



GUIDELINES

ASSESSMENT OF RDP RESULTS: HOW TO PREPARE FOR REPORTING ON EVALUATION IN 2017

FINAL DRAFT – July 2016

Disclaimer: This draft version has not yet been subject to a language and editorial check.

Copyright notice

© European Union, 2016

Reproduction is authorised provided the source is acknowledged.

Recommended citation: (DRAFT VERSION NOT YET TO BE QUOTED) EUROPEAN COMMISSION – Directorate-General for Agriculture and Rural Development – Unit E.4 (2016): Guidelines. Assessment of RDP results: How to prepare for reporting on evaluation in 2017. Brussels

Disclaimer:

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this report. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.



The Evaluation Helpdesk is responsible for the evaluation function within the European Network for Rural Development (ENRD) by providing guidance on the evaluation of RDPs and policies falling under the remit and guidance of DG AGRI's Unit E.4 'Evaluation and studies' of the European Commission (EC). In order to improve the evaluation of EU rural development policy the Evaluation Helpdesk supports all evaluation stakeholders, in particular DG AGRI, national authorities, RDP managing authorities and evaluators, through the development and dissemination of appropriate methodologies and tools; the collection and exchange of good practices; capacity building, and communicating with network members on evaluation related topics.

Additional information about the activities of European Evaluation Helpdesk for Rural Development is available on the Internet through the Europa server (<http://enrd.ec.europa.eu>).

GUIDELINES

ASSESSMENT OF RDP RESULTS: HOW TO PREPARE FOR REPORTING ON EVALUATION IN 2017

FINAL DRAFT – July 2016

CONTENT

INTRODUCTION.....	7
PART I – FOCUSING, MANAGING AND REPORTING ON EVALUATION	9
1 Reporting requirements in 2017	11
1.1 Monitoring and evaluation requirements in the AIR 2017	11
1.2 Evidence used in the evaluation in 2017?	14
1.3 Reporting in case of low uptake and small programmes.	15
1.4 Reporting on the evaluation of RDP specific elements	16
2 Links to other EU reporting requirements and reporting requirements beyond 2017	18
2.1 Links and synergies between the AIR 2017 and other EU reporting requirements	18
2.2 Links between reporting in 2017 and the overall reporting requirements on rural development in 2014-2020	22
3 Ensuring good quality evaluation in 2017	23
3.1 Careful planning and preparation of evaluation	23
3.2 Quality assurance and quality control across the evaluation process	23
4 Dissemination, communication and follow-up of evaluation results	26
4.1 Dissemination and communication of evaluation in 2017	26
4.2 Follow-up of evaluation findings.....	26
PART II SETTING UP THE SYSTEM TO ANSWER EVALUATION QUESTIONS AND ASSESSING RDP ACHIEVEMENTS	28
5 PREPARING THE EVALUATION	30
5.1 Revisit the RDP intervention logic.....	31
5.2 Link intervention logic to evaluation elements	37
5.2.1 Define evaluation elements: evaluation questions, judgement criteria, indicators ..	37
5.2.2 Check consistency of evaluation questions and indicators with RDP intervention logic	42
5.2.3 Develop RDP specific evaluation elements	47
6 STRUCTURING THE EVALUATION	54

6.1	Set up a consistent evaluation approach	55
6.1.1	Selection of a suitable evaluation approach to the evaluation of RDP results in 2017.....	55
6.1.2	Select evaluation methods and their combination.....	59
6.2	Establish the evidence for evaluation	64
6.2.1	Review data requirements for the chosen methods	64
6.2.2	Manage and collect data for evaluation	71
7	CONDUCTING THE EVALUATION	74
7.1	Observing.....	74
7.2	Analysing.....	76
7.3	Judging.....	80
PART III - Annexes		81
Annex 1 – Proposed SFC template for point 7 of the AIR submitted in 2017		82
Annex 2 - Elements of the CMES and their use in evaluation		95
Annex 3 - Overview of reporting requirements on RDP implementation and evaluation in the programming period 2014-2020		98
Annex 4 – Working Steps in Setting up the System to Answer the evaluation questions in 2017		101
Annex 5 - Check-list for assessing the quality of the evaluation report		106
Annex 6 - The role of rural development stakeholders in the dissemination and communication of evaluation results		108
Annex 7 - Minimum requirements and good practice in reporting on evaluation in the AIR		109
Annex 8 - Timing for preparing and implementing the system to answer the CEQ, and answering CEQ		111
Annex 9 – Criteria for selection of evaluation approach		113
Annex 10 - Tools for qualitative appraisal of RDP operations ´ primary and secondary contributions to focus areas and synergies between measures, between FAs and between RD priorities		114

TABLES, FIGURES AND TOOLS

Table 1. Links between AIR 2017 and other EU reporting requirements	20
Table 2. Example of judgment criteria for focus area-related CEQ and PSEQ	38
Table 3. Evaluation task and data needs in 2017	65
Table 4. Types of data used in rural development evaluation	66
Table 5. Methods and related data needs (example)	67
Table 6. Evaluation stakeholders and their role in ensuring data availability and quality	69
Table 7. SFC template for point 7 on Common Evaluation Questions no.4 (empty template)	83
Table 8. SFC template for point 7 on Common Evaluation Questions no.4 (filled example)	87
Table 9. Timing for preparing and implementing the system to answer the CEQ, and answering CEQ	112
Table 10. Primary and secondary contributions of operations within measures/sub- measures to rural development focus areas – example	115
Table 11. Synergies and negative transverse effects between focus areas and between priorities	20
Figure 1. Content of Annual Implementation Reports	9
Figure 2. The common and programme specific monitoring and evaluation system ...	16
Figure 3. Links between AIR submitted in 2017 and other EU reporting requirements in 2017 and 2018	18
Figure 4. Overview of reporting requirements and links 2016 – 2019	19
Figure 5. Overview of the evaluation process	23
Figure 6. Major steps in setting up the system to answer evaluation questions	29
Figure 7. Different types of programme effects	34
Figure 8. Relation between evaluation elements	38
Figure 9. Horizontal and vertical consistency between objectives, evaluation questions and indicators	43
Figure 10. The intervention logic and evaluation elements for communication and information exchange	46
Figure 11. LEADER as an evaluation topic	62
Figure 12. Decision-making process for data requirements and data management. ...	73
Tool 1. Matrix for checking horizontal consistency at focus area level (example) .	44

Tool 2.	Matrix for vertical consistency check (example)	45
Tool 3.	Working procedure for filling the gaps in evaluation elements using the SWOT analysis	50

INTRODUCTION

Starting in June 2016, and each year until 2024, Member States shall submit to the Commission an annual implementation report (AIR). The AIR provides information about the implementation of the rural development programme, as well as the evaluation plan.

The AIR submitted in 2017 (hereinafter referred to as 'AIR 2017') shall also include the quantification of programme achievements, in particular through the assessment of the result indicators (including complementary result indicators), and answers to relevant evaluation questions.

The Technical Handbook of the CMEF of the CAP 2014 – 2020 and its annexes provide general guidance as well as detailed fiches for each result indicator to be reported in the AIR 2017 and 2019. Each fiche contains the link to the RD priority and the focus area, the definition of indicator, unit of measurement, the methodology for calculation, data requirements and sources, including those collected via the operations database, point and frequency of data collection, and the means of data transmission to the Commission. The definition and detailed description of common indicators are provided in the legal framework and the above-mentioned fiches.

However, in order to support the managing authorities and evaluators in using them, and to ensure consistency across all Member States and RDPs, it was agreed that additional guidance would be provided. It should address questions which are frequently asked by evaluation stakeholders in the Member States, such as:

- How to use complementary result indicators in the assessment of RD interventions in the AIR 2017? How to attribute results to the RDP interventions?
- How to address the assessment of secondary contributions of RD operations to focus areas under which they have not been programmed but to which they also make a contribution?
- How to address challenges caused by the flexibility in programming of EAFRD interventions in individual RDPs?
- How to deal with gaps in the monitoring and evaluation system linked to each individual RDP caused by a flexible programming approach?
- How to develop programme-specific result indicators and related evaluation questions?
- How to report on the assessment of RDP achievements in 2017?
- How to communicate and disseminate evaluation results?

With a view to comply with legal requirements and enable Member States to establish a robust monitoring and evaluation system, capable of generating the required information to be reported in the AIR 2017, the above-mentioned questions should be answered at an early stage of programme implementation.

Against this background the Annual Work Programme 2015 of the Evaluation Helpdesk foresaw to build up on existing legal framework and guidance, and develop support for Managing Authorities, paying agencies and evaluators in reporting on the RDP achievements in the enhanced AIR 2017, and beyond.

Under the guidance of DG AGRI Unit E.4 “Evaluation and studies”, a Thematic Working Group consisting of qualified evaluation experts was established in order to:

- Examine the challenges related to the reporting on evaluation in the AIR 2017, in consultation with stakeholders;
- Discuss possible solutions to overcome these challenges considering methodological and practical issues;
- Identify effective approaches to assess the progress in achieving the objectives of the RDP, in particular through the assessment of the result indicators, including complementary result indicators, and of secondary contributions and answering related evaluation questions;
- Develop guidelines for preparing and drafting the evaluation component of the enhanced AIR 2017.

Work has been carried out in a series of thematic workshops, desk-research, drafting, editing and lay outting. The draft guidance document was discussed with the Expert Group on Monitoring and Evaluating the CAP in several stages: during its 7th meeting on 25th June 2015 the Expert Group discussed the outline of the guidelines; a Sounding Board composed of Expert Group members discussed the 2nd draft of the guidelines in September 2015; during its 8th meeting on 12 November 2015 the Expert Group discussed the final draft of the document. The present guidelines have been structured in several parts:

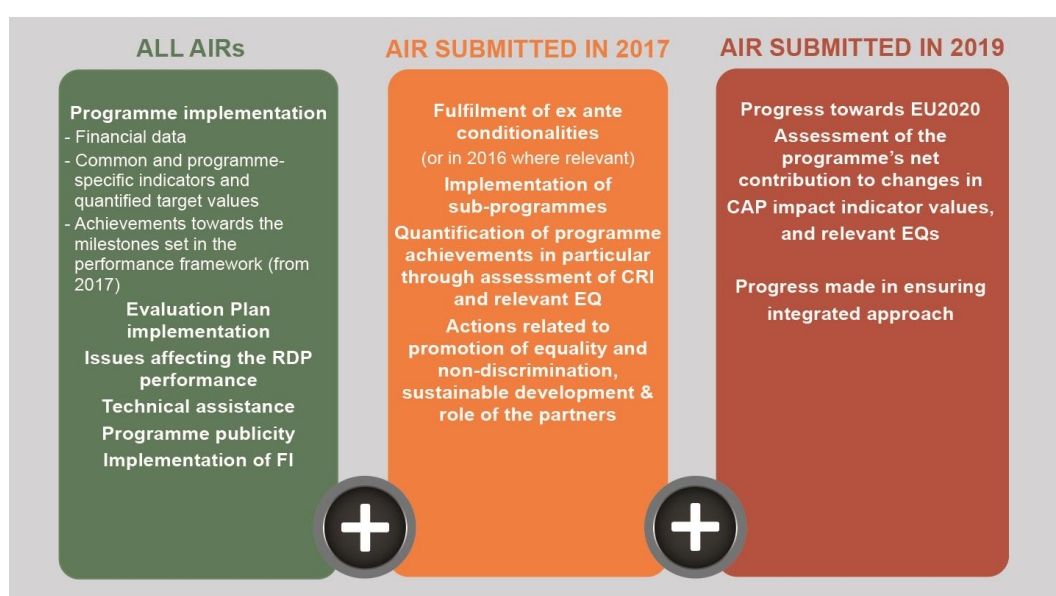
- **PART I** (mainly for managing authorities) provides information and recommendations on what needs to be reported in the AIR 2017. Specific attention was given to the governance and management of evaluation, including the involvement of various actors and communication of evaluation results.
- **PART II** (for managing authorities, paying agencies and evaluators) provides detailed guidance on the preparation, structuring and conduction phases of evaluation, including methodological guidance and recommended practices. This part leads the reader through the process of answering all common and programme-specific evaluation questions for the entire programming period (with a specific focus on the information to be reported in the AIR 2017).
- **Part III – Annexes** consisting of several practical tools such as the SFC template for point 7 of the AIR 2017, an overview of reporting requirements, check-lists for assessing the quality of the evaluation report, terms of reference, etc.
- **A separate annex** is providing fiches with detailed methodological guidance on how to address each CEQ (no. 1-21) to be answered in 2017.

PART I – FOCUSING, MANAGING AND REPORTING ON EVALUATION

Overview of Annual Implementation Reports

The 2014-2020 programming period will include two enhanced AIRs, which combine both monitoring and evaluation elements and will be submitted in 2017 and 2019. The monitoring elements of the 2017 and 2019 enhanced AIRs are identical to previous standard AIRs. However, more evaluation elements will be included as the programming period advances.

Figure 1. Content of Annual Implementation Reports¹



Source: European Evaluation Helpdesk for Rural Development, 2015

Common monitoring and evaluation framework as basis for reporting

The legal framework sets up a common monitoring and evaluation system (CMES) for rural development² as part of the common monitoring and evaluation framework for the whole CAP (CMEF). The CMES ensures that a common evaluation approach is applied across the EU in **conducting and reporting** the evaluation of rural development programmes, namely:

¹ Annex VII to the Commission Implementing Regulation (EU) No 808/2014 of 17 July 2014 laying down rules for the application of Regulation (EU) No 1305/2013 of the European Parliament and of the Council on support for rural development by the European Agricultural Fund for Rural Development (EAFRD) (OJ L 227/18 31.7.2014)

² Article 14 and 67 of Regulation (EU) No 1305/2013 and Annexes IV, V and VI to the Commission Implementing Regulation (EU) No 808/2014

- Demonstrating the progress and achievements of the EU and Member State rural development policy and assess the policy's impact, effectiveness, efficiency and relevance,
- Supporting a common learning process on how to prepare and implement rural development policy better,
- Contributing to better targeted support for rural development.

The CMES is designed to enhance the accountability and transparency of the EU rural policy and to facilitate the **reporting on evaluation findings**, as well as their synthesis at the EU level.

1 REPORTING REQUIREMENTS IN 2017

1.1 Monitoring and evaluation requirements in the AIR 2017

The overall reporting requirements for the AIR 2017 combine monitoring and evaluation elements.

- **Monitoring:** The MA will gather and report information on various issues, such as progress on implementing the Evaluation Plan³, financial commitments and expenditures⁴, data on indicators⁵ and progress towards targets as in standard AIR. The MA will also ensure that monitoring information and other available data will be used as input to evaluation.
- **Evaluation:** The MA will report in the AIR 2017 information resulting from evaluator's work, namely on the assessment of programme's achievements through quantification of common, additional and programme-specific result indicators and answers to evaluation questions.

By whom?

The AIR 2017 must include findings from RDP evaluation⁶ carried out by internal or external experts that are functionally independent from the MA.⁷ The reporting itself, i.e. summarising the main findings and recommendations should be done by the MA, who is the "owner" of the AIR.

What is required to be reported on evaluation in the AIR 2017?

The **main focus of the evaluation to be reported in 2017** is on results achieved so far. In this respect the AIR 2017 should include the following information stemming from evaluation activities:

- Summary of evaluation through the quantification and assessment of programme achievements denoted by result indicators (including complementary result indicators⁸, programme-specific indicators and additional indicators). The quantification and assessment of indicators will be based on primary and secondary contributions of RDP operations⁹. Values of result indicators should relate to completed operations. If there are no completed operations, it may be necessary, depending on the stage of implementation, to assess selected ongoing operations.

³ Annex VII, point 2, to the Commission Implementing Regulation (EU) No 808/2014

⁴ Annex VII, point 1(a) to the Commission Implementing Regulation (EU) No 808/2014

⁵ Annex VII, point 1(b) to the Commission Implementing Regulation (EU) No 808/2014

⁶ Annex VII, point 7 to the Commission Implementing Regulation (EU) No 808/2014

⁷ Article 54 (3) of Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013 laying down common provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development and the European Maritime and Fisheries Fund and laying down general provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund and the European Maritime and Fisheries Fund and repealing Council Regulation (EC) No 1083/2006 (OJ L 347/320 20.12.2013)

⁸ Article 14(b) and Annex VII, point (7) to Commission Implementing Regulation (EU) No 808/2014

⁹ Additional contributions of operations to focus areas, other than those under which they are programmed.

- Validation of values of target indicators which have been set up at the time of programme design.
- Answer relevant (focus area related and related to other aspects of the RDP) evaluation questions¹⁰ based on the assessment of the above indicators.

Recommended practices:

- Different forms of support (financial instruments) applied in the implementation of RDP measures can also affect the programme's achievements and the net values of result indicators. Therefore, the **comparison of the effects of various financial instruments** (in cases where they are applied) may be part of the assessment.
- The **assessment of delivery mechanisms** (as part of the assessment of performance of the programme), is considered to be a good practice. Delivery mechanisms, i.e. the set of processes and procedures that ensure that policy objectives become concrete actions¹¹, is seen as one of the factors which affect the performance of the programme and the measures taken¹². Among the issues to be analysed with respect to the RDP delivery mechanism and its effectiveness and efficiency, are:
 - Targeting the EAFRD support to selected groups of beneficiaries,
 - Selected types of support (financial instruments, simplified cost options) with respect to measures,
 - Information dissemination of EAFRD support to beneficiaries,
 - Application, selection, contracting and payment procedures,
 - Monitoring mechanisms and feedback from beneficiaries on the application of selected measures etc.

What is the reporting period for the AIR 2017?

The information to be reported in 2017 should cover the years 2014-2016. The compilation of the report and the evaluation tasks (hereinafter: evaluation in 2017) has consequently to take place in the first half of 2017.

In order to have a global picture of policy results/achievements in relation to the 2014-2020 Multiannual Financial Framework expenditure, Member States should take into account in their evaluations the portion of the 2014-2020 budget implemented according to 2007-2013 rules ("transitional expenditure").

How should Member States report on evaluation in 2017?

The reporting on evaluation in the AIR 2017 will be done through the AIR SFC template for point 7, which will be composed of tables prepared separately for:

- a) each of focus area-related common evaluation questions (CEQ), number 1 – 18,

¹⁰ Article 50 of Regulation (EU) No 1303/2013 and Annex VII, point 7 of Commission Implementing Regulation (EU) No 808/2014

¹¹ http://enrd.ec.europa.eu/enrd-static/policy-in-action/improving-implementation/delivery-mechanisms/en/delivery-mechanisms_en.html

¹² Annex VII, point 3 of Commission Implementing Regulation (EU) No 808/2014

- b) common evaluation question related to other aspects of RDP
 - TA and NRN, CEQ number 20 and 21
 - Programme synergies, CEQ number 19
- c) programme-specific evaluation questions (PSEQ) linked to assessment of:
 - programme specific focus areas (to be multiplied in accordance to number of specific focus areas included in the RDP)
 - RDP specific topics (e.g. assessment of delivery mechanism)

The information collected via tables of the SFC template for point 7 summarises key points important for the EU and the Member States. This information on evaluation submitted through the SFC-template point 7 does however not include the full evaluation report of Member States which in contrast to the last programming period is not anymore submitted to the EC.

The template contains the mandatory parts, namely:

- **Title of the evaluation question** (applicable for all types of evaluation questions),
- **Rationale** explaining why the programme specific evaluation question has been employed in the RDP evaluation (applicable for evaluation questions mentioned in point c) i and c) ii),
- **Intervention logic of the FA** to which is the evaluation question linked: primarily programmed measures/sub-measures and measures/sub-measures programmed under the other FA but showing the secondary contribution to this FA (applicable for evaluation questions mentioned in point a) and c) i),
- **Links between judgment criteria and indicators used to answer the evaluation question** (applicable for evaluation questions mentioned in point a), in which case the links between judgment criteria and common and additional indicators are shown, and for evaluation question mentioned in c) i, where links between judgment criteria and programme specific indicators are shown),

Indicators:

Common CMES indicators¹³ which are used to answer the common evaluation questions no 1 – 18 cover common result/target indicators and complementary result indicators. If relevant also common context indicators can be used for answering CEQ.

Additional indicators (developed by Member States) are used in case the common indicators are not sufficient to provide robust answers to common evaluation questions specified with the commonly defined judgment criteria. Additional indicators can be also developed if Member States employ their own judgments criteria.

Programme-specific indicators (developed by Member States) are used to answer programme-specific evaluation questions, in case common and additional indicators cannot be used for this purpose.

- **Description of the quantitative and qualitative methods** applied to assess the result indicators, including explanation why the method was used, description of

¹³ Article 14(1)(b) and Annex IV to Commission Implementing Regulation (EU) No 808/2014

steps and highlighting the challenges of using the method and solutions to overcome them (applicable for all types of evaluation questions),

- **Quantitative findings** in values of all indicators as the outcome of quantitative analysis, used in answering the evaluation questions and data sources (applicable for evaluation questions mentioned in point a), b) i, and c) i),
- **Answers to the evaluation question** (applicable for all types of evaluation questions),
- **Conclusions and recommendations.** Conclusions are based on evaluation findings and recommendations are linked to conclusions (applicable for all types of evaluation questions).

Furthermore the SFC template also contains the non-mandatory parts, namely:

- **Qualitative findings, provided** in narrative text based on the collected qualitative information. This includes the findings of the qualitative analysis, if used as introductory assessment, or as triangulation for quantitative assessment or as self-standing qualitative assessment (applicable for all types of evaluation questions),
- **Problems encountered influencing the validity and viability of evaluation findings** includes the explanation of problems which evaluators have faced during the assessment (applicable for all types of evaluation questions).

An example of an empty and a filled SFC template can be found in Annex 1 of PART III of the Guidelines.

1.2 Evidence used in the evaluation in 2017?

The evaluation to be reported in 2017 will rely on a variety of information, including

- Monitoring data on beneficiaries.
- Additional data collected by evaluators from a sample of beneficiaries.
- Data on non-beneficiaries from national/regional statistics, FADN, annual accounts, etc.
- Qualitative information complementing insufficient quantitative data,
 - e.g. to answer evaluation questions with respect to Technical Assistance, NRNs, delivery mechanisms, etc., or
 - to answer Evaluation Questions in case of low or no uptake.

Using data from monitoring systems (application forms before the project start and payment request after the project is finalised) to **select sample populations of beneficiaries for the assessment of secondary contributions** of operations on other Focus Areas. The decision on samples¹⁴ (choosing random, systematic, stratified sampling, size of samples etc.) depends on the secondary contributions to assess (significant, marginal, etc.) and should be taken by the evaluators.

¹⁴ <https://www.statpac.com/surveys/sampling.htm>

1.3 Reporting in case of low uptake and small programmes.

Relevant common and programme-specific result indicators should be calculated for all those RDP measures and focus areas, which have shown sufficient uptake.

However, if a programme approval or start was delayed severely, there will be little or no monitoring data on completed operations to assess result indicators. In case of low uptake, it is necessary to take into consideration any information available on potential beneficiaries (e.g. assess the total available population of beneficiaries: applications, existing/ongoing contracts) and explain the situation why result indicators could not be calculated as required.

In case of no uptake, methods based on the theory of change or qualitative assessments can be used to get evidence on potential RDP achievements.

The legal acts also require reporting in the AIR on evaluations that have been done in the programme area during the previous years. This covers all studies relevant for a given RDP area. For example, there could be studies conducted by research institutes or universities in the area of climate change, biodiversity, local development, or business development, which provide useful information on RDP beneficiaries and territories. Findings of such studies have to be summarised in the AIR. They can be further used for evaluation purposes in case of low/no uptake.

Small programmes with a small population of beneficiaries (typically in RDPs of multi-regional Member States) may find it difficult to quantify result indicators and answer the relevant evaluation questions using quantitative methods, due to a lack of data. However, the CMES requirements are minimalistic and must be applied in the evaluation of small programmes as well. In this situation, the whole population of beneficiaries must be taken into consideration when calculating the indicators and answering EQ. To comply with these requirements qualitative research may be also conducted to answer the EQs. Furthermore, the RDP evaluation should in each case focus on the assessment of the level of achievement of RDP targets and milestones.

Further reading

Guidelines: Establishing and implementing the Evaluation plan of 2014-2020 RDPs, Evaluation Helpdesk 2014-2020, PART I, Chapter 3.4 and PART II, Chapter 6.4, Brussels, 2015, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDPs, PART II, Chapter 1, 4 and 5 and Part III, Annexes 5 and 6, Evaluation Helpdesk 2007 – 2013, Brussels, 2015, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

1.4 Reporting on the evaluation of RDP specific elements

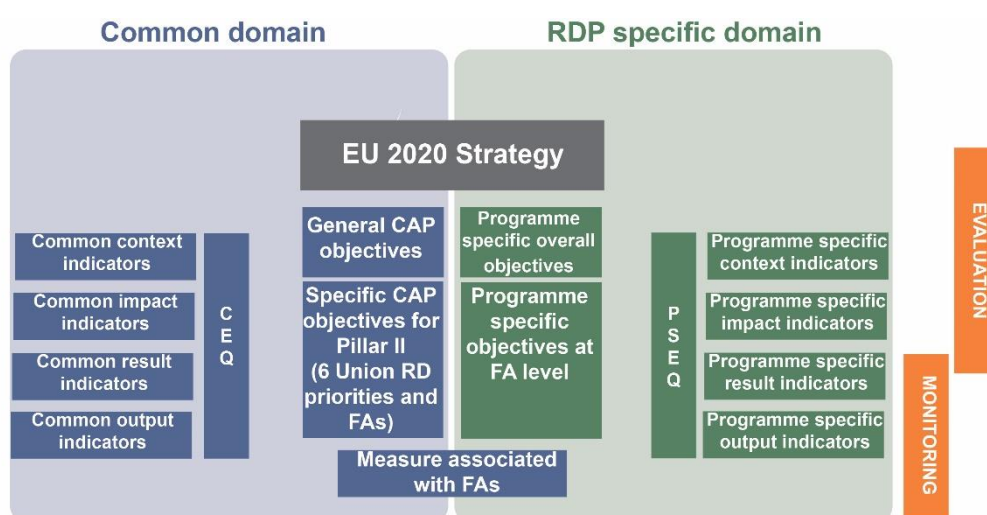
Flexibility in programming and its implications on evaluation

Flexibility in programming is one of the new characteristics of the 2014-2020 programming period. It aims to strengthen the strategic approach and to increase the effectiveness, efficiency and performance of rural development policy. In practical terms, Member States have flexibility to design the RDP specific intervention logic, composed of common and programme-specific objectives, and to adapt the combination of measures. Measures are no longer attributed to specific “axes” as in the past, but can be flexibly programmed and combined under the Union priorities/focus areas and programme-specific objectives. The underlying rationale is that Member States should be enabled to mix and combine measures under focus areas in a way that better reflects their specific rural development needs.

Flexibility in programming mirrors the flexibility of the monitoring and evaluation system. In this respect, the CMES provides, as a minimum common evaluation elements (e.g. common evaluation questions, common indicators, minimum requirements for evaluation plan, etc.), and allows space for MA to develop programme-specific evaluation elements tailored to their specific RDP (e.g. programme-specific evaluation questions and indicators, internal evaluation planning documents, etc.) in order to complete a full evaluation.

The figure below illustrates the linkages between objectives, and common and programme-specific evaluation elements connected with the programme intervention logic.

Figure 2. The common and programme specific monitoring and evaluation system



Source: European Evaluation Helpdesk for Rural Development, 2015

The MA may need to develop programme-specific evaluation elements to assess aspects that are of particular interest for them. These aspects could include the assessment of programme specific focus areas, the NRN, RDP delivery mechanisms, issues supported via technical assistance, such as administration and management, communication, etc. The specific evaluation topics also require the development of programme specific EQs, judgment criteria and indicators to collect evidence to answer the EQ.

Programme specific indicators these should not be mixed up with additional indicators which are developed in Member States in addition to common indicators used to answer the common evaluation questions linked to focus areas. Additional indicators can be also developed in case programme authorities define their own judgment criteria to specify the common evaluation questions.

It is better to define the programme-specific evaluation elements already in the programme design phase and include them in the evaluation plan. As an alternative, the programme-specific elements can be designed later, e.g. before or at an early stage of programme implementation to ensure suitable data will be available early on. Further programme-specific elements can also be developed when conducting the actual evaluations, in case gaps are still detected by the evaluators.

The principle of proportionality should be respected when developing and reporting on programme-specific monitoring and evaluation elements. This implies that the utility of additional information reported should be balanced against the resources required to provide this additional information.

PART II, Chapter 5, provides detailed guidance on how to develop programme-specific evaluation questions and indicators.

Reporting on programme-specific elements

Reporting on evaluation requires the use of both common and programme-specific evaluation questions and common, additional and programme specific indicators¹⁵. The standard AIRs are the main channel for MA to report on the achievements of the RDP by referring to financial data, quantified values of indicators, including target indicators. In the AIR 2017, the RDP achievements should be quantified through the assessment of common, additional and programme-specific results indicators¹⁶. The evidence collected via result indicators helps to answer focus area related common and programme-specific questions.

¹⁵ Article 50(2) of Regulation (EU) No 1303/2013

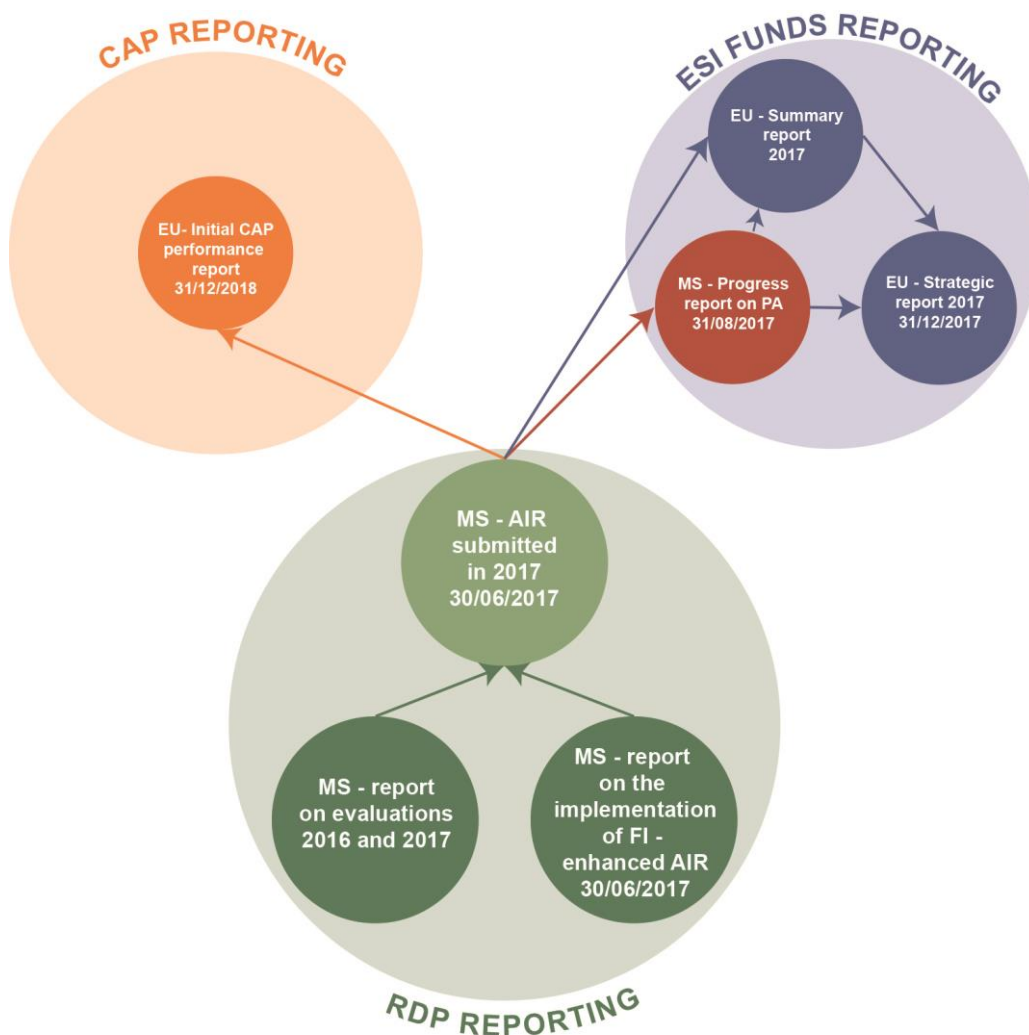
¹⁶ Annex VII, point 7 to Commission Implementing Regulation (EU) No 808/2014

2 LINKS TO OTHER EU REPORTING REQUIREMENTS AND REPORTING REQUIREMENTS BEYOND 2017

2.1 Links and synergies between the AIR 2017 and other EU reporting requirements

The AIR 2017 has various links with other reports at Member State and EU levels to be submitted in 2017 and in 2018. Together these reports inform on the implementation and progress of programmes financed by the ESI Funds and the CAP (See Figure 3).

Figure 3. Links between AIR submitted in 2017 and other EU reporting requirements in 2017 and 2018



Source: European Evaluation Helpdesk for Rural Development, 2015

The links between AIRs and other reporting requirements in the period 2016 – 2019 are illustrated in figure 4 and summarised in table 2.

Figure 4. Overview of reporting requirements and links 2016 – 2019

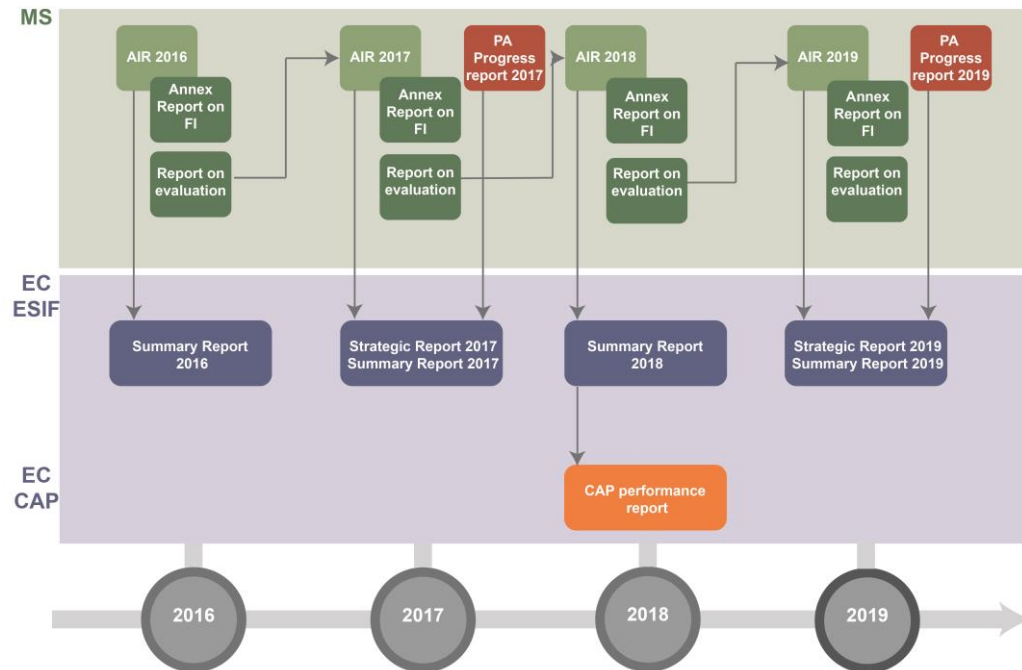


Table 1. Links between AIR 2017 and other EU reporting requirements

Reporting requirements	Legal source	Links and synergies with AIR 2017
RDP related reporting		
Reporting on evaluation (RDP level) The MAs shall ensure that evaluations are carried out according to the Evaluation Plan (in particular those to assess effectiveness, efficiency, and results. At least once during the programming period, an evaluation shall assess how support from the ESI Funds has contributed to the objectives for each priority. All evaluations shall be examined by the Monitoring Committee and sent to the Commission.	1303/2013, Art. 50.2 and 56.1 Commission Implementing Regulation No 808/2014, Annex VII, point 2 d) and e)	AIR must contain the synthesis of the findings of all RDP evaluations that may have become available during the previous financial year. AIR 2017 has to include evaluation elements, such as answers to EQs and assessment of progress in terms of result indicators related to focus areas.
Report on the implementation of financial instruments (RDP level) The MA shall send to the Commission each year a specific report covering the operations comprising financial instruments. This report includes information inter alia on implementation arrangements of the financial instrument, identification of bodies implementing financial instruments, financial information on the financial instrument, the performance of the financial instrument and contribution to the achievement of indicators of the priority or measure concerned.	1303/2013, Article 46.2 Commission Implementing Regulation No 808/2014, Annex VII, point 10	This report on implementation of financial instruments is annexed to the AIR. The AIR 2017 should additionally contain information on the progress of achieving the expected leverage effect of investments made by the financial instrument and the contribution to the achievement of indicators of the priority or measure concerned.
Partnership Agreement related reporting		
Progress report on PA (Member States level) By 31 August 2017, the Member States shall submit to the Commission the progress report on the implementation of the Partnership Agreement as at 31 December 2016.	1303/2013, Article 52.1	The progress report relates to the implementation of ESI Funds via operational programmes under the Partnership Agreement. The AIR 2017 provides the input in relation to the implementation of EAFRD RDP results and the RDP contribution towards policy objectives (where appropriate).
Strategic report (EU level) In 2017 the Commission shall prepare a strategic report summarising the progress reports of the Member States. The strategic report shall be submitted to the European Parliament, the Council and the European Economic and Social Committee and the Committee of the Regions by 31 December 2017.	1303/2013, Art. 53.2	The information from the AIR 2017, including reporting on evaluation, is fed into the progress report on the PA submitted in 2017, which is synthesized in the Strategic report at the EU level.
ESI Fund related reporting		

Reporting requirements	Legal source	Links and synergies with AIR 2017
Summary report (EU level) The Commission shall submit a summary report in relation to ESI Fund programmes each year from 2016 to the European Parliament, the Council and the European Economic and Social Committee and the Committee of the Regions.	1303/2013, Art. 53.1	The AIR 2017 also provides the input to the summary reports on the implementation of ESI Funds, submitted each year. In 2017 the summary report will form part of the strategic report.
CAP related reporting		
CAP performance report (EU level) The Commission presents the initial report on the implementation of Article 110 of Regulation 1306/2013. This report concerns the monitoring and evaluation of the CAP and the first results of its performance. This report shall be presented to the European Parliament and the Council by 31 December 2018.	1306/2013, Article 110.5	The AIR 2017 will include the synthesis of the findings of all evaluations of the RDP as well as changes in the values of result indicators. This body of information will therefore form one of the information sources on the performance of rural development measures for the preparation of the CAP report.

2.2 Links between reporting in 2017 and the overall reporting requirements on rural development in 2014-2020

Reporting requirements on the evaluation in 2017 relate closely with the reporting on evaluation across the entire programming period. Since evaluation tasks become more sophisticated towards the end of the programming period, it is necessary to properly set up and develop the monitoring and evaluation system. In this way, it will be possible to provide information for the evaluation across the programming period and the ex post evaluation.

The summary of reporting requirements across the programming period 2014-2020 can be found in the Annex 2 to PART III of the guidelines. Annex 6 of PART III explains what is required (mandatory) and what is recommended practice when reporting on evaluation in the AIRs over the programming period. It also proposes how various sections of AIR (standard and enhanced) can be linked together to provide more comprehensive information on the evaluation.

Further reading

Guidelines: Establishing and implementing the Evaluation plan of 2014-2020 RDPs, Evaluation Helpdesk 2014-2020, PART I, Chapter 5.5 and PART III, Annex 14, Brussels, 2015, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

3 ENSURING GOOD QUALITY EVALUATION IN 2017

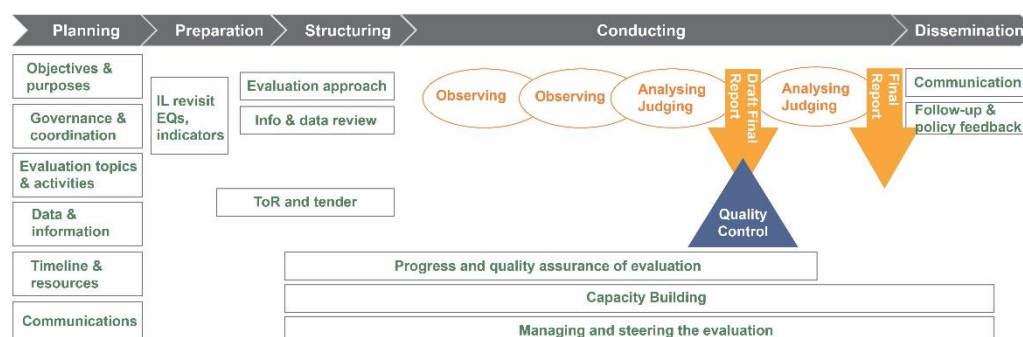
3.1 Careful planning and preparation of evaluation

There is limited time to conduct the evaluation tasks to be reported in 2017. Hence, prior planning and preparation is vital for the MA to be able to provide the necessary information in the enhanced AIR and to submit it to the Commission by the end of June 2017.

In the current programming period, the MAs are **required to plan evaluations** from the beginning, with the help of an evaluation plan, which is part of the rural development programme¹⁷. As the evaluation plan was developed early during the programming phase, it often covers only minimum requirements at a general level. However, the evaluation plan may still be further specified in the course of programme modifications and/or complemented with an internal, more detailed planning document. Such a document may contain more comprehensive information on the planned evaluation activities, topics and their timing.

In order to facilitate evaluation in 2017, it is vital to identify the evaluation needs and activities related to the RDP early on. Therefore, relevant stakeholders need to dedicate resources and prepare the application of evaluation methods at early stages of implementation.

Figure 5. Overview of the evaluation process



Source: European Evaluation Helpdesk for Rural Development, 2015

3.2 Quality assurance and quality control across the evaluation process

It is recommended to build quality assurance and quality control components into the evaluation process. **Quality assurance** focuses on the process and is a proactive way of preventing low quality outcome. Quality assurance includes a notion of observation, development of quality standards, and continuous improvement. **Quality control**, on the other hand, is product-oriented and ensures that the outcome is what was expected. It is typically performed at the end of the process.

¹⁷ Article 8(1)(g) of Regulation (EU) No 1305/2013

The overall responsibility for quality control of evaluations lies with the **MA**. MAs have different means to safeguard quality, e.g. drafting of precise terms of reference, selecting qualified evaluators, setting up evaluation steering groups, keeping regular contact with data providers, communicating regularly with evaluators, and requiring a high standard of evaluation reporting, etc. The capacity of the **evaluator** to use advanced evaluation methods and to bridge existing data gaps also influences the quality of evaluation. **Other parties** who may contribute to the improvement of the quality of evaluation are:

- Steering groups (if established) can ensure both the professional capacity of evaluators and the quality of evaluation reports;
- Paying Agencies can assist in providing higher quality data for evaluations;
- Programme beneficiaries, their associations and the general public (taxpayers), who show interest in evidence-based recommendations and high quality evaluation.;

The most typical **quality assurance tools** used in the evaluation process are checklists, content lists, general and detailed process maps. These tools can be developed separately, but it may be useful for the MA and other stakeholders involved in the evaluation (e.g. evaluation steering group members) to develop a complete **evaluation quality assurance handbook**, with detailed process maps (sequencing tasks and responsibilities of main stakeholders to be followed at each phase) and checklists for each phase. A complete handbook for assuring the quality of evaluation can also serve as a capacity-building manual. Maintaining institutional memory with regard to evaluations is vital.

Quality assurance in planning the evaluation

Once the overall process and timetable for the evaluation are outlined, evaluation needs must be examined. Typically, a **concept note** is written to improve the focus of the evaluation planning and to prepare for the terms of reference. A typical content list for a concept note includes: the topic, timing, scope, and key areas of focus, as well as stakeholder roles.

Communication and capacity building plans for the evaluation should be drafted at the planning phase to maximise the quality of the evaluation and the use of evaluation results. A content list or a checklist can be used to ensure the inclusion of major stakeholders, usage of correct communication channels and tools, identification of communication responsibilities, and optimal timing of communication.

Quality assurance in preparing evaluation

The most vital issues are to attain a high quality of the intervention logic, evaluation questions, indicators, evaluation approach, information and data review, evaluation focus, data gaps, and the extent to which the topic can be evaluated. These issues influence the drafting of the **terms of reference** and the contract drawn up with the evaluator. A systematic, critical scrutiny of the proposed evaluation questions may be done through a checklist with guiding questions such as: “Do the EQs cover all the

objectives and evaluation needs? Are questions clearly formulated? Can these EQs be answered using the data that is available or which can be gathered?”

When drafting the terms of reference, it is important to include a **quality assessment grid** against which the quality of the final report will be judged. The checklist can also include the quality criteria related to the content (accuracy, adequacy, relevance, and clarity), as well as the assessment of the proposed evaluation process and timeline. (see Part III, Annex 5: Check-list for assessing the quality of the evaluation report).

Quality assurance in structuring and conducting the evaluation

The quality assurance process starts with the implementation of evaluation. The terms of reference shall be used as a reference point.

It is recommended to develop **quality standards** for the inception report (including criteria on content such as methodological approach and data collection methods, as well as criteria for the operational plan and the process of evaluation). The inception report may also include a section on quality assurance drafted by the evaluator. Furthermore, it is equally important for the client to ensure that internal and external processes are in place for the timely delivery of materials to the evaluator.

The client and/or steering group can develop and use checklists to scrutinise the content and quality of the progress and draft final reports. These tools can help to improve the quality and focus of the final report, as well as to give systematic and constructive feedback to the evaluator.

The final report should be subject to a quality assessment, preferably by using the quality assessment grid included in the terms of reference¹⁸.

Quality assurance in dissemination of evaluation findings

The MA should draw up a plan, timetable and a **checklist for disseminating and following up recommendations** of the evaluation. The fulfilment of this plan should be assessed at regular intervals. It is good practise for the MA and the evaluators to give mutual feedback at the end of the evaluation. The MA may also wish to assess the quality of the evaluation communication at the end of the process.

Further reading

Guidelines: Establishing and implementing the Evaluation plan of 2014-2020 RDPs, Chapter 5.3, Evaluation Helpdesk 2014-2020, Brussels, 2015, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDPs, PARTI Chapter 1.2 and Part III, Annex 4, Evaluation Helpdesk 2007 – 2013, Brussels, 2015, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

¹⁸ Annex 6 to 'Quality Assessment Form', DG Market Guide to Evaluating Legislation, pages 87-97 http://ec.europa.eu/dgs/internal_market/evaluation/evaluation/index_en.htm

4 DISSEMINATION, COMMUNICATION AND FOLLOW-UP OF EVALUATION RESULTS

4.1 Dissemination and communication of evaluation in 2017

The value of an evaluation depends on its dissemination, communication and finally the use of its findings to improve the policy.

In terms of dissemination, it is fundamental that the reporting on evaluation in the AIR 2017 is made public, e.g. on the website of the MA. Furthermore, to increase transparency, evaluation reports should be disseminated to the participants of the evaluation process and all the relevant stakeholders. As a recommended practice, a **citizens' summary** of the main findings of the evaluations should be drafted. It might also be useful to translate the citizens' summary into English.

Communication occurs throughout the evaluation process, but the main communication effort comes at the end, after the results and recommendations have been finalised. The communication actions should follow the **evaluation communication plan** developed at the beginning of the evaluation process. The main focus of the communication should be on results and achievements of the RDP. In case of low uptake of measures, results of other studies linked to the RDP (e.g. on water efficiency) could be communicated.

The evaluation communication plan shall be monitored and assessed to check its efficiency and effectiveness in delivering the key messages to the target audiences.

4.2 Follow-up of evaluation findings

Evaluation as part of the governance of the EU programmes constitutes a strategic management tool. When used effectively, the follow-up on evaluation findings results in:

- improving programme design and implementation,
- better and more programme results and impacts
- strengthening the use of evaluations;
- stimulating an evaluation culture based on the organizational learning and enhancing the liability for results;
- facilitating the discussion about programme evaluation;
- motivating stakeholders and programme managers to actively support and participate in the RDP performance improvement; and
- enhancement of public policies.

If the evaluation fulfils this role, all involved stakeholders, and particularly Managing Authorities and policy makers, need to pay considerable attention to evaluation findings and recommendations.

Further reading

Guidelines: Establishing and implementing the Evaluation plan of 2014-2020 RDPs, Chapter 3.6, Evaluation Helpdesk 2014-2020, Brussels, 2015, enrd.ec.europa.eu/en/evaluation-helpdesks-publications

Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDPs, Chapter 1.2.3, Evaluation Helpdesk 2007 – 2013, Brussels, 2015, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

PART II SETTING UP THE SYSTEM TO ANSWER EVALUATION QUESTIONS AND ASSESSING RDP ACHIEVEMENTS

Evaluation questions (EQ) are an important component in the evaluation of rural development policy. They define the focus of evaluation in relation to EU and programme-specific policy objectives and help to demonstrate the progress, impact, achievements, effectiveness, efficiency and relevance of rural development programme interventions¹⁹.

The common evaluation questions (CEQ)²⁰ were designed by the Commission in collaboration with the Member States. They are part of the common monitoring and evaluation system for rural development and represent a minimum common set of EQ established to enhance the comparability of evaluation results across the EU. CEQ provide support to Member States in the evaluation of achievements of EU level policy objectives and expected results and impacts.

Programme-specific evaluation questions (PSEQ) may be developed by Member States to capture programme-specific achievements towards RDP policy objectives, results and impacts. There are three types of CEQ: 1) focus area-related CEQ, 2) CEQ related to other aspects of the RDP, and 3) CEQ related to Union level objectives.

Legal requirements related to evaluation questions

It is a **legal requirement** to answer CEQ and PSEQ²¹ at different points in time during the programming period and also after programme implementation. Member States shall provide evidence-based answers:

- In the annual implementation report (AIR) **2017**, EQ related to the RDP focus areas and evaluation questions related to other RDP aspects (synergies, TA and NRN) shall be answered by means of common, additional²² and programme-specific result indicators²³,
- In the AIR **2019 and in the ex post evaluation report**, all CEQ and PSEQ shall be answered by means of common, additional and programme-specific result and impact indicators.

Steps in setting up the system to answer evaluation questions

Once the evaluation is carefully planned, the major steps in setting up the system to answer the evaluation questions are taken in the preparing, structuring and conducting phase of the evaluation.

¹⁹ Article 54(1) of Regulation (EU) No 1303/2013; Article 68(a) of Regulation (EU) No 1305/2013

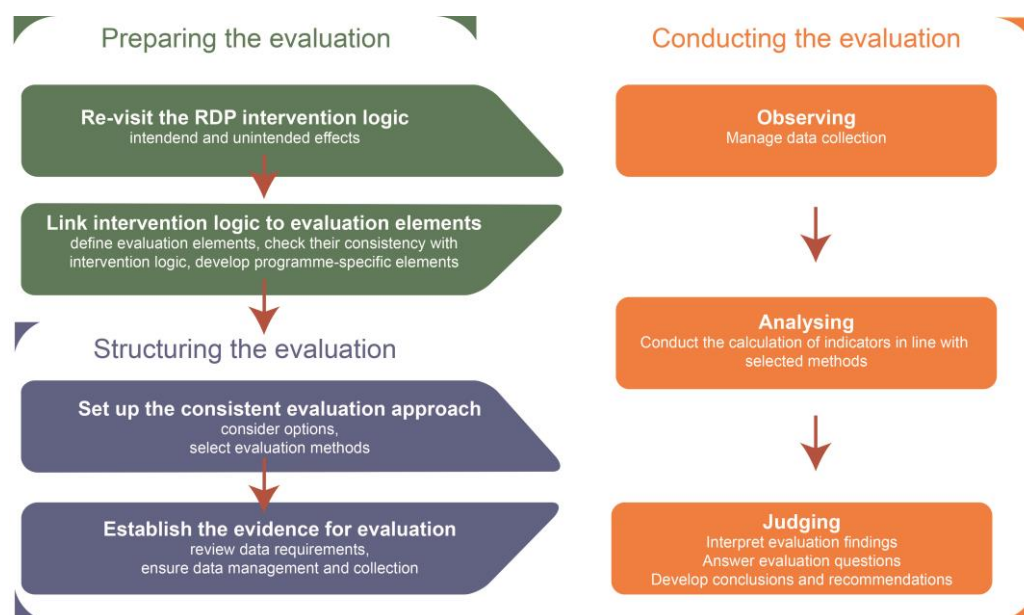
²⁰ WD: Common Evaluation Questions for Rural Development programmes 2014-2020, <https://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

²¹ Article 14(1)(c) and Annex VII, point (7) to Commission Implementing Regulation (EU) No 808/2014,

²² Additional indicators are applied **additionally** to common indicators if the latter are not sufficient to answer the CEQ in a satisfactory manner

²³ Programme-specific indicators are applied to answer PSEQ

Figure 6. Major steps in setting up the system to answer evaluation questions



Source: European Evaluation Helpdesk for Rural Development. 2015

Preparing the evaluation (see chapter 5): In the preparation phase the MA sets the cornerstones for a successful evaluation. It is necessary to revisit the RDP intervention logic taking also into consideration the findings of the ex ante evaluation. The intervention logic must be linked to the evaluation elements (evaluation questions, judgement criteria, indicators) and their consistency be checked. If necessary, programme-specific evaluation elements may be developed.

Structuring the evaluation (see chapter 6): In the structuring phase, the MA together with evaluation stakeholders (evaluators, data providers etc.) set up the evaluation approach and establish the basics for collecting the evidence necessary to answer the EQ. For this purpose, the evaluation methods are selected and combined; information needs are identified; data sources are screened and provisions are made in order to get data and information in the required format for RDP evaluation.

Conducting the evaluation (see chapter 7): The conduction phase is led by the evaluators independent from the MA. It starts with observing the evidence (collecting available data and filling the data and information gaps) and analysing it with the foreseen methods. For this purpose, the indicators are calculated and assessed. Finally, the evaluator provides judgements on evaluation findings, answers the EQ and drafts conclusions and recommendations for the improvement of RDP design and implementation.

A summary of all working steps necessary for setting up the system to answer the evaluation questions is presented in Annex 4, Part III. The table includes also the responsibilities of the stakeholders and the relevance for the Terms of Reference.

5 PREPARING THE EVALUATION

Member States have started to plan evaluations with the help of an **evaluation plan**, which is part of the rural development programme²⁴. Some Member States have moreover complemented the evaluation plan with an internal, more detailed planning document, which contains more comprehensive information on the foreseen evaluation activities, topics and their timing.

Once the RDP has been approved, the MA in collaboration with other relevant evaluation stakeholders should start to prepare the evaluation to be reported in 2017 and later on. The main emphasis should be put on setting up the system to answer the EQ.

In the preparation phase for reporting on evaluation in 2017 (and for the entire programming period), it is important for programme authorities to ensure that:

- The RDP intervention logic, the EU common and programme-specific objectives and expected effects are clearly understood and validated by relevant stakeholders (see PART II, chapter 5.1)
- EU common and programme-specific evaluation elements (EQ and indicators) and their relation to RDP intervention logic (overall and focus area-related objectives, measures, sub-measures) are identified and known by relevant stakeholders and that they are able to judge on the sufficiency of evaluation elements to capture all RDP effects (see PART II, chapter 5.2),
- All the terms used in the RDP objectives, focus areas, measures, EQ (and their judgment criteria) and indicators are defined (see PART II, chapter 5.2),
- The relevant evaluation stakeholders are familiar with evaluation approaches²⁵ and their suitability to serve the purpose of evaluation, respecting the data availability, quality and frequency (see PART II, chapter 6.1),
- The monitoring system is adjusted to the evaluation needs, e.g. data on result indicators and additional information are collected from beneficiaries (see PART II, chapter 6.2),
- Existing data sources, providers, arrangements to adjust data to RDP needs and data gaps are recognised and procedures how to fill the gaps defined (see PART II, chapter 6.2),

In order to accomplish these tasks, the management of evaluation should be established, staffed, equipped with necessary knowledge and skills. If possible, all relevant stakeholders²⁶ should be present in evaluation working groups or steering groups (see PART I, chapter 3).

²⁴ Article 8(1)(g) of Regulation (EU) No 1305/2013

²⁵ Theory-based, quantitative, qualitative, mixed approaches

²⁶ Managing Authority, paying agency, evaluation experts, evaluators, data providers, monitoring committee members

Whenever this preparatory phase is completed, the actual evaluation exercise can start or, in case the MA wishes to contract an external evaluator, the terms of reference can be drafted.

Further reading

Guidelines: Establishing and implementing the Evaluation plan of 2014-2020 RDPs, Evaluation Helpdesk 2014-2020, PART I, Chapter 3.4 and PART II, Chapter 6.4, Brussels, 2015, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>
 Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDPs, PART II, Chapter 1, 4 and 5 and Part III, Annexes 5 and 6, Evaluation Helpdesk 2007 – 2013, Brussels, 2015, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

5.1 Revisit the RDP intervention logic

To revisit the intervention logic may become necessary in case ex ante findings have not been sufficiently robust, or not fully taken into consideration when building the RDP, or following an amendment of the RDP. Moreover, changes in the RDP context may have affected the relevance of the intervention logic with regard to the needs that had been identified at the time of programme design. Other needs may also have become more relevant at the time of RDP evaluation.

Recommended working steps:

Only if the above-mentioned changes have been identified, it is recommended to:

- **revisit the ex ante findings** on the RDP's intervention logic's coherence and relevance, and
- **appraise if intended direct RDP effects** on beneficiaries are still valid and expected synergies between priorities at the level of the entire programme can be achieved²⁷ (→ **see tool to appraise various programme effects in Annex 10, PART III**)

Expected outcome: Revisited intervention logic

Revisit ex ante findings and indirect programme effects

Already during programme design the ex ante evaluator had examined:

- the intervention logic's²⁸ external coherence with EU policy objectives (the RDP contribution to EU 2020 strategy²⁹ and CAP³⁰).

²⁷ This appraisal is also very important for the assessment of programme synergies and answering the CEQ 19: "To what extent have the synergies among priorities and focus areas enhanced the effectiveness of the RDP?"

²⁸ Also read Part II, Chapters 1 and 2 of the „Getting the most of your RDP: guidelines for the ex ante evaluation of 2014-2020 RDPs" http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

²⁹ Union Strategy for smart, sustainable and inclusive growth http://ec.europa.eu/europe2020/index_en.htm
³⁰ <http://ec.europa.eu/agriculture/cap-post-2013/>

- the relevance in addressing the most important needs of the programme territory, deriving from the SWOT analysis and needs assessment, and
- the internal coherence between programme objectives, planned inputs (budget and forms of support), combination of measures, and expected RDP's outputs, results and impacts.

The intervention logic should not only be appraised for direct and intended programme effects, but also for several types of other effects which may play an important role in the programme's performance. These effects may influence the intended RDP achievements and expected results and impacts in a positive or negative way. Although it will later be the task of the evaluator to net out programme effects, programme authorities need to develop an understanding of what their programme may "produce" apart from what has been planned. The RDP intervention logic, namely its priorities, focus areas and measures can produce indirect effects³¹, which may be intended (identified in the context analysis) or unintended (unforeseen in the context analysis and not flagged with the context indicator).

Indirect programme effects to be looked at during revisiting the RDP intervention logic:

- **Leverage effects** are the propensity for public interventions to induce private spending among direct beneficiaries (e.g. farmer receiving subsidy can better invest in property, which is not part of business). Leverage effects are usually unintended.
- **Deadweight loss effects** are changes observed in the economic, environmental or social situation of programme beneficiaries which would even have occurred without the intervention (e.g. farmers would invest anyway without subsidy later or with their own money, or they would use loans). Deadweight losses are usually unintended effects.
- **General equilibrium effects** occur when programme interventions positively or negatively affect RDP non-participants. They usually play a more important role in the evaluation of large programmes than in the evaluation of small programmes and include:
 - *Multiplier effects* resulting from increased income and consumption generated by the RDP. Multiplier effects are cumulative and take into account the fact that a part of the income generated is spent again and generates other income, and so on in several successive cycles. In environmental terms, the intended support of selected species may lead to support of other species as well. Multiplier effects are positive, often intended and expected.
 - *Displacement effects* occur in a programme area at the expense of other areas. For example, the programme support is affecting positively the employment at the expense of increasing unemployment in neighbouring areas. In environmental terms, the intended support of HNV farmland may lead to deterioration of farmland in neighbouring areas. Displacement effects might be unintended (if they cause further regional disparities) or intended (if they contribute to balancing disparities among regions).

³¹ Detailed description of programme effects can be found in Guidelines for ex post evaluation of 2007-2013 RDPs, Chapter 4.2.2 Challenge: identification of programme effects, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

- *Substitution effects* are obtained in favour of direct programme beneficiaries but at the expense of units that do not qualify or participate in a given programme. For example, support to small farmers for irrigation technology may increase their competitiveness in relation to non-supported large farms. Substitution effects may be unintended (if support goes to those already reached), or intended (if they contribute to balancing the socio-economic situation in the programme area).

Moreover, there are effects between measures, focus areas and priorities, which have been taken into consideration already during the programme design and ex ante evaluation. If necessary, these effects should be appraised when preparing the evaluation. These effects are:

- **Secondary contributions** of operations to focus areas other than those under which they have been programmed. The legal framework requires flagging the intended secondary contributions during the programme design/ex ante evaluation and identifying potential secondary contributions for each individual operation when it is included in the operations database. The validity of this flagging might be revisited during the preparation of the evaluation, and corrected if necessary.
- **Transverse effects**, which are horizontal effects between measures, between focus areas or between priorities. Positive transverse effects are often called **synergies**. Transverse effects occur if measures, focus areas or priorities are weakening or fostering each other in their effects. Transverse effects might be intended or unintended. Within the RDP there is often scope for a policy instrument in one area to impact on another. For example, the policy instrument vis-à-vis areas facing natural constraints can enhance the performance of local firms and (possibly) lead to rural economic growth and job creation. As for synergies, an example may be that investing in off-farm diversification might foster the competitiveness of the farming sector

Figure 7. Different types of programme effects

	RD priority/ Impacts	RD focus area/ Results of measures and measure combinations	RD measure
RD priority/ Impacts	Transverse effects (positive = synergies and negative) Intended/unintended		
RD focus area/ Results of measures and measure combinations	Primary/secondary Direct/indirect Intended/unintended	Transverse effects (positive = synergies and negative) Intended/unintended	
RD measure	Primary/secondary Direct/indirect Intended/unintended	Primary/secondary Direct/indirect Intended/unintended	Transverse effects (positive = synergies and negative) Intended/unintended

A tool for the qualitative appraisal of RDP operations' primary and secondary contributions as well as to trace transverse effects/synergies can be found in Part III, Annex 10.

Specificities of LEADER

LEADER is programmed under focus area 6B (focus area 6A in the United Kingdom). This is also the focus area for which primary contributions are expected, and where job creation and proportion of the rural population affected are included as target indicators.

However, LEADER is implemented via local development strategies (CLLD strategy), which are in fact similar to small programmes that support a broad range of operations. The scope varies among Member States/regions³². This implies that the projects financed within the CLLD strategy measures will contribute to a range of focus areas, beyond FA 6B (secondary contributions of LEADER).

Two types of secondary contributions of LEADER projects to FA (Data items relevant for LEADER to be collected for each project) can be distinguished³³:

- Predominant contributions to FA

³² In some cases, the CLLD strategy may deliver only RDP measures, or measures listed in the Regulation (EU) No 1305/2013, in other cases they encompass their own measures.

³³ See Working Document Data items list for Pillar II operations database.

- Additional contributions to FA(s)

This implies to look at LEADER secondary contributions to focus areas other than FA 6B, under which LEADER is typically programmed, when appraising the RDP intervention logic. Where contributions via the CLLD strategy are expected, their validity should be confirmed. For this purpose, a LAG level operations database shall be established, similarly to the RDP operations database, to capture all operations with primary and secondary CLLD strategy' contributions to the values of result indicators used in the assessment of achievements under the focus areas. The estimation of LEADER contributions may be calculated by using **samples** of beneficiaries which have flagged the related FAs as predominant or additional contributions in the LAG operations database.

Another specificity of LEADER is linked to the **LEADER method** translated into 7 principles (partnership, bottom-up, multi-sector and area-based strategies, innovation, networking and cooperation) and the **LEADER added value**, e.g. increased social capital or improved local governance. In many cases the LEADER principles and the LEADER added value are neither articulated as LEADER-related programme-specific objectives, nor are they mentioned as evaluation topic in the Evaluation Plan. In each case these effects should be considered when revisiting LEADER in the RDP intervention logic. This helps to look at these effects (positive and negative) and to articulate them. It may also lead to the identification of additional evaluation topics with respect to LEADER.

Specificities of technical assistance (TA)

Technical assistance, financed at up to 4 % of the total amount of the rural development programme, is a horizontal measure which at the initiative of the Member State³⁴ supports actions for:

- preparation, management, monitoring, evaluation, information and communication, networking, complaint resolution, and control and audit,
- reduction of the administrative burden on beneficiaries, including electronic data exchange systems, and
- reinforcement of the capacity of
 - Member State authorities and beneficiaries to administer and use the EAFRD
 - Relevant partners in line with the indicative areas, themes and good practices concerning how the competent authorities of the Member States may use the EAFRD to strengthen the institutional capacity of relevant partners in accordance with the legal framework and to support exchange of good practices between such partners³⁵.

³⁴ Article 51 of Regulation (EU) No 1305/2013

³⁵ Article 5 and 59 of Regulation (EU) No 1303/2013

Although technical assistance is not necessarily a “visible” part of the RDP intervention logic it still supports the implementation of the RDP and contributes to the achievement of RDP objectives.

In case more specific topics in the evaluation of actions for TA are envisaged at a given time (2017, 2019, ex post), it is considered a recommended practice to set up an intervention logic for these topics, formulate objectives, expected outputs, results and impacts. While revisiting the intervention logic, it is recommended to re-check if the objective linked to the TA evaluation topic and expected effects of planned activities and budgets are still valid.

Specificities of national rural network (NRN)

The NRN is one important TA action and its evaluation is either carried out as part of RDP evaluation or as self-standing evaluation topic.

The NRN groups the organisations and administrations involved in rural development with the aim: a) to increase the involvement of stakeholders in the implementation of rural development; b) to improve the quality of implementation of rural development programmes; c) to inform the broader public and potential beneficiaries on rural development policy and funding opportunities; and d) to foster innovation in agriculture, food production, forestry and rural areas³⁶. The NRN, as other TA actions, constitutes a horizontal support contributing to all RDP objectives. As the NRN has got an action plan and a structure (NSU and network) it is considered a recommended practice to specifically articulate the NRN intervention logic. Moreover, specific NRN Programmes, applied in some multi-regional Member States, must be designed around a proper NRN intervention logic³⁷.

The practice shows that some Member States have formulated an own intervention logic for NRN already during the programme design. In these cases, the intervention logic should be revisited together with the RDP intervention logic. The same applies for the NRNPs.

In the absence of an own NRN intervention logic, it is recommended that the MA or the NSU, in co-operation with the MA, formulate the NRN intervention logic during the preparation phase of the evaluation. This requires the formulation of NRN specific objectives, expected results and impacts (NRN added value going beyond the envisioned NRN common and programme-specific objectives).

Further reading

Getting the most from your RDP: Guidelines for the ex ante evaluation of 2014-2020 RDPs, PART II, Chapter 1 and 2, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>.

³⁶ Article 54 of Regulation (EU) No 1305/2013

³⁷ Getting the most from your RDP: ex ante evaluation guidelines of 2014-2020 RDPs, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDP, PART II, Chapter 1, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html.

Metis/WIFO/AEIDL (2014). Investment Support under the Rural Development Policy. Final Report. European Commission, Brussels.

Psaltopoulos, D., Phimister, E., Ratinger, T., Roberts, D., Skuras, D., Balamou, E., Bednarikova, Z., Espinosa, M., Gomez y Paloma, S., Mary, S., Nohel, F. and Santini, F. (2012). Ex ante Spatial Policy Impact Analysis of the Rural Development Policy in European Rural Areas.

(RURAL ECMOD), JRC Scientific and Technical Reports, Office for Official Publications of the European Communities, Luxembourg

5.2 Link intervention logic to evaluation elements

5.2.1 Define evaluation elements: evaluation questions, judgement criteria, indicators

In order to provide robust answers to EQ it is important to ensure that all common, additional and programme-specific evaluation elements (EQ, judgement criteria, and indicators) are complete, clear and well defined. Furthermore, all terms used in these evaluation elements should be clear and understandable for the involved stakeholders.

Recommended working steps:

- **Examine if all terms used in the formulation of CEQ, their judgment criteria and common indicators are clear and understandable:** Appraise the clarity of all terms proposed in the EC Working Document: “*Common evaluation questions for rural development programmes 2014-2020*” and check if additional judgment criteria and definitions are still needed. Develop additional indicators if common indicators are not sufficient to answer the common evaluation questions. Check if all terms used in common indicators are clear and sufficiently defined.
- **Examine if all terms used in the formulation in PSEQ, their judgment criteria, and programme-specific indicators are clear and understandable:** This step is conducted if the RDP contains PSEQ and programme-specific indicators. For PSEQ it is important to check if they are equipped with judgement criteria and well in line with the expected RDP success. The clarity of all terms used in PSEQ and judgment criteria shall be examined and their definitions provided, if needed. Judgment criteria for PSEQ should be developed in case they do not exist.
- **Develop fiches for all additional and programme-specific indicators:** In case fiches already exist, check the clarity of all terms, e.g. with respect to measurement unit/formula, suitability of calculation methods, and accessibility of data in the required format. If this is not the case, develop fiches for all programme-specific indicators.

Expected output: Revised intervention logic, indicator fiches, programme-specific indicators and PSEQ, and additional indicators (if needed)

Developing judgement criteria for evaluation questions

Judgment criteria further define EQ and help to specify the expected success of the rural development policy interventions at the level of focus areas, EU objectives and other specific RDP aspects. Judgment criteria should be formulated in a clear manner and all terms used should be well defined.

Figure 8. Relation between evaluation elements



Source: European Evaluation Network for rural development 2007- 2013

For CEQ, the judgment criteria are proposed in the *Working paper: Common Evaluation Questions for rural development programmes 2014-2020*³⁸. Member States may decide to specify further the success of RDP interventions with additional judgment criteria for CEQs. Terms used in the CEQ are usually defined in the DG AGRI Glossary³⁹. If not, other official glossaries – at EU, national or regional level - can be used for this purpose.

PSEQ are formulated either by Managing authorities (in the Evaluation Plan during the programme design or later) or by evaluators (during the RDP evaluation). Similar to CEQ, all terms used in PSEQ should be clear and well defined. As for CEQ, it is recommended to define judgment criteria also for PSEQ.

Table 2. Example of judgment criteria for focus area-related CEQ and PSEQ

CEQ	The success of intervention will be judged with (judgment criteria):
To what extent have RDP interventions supported innovation, cooperation and the development of the knowledge base in rural areas? ⁴⁰	<p>Commonly proposed judgment criteria⁴¹:</p> <ul style="list-style-type: none"> • RDP projects have been innovative and based on developed knowledge • Operational groups have been created • Variety of partners involved in EIP operational groups • Innovative actions have been implemented and disseminated by the EIP operational groups <p>Example of additional judgment criteria (added by Member States, if needed):</p> <ul style="list-style-type: none"> • The share of partners in operational groups from research and academia represents at least 50%.

³⁸ <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

³⁹ http://ec.europa.eu/agriculture/glossary/index_en.htm

⁴⁰ Annex 1, CEQ No 1 to Working paper: Common evaluation questions for Rural Development Programmes 2014-2020

⁴¹ Annex 1 to Working paper: Common evaluation questions for Rural Development Programmes 2014-2020

CEQ	The success of intervention will be judged with (judgment criteria):
PSEQ	The success of the intervention will be judged with (judgment criteria):
To what extent have RDP interventions prevented youth exodus from rural areas?	<ul style="list-style-type: none"> • RDP projects encouraged the establishment of businesses by young people • RDP projects facilitated the employment of young people • RDP projects supported the participation of youth in rural development activities

Developing additional and programme-specific indicators

EQ are answered by means of indicators. The legal acts and the CMES distinguish between:

- **Common indicators**⁴² for context (describe the programme environment, facilitate the SWOT and needs assessment), output (measure the RDP outputs at measure level), results (measure RDP results at focus area level) and impact (measure RDP impacts at programme level) are elaborated in detail in various EC working documents in the form of indicator fiches⁴³. Common indicators are used to answer CEQ.
- **Additional indicators** are developed “in addition” to common indicators whenever the common indicators cannot answer the CEQ in a satisfactory manner. They should not be mixed with programme-specific indicators which are only used to answer PSEQ. Additional indicators shall also meet the RACER criteria.

The Commission encourages the Member States where support under a given measure/sub-measure is significant, such as sub-measure 4.4 "Non-productive investments", to define additional result indicators in order to ensure better monitoring and assessment of those measures/sub-measures' contribution to achieving the Union objectives.

- **Programme-specific indicators** are developed by stakeholders in the Member States (Managing authorities, evaluators) as context, output, result and impact indicators in order to answer PSEQ. These indicators shall meet the RACER criteria.

Each indicator is ideally defined in the form of a fiche. Fiches for all common indicators are developed by the European Commission. Indicator fiches for programme-specific indicators and additional indicators should be defined by the Member States. It is recommended to follow the structure of the indicator fiches designed for the common indicators. The fiche shall provide at least:

- The name and definition of the indicator
- The link to the RDP objective/focus area, and EQ
- The measurement unit or formula of the indicator,
- The method for calculation

⁴² Article 14(1)(b) and Annex IV to Commission Implementing Regulation (EU) No 808/2014

⁴³ Working documents: Impact indicators, Common context indicators fiches, Target indicators for Pillar II, Complementary result indicators for Pillar II.

- The types, sources and accessibility of data to calculate the indicator.

The evaluation stakeholders (Managing authorities, evaluators) shall examine and understand the indicator fiches, including definitions used, clarity of definition, accessibility of the indicator with respect to data sources and collection, as well as the method proposed for the calculation of the indicator. In case of programme-specific and additional indicators, programme authorities or other evaluation stakeholders (evaluators) are responsible for the clarity of all terms used in the respective indicator fiches.

Specificities of LEADER

The CMES provides one common evaluation questions which applies to LEADER: “To what extent has LEADER interventions supported local development in rural areas? (adapted CEQ number 17 to LEADER). The CMES further contains the following common indicators for LEADER at RDP level (output and target)⁴⁴:

Indicator	Output	Target
Population covered by LAG	O18	T21
Number of LAGs selected	O19	
Number of LEADER projects supported	O21	
Number of cooperation projects supported	O21	
Number and types of project promoters	O22	
Unique identification number of LAG involved in cooperation project	O23	
Number of jobs created		T23

Considering the flexibility in applying LEADER in the Member States, the above mentioned common evaluation question and indicators might not capture all specificities of LEADER and its expected/intended effects (including the added value of LEADER). Therefore, Member States may have developed LEADER-related PSEQ and programme-specific indicators in the RDP’s Indicator Plan or the Evaluation Plan. In such a case, all terms used in PSEQ and programme-specific indicators should be well defined and a dedicated indicator fiche should be elaborated. If neither PSEQ nor specific indicators for LEADER have been developed but specific evaluation topics are planned, they may still be defined at later stages (see chapter 5.2.3 ‘Develop RDP specific evaluation elements’).

In addition, the operations implemented under each CLLD strategy will be encoded in the LAG level operations database, which will enable to collect data and information for the assessment of LEADER contributions to the achievements of RDP objectives, results and impacts, and effectiveness and efficiency of RDP interventions implemented via LEADER.

⁴⁴ Annex IV to Commission Implementing Regulation (EU) No 808/2014

Specificities of TA

In case the Member State included in the evaluation plan evaluation topics which relate to actions under technical assistance, these topics may be accompanied by clearly defined PSEQ and related programme-specific indicators. However, if this is not the case, the above elements can be formulated later. Chapter 5.2.3 'Develop RDP specific evaluation elements' provides more guidance in this respect.

Specificities of NRN

A similar situation applies to common evaluation elements for NRN evaluation. There is only one CEQ for NRN „To what extent has the national rural network contributed to achieving the objectives laid down in Regulation (EU) No 1305/2013, Article 54 (2)?“. This EQ is supported by three common output indicators⁴⁵:

Indicator	Output
Number of thematic and analytical exchanges set up with the support of NRN	O24
Number of NRN communication tools	O25
Number of ENRD activities in which the NRN has participated	O26

Due to this limited set of common evaluation elements, Member States are advised to define additional indicators which will allow to answer this NRN-related CEQ. If desired, Member States may also develop further NRN-related PSEQ and programme-specific indicators to answer them.

It is evident that PSEQ and programme-specific indicators (including result and impact indicators) should be developed for NRNPs in order to measure the expected added value. As in the case of LEADER, all programme-specific indicators should be developed in form of indicator fiches (see chapter 5.2.3 'Develop RDP specific evaluation elements').

Further reading

Getting the most from your RDP: Guidelines for the ex ante evaluation of 2014-2020 RDPs, PART II, Chapter 3, link (to be added after revised version will be put on the web). Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDP, PART II, Chapter 1, 2 and 3 link: http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html.

Working paper: Common Evaluation Questions for rural development programmes 2014-2020, link: <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

Working document: Common context indicators fiches

Working document: Impact indicators fiches

Working document: Target indicators fiches for Pillar II

Working document: Complementary result indicators fiches for Pillar II

⁴⁵ Annex IV to Commission Implementing Regulation (EU) No 808/2014, and WD, Data items list for Pillar II operations database'

Working document: Data items list for Pillar II operations database (Outputs and targets)
 Working Document: Rural Development Monitoring (2014-2020)-Implementation Report
 Tables
 Working document: Rural development programming and target setting (2014-2020)

5.2.2 Check consistency of evaluation questions and indicators with RDP intervention logic

In order to answer CEQ and PSEQ in a correct way and in line with their judgment criteria, their consistency with the intervention logic and indicators must be ensured. The assessment of the horizontal and vertical consistency between objectives, EQ and indicators helps to judge on the adequateness of the defined indicators and judgement criteria to measure the achievements against objectives and to answer the EQ. This step furthermore helps to identify potential gaps and to ways to overcome them.

Recommended working steps:

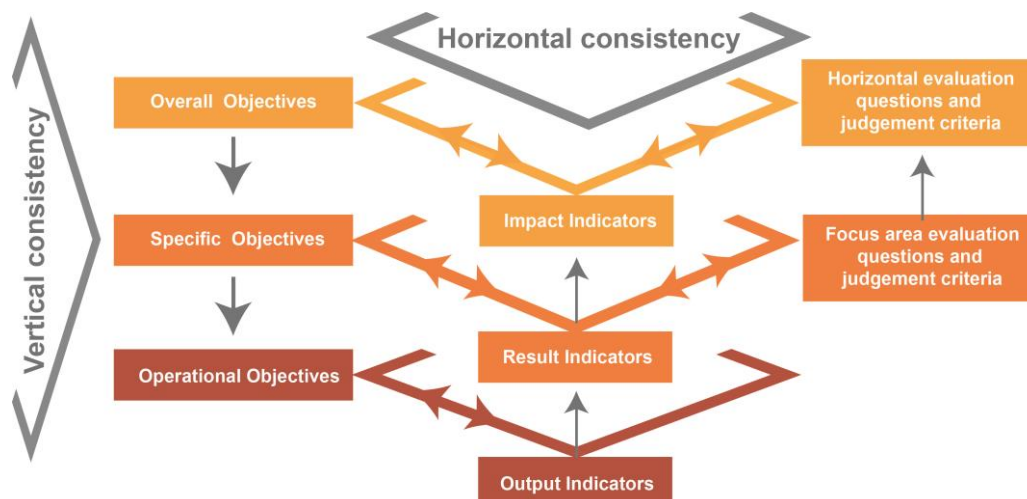
- **Revisit the ex ante findings with respect to the consistency between RDP objectives, EQ/judgment criteria and indicators:** If gaps are identified, proceed with the following working steps.
- **Check the horizontal consistency between RDP objectives, EQ/judgment criteria and indicators:** The triangular consistency between objectives, EQ/judgment criteria and indicators should be well established at each level of the intervention logic. Identify gaps in the ability of EQ to capture RDP effects towards objective achievements and in the ability of indicators to answer EQ. Propose solutions to bridge gaps. (→ **Tool 1 Matrix for checking horizontal consistency at focus area level (example)**).
- **Examine the vertical consistency in the hierarchy of objectives, EQ and indicators:** Check if the answers to lower level (related to focus area) EQ can provide useful and sufficient information on programme results in order to allow to answer higher level EQ. Appraise if indicators at lower level provide sufficient evidence to conduct the analysis at higher level, e.g. if proposed common and programme-specific result indicators are able to provide sufficient information to assess programme's impacts. In case of identified gaps make suggestions how to bridge them. (→ **Tool 2 Matrix for vertical consistency check (example)**)

Expected outcome: consistent evaluation framework

Using ex ante evaluation as starting point

As a first consistency check had already been part of the ex ante evaluation it should be used as a starting point. However, the ex ante evaluation's findings may not have been fully considered by programme authorities and the programme context may also have changed. It is therefore considered good practice to repeat this assessment.

Figure 9. Horizontal and vertical consistency between objectives, evaluation questions and indicators



Source: European Evaluation Helpdesk for Rural Development, 2015

Checking the horizontal consistency

The horizontal consistency reflects the relation between EU and RDP policy objectives, CEQ and PSEQ and indicators at result and impact level and covers:

- Consistency of overall programme objectives with EU objectives, related evaluation questions/judgment criteria and impact indicators, which are used in the assessment of RDP achievements towards EU objectives and programme impacts. In case the RDP contains RDP-specific overall objectives their consistency with PSEQ and programme specific impact indicators should be examined as well.
- Consistency between RD focus areas' objectives, focus area-related EQ/judgment criteria and result indicators. These indicators help to assess to what extent the specific objectives have been achieved within the group of programme's beneficiaries – programme results. Any gaps in the consistency between CEQ and common indicators should be identified and bridged by proposing additional indicators (see fiches for CEQ published in separate annex). In case the RDP contains programme-specific focus area-related objectives, PSEQ and programme-specific indicators (e.g. mentioned in the evaluation plan/indicator plan), their consistency needs to be examined as well.

When testing the horizontal consistency between objectives, EQ and indicators the following guiding questions should be verified:

- To what extent do the EQ allow a sound assessment of achievements of RDP objectives?
- To what extent does the evidence collected by means of indicators enable to answer the EQ?
- Which PSEQ and programme-specific indicators still need to be developed to fill gaps in the consistency?
- Which additional information needs to be collected to make the "additional" programme-specific indicators operational?

For an effective identification of potential gaps in the horizontal consistency, the following tool can be used:

Tool 1. Matrix for checking horizontal consistency at focus area level (example)

Objective (FA related)	Common evaluation question	Judgment criteria	Indicator(s)	Identified gaps ⁴⁶	Additional indicators for filling identified gaps
6B: Fostering local development in rural areas	CEQ 17: To what extent have RDP interventions supported local development in rural areas?	<p>Services and local infrastructure in rural areas have improved</p> <ul style="list-style-type: none"> • Access to services and local infrastructure has increased in rural areas • Rural people have participated in local actions • Rural people have benefited from local actions • Employment opportunities have been created via local development strategies • Rural territory and population covered by LAGs has increased 	<p>% of rural population covered by local development strategies (FA 6B – Result indicator)</p> <ul style="list-style-type: none"> • Jobs created in supported projects (LEADER) (FA 6B – Result indicator) • % of rural population benefiting from improved services/infrastructures (FA 6B – Result indicator) 	Evidence collected via common indicators does not show the LEADER contribution to improved services/infrastructure	<ul style="list-style-type: none"> • Number of projects/initiatives supported by the Local Development Strategy • % of RDP expenditure in LEADER measures with respect to total RDP expenditure

Checking the vertical consistency

The vertical consistency check follows the hierarchy of objectives and assesses in particular:

- if the achievement of operational objectives leads to the achievement of specific and, consequently, of overall objectives.
- if answers to EQ at lower level (related to focus areas) provide useful and sufficient information on programme results and can be used to answer EQ at higher level (related to EU objectives).
- if the information collected via result indicators can be used in the assessment of impact indicators.

For the appraisal of the vertical consistency, the following guiding questions can be used:

- To what extent can the evidence collected by means of common and programme-specific indicators at lower/micro level (output for results and result for impacts) provide sufficient evidence to carry out an assessment of impacts at higher/macro level?
- Which gaps can be identified in the vertical set of common and programme-specific indicators?
- Which indicators/additional data collection shall be proposed to fill these gaps?

⁴⁶ This is the case when the common elements are neither able to answer the relevant EQ nor to be used later in the assessment of impacts

To check the vertical consistency between various levels of indicators and to fill the identified gaps, the following table can be proposed:

Tool 2. Matrix for vertical consistency check (example)

Impact indicator	Result indicators related to the impact indicator	Identified gaps	Filling the identified gaps (programme specific indicators and additional information)
I14: Rural employment rate	6A: Jobs created in supported projects (including secondary contributions to job creation from operations programmed under other FA, among others also under 6B (for LEADER))	No means to collect information on the employment rate for beneficiaries in supported sectors	Employment rate in supported sectors

Specificities of LEADER

In the case of LEADER, the consistency check between the RDP intervention logic and evaluation elements does not only cover focus area 6B but also those focus areas to which the CLLD strategy operations are likely to contribute. This consistency check will allow among others to follow contributions (predominant and additional) as flagged by beneficiaries of projects implemented under the CLLD strategy via the LAG operations database.

If the programme contains programme-specific objectives for LEADER (e.g. concerning specific effects of LEADER, its added value, implementation of LEADER method, etc.), their consistency with PSEQ and programme-specific indicators should also be checked. If there are no PSEQ and programme-specific indicators, they can still be developed during the programming period.

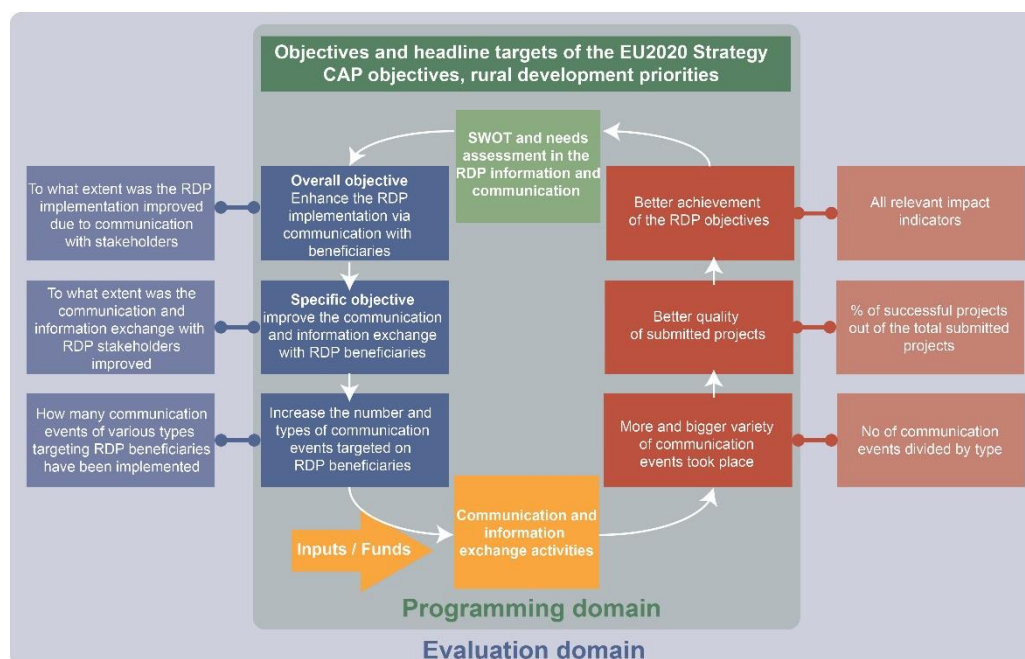
Specificities of TA

If Member States included in the evaluation plan also evaluation topics which relate to actions supported by TA, such as management and administration (including the delivery mechanism), RDP communication, capacity building, etc. and these are equipped with PSEQ and programme-specific indicators, the consistency check has to cover them as well. They can be outlined already in the evaluation plan, or evaluation stakeholders may propose them during the programming period.

In case evaluation topics linked to actions supported by technical assistance are not equipped with PSEQ and programme-specific indicators, these can be developed also later during the programming period.

As an example, the intervention logic for communication and information exchange with evaluation stakeholders and related evaluation elements is shown in the figure below:

Figure 10. The intervention logic and evaluation elements for communication and information exchange



Source: European Evaluation Helpdesk for Rural Development, 2015

Specificities of NRN

In case the programme contains PSEQ and programme-specific indicators in relation to NRN, their consistency with NRN objectives should be checked as in the case of the RDP. This will be done in each case for NRNP.

If the RDP does not contain NRN- specific EQ and indicators, chapter 5.2.3 'Develop RDP specific evaluation elements' provides general guidance. Concrete advice with respect to NRN evaluation elements will be provided in separate NRN evaluation guidelines⁴⁷.

⁴⁷ Information can be also found on: enrd.ec.europa.eu/enrd-static/evaluation/good-practices-workshops/national-rural-networks/en/national-rural-networks_en.html

Further reading

Getting the most from your RDP: Guidelines for the ex ante evaluation of 2014-2020 RDPs, PART II, Chapter 2 and 3,
 Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDP, PART II, Chapter 1, 2 and 3, link: http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

Working paper: Common Evaluation Questions for rural development programmes 2014-2020, link: <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

5.2.3 Develop RDP specific evaluation elements

Whenever the scrutiny of the consistency between the intervention logic and evaluation elements identifies any major gaps, it is necessary to develop PSEQ and programme-specific indicators, as well as additional indicators. This will enable a correct and comprehensive assessment of programme results and impacts, since the CMES only provides the necessary minimum elements to accomplish the evaluation tasks

Recommended working steps

- **Development of PSEQ:** If the previous consistency check (chapter 5.2.2) between the SWOT analysis, intervention logic and evaluation elements has identified very important programme-specific issues to be evaluated and not covered by already existing PSEQ, new PSEQ/judgment criteria can be developed. (→ **Tool 3 Working procedure for filling the gaps in evaluation elements using the SWOT analysis**)
- **Development of programme-specific indicators:** Newly developed PSEQ can require the formulation of new programme-specific indicators, to capture the primary, secondary, intended and unintended effects within the area of the newly identified evaluation topic. However, new programme-specific indicators should be developed only if already existing programme-specific indicators are not sufficient to answer the newly developed PSEQ.
- **Re-checking consistency between the complete set of evaluation elements and the intervention logic:** The consistency of newly developed programme-specific evaluation elements with the RDP intervention logic shall be re-checked and verified.

Expected outcome:

- Comprehensive list of common and programme-specific result and impact indicators able to capture all direct and indirect RDP effects. This list is the basis for identifying data needs and establishing data management.

Focus area-related CEQ and additional indicators

Several CEQ related to focus areas are linked with common result (target) indicators which measure the percentage of all supported units under a given RD support scheme.

In most cases the above mentioned CEQ cannot be answered in a satisfactory manner with this type of indicators (either they are not fully consistent with the judgment criteria proposed in the *Working Document Common evaluation questions for RDPs 2014-2020*, or stakeholders in Member States have suggested additional judgment criteria). For some of these CEQ, complementary result indicators exist to provide a better assessment of results achieved⁴⁸. For the other CEQ, additional indicators may be developed already for the RDP evaluation to be reported in 2017 to improve the robustness of answers⁴⁹. Additional indicators should not be mixed with programme-specific indicators, which are used to answer the PSEQ and measure programme-specific effects.

Concrete advice how to develop additional indicators in answering focus area-related CEQ can be found in templates for CEQ, which will be published in a separate document.

Programme-specific evaluation questions and indicators

PSEQ and programme-specific indicators are developed in order to capture programme-specific effects, particularly in case the RDP contains specific objectives and priorities. PSEQ can also be developed in order to break down the rather general CEQ in order to provide more specific evidence. Moreover, PSEQ can be formulated for LEADER and specific evaluation topics presented in the evaluation plan (e.g. TA, NRN, delivery mechanisms, etc.).

Practice has shown that Member States have so far developed programme-specific indicators rather than PSEQ. If there are gaps in the consistency between the RDP intervention logic and evaluation elements (see chapter 5.2.2), PSEQ and programme-specific indicators should be formulated.

Programme-specific elements in RDPs

The review of approved RDPs shows that Member States often use programme-specific context indicators in the description of the initial sectoral, environmental and socio-economic conditions characterizing a given programming area (see: SWOT analysis included in RDP⁵⁰), whenever the common context indicators are not sufficient to provide a realistic picture. Also, some programme-specific result indicators are already formulated (lists of these indicators are available in annexes of the approved RDP⁵¹). Yet, from the perspective of evaluation it may appear that these indicators are incomplete and/or internally inconsistent and have to be re-examined during the consistency check between the intervention logic and evaluation elements, adjusted and complemented, e.g. in order to assess all important programme-specific effects, including potential programme-specific indirect, secondary, unexpected or negative

⁴⁸ Focus area-related CEQ number 4 (FA 2A), 11 (FA 5A), 12 (FA 5B), 13 (FA 5C), and 14 (FA 5D). In case of CEQ 14, there are two complementary result indicators to answer it

⁴⁹ For example, focus area-related CEQ number 6 (FA 3A), 8 (FA 4A), 9 (FA 4B), etc.

⁵⁰ Article 8(1)(b) of Regulation (EU) No 1305/2013, and Annex I, part 1(4) to Commission Implementing Regulation (EU) No 808/2014

⁵¹ For example, some RDPs have focus area 2C for forestry or food industry

effects. In all these cases relevant PSEQ and a set of consistent programme-specific indicators have to be formulated.

Development of programme-specific evaluation questions

PSEQ are linked to programme-specific objectives in precise terms. The more precise the objective is, the easier is the formulation of the EQ and the more straightforward it becomes to link individual programme interventions to specific outcomes. In RDPs themselves, programme-specific objectives are not necessarily accompanied by PSEQ. When developing PSEQ the task is to identify to what extent:

- CEQ reflect the programme-specific objectives and expected effects of the intervention logic;
- CEQ, related judgement criteria and indicators will enable the evaluator to capture the full range of achievements of programme-specific objectives of the particular RDP and the programme-specific effects;

The formulation of PSEQ should:

- Reflect changes over time resulting from implementation of specific programme elements (e.g. specific programme measures) or specific characteristics of beneficiaries or programming area (e.g. age, gender, type of farms, environmental conditions, etc.).
- Specify an activity and an area of interest that can be clearly measured or observed.
- Be formulated in a manner that points to only one distinct aspect or change and establishes a clear causal relationship between the programme and a desired change in outcome (effect) (i.e. "To what extent did the change happen due to the intervention?").
- Be clear, specific, straightforward and phrased in terms of capturing the contribution of a given programme to the programme-specific objectives sought in terms of the programme results and impacts.
- Take into consideration individual programme objectives and specificities of a given programming area.

It can be expected that in each individual RDP the consistency-check of its intervention logic with evaluation elements (including its linkages to the SWOT analysis) will lead to:

- A breakdown of the CEQ in more specific PSEQ,
- The formulation of new PSEQ, which focus the evaluation on newly identified evaluation topics, and
- a fine-tuning of already existing PSEQ.

As regards the process of developing PSEQ, it is recommended that the evaluator involves more evaluation stakeholders, e.g. in the form of systematic structured interviews or a workshop with programme stakeholders (including practitioners, beneficiaries, partners and policy makers). This will help to formulate relevant PSEQ by capturing basic trends in "possible" and "experienced" programme effects. By suggesting areas where the programme includes a range of unintended or uncertain effects, the evaluator and the stakeholders can draw on an earlier developed theory of change.

Clearly, at the first stage of evaluation, a preliminary qualitative assessment is essential because it can provide valuable insiders' perspectives and lead to the formulation of important PSEQ focused on programme performance, especially regarding its positive, negative, intended, and unintended effects.

Development of programme-specific indicators

PSEQ should be answered with programme-specific indicators, by providing measurable information of individual aspects of the programme performance at an appropriate level of accuracy. Programme-specific indicators should be developed in line with RACER criteria⁵² (relevant, accepted, credible, easy, and robust). Programme-specific indicators can be formulated as context, output, and result indicators and should be coherent with PSEQ, the intervention logic and the SWOT analysis (see the section below). Programme-specific indicators should show unintended and indirect programme effects. Their development should follow the procedures described in Chapter 5.2.3.

To sum up, the newly developed programme-specific indicators may therefore comprise:

- result and impact indicators used to answer newly developed PSEQ, which derive from:
 - Programme-specific objectives,
 - SWOT analysis,
 - Identified unintended and indirect effects,
 - Breakdown of “old” PSEQ and CEQ.
- context indicators enabling analysis of programme impacts.

Tool 3. Working procedure for filling the gaps in evaluation elements using the SWOT analysis

The **gaps in evaluation elements can be identified** via checking the consistency of existing PSEQ and programme-specific indicators with individual aspects of the SWOT analysis included in the RDP. The SWOT analysis is based on judgments, and it is therefore subjective and qualitative by nature. The SWOT analysis can also be used for evaluation, e.g. as a tool/instrument to identify those areas where key improvements due to the programme can be expected.

The analytical procedure can be as follows:

First, it is important to understand that weaknesses and strengths identified by the SWOT analysis can be controlled by the programme, i.e. the RDP is always expected to affect them. By contrast, opportunities and threats are mostly external aspects, which are usually out of the control of the RDP and are determined by its general socio/economic/environmental endowment (i.e. they usually remain unaffected by the programme).

Furthermore, it is important to bear in mind that one of the main reasons why a given RDP was installed was to:

Reduce identified sectorial, environmental and socio-economic weaknesses, and
Maintain and enforce strengths.

⁵² Technical Handbook on the Common Monitoring and Evaluation Framework of the CAP 2014 - 2020

Second, the PSEQ can be formulated on the basis of the SWOT analysis in the form of cause- (given RDP) and effects- (results and impacts) relations, by asking: "To what extent has the RDP contributed to reduction of weaknesses stated in the SWOT analysis, and NOT affected the strengths mentioned in the SWOT analysis?"

While the first part of the PSEQ focuses on intended programme effects, the second part of the question stresses the avoidance of negative unintended effects. Given the above, answering these types of questions will most likely require the construction and use of a set of indicators, which will allow measuring the above mentioned intended and unintended effects.

In specific cases, an extended evaluation may seek to provide an answer to the question: to what extent have opportunities and threats mentioned in the SWOT analysis been affected by a given programme/measure. Yet, this will only be possible if the magnitude and the scale of the programme had been assessed as substantial from a macro-economic perspective.

Third, translate weaknesses and strengths mentioned in the SWOT into specific result and impact indicators enabling answering the PSEQ.

The SWOT analysis is usually based on the analysis of context indicators reflecting the state of the economic, social and environmental situation in a given territory prior to a RDP intervention. However, as the RDP is implemented, the performance of context indicators representing sectorial weaknesses and strengths will be affected by the programme itself as well as by other exogenous factors. While performance of specific context (and result) indicators measured at a micro-level (e.g. labour productivity in agriculture) will reflect inter alia the micro-effect of a given programme and measure, the performance of similar context (and impact) indicators, measured for a given territory (e.g. labour productivity in agriculture), can be expected to change, inter alia, due to programme impacts.

The context indicators should therefore serve two purposes:

- Contribute to the identification and measurement of strengths and weaknesses within the region, as basis for the SWOT analysis, and
- Help to analyse impacts achieved within the programme in light of the general economic, social, structural or environmental trends.

Context indicators showing weaknesses and strengths of rural areas targeted by the specific RDP are expected to improve, or at least not to deteriorate. The task is therefore to identify all the main weaknesses and strengths and express them in the form of relevant programme-specific result and impact indicators.

Fourth, compare the list of common and already existing programme-specific result and impact indicators with newly developed programme-specific result and impact indicators (with indicators derived from the SWOT analysis).

Existing programme-specific indicators, already included in the RDP, should be the subject to scrutiny of their suitability to assess programme-specific effects during the preparatory and structuring stage of the evaluation (Chapter 5.2).

Re-checking consistency between the complete set of evaluation elements and the intervention logic

The newly developed PSEQ and programme-specific indicators extend the scope of the evaluation by asking inter alia how effective and efficient the RDP has been in addressing needs and priorities of a given programme area, expressed in newly

identified programme-specific objectives/topics. They should therefore be consistent with the intervention logic and the already existing evaluation elements and show in detail the interactions between the priorities, focus areas and measures, on one hand, and the linkages between expected programme-specific outputs, results and impacts, on the other hand. Newly developed PSEQ and indicators allow collecting data and information to capture the primary, secondary, expected/unexpected, intended/unintended effects of the RDP interventions (see Part III, Annex 10, tool for qualitative appraisal of RDP operations' primary and secondary contributions).

The final set of common, additional and programme-specific indicators for the RDP should allow to answer all CEQ and PSEQ and reflect the RDP national/regional and EU priorities and objectives⁵³.

Specificities of LEADER

Often the LEADER-related programme-specific evaluation topics are formulated already in the evaluation plan or at later stages during programme implementation. Less often programme authorities formulate at the same time also LEADER-related PSEQ and indicators. Usually, the LEADER-related questions and indicators are developed additionally either by managing authorities or evaluators. Similar rules, described above, are applied also in the case of formulation of LEADER-specific evaluation questions and indicators. LEADER-related evaluation topics influence what data should be collected via LAG operations database mentioned in previous chapters.

LEADER-specific evaluation topics may have been formulated also at LAG level in relation to the CLLD strategy.

Specificities of TA

Managing authorities may decide to assess specific topics in relation to actions supported by TA. As mentioned in the previous steps, each time such a topic is identified, the programme-specific objective, EQ and indicators should be formulated either by the managing authority itself or by evaluators.

Specificities of NRN

Often the NRN-specific evaluation topics are formulated in the RDP's evaluation plan. Once the NRN-related evaluation needs, topics and objectives have been defined, the NRN-specific evaluation questions are formulated in consistency with NRN-related objectives and shall be further specified with judgment criteria. NRN-specific indicators are formulated in consistency with the above judgment criteria. Concrete advice with respect to NRN evaluation elements will be provided in the NRN evaluation guidelines, to be published in 2016.

⁵³ Capturing the success of your RDP: Guideliens for the ex post evaluation of 2007-2013 RDPs

Further reading

Getting the most from your RDP: Guidelines for the ex ante evaluation of 2014-2020 RDPs, PART II, Chapter 2 and 3.

Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDP, PART II, Chapter 1, 2 and 3

http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

Guidelines: Establishing and implementing the evaluation plan of 2014 - 2020 RDPs, Part II, Chapter 6, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

WD: Common Evaluation Questions for Rural Development Programmes 2014-2020, <http://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

6 STRUCTURING THE EVALUATION

In the structuring phase, the evaluation stakeholders set up the **evaluation approach** and establish the basics for collecting the **evidence** necessary to answer the EQ. For this purpose, the evaluation methods are selected and combined; information needs for the assessment of common and programme-specific indicators are identified; data sources are screened and provisions are made in order to get data and information in the required format for RDP evaluation.

There is however no sharp borderline between preparing and structuring the evaluation. In some Member States, managing authorities define already in the terms of reference a certain evaluation approach or even an evaluation method, whereas in other Member States it is up to the evaluators to propose the approach.

The decision on the selection of the evaluation approach ideally stays with the Managing Authority or evaluation experts within the ministry. In case the evaluators propose the evaluation methods, it is recommended that the Managing Authority closely follows the selection of the evaluation approach and methods. This is important also with a view to later judgement on the quality of the outcomes of the evaluation.

As for conducting and reporting on the evaluation in the AIR 2017 the structuring phase should preferably focus on the legal requirement⁵⁴ linked to this particular report: “reporting and quantification of programme achievements, in particular through assessment of the complementary result indicators, and relevant evaluation questions”, meaning those related to RD focus areas and other RDP aspects (synergies among priorities and focus areas, NRN and TA).

In spite the focus of these guidelines is on reporting of the evaluation in 2017, it is recommended not to limit the structuring phase to CEQ No 1 – 21, but to prepare for all RDP-related EQ. The procedure is similar. Such an approach pays back later, when data will be needed to calculate programme impacts and answer horizontal EQ – common EU objectives and overall RDP-related EQ (in 2019 and in the ex post).

⁵⁴ Annex VII, point 7 to Commission Implementing Regulation (EU) No 808/2014

6.1 Set up a consistent evaluation approach

6.1.1 Selection of a suitable evaluation approach to the evaluation of RDP results in 2017

The decision on the evaluation approach (e.g. theory of change, quantitative, qualitative and mixed approaches) is one of the key steps in preparing the evaluation. The selected approach has to capture and quantify programme achievements, in particular through the assessment of result indicators and answering focus area-related EQ. The approach has implications on data requirements, but also on the quality and robustness of evaluation findings.

Recommended working steps:

- **Review what the evaluation approach needs to be able to capture for the specific evaluation in 2017.** List those aspects that the evaluation approach needs to fulfil in 2017: quantification of programme achievements, assessment of programme results (direct programme effects, indirect programme effects, secondary contributions, synergies and transverse effects, proportionality of analysis, assessment of counterfactual situation).
- **Review various evaluation approaches based on quality criteria:** Different evaluation approaches (theory of change, quantitative, qualitative and mixed approaches) should be critically reviewed by taking into consideration the criteria for selection of evaluation approaches. This can be done, e.g. by assigning scores (1-5) to each separate criterion (e.g. causality, selection bias, etc.) and by summing up the obtained results. Furthermore, scores might be adjusted by assigning to each of the above criteria individual weights (the latter show the “theoretical importance” of an individual criterion versus other criteria). **(0 - Criteria for selection of evaluation approach)**
- **Assess adequateness of approaches for assessment of programme effectiveness and efficiency:** Under this step the Managing Authority and evaluators may consider various aspects of practicability of each evaluation approach by taking into consideration their suitability for the analysis of programme effectiveness and efficiency at various levels (i.e. micro-level, regional-level, macro-level), as well as their ability to provide results at various descriptive scales (i.e. nominal scale, ordinal scale and cardinal scale). Here, evaluators may assign scores to individual approaches by combining individual criteria, e.g. ability to assess programme effectiveness at micro-level using cardinal scale (score=5), etc.
- **Assess adequateness of approaches for analysing achievements of RDP objectives:** Under this step, evaluators may consider other aspects of the evaluation approach’s practicability by answering the question: which evaluation approach is most suitable for the analysis of common, horizontal and specific objectives of individual RDPs. Or, which approach appears as the most advantageous concerning the analysis of a particular type of objective, e.g. horizontal objective? This can also be done, e.g. by assigning scores (1-5) to each criterion mentioned above.

Expected outcome: Decision on evaluation approach.

What needs to be captured by the evaluation approach?

The main focus of the RDP evaluation in 2017 will be the quantification of programme achievements, in particular through the assessment of result indicators and answering focus area-related EQ. This quantification will reflect the real uptake of the RDP and will be based on all operations collected through the operations database by the end of 2016.

The assessment of programme results⁵⁵ requires taking into consideration:

- Direct programme effects which occur at the level of programme beneficiaries (to see them involves in most cases comparison of effects between programme beneficiaries and non-beneficiaries),

Indirect programme effects (e.g. deadweight loss, leverage effects, etc.), which also occur at the level of programme beneficiaries; they have to be assessed on a case by case basis, depending on the indicators and likelihood of indirect effects to appear. More information on programme direct and indirect effects can be found in Chapter 5.1 'Revisit the RDP intervention logic'.

In assessment of RDP results the calculation of result indicator covers the operations:

- Primarily programmed and implemented under the FA which relates to the result indicator.
- Operations which are programmed under other FAs but which contribute secondarily to the above result indicator.

To assess **secondary contributions** of operations to focus areas other than those under which they are programmed, Member States can use sampling techniques. Samples are based on information collected via operations database, namely from application forms and payment requests, where beneficiaries flag all focus areas to which implemented operations contribute in addition to the one under which they are programmed, while estimating the extent of their contribution (in ratio). If massive contributions are expected, samples should be representative of the identified population. In case of minor contributions of certain operations to other focus areas, the total population of beneficiaries could be used for calculation. If, however, the contributions are minimal, they can be neglected. More information on the assessment of secondary contributions with respect to concrete result indicators and related CEQ can be found in the Annex 9, published as a separate document. Q.

The assessment of secondary contributions also shows where the programme is generating synergies (positive transvers effects) among focus areas and rural development priorities. For preliminary analysis of synergies or potential negative transverse effects, the evaluators may use the qualitative assessment tool of Annex 8 of PART III.

⁵⁵ Assessment of impacts of a given RDP at territorial or macro level (which should be reported in AIR submitted in 2019 and ex-post evaluation) will require in addition to the above effects also consideration of other indirect programme effects (e.g. displacement effects, multiplier effects and other general equilibrium effects) occurring at the level of regions affected by a given RDP

Challenges in the assessment of programme results

The **major challenge** in the assessment of RDPs is to attribute observable changes in the programme area (as values of indicators or qualitative changes) to RDP's intervention. The robustness and rigour of the identified attribution will depend on the selected evaluation approach. A number of challenges must be addressed in the selection of the evaluation approach⁵⁶:

- The scope of socio-economic and environmental objectives is very broad because of the heterogeneity of rural areas and their specific strengths and weaknesses; Availability of data is the key factor in the selection of methods to assess the programme effects: for quasi-experimental evaluation, micro-data at single farms/firms, municipality, district, NUTS III, NUTS II are needed, which are aggregated and extrapolated up to the programme area. In case of lack of data, the input-output models can be used. However, such methods are not ideal from an analytical point of view because they do not provide robust findings⁵⁷.

The main methodological challenge in the assessment of programme results (at a micro-level) in 2017 will be to answer the question: "What would have happened to the respective programme beneficiaries /area without the programme?" by providing the evidence of a true cause-and-effect link between the values of observed indicators and the RDP.

This question cannot be answered straightforwardly: It is not easy to establish whether causal relationships between the programme and the observable values of the indicators exist as also other factors (independent from the RDP) may simultaneously affect it. For example, the change in agricultural output/AWU observed for the group of programme beneficiaries can be affected by the change in input and output prices, managerial skills of beneficiary farmers, factor endowment, etc.

This causal inference problem can be solved empirically by finding an appropriate counterfactual. While finding a suitable counterfactual is more easily achieved for RDP measures focused on sectorial and socio-economic effects, it proves to be more complicated for most of the environmental measures affecting biodiversity, water quality, HNV, climate change, etc. Due to the complexity and site specificity of programmes oriented particularly at the environment, the identification of control groups/areas and the establishment of a situation with and without the programme may become especially difficult.

⁵⁶ See EUROPEAN COMMISSION - Directorate-General for Agriculture and Rural Development – Unit E.4 (2014): Investment Support under Rural Development Policy. Final Report. Brussels. ISBN 978-92-79-35314-7

⁵⁷ For more information see Guidelines for ex post evaluation of 2007-2013 RDPs, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

Selection of the most advantageous evaluation approach

There are four basic evaluation approaches: a) theory of change, b) quantitative, c) qualitative and d) mixed approaches (see [Guidelines for ex-post evaluation of RDPs 2007-2013](#)).

The evaluation practice shows that a mixed-method approach involving the integration of rigorous (e.g. quasi-experimental) quantitative and qualitative methodologies is the most adequate empirical approach to evaluate RDPs provided that sufficient data on beneficiaries can be collected. Mixed-methods evaluations seek to integrate social science disciplines with quantitative (counterfactuals) and qualitative approaches to theory, data collection, data analysis and interpretation.

Mixed-methods approaches can help to develop more comprehensive evidence of programme results (and impacts) and, for example, can be used to distinguish between implementation failure and theory failure in the event of a lack of intended results/impacts. The key strength of this approach is the ability to provide a triangulation of specific methods and data.

One of the main areas in the RDP evaluation where mixed-methods are especially applicable is the verification and in-depth analysis of the main reasons for high, medium or low effectiveness of programme support. If the results from different methods converge, then inferences about the character and magnitude of these impacts will be stronger. If they diverge, mixed-methods can provide a more objective explanation of factors behind it.

The selection of a robust evaluation approach must take into account the limitations and requirements of different methods. For testing evaluation approach and methods, the evaluators should apply the criteria for basic evaluation standards, such as credibility, rigour, reliability, robustness, validity, transparency and practicability (see also the table of criteria for selection of evaluation approach in Part III, Annex 9)⁵⁸.

⁵⁸ Guidelines for the ex post evaluation of 2007-2013 RDPs, chapter 4, enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

Further reading

Capturing the success of your RDP: guidelines for ex post evaluation of 2007-2013 RDPs, Chapter 4, Part II, enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

6.1.2 Select evaluation methods and their combination

Although the main responsibility for the choice and use of evaluation methods lies on evaluators, the Managing Authorities may however express a preference for certain methods in the terms of reference. In each case, the choice of evaluation methods should meet the criteria for conducting a high quality evaluation. The application of different methods, even using the same data set, may lead to different results, including the estimated magnitude of programme effects and even its sign (+/-). In consequence, the application of inadequate evaluation methods and techniques which are unable to control the selection bias and eliminate other systematic errors, may considerably obstruct an evidence-based policy-making.

Recommended working steps:

- **Verify the applicability of a given evaluation method in the context of a pre-selected evaluation approach:** The recommended mixed evaluation approach combines various quantitative and qualitative approaches with a theory based approach. It is crucial that at least some of the selected evaluation methods enable to assess the true programme effects and attribute observed changes to the intervention.
- **Review the ability of a method to meet basic evaluation standards:** Conducting rigorous evaluations of impacts (a mix of quantitative methods based on counterfactual analysis and credible qualitative methods) is crucