, data re-classification, data extra- or interpolation for missing years, data weighting (in case of composite indicators), and the like;
- **Data comparability** should also be assessed, particularly if there are issues that affect the comparability of the indicator across the programme geography;
- **Regional units** utilised are important to identify particularly in case there are deviances from the actual programme area (e.g. have data for NUTS 2 regions been used even if the programme area only includes some related NUTS 3 regions);
- **Any other information of relevance** not listed above.

In principle, meta data should include all necessary information so as to enable anyone (with basic indicator skills) utilising the recorded data sources to re-construct the indicator and end up with precisely the same figure as the original constructor did. It should also include all required information for enabling a critical external evaluation of the indicator, including its drawbacks, pitfalls and limitations.

In the course of the programming process, most programmes will in all likelihood conduct a more or less extensive data mining exercise in order to identify a broader spectrum of tentative data sources and likely indicator candidates, based upon which a smaller number of indicators will be then subsequently selected. It is preferable also to include a brief description of data sources that have been checked but eventually not used, and the reason therefore. This issue is helpful both if it turns out that an already chosen indicator for one reason or the other would have to be replaced, as well as for the purpose of the programming exercise beyond 2020. Finally, such documentation also demonstrates that "no stone has been left unturned", which acts as a partial justification (e.g., toward stakeholders and the like) for tentative limitations of an eventually chosen indicator.

## Setting baseline and target values

Each result indicator must have a fixed baseline value as well as a target value.

The **baseline value** is the value before the effects of the programme start taking place. As most (second-hand source) indicators generally tend to lag a few years behind, obtaining a baseline value for the actual kick-off of a programme is seldom feasible. Hence the latest available information will serve the need, provided that the time lag is "acceptable". If the time lag is "considerable", the appropriateness of the chosen indicator could be questioned.

The **target value** is the foreseen level of the indicator at the end of the programme[18].The same time lag as mentioned above should be taken into consideration when setting up the target value. For example, if the value of the baseline at the onset of the programme is a number that is two years old, then it is likely that a two year old figure will be available also at the end of the programme, which should be accordingly noted.

[18] 'Until the end of the programme' meaning at the latest at the moment of submission of the programme's final report.

The **baseline value for an output indicator** is by definition zero, as none of the foreseen actions have yet taken place. Setting **the target value for an output indicator** requires careful thought as that value will in effect be an ex ante quantitative estimation of what the programme is expected to produce. Beyond the "best possible estimate", there are few general guidelines as to how this target should be set. Such an estimate should however be motivated, and if proven incorrect during the course of the programme, may call for a revision afterward.[19]For interventions that have been in place previously (i.e., during the programme period 2007-2013), the process may be easier, as the estimation can be based on what projects in this period have reported.

Setting **the baseline for a quantitative result indicator** generally implies giving it the value of the latest available data. Setting **the target value for a quantitative result indicator** implies a need to analyse the past trend of the indicator, and relate this trend to the scope of the intervention. The basic question to pose would be: if output targets are met, how large would their expected effect on the result indicator be? However, one also needs to consider in which direction other external factors do point to.

| | | | | |
|---|---|---|---|---|
| **Change in the result indicator** | **=** | **Change resulting from the intervention** | **+** | **Change resulting from other (external) factors** |

Please refer to annex 1 and 2 for detailed answers to these questions.

When collecting data for setting up the baseline value for a result indicator, it is preferable to aim for collecting as long as possible time series. These aid considerably the setting of realistic target values. Relying on just one particular year without knowledge of the direction into which the indicator is moving could be hazardous. Collecting time series is also a preferable conduct in cases where data needs to be assembled for several geographic entities; identifying the newest "common data year" becomes then much easier.

Time series are also helpful in identifying volatile result indicators. This is, indicators that for one reason or the other display large variations from one year to the next. In such cases, it is an option to set the baseline as a moving average of several consecutive data years, thereupon eliminating some of the volatility. In that case, also the corresponding target value could be expressed as a comparable moving average.

Setting **targets for qualitative result indicators** could for example include denoting an expected pace and direction of change involving a change in a normative direction that is generally acknowledged (for example, an "improvement" of the current status). In case of categorical data, spelling out a range of expected outcomes is one option.

Setting targets in a relative manner could be equally relevant also for quantitative result indicators. For example, a programme might for the specific priority of "enhancement of competitiveness of SMEs" spell out the target as "a higher productivity increase than the corresponding weighted national increase for SMEs". Spelling out the target in this way takes into account economic fluctuations that lay beyond the scope of the programme itself.

Setting the baseline for an own survey requires that the survey is conducted before the operational programme is fully written, or at least at the very onset of the programme.[20]

[19] In situations of uncertainty, it is recommendable to include a contingency plan in the operational programme, addressing the possible future revision of the monitoring indicators during the course of the programme. This should be part of the Monitoring and Evaluation plan foreseen in Article 78.2 of the ENI CBC Implementing Rules.

[20] Please refer to the communication letter from DG DEVCO accompanying this guidance.

## Disaggregation of national data

As all ENI CBC programmes include territorial units smaller than the nation state, one of the most common bottlenecks is the scarce availability of regional data in comparison to what is available at the national level. In many cases a disaggregation (i.e., a regional breakdown) of national data could be a viable option providing – in the absence of more accurate information – at least an acceptable base for monitoring the achievements of a programme.

Typically utilised disaggregation denominators include relative shares of the national population, area, economy (e.g., volume of GDP), the labour market (e.g., number of people employed), as well as combinations thereof. The most important factor to consider is that the break down denominator could be argued to reflect regional differences and bear relevance for the actual indicator that is being broken down.

For example, if attempting to disaggregate available national estimations on $CO_2$ emissions produced by road transport to the regional level, it would not be defendable to break down the national figures by the relative share of national population in the concerned regions. Instead – if available – the registered number of cars per region would be a more viable option despite that a) car ownership and car usage are not the same, and that b) also other road vehicles than cars (trucks, buses, etc.) omit $CO_2$.

Similarly, if for example wishing to regionalise available national estimations on the economic volume of the tourism industry, utilising the relative share of the national economy as a disaggregation denominator would not be the most optimal choice, as regional differences in the economic structure are expectedly substantial. Using for example the number of available hotel beds would be a far better option, regardless of the fact that a) occupancy rates most likely vary, and that b) some sections of the tourism industry do not involve overnight stays.

If possible, it is wise to statistically check whether or not the chosen denominator has an acceptable statistical relationship with what is being disaggregated[21]. If nothing else is available, such a check can be performed utilising national data only. However, in many cases regional data is available for some but not all parts of a CBC programme, whereupon those parts where data are available can be used as test cases.

At the end of the day it is important to acknowledge that any information obtained by disaggregating national data is to be considered indicative at best. In the absence of available information it is however better to have something than nothing at all. In order to increase the credibility of such an indicator, an explicit documentation of the disaggregation procedure is paramount (see subchapter on “Data documentation” on page 17). All coming updates of the indicator should also be performed with the same disaggregation method. A further hypothetical example of disaggregation of national data is provided in Annex 3 (Case study 1).

[21] Depending on the type of data, correlation coefficients (e.g., Pearson, covariance, SSCP), linear regression, and the like are generally suitable methods. In case of very small samples (i.e., few regions in reference countries), using panel data (data for several years) is one option to increase the testing accuracy.

# 4. Condensed checklist for result indicators

Underneath is a brief checklist that covers the most essential features of result indicators described in the previous chapters. If, when applied by a programme, several conditions appear not fulfilled, a revision of the indicator should be seriously considered.

**The result indicator six-pack: a condensed checklist**

1. At least one result indicator for each programme priority (with corresponding output indicator[s])
2. A baseline and a target value for the result indicator
3. A result indicator responsive to policy and reflecting the foreseen results of the intervention
4. A plausible causal connection between the output and the result indicator so that a change in the output indicator is expected to induce a change in its corresponding result indicator
5. An objectively verifiable result indicator with sufficient definition and documentation so as to enable (i) a critical judgement and (ii) a repetitive calculation of it
6. A target value that is realistic in relation to the scope of the intervention

**Note:** this list is not exhaustive and there may also be other factors affecting the quality, relevance, and usefulness of the proposed result indicator.

## Annex 1: Causality between output and result indicator

The requirement of an intervention logic stresses the fact that a causal chain of reasoning between the output and the result indicator should be present in a similar way as the causal link between the identified problem and the chosen intervention. This places certain restraints and conditions for the design of suitable result indicators, as they cannot be developed in a complete vacuum. **A change in an output indicator should hence in some form be expected to induce a change in its corresponding result indicator.**

In an ideal situation, such a linkage has in the context of programme monitoring three principal conditions of causality to fulfil, which are:

1. that the output and the result indicator are associated (**relationship condition**);
2. that there is a temporal and logical precedence between the indicators (**antecedent condition**); and
3. that the observed association should not be explained by external factors not included in the measurement (**lack of alternative explanation condition**).

The hypothetical example below illustrates a practical application of these conditions and the related challenges that may occur.

A hypothetical programme utilises an output indicator named "additional population served by improved water supply" in order to monitor the related actions of the programme. It intends to use "decrease in the number of infectious drinking water born diseases" as a result indicator. The programme wishes to secure that there exists a plausible causal relationship between the two variables. It subsequently argues in the following manner:

(1) The **relationship condition** (a negative relationship in this case) could be expected to be fulfilled, as the more households will be connected – the less recorded diseases could logically be expected. This condition is also possible to validate statistically during the implementation of the programme.

(2) The **antecedent condition** stipulates that the reduction in diseases is a consequence of the increase in the number of connected households, not that the reduction of diseases is the cause for additional households being connected. It further stipulates that the reduction in diseases must be expected to happen after the households have been connected, not before. The programme concludes that also the antecedent condition could logically be expected to be fulfilled. Also this condition is possible to validate statistically during the implementation of the programme.

(3) Finally, the **lack of alternative explanation condition** requires that the reduction in recorded diseases is not the result of some third variable, but only the result of more households having access to safe drinking water as a result of the action. Whether or not this third condition is fulfilled in this hypothetical case can be questioned. The programme is aware that, for example, an increase in the general sanitation status (such as improved waste water treatment, improved food handling hygiene and other issues external to the programme's actions) is likely to also affect the result indicator, probably even to a much larger extent than the programme actions as such. Validating this condition statistically during the implementation of the programme is possible, but remains technically beyond the capacity of the programme.

At the end of the day, the programme concludes that despite this apparent partial deficiency in the causal relationship between output and result indicator, the chosen result indicator is nonetheless viable enough for monitoring the effects of the action.

In setting up the entire monitoring indicator framework, attention should thus be given to a selection of output and result indicators that are expected to be causally linked by fulfilling the three principal **conditions of causality**. Fulfilling the final lack of alternative explanation condition is (as the example above illustrates) difficult to manage due to the limited financial volume of the CBC programmes and the wide scope of actions in general, which implies that also other factors beyond the programme will have an effect on the result indicator. This challenge is further aggravated in a multi donor environment such as ENI CBC. If at all possible, this issue should hence be addressed by other means (see Annex 2).

Another common challenge in measurement is **spurious correlation**, which involves correlation via confounders. A confounder is a third variable that acts as a "transmitter" between the two variables of actual interest. A classical example from northern Europe is the recorded relationship between ice cream sales and drowning accidents. The more ice cream is sold, the more people drown. The confounder variable in this example is warm summer weather, which implies that people buy more ice cream and go swimming more frequently, which in turn increases the propensity for drowning accidents.

In CBC circumstances, spurious correlation could for instance be observed between cross-border investments in the fight against organised crime such as trafficking (output indicator) and an increase in *recorded (!)* crime levels (result indicator). The spurious correlation could stem from improvement of police efficiency that eventually leads to more trafficking cases being identified, recorded, and eventually prosecuted.

In this simple example the increase in police efficiency is the confounder variable responsible for more crimes being recorded. Identifying that confounder implies that – regardless of what the statistics have to tell – one will *not* arrive at the conclusion that the more programme money is being spent on the fight against organised crime, the more crime this will result in.

**Bidirectional relations** (i.e., relations in which the output and the result indicator affect each other simultaneously) are in practice rather frequent when it comes to programme monitoring. For example, improving the cross-border language skills (output indicator) may be expected to result in increased overall integration across the border (result indicator), but likewise could increased integration be expected to facilitate better language skills.

Although problematic from a purely scientific point of view, the use of such bi-related indicators is nonetheless defendable; provided that they meet the other relevant criteria for causality and that this bi-directionality is accordingly considered in the interpretation of the indicator.

# Annex 2: Measuring only the effect of the intervention

Irrespective of how successful a programme is, it can only to a limited extent affect overall societal trends. Ideally a programme management would want to measure the precise effect of the programme by comparing what would have happened *without* the programme to what has happened *with* the programme being implemented. Generally this is referred to as **measuring the counterfactual impact:** separating which part of the change the programme has directly contributed to and which part of the change is due to other external factors that lay beyond the effect of the programme.

Translated to the terminology of programme monitoring, any change in a result indicator is therefore a sum of what the programme has contributed to, as well as the additional effect of factors (which in most cases probably are larger) that lay beyond the scope of the programme. In other words:

| | | | | |
|---|---|---|---|---|
| **Change in the result indicator** | **=** | **Change resulting from the intervention** | **+** | **Change resulting from other (external) factors** |

In an ideal case, a result indicator should be able to measure the societal influence of the intervention alone and disregard effects produced by happenings outside the sphere of the intervention. Such is of course seldom the case, and in a majority of settings other programme external factors usually exert far more (either positive or negative) influence on the result than does the programme itself.

From a monitoring point of view this is alarming, as many commonly utilised result indicator frameworks are completely insensitive to this difference. Simple "before-and-after" analyses usually include the effects of both the programme as well as the programme-external effects without their differentiation.

Separating the effect of the programme from other factors is a technically challenging task to accomplish and considerable attention is usually required when constructing a bias-free framework for measuring results. How precisely this should be performed, depends entirely on the case.

More generally, as one can only (in retrospect) examine what happened in a situation when the intervention was carried out and not what would have happened without the intervention, other methods need to be utilised. By and large this involves some sort of comparison between those that received support ("the treatment group") and those that did not receive support ("the control group"). The difference in outcome between the two groups serves as an estimate of the effect of the intervention. The big challenge herein lays in finding a suitable control group against which to measure.

Both experimental and quasi-experimental (non-experimental) methods can be used for this. The former in broad terms includes techniques where one is able to statistically establish that there exist no differences between the two compared populations[22]. Data that actually allows for such techniques are hardly ever available in CBC programme circumstances.

The latter group of quasi-experimental techniques primarily involves comparing slightly differing populations to each other, thereby acknowledging and accepting a reduction of the accuracy of the

[22] "Population" here referring to whatever entities are being analysed, i.e. the population of enterprises, the population of protected areas, or the population of border crossing points, etc.

measurement of the effect. In terms of CBC programme result monitoring, the following techniques could tentatively be of relevance:

- **Difference-in-difference-estimation** (or Double Difference), where the treatment group and the control group are compared to each other at face value. If the supported population fares better than the non-supported one, then one could under certain conditions conclude that the intervention has had a positive effect. However, what if the intervention is by default directed primarily to high performers (for example those with a capacity to apply for support) that in any case would have fared better also without the intervention? In that case;
- **Discontinuity design** could be one solution. In a situation where support is directed to beneficiaries based on some sort of cut-off rate (e.g., "Small growth enterprises with maximum 49 employees and a turnover growth rate of more than 20 % per year"), then one can compare the development between those that barely fulfilled the criteria (e.g., supported enterprises with 45-49 employees) to those that were on the brink of fulfilling them (e.g., non-supported enterprises growing fast enough, but having 50-54 employees). When the two compared populations in this way are narrowed down, their development could be expected to be rather similar also without the intervention, which makes the comparison much more justified. This however generally requires that the overall populations are rather large;
- **Propensity score matching** is a further method of narrowing down the two compared populations so that they as far as possible resemble each other. In short, this statistical technique estimates the probability (propensity score) of an entity (e.g., an enterprise) being in the treatment group in comparison to the corresponding probability for all entities in the treatment group and the control group. Supported entities are then matched (either to only one non-supported one, or to many non-supported ones) on the basis of the propensity score, thereby creating fare pairs for comparison. Such techniques however usually require access to data rather sophisticated in ENI CBC programme circumstances; and
- **Judgemental matching** is tentatively a more feasible mode of conduct. It involves making an expert assessment as to what constitutes a suitable control group.

Other more sophisticated approaches include the instrumental variables method, or developing a statistical (often regression-based) model for measuring the counterfactual impact[23]. Case study 6 in Annex 3 describes the construction of an indicator that to a certain degree provides counterfactual evidence.

Common for all counterfactual methods is that they generally require longitudinal data for the treatment group as well as the control group both before and after the intervention. Such information may be challenging to obtain and programmes are in many cases likely to be confounded to more simple measurement endeavours involving being unable to separate the programme effect from all other external effects. To the least then, this fact should be made explicit and accounted for both when the result targets are set as well as when the progress of the programme in result terms is monitored and described.

[23] For a slightly more detailed, but still largely non-technical, description of counterfactual impact techniques and practices, see e.g. European Commission, Directorate-General for Employment, Social Affairs and Inclusion (2012): Design and Commissioning of Counterfactual Impact Evaluations. A Practical Guide for ESF Managing Authorities, Luxembourg: Publications Office of the European Union: 2013.

# Annex 3: Case studies on the development of result indicators

In order to illustrate the application of the theoretical considerations raised in the main document, seven case studies explaining the process of development of result indicators are herewith provided. Most of these cases are hypothetical, but a few of them are wholly or partially based on real-life examples from EU CBC programmes. The case studies cover a wide array of practical, day-to-day issues likely to be encountered by the programmes when developing their result indicators. They also depict the possible difficulties in obtaining suitable instruments for measuring the results of a CBC programme. The case studies aim at providing inspiration for innovative development of indicators in a situation where, at first sight, "no data are available".

**IMPORTANT!**

Please note that the **case studies should be beheld from the point of view of their methodological representativeness** rather than from their CBC ditto. The examples provided are thus explicitly chosen based on their ability to highlight critical attention points in general result indicator development alone and may in terms of their background narration not always bear great relevance for "pure" cross-border cooperation.

## Case study 1:

### Promotion of a low carbon economy -reduction of fossil energy consumption

❖ *Chosen result indicator*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| Reduction of fossil energy consumption | Estimated fossil energy consumption | Ktoe (1000 ton oil equivalents) | Average for last 3 data years | No more than 20 % increase* |

*In a situation where the increase *ceteris paribus* is expected to be substantial.

❖ *Keywords*

Hard measures, secondary data source, quantitative indicator, intervention logic, data disaggregation, volatile indicator, normative interpretation of indicator, missing counterfactual impact measurement, ***TO 6 "Environmental protection, climate change mitigation and adaptation"***

❖ *The context*

A land border crossing programme with regions from three countries aims within thematic objective (TO) 6 to promote the transformation to renewable energy by means of a programme priority focussing on a reduction of fossil energy consumption. The foreseen actions relate to investment support for renewable energy production.

The programme chooses "Increase in energy production capacity in MW of facilities using renewable energy resources, built/equipped by the projects, including electricity and heat energy" as a suitable output indicator, as it will be able to directly measure the concrete actions of the priority. The programme will obtain this data from the project data base.

The programme wishes to obtain a result indicator that as closely as possible matches the objectives of the priority ("reduction of fossil energy consumption") and is similarly linked to the output indicator. They opt for "Fossil energy consumption in ktoe" as the primary result indicator for this programme priority.

❖ *The method*

The programme performs a data mining exercise scanning the data bases of the NSIs, the ministries responsible for energy, as well as two university research institutes focused on the renewable energy economy. From one of the latter they learn that it is not possible to obtain data on fossil energy consumption from the producers on the market (as they only have data on production, of which some is exported, and some imported) and that only approximate estimations performed by the NSIs are available. This data is updated annually with a time lag of two years.

Regarding fossil energy consumption, data are available at the regional level for only one of the programme's three participating countries (Country A), hence leaving the programme no choice but to disaggregate the national data to the regional level for the regions of the two remaining countries (B and C). The programme decides that "having something is better than having nothing", and pursues this track. The discussion now focuses on what variable to use as a denominator. Population as such is deemed inappropriate because for example urban-rural differences in energy consumption are substantial.

As a denominator for disaggregating national data on energy consumption for countries B and C, the programme chooses a combination of the relative share of the participating region's volume of national GDP and the corresponding share of the participating regions area. The logic behind these chosen disaggregation variables is that energy consumption is reflected in at least the volume of the

economy (large economy = much energy consumed) as well as in the settlement pattern (large area = longer distances, generally sparse populations). For the sake of simplicity and transparency, it gives equal weight to both denominators.

The programme now tests this approach in country A, where actual regional data are available. They do it for all regions in that country by disaggregating the national data with the proposed method, and compare the outcome of this disaggregation to the actual situation by means of a correlation analysis, which indicates that the actual and the disaggregated estimated data are moderately correlated having a Pearson correlation coefficient of 0.76. The programme deems the disaggregation method accurate enough to provide a rough picture of fossil energy consumption in the regions of countries B and C.

The programme collected longer time series to be able to check the volatility of the indicator. The last five years for which common data were available displayed that there have been rather large variations in the past years. The programme consequently chooses the average for the last three commonly available data years as the baseline value.

In setting up the target value, the programme notes that a linear extension of fossil energy consumption points towards a general trend of rapid increase, whereupon the programme sets the target value at a "20 percent increase by 2023" compared to the baseline. The target value is equally a three-year average. The entire method is thoroughly documented including all data and references to their original sources.

- ***Additional aspects reflected upon***

The programme acknowledged that an increase in renewable production (or its capacity) may not necessarily lead to a reduction of fossil energy consumption, particularly within its programme area, where energy consumption *per se* has been on the rise for more than two decades. The programme nonetheless chose to pursue this path despite the minor deficiency in the programming logic, which in all likelihood also will spill over to the chosen indicators (fossil energy consumption increases despite additional capacity of renewable energy production).

The programme also discussed the fact that the limited accuracy of the disaggregation procedure will entail that measured changes within the programme area may entirely be the result of happenings outside the programme area. This holds especially true for country C, which has a rapidly expanding second-tier metropolitan area (located outside the programme area), where recent investments in aluminium smelters entail huge increases in energy consumption, of which a majority is still fossil.

The programme further debated the normative interpretation of the chosen result indicator. During the past two decades, all energy consumption (including fossil) has substantially been reduced during each economic recession. The normative interpretation of a reduction in fossil fuel consumption in such a situation could be questioned. The programme discussed whether it should relate the fossil energy consumption to the GDP in order to overcome this challenge. However, as they are aware of the technical challenges required for harmonising GDP (they all have separate national currencies), they decide to opt for the current indicator nonetheless.

The programme finally discussed the fact that they have no way of finding out how much the investments of the programme actually are contributing to reduced consumption of fossil fuels. They are aware that there are a multitude of related parallel efforts by the private sector underway as a consequence of increased world energy prices. They also discuss the inability for counterfactual impact measurement but see no realistic choice to overcome that challenge.

# Case study 2:

## Improving the mobility of persons and goods - road upgrading

### ❖ *Chosen result indicator*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| Upgrading of secondary roads to primary road standards | Estimated average speed on secondary roads | Km/h | 48 km/h | 5 km/h average increase |

### ❖ *Keywords*

Secondary data source, hard measures, quantitative indicator, distinction between output and result, data comparability & harmonisation, communicative value of indicator, ***TO 7 "Improvement of accessibility to the regions, development of sustainable and climate-proof transport and communication networks and systems"***

### ❖ *The context*

A land border crossing programme covering only the border regions of two countries is in a situation where its primary road network is in an acceptable condition and the area's main bottleneck is the poor condition of its secondary road network. The need for new road construction is not as acute, as the secondary road network spans the region acceptably; it is only the condition of it that is in need of improvement. It consequently aims within TO 7 to improve the mobility of persons and goods by means of a programme priority supporting the upgrading of secondary roads to primary road standards, thereby improving the accessibility of virtually all corners of the eligible area. The cross-border impact of such activities is expected to be significant. The planned activities relate primarily, but not exclusively, to hard investment support for such upgrading.

The programme chooses "Total length of reconstructed or upgraded secondary roads in km" as the primary output indicator for the priority. The programme considers this indicator being able to directly measure the concrete actions of the programme priority. The programme will obtain this data from the project data base.

The programme is now aiming at identifying a result indicator that as far as possible would reflect the expected results. Some in the programme argue that the upgrading of the roads in itself is the primary goal that should be reflected in the result indicator (which should hence be the same as the output one). Other voices instead claim that the primary objective should be the wider societal aim of improved mobility of persons and goods. The programme opts for the latter interpretation, adjusts the operational programme accordingly, and commences the search for an indicator reflecting this.

Modelled road accessibility would be an ideal indicator for this priority, but such data are only available for the EU Member State of the programme. Since the effort and cost of extending such data to the neighbouring country (and updating it for both) would be considerable, the programme opts for a more simple approach, and defines the indicative result indicator for this programme priority as "average speed in km/h on secondary roads".

### ❖ *The method*

The primary source for data on road usage lies with the related national ministries responsible for transport and communication. For one country, data on average road speed are available annually, from the other country only bi-annually. The programme will hence monitor this issue every two years only. As the programme at this stage has no conception of precisely which secondary road

stretches will be receiving support, it chooses to monitor all secondary roads within the eligible area.

The programme is now faced with the fact that the classification of roads differs between the two countries. After discussing the issue with experts in the concerned transport ministries, they opt for a joint definition that in country A involves including road stretches that in country B would be classified as tertiary roads. The experts assess this deviation as only increasing the total km of surveyed road stretches by some 15 %, which the programme deems acceptable[24].

The programme agrees upon the method in which the average speed will be calculated. Out of several possibilities they end up with "the average annual estimated speed in km/h per measured road stretch weighted with the length in km of each measured road stretch".

The programme sets the baseline value at the situation two years ago when the latest data for country B were available (even though country A had fresher data available), when the accordingly weighted average speed on these surveyed road stretches was 48 km/h. The corresponding target rate is set at an average 5 km/h increase by 2023.

❖ ***Additional aspects reflected upon***

The programme was pleased with the fact that the chosen result indicator is by and large able to measure the true impact of the programme itself, as they do not foresee any other major secondary road upgrading to take place in the regions concerned. The causal linkage between the output indicator (km roads upgraded) and the result indicator (average road speed increase) is also fairly straightforward.

They were however also aware that the effects of this road upgrading stretch beyond the upgraded roads hence implying both positive and negative effects for these areas. The former includes for example issues such as improved connectivity from areas adjacent to the supported road, whereas concerning the negative effects, the upgrading of the roads also results in increased pressure on the primary road network to which they connect, for which the average speed may actually be reduced due to increasing congestions. The programme nonetheless concluded that these issues are minor in comparison to the overarching results and need hence not be included in the measurement.

The programme also noted that the average speed *per se* does not take into account the actual accessibility of the residents. This in particular because this remote border region has an ageing population, many of which do not have access to a private car.

The programme further discussed whether to set the target value in absolute numbers (5 km/h average increase) or as relative ones (i.e. an increase of 10 %). Despite that there is no real difference between the two measurements the programme argued that an absolute number carries a far better communicative value for the general public and opted for that choice.

The programme finally debated the realism in being able to increase the average speed by as much as 5 km/h. The scope of the programme implies that tentatively less than a tenth of the roads will in the best case be concerned, and even these only partially. However, as the decisions to support will be made on a case by case basis, the programme has the possibility to at least include 5-6 strategically notorious bottleneck stretches (incl. two tunnels and one bridge) in the support; stretches that will have an impact also on other, non-supported, secondary road stretches in the region.

[24] As it is foreseen that the supported road stretches will in the course of the programme be decided upon on a case by case basis, the programme also makes a note of this deviation in the road classification; something which might have an effect on actual support decisions to be made during the implementation of the programme.

# Case study 3:

## Improving the mobility of persons and goods – development of strategies for connecting secondary rail networks

### ❖ *Chosen result indicator*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| Development of strategies for connecting secondary rail networks to main national rail grid, harmonising disparate national and regional rail strategies | Train/car travel time ratio, sample of road stretches | Travel time by car / travel time by train on Tuesday week 19, 12:00 hours ± 2 hours | 1.44 | 1.35 |

### ❖ *Keywords*

Soft measures, quantitative indicator, primary data collection, sample data, weak output indicator, ***TO 6 "Environmental protection, climate change mitigation and adaptation"***

### ❖ *The context*

A land border crossing programme aims within TO 6 to promote changes to a low carbon economy by supporting actions aimed at increasing local rail accessibility to the national rail grid in the three participating countries, thereby reducing the necessity for use of private car. As resources are not sufficient for hard investment support, the programme decides on soft actions related to development of strategies for connecting secondary rail networks in the eligible regions to the main national rail networks as well as harmonising disparate national and regional strategies in order to make optimal use of possibilities for cross-border connections. The programme also discusses whether such actions could have been executed within TO 7, but since the programme also wishes to work on a reduction of fossil energy consumption, they decide to carry out the here depicted actions under TO 6 instead.

The main target group for activities under this programme priority are national and regional authorities responsible for transport infrastructure planning, national rail network authorities responsible for rail operation, as well as local and regional authorities responsible for economic development and housing. As an output indicator for this priority the programme developed a programme-specific one, namely "Number of institutions using programme support for cooperation in strategic rail transport projects".

The ideal result indicator should be able to reflect the two principal objectives of the programme priority, namely: 1) an increase in secondary rail network usage as a consequence of the executed tasks; and 2) a reduction of private car usage for the same reason.

After a rather thorough data mining, the programme is faced with the fact that even though data on rail use would in principle be available for all three countries, no such corresponding information exist for private car use. The options are thereafter twofold: 1) either concentrate result measurement on rail use only; or 2) try to develop an alternative measurement that encapsulates both aspects.

The programme discusses the options and decides to pursue another track. They argue that even though they cannot measure road *and* rail usage as such, they could measure the propensities for

road and rail use, since inhabitants in the region tend to choose whether to use road or rail based on time consumption. (The cost is in this respect a secondary issue.)

The programme therefore concentrates result measurement on the difference in time between driving a personal car and using train services instead. As no data on driving time are available from national statistics in the concerned countries, the programme decides to develop an own indicator.

❖ ***The method***

The programme sends through the JPC out a simple survey to all 12 regional authorities and asks them to identify the three most strategic secondary rail transport lines in their corresponding region that connect to the national rail grid. By combining these answers, the programme constructs a list of 15 strategic secondary rail network connections where the start- and the end nodes are clearly defined.

The programme then commissions a local university to conduct a simple analysis identifying the driving time between these 15 pairs of nodes as well as the corresponding time it takes to perform the same trips using the train services. The survey is conducted for a particular Tuesday at 12:00 o'clock (± ca. 2 hours) for week # 19 that is not a national holiday in any of the three countries.

The train service times are obtained from the national train operators in each country. The car driving times are obtained from Google maps (where the "fastest driving time" option is chosen). The programme ends up with the following table:

| Route | | Travel time in minutes by | |
|---|---|---|---|
| From | To | Car | Train |
| Location A | Node 1 | 36 | 55 |
| Location B | Node 1 | 48 | 56 |
| Location C | Node 1 | 31 | 37 |
| Location D | Node 2 | 45 | 62 |
| Location E | Node 2 | 55 | 75 |
| Location F | Node 3 | 53 | 89 |
| Location G | Node 3 | 49 | 60 |
| Location H | Node 4 | 8 | 12 |
| Location I | Node 4 | 6 | 14 |
| Location J | Node 5 | 24 | 41 |
| Location K | Node 5 | 23 | 36 |
| Location L | Node 5 | 19 | 24 |
| Location M | Node 6 | 57 | 109 |
| Location N | Node 7 | 48 | 54 |
| Location O | Node 7 | 41 | 56 |
| *In total for all 15 routes* | | *543* | *780* |
| **Train/car travel time ratio: 1.44** | | | |

The train to car travel time ratio for these stretches is 1.44. In other words, it would take 44 % longer to travel these 15 routes by train than it would take to travel them by private car. The programme sets this as the baseline value. It sets the corresponding target rate for 2023 at 1.35, arguing that such a small reduction is a realistic goal taking into account the long time span for strategic infrastructure plans. The programme plans to conduct a similar survey every second year, utilising the same weekday of the same week each year (i.e. always the Tuesday of week # 19) and the same rail stretches.

❖ ***Additional aspects reflected upon***

The programme discussed the appropriateness of the output indicator, which in essence could be boiled down to "nr of projects × nr of participants in these projects". As the possibilities for better output indicators such as e.g. estimating the population benefiting from increased accessibility were rather limited (in only one of the three countries data on population were available at a lower territorial level than the region, which is too general), the programme nonetheless decided to go for the current output indicator.

The programme further discussed the feasibility of the result indicator in terms of it being extremely dependent on the travel time parameters of Google maps, which are largely reflections of the road hierarchy classification used as a basis for these calculations. No alternative solutions were however identified.

The programme finally discussed the fact that the strategic planning efforts foreseen within this programme priority will also concern other rail stretches than merely the listed 15 ones. The programme nonetheless argued these chosen 15 ones are identified as in this respect being strategically important for the 12 concerned regions and are able to act as indicative enough mirrors for the wider results of the programme.

# Case study 4:

## Improvement of border management operations – improving border staff competence

### ❖ *Chosen result indicators*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| Improved competence level and skills of land border crossing personnel | a) Annual nr of persons crossing the border as a ratio to nr of customs personnel directly employed at the border crossing points | Ratio of persons to personnel | Average of latest 2 data years | A five percentage increase of the ratio |
| | b) Annual nr of private cars crossing the border as a ratio to nr of customs personnel directly employed at the border crossing points | Ratio of cars to personnel | | |
| | c) Annual nr of trucks crossing the border as a ratio to nr of customs personnel directly employed at the border crossing points | Ratio of trucks to personnel | | |

### ❖ *Keywords*

Soft measures, quantitative indicator, primary data collection, partial data sources, external impact on indicator, ***TO 10 "Promotion of border management and border security, mobility and migration management"***

### ❖ *The context*

A land border crossing programme covers the entire stretch of border between two participating countries. It aims within TO 10 at improving the competence level and skills of land border crossing personnel, something that in the past has proven a bottleneck for increased cross-border trade and integration and in the extension, as a barrier for increased economic integration between the concerned countries. They plan to achieve this goal by supporting education of both customs and military border control staff. This among other involves language training, training in Schengen customs procedures, visa procedures, training in safety procedures, and the like. In order to optimise the integrative aspects of the actions, the programme primarily foresees joint education of staff on both sides of a border crossing point, whereupon such networking could additionally aid more effective integration of procedures and protocol.

The programme chooses "nr of staff directly involved in training" as a primary output indicator. In addition to this, four further output indicators are chosen: "Increased throughput capacity of private cars/24 hours"; "Increased throughput capacity of trucks/24 hours"; "Increased throughput capacity of persons/24 hours"; as well as "Number of border crossing points with increased throughput capacity" in order to be able to track in how many of them activities have been conducted.

The programme is looking for a result indicator that would be able to capture the effects of increased border staff efficiency and hence be directly related to the output indicator.

❖ *The method*

The programme discusses two separate broad measurement trajectories. On the one hand border staff efficiency could be viewed in terms of their capacity at handling the procedures required to operate the flows across the border.

On the other hand the wider CBC goal of increased integration should also tentatively be reflected. As virtually all commodity trade between the two participating countries is conducted by land transport (by and large only the bulk trade is performed on rail), measuring cargo flows (for other than bulk commodities) either as export or as import between the two countries could be one option. The programme discusses this possibility but eventually decides to discard it because the external impacts on changes in trade flows are too substantial to neglect. Trade agreements, global economic trends, and the like, all have a large effect on trade flows in comparison to border control staff efficiency alone. The causal link to the output indicators would be too weak and the trade in services, which is the most rapidly expanding export branch, would also not be included. Additional options at quantifiably measuring increased integration (as a consequence of more efficient borders) are limited, as e.g. cross-border commuting and cross-border employment are not major issues at this land border. The programme consequently investigates the first trajectory.

The programme argues that for example the number of staff required to operate border procedures as a ratio to throughput could constitute a viable option. Relating the throughput to personnel required would also partially omit the considerable annual fluctuations in trade flows.

They discuss whether to relate the nr of staff to throughput capacity (which would be available as it is an output indicator) or to actual throughput (i.e. not merely capacity). They decide to go for the latter one instead.

From the state agencies responsible for customs operations they learn that there are records of the actual number of people, cars and trucks crossing each border point (as it is a Schengen border) on an annual basis and that such data is in principle accessible to the programme, but with one year time lag. In principle corresponding data on train border crossings would be available, but the programme decides for the sake of simplicity to omit the single connecting train line between the countries from the measurement.

The programme is interested in measuring all three principal modalities on land borders (persons, cars, trucks) and discusses whether these could be merged into an index depicting all land cross-border traffic. They however decide that for communicative purposes (as well as for the sake of transparency) it is better to keep the modalities separate, and decide upon three parallel result indicator numerators for the programme priority: “Annual nr of persons crossing the border”, “Annual nr of private cars crossing the border” and “Annual nr of trucks crossing the border”.

As for the denominator (i.e. “nr of staff”) the programme learns that data on nr of customs staff employed per crossing point is available. However, the required regionalised (per crossing point) data implies that all supportive personnel (located elsewhere than at the actual crossing points) cannot be included in the measurement. Additionally, all data on military border control staff are classified information and hence not available. The programme decides nonetheless to pursue this path and defines the denominator as “nr of customs personnel directly employed at the border crossing points”.

Due to recent fluctuations, the baseline is set at the average value for the past two data years. By examining the major trend over the past six-seven years, the programme sees that a slight improvement in efficiency (measured in this way) is discernible in the data, on average some half a percentage per year. The target rate is hence set at a five percentage increase of the ratio, i.e. slightly over the general trend. This implies that either the same amount of throughput can be

handled with less personnel, or that an increased throughput can be handled with the current level of personnel, both cases indicating a normative improvement.

❖ ***Additional aspects reflected upon***

The programme discussed the impact of a situation with decreased throughput for example as a consequence of a global economic recession. In the short term at least, such a situation would not automatically lead to a corresponding reduction of personnel. Such sub-optimising of personnel would – despite all training efforts – consequently be directly visible in a worsening result indicator. The programme however concluded that this is an issue to be dealt with in case it emerges, and made a point of tentative needs for revising the indicator, should this happen.

More generally they discussed the fact that the result indicator has no possibility to account for the counterfactual impact of the activities of the programme. Fluctuations in both trade volumes and trade flows (which are reflected in the traffic) as well as customs personnel policy lie largely outside the scope of the programme activities; in particular since private operators are increasingly awarded contracts for the security controls of the border, operators for which no data are available.

The programme is nonetheless pleased with the fact that, apart from the tentatively substantial external influences, there is a partial causal linkage between one of the output indicators ("number of staff directly involved in training") and the three separate result indicators, albeit effects of external factors are expectedly substantial.

Finally the programme further discussed the fact that a considerable fraction (a rough estimate lands at half) of the staff operating the border are not accounted for due to the military border control staff as well as the supporting customs staff being omitted from the equation. This issue can however not be solved, and measuring some, they argue, is better than measuring nothing.

# Case study 5:

## Supporting the integration of vulnerable population groups - fostering activities aiming at reduction of poverty and social marginalisation

### ❖ *Chosen result indicators*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| a) Overall reduction of poverty | Population with equivalised disposable income below 60 % of the national median equivalised disposable income after social transfers*<br>or<br>Population below national poverty line | Persons | 26.3 mill. | Reduction to 25.8 mill.<br>(of which 0.08 mill. the expected programme contribution) |
| b) Reduction of the material welfare gap between the two shores of the programme area | Ratio of GDP/capita between the rich and poor shore of the programme area | Ratio in GDP/capita in PPP current international dollars | 2.41 | 2.35 |

*Official EU 2020 Strategy Headline Indicator

### ❖ *Keywords*

Sea basin programme, quantitative indicator, data disaggregation, rough estimations, use of monetary units in indicators, large external impact on indicators, output-result indicator congruence, ***TO 4 "Promotion of social inclusion and fight against poverty"***

### ❖ *The context*

A sea basin programme aims within TO 4 to support the integration of vulnerable population groups by fostering activities aiming at reduction of poverty and social marginalisation. Deeming it of crucial importance, the programme has decided to allocate a substantial share of its resources for this particular objective.

Foreseen actions cover a wide array of parallel endeavours, among other: educational activities aimed at raising the level of skill of marginalised groups, fostering start-up incubators targeted at vulnerable groups, as well as awareness raising activities regarding existing national social service options, etc.

The wide variety of actions call either for numerous parallel output indicators targeted at each foreseen type of action, or a more broadly applicable indicator that enables capturing roughly the essence of all activities, albeit with a more universal approach. The programme opts for the latter approach and identifies "number of participants in activities implemented by projects promoting social inclusion" as the principal output indicator.

The concerned programme area is in terms of social inclusion in practice divided into two halves. In the less affluent Partner Countries poverty levels remain high throughout and show only slow signs of being reduced. Poverty in the more wealthy EU Member States of the area is after a long period of reduction again on the rise due to increasing income differences and social stratification. Despite this turn in the trend, the material wealth gap in the region is still a primary cause for the substantial (both legal and illegal) flows of migrants in the region.

The programme hence identifies two separate but thematically connected objectives for the activities: reduction of poverty overall; and reduction of the material welfare gap between the two shores of the programme area. That calls for at least two separate result indicators. The programme realises that the area is so large, that the possibilities for conducting e.g. surveys in the entire eligible area are rather limited due to their cost. Hence, the result indicators should primarily be obtained from secondary data sources.

- ***The method***

The programme first discusses the benefits and drawbacks of absolute vs. relative poverty measurements. They agree that it would be more relevant to measure poverty in an absolute sense (e.g. through material deprivation measures or the like) but soon learn that no such corresponding data exist outside the EU countries. Hence only a relative measurement of poverty can be used. Apart from the EU regions, no such data are at the regional level available for the Partner Countries of the programme area. This data will hence for these countries have to be disaggregated from the country level.

For the regions in the Member States they utilise data on "persons with an equivalised disposable income below 60 % of the national median equivalised disposable income after social transfers". This data is available at Eurostat.

For the Partner Countries, they first collect national data on the "share of population below the national poverty line", which is being measured slightly differently in each country. Such data are available at the website of the United Nations Millennium Development Goals Indicators.

They then divide the number of these persons with the eligible region's share of total population in each corresponding country (obtained from respective NSI), ending up with a crude estimate on the number of persons below the national poverty line in the eligible area. They sum up the figures for the two shores of the programme area and end up with a rough estimate of 26.3 million impoverished persons two years ago. They set this as the baseline value.

By studying the time series of the data they see that the trend of poverty indicates roughly a quarter of a percentage unit decrease over the years on average. They set the target at a slightly faster pace than that, ending up with a target value of 25.8 million, which implies a reduction of poverty for some 500 000 persons. In light of the scope of the programme's activities and allocated funds, the expected "additional" effect (i.e. outside the past general trend) would constitute some 80 000 persons.

The programme then commences to the operationalisation of the measurement for the material welfare gap dividing the region. They realise that the collected data on poverty does not lend itself for measuring the overall welfare gap for two reasons. Firstly, because relative measure of poverty (as opposed to absolute ones) do not lend themselves to just comparison across the entire area (since they are relative) and secondly, because the poverty measures address only those classified as "poor" and exclude the remains of the population. A need for an additional approach is therefore needed.

The programme agrees that for this geography, GDP per inhabitant constitutes a good enough measurement of the level of material welfare. Regionalised figures are only available for the EU-parts of the programme area, where Eurostat carries time series on GDP in PPP (Purchasing Power Parities).

For both the Member States as well as the Partner Countries, the programme then collects national data from the World Bank on GDP in PPP in current international dollars. They first convert the Eurostat data to international dollars by relating the regionalised data to the national total, hence ending up with regionalised figures on GDP in international dollars for the eligible EU-internal regions.

They then for the eligible non-EU regions assess the eligible region's share of respective country total with data available regional accounts, ending up with estimates of regionalised figures on GDP in international dollars for the eligible regions in the Partner Countries.

However, for three countries of the programme area, no regionalised national GDP data are available at all. The programme hence roughly estimates this by simply using these regions' share of population of the national total as a denominator for dividing the total national World Bank figures on GDP in international dollars. They are well aware that this denominator is grossly misleading, but in light of the overall volume of GDP in the entire programme area, this over- or underestimation appears if not negligible, then at least minor.

Finally the programme sums up the obtained GDP data for the two shores of the programme area, divides this by the eligible population in these two areas, and relates these two figures to each other. They hence arrive at an estimated ratio of GDP/capita in PPP current international dollars between the rich and poor shore of the programme area of 2.41, which implies that GDP per capita in the richer part of the programme area is 141 % higher than the corresponding level in the less affluent part of the region. This is set as the baseline value.

Once more examining the trend for the past decade, the programme learns that this welfare gap has been reduced less than one tenth of a unit per year on average. The programme consequently sets the target rate at a modest level of 2.35, a rate slightly faster than the current trend.

- ***Additional aspects reflected upon***

The programme discussed the somewhat weak linkage between poverty reduction and overall material welfare levels. Albeit reductions in poverty as such have an extreme impact on the persons concerned, they tend to induce increases in overall levels of welfare to a much smaller extent than would for instance a similar welfare increase for the middle class or indeed, for the very richest. This challenge is consequently noted, although practically nothing can be done to address it. The programme agrees to monitor income stratification in each concerned country in parallel so as to assess where tentative reductions in difference in material welfare levels between the two shores of the area stem from.

The programme further discussed that the expected additional reduction in poverty (i.e. more than the general trend line reduction) of 80 000 persons is far above the target group of the programme, which in the best case will not cover more than 35 000 to 45 000 persons. They however argue that if the programme manages to aid a partial reduction of poverty in this target group, the overall effect will most likely be multiplied as a consequence of other members of the target population's households (i.e. children, elderly) benefiting from this as well. In light of that, the target rate of the programme is by and large congruent with the expected outputs of it.

On a more general level the programme noted that the external effects on the result indicators are substantial and largely outside the scope of the programme. Global economic trends, national shifts in social policy, an imminent risk of civil unrest, and the like, probably exert far more influence on the chosen result indicators than the programme is able to do. No solution to this measurement problem was identified, however.

# Case study 6:

## Promotion of entrepreneurship - support development of small enterprises

### ❖ *Chosen result indicator*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| Development of small enterprises | Turnover, average for whole enterprise population | Euro | Control group value: annual growth of 28 % | A higher annual growth on average for the treatment group than for the control group |
| | Nr of employed, average for whole enterprise population | Employed persons | Control group value: annual growth of 21 % | |

### ❖ *Keywords*

Counterfactual impact measurement, quantitative indicator, use of monetary units in indicator, ***TO 1 "Business and SME development"***

### ❖ *The context*

A land border crossing programme between two countries aims within TO 1 at stimulating increased entrepreneurship by supporting business development organisations. The foreseen actions relate to financial support of enterprise incubators, support that is to be channelled to a secondary target group. The secondary target group for these development organisations are small enterprises with a high growth potential, which are defined as enterprises no more than 5 years old, with 10-49 employees, and an average annual growth in turnover and employment of more than 20 %.

As the programme has performed similar activities also in the 2007-2013 programming period, they now wish to further refine their measurement by attempting to separate the effects of the programme from other external effects by means of a result indicator containing a measurement of the counterfactual impact.

As the principal output indicator for this priority the programme chooses "nr of business development organisations receiving support". The programme agrees that each supported development organisation must collect data on the enterprises that they in turn support. A minimum level of collected information concerns the number and change in number of persons employed by the company as well as the same for its annual turnover. This data is fed into a programme data base.

### ❖ *The method*

The programme aims at aggregating the employment and the turnover data on the supported enterprises that is recorded in the programme data base. This will constitute the treatment group (see Annex 2). The aggregation will involve translating the figures on turnover for one of the countries to euro, which is done using the official exchange rates of the European Central Bank, hence in effect disregarding the existing differences in purchasing power.

For both of the concerned countries, there are national (but not regional) statistics on the average change in employment and turnover of gazelle enterprises[25] available at the OECD. These two data

[25] Gazelles are enterprises that have been employers for a period of up to five years, with average annualised growth in employees (or in turnover) greater than 20% a year over a three-year period and with ten or more employees at the beginning of the observation period (OECD/Eurostat definition).

sets will constitute the control group. As the data for the control group also contain number of enterprises, the programme is able to aggregate the information for the two countries concerned (similarly as will be done with the treatment group data).

The programme sees that the control group has jointly for the two countries had an average annual increase in turnover of some 28 % and a related increase in employment of 21 %. The programme has no conception of what these figures could be in the treatment group[26] and consequently sets the baseline value as that of the control group.

The programme wishes *a priori* to account for possible economic downturns or rapid upswings that may affect the performance of these enterprises and therefore sets the target value as "a higher annual average growth in employment and turnover for the treatment group than for the control group". The observed difference in performance would, considering the limits of the measurement, constitute a rough indication of the true effect of the intervention.

❖ ***Additional aspects reflected upon***

The programme discussed the applicability of using the entire national populations of high growth enterprises as the primary reference group of the result indicator. In particular since the capital metropolitan areas in both countries are rather dynamic (in contrast to the border regions) and expectedly affect the national figures to a large degree. No solution to this challenge was however identified.

The programme further discussed the possibilities of utilising a discontinuity design (see Annex 2) in the selection of the control group, something which in light of available data was not possible. Purchasing data on high growth enterprises e.g. just below the gazelle threshold would have constituted a too large financial burden, in particular in contrast to the freely available data currently utilised.

The programme also discussed the effects of the control group containing also data on gazelles with more than 49 employees, something which the treatment group does not contain. In absolute terms, growth in both employment and turnover could for some of these larger enterprises be rather substantial in comparison to the corresponding summated growth in the smaller (i.e. 10-49 employees) enterprises, hence affecting the overall average figures for the control group. A telephone conversation with the two national chambers of commerce however revealed that the number of these enterprises could be expected to be rather limited, whereupon the measurement problem in this respect is likely to remain rather limited and should cause no alarm.

In addition, the programme reflected upon the fact that the control group by definition also includes all enterprises from the treatment group which affects the control group's summated performance. It discussed whether it would be feasible to subtract the performance of the treatment group from the control group's ditto, but decided that since the treatment group nonetheless is expected to be comparatively small, this would not be necessary to perform, hence also improving the overall transparency of the result monitoring exercise.

The programme also reflected upon the fact that their chosen result indicator in practise does not span beyond the direct beneficiaries of the support. The result in terms of efficiency in other (non-supported) gazelle enterprises within the eligible area remains unknown. No practical solution to overcome this issue was however identified.

The programme further discussed the challenge that while the treatment group of enterprises in principle is a stable one (only with more enterprises included in it while the programme emerges, but none dropped out of it), the control group is not stable. It only includes enterprises that meet

[26] Apart from the minimum values of the enterprise selection criteria.

the set criteria of gazelles at a given point in time, and if an enterprise does not anymore conform to these criteria, it is not anymore included in the measurement. No solution to this challenge was however identified, since obtaining longitudinal control group data would require specific data runs and consequently be very costly.

Finally the programme noted that the currency unit of the measurement of the increase in turnover is based on euro in current prices, which provides a skewed picture of the actual increase since it does not take into account inflation. The programme however assessed that a transformation of the aggregated treatment and control group turnover into fixed prices (not affected by inflation) is technically too challenging for them. In the current situation this deviation is nonetheless expected to be fairly similar across the entire programme area. The programme however makes a note of this issue and agrees to take the matter into consideration in a tentative future situation where changes in inflation would be substantial between the participating countries.

# Case study 7:

## Increasing capacity of local authorities - development of local governance

### ❖ *Chosen result indicator*

| Objective | Result indicator | Unit | Baseline | Target |
|---|---|---|---|---|
| Improvement of local level governance practices | Expert panel statement on the status of local level administrative and governance practices and skill levels | Qualitative, but contains 3 categories, i.e. issues:<br>i) in need of acute action;<br>ii) in need of tentative attention; and<br>iii) satisfactory for the time being. | A priori assessment | Qualitative improvement implying among others a shorter list nr i) |

### ❖ *Keywords*

Qualitative indicator, ***TO 5 "Support to local & regional good governance"***

### ❖ *The context*

A land border crossing programme aims within TO 5 at increasing the capacity of local level authorities in the general field of governance, where there exist well-documented challenges in all three concerned countries. This somewhat vaguely stated objective is paralleled by a large variety of foreseen actions, stretching from (a cross-border) transfer of best practices to joint development projects for improved governance and other administrative practices.

Possibilities for good output indicators remain limited. Well aware of the vagueness of it, the programme nonetheless opts for "nr of participating organisations cooperating for improved governance". The discussion on how to obtain a matching result indicator then commences.

### ❖ *The method*

The programme discusses the options for obtaining quantitative data on governance, improvement of which is the principal objective of the programme priority. Local level adaptations of e.g. the underlying subjective assessments of the World Governance Indicators are discussed. These are however deemed unfeasible from a thematic point of view and unrealistic due to the number of participating local authorities being fairly large (and hence costly to evaluate). The programme decides to abandon the quantitative approach and strives for a qualitative indicator instead.

The programme opts for an expert panel judgement on the level of local level administrative and governance practices and skill levels and their development during the course of the programme. The programme establishes this expert panel before the programming phase is finalised in order to obtain a baseline assessment of the current status.

The JPC accordingly contacts a collection of university scientists, regional level practitioners (that are not involved with the programme), a limited representation of stakeholders (NGOs, chambers of commerce), as well as one representative from the national ministries responsible for local level administration. Simultaneously the programme obtains loose commitment from the same persons to participate in the panel two times during the course of the programme (mid-term and ex-post)

During the programming stage, the panel meets two times, where the second meeting results in a collective statement on the status of local level administrative and governance practices and skill

levels also containing a list of issues in need of acute action, issues in need of tentative attention, as well as issues deemed satisfactory for the time being.

The same procedure is repeated for all three parts of the programme area, whereupon the (internally elected) chairmen and vice chairmen of the panels from each country meet once, and discuss a joint assessment for the whole eligible area. This assessment is then set as the baseline status.

Bearing in mind that this is a CBC programme, it is agreed that this joint assessment should be limited to issues that are common to all three participating countries and hence of cross-border relevance. No definite upper "size limit" is set for the list. The programme argues that the larger the pool of tentative aspects in need of action or attention, the better are the possibilities for finding relevant issues that could be addressed by cross-border cooperation.

The target rate is set as a qualitative improvement implying among other that the list of issues in need of acute action would get shorter, and that the corresponding list of issues in need of tentative attention has changed (i.e. acute issues transferred here, tentative attention issues transferred to next level).

- ***Additional aspects reflected upon***

The programme discussed the challenges in the composition of the expert panel. In order to safeguard an unbiased assessment it decided that a large fraction of its members should be either not connected to local or regional administration, or alternatively should primarily come from outside the regions to be assessed.

The programme further discussed how to overcome that the expert panel does not get stuck on issues stemming from the considerably varying governance and administrative traditions between the three participating countries. It concluded that the expert panel should consist of persons that have been exposed to international settings.

The programme finally discussed methods of obtaining the forthcoming assessments from the panel in such a mode, that qualitative improvements observed could be attributed to the activities of the programme. As a partial solution for this, the programme decided to keep the expert panel thoroughly informed of the concrete activities of the programme with separate information meetings prior to each scheduled expert panel meeting.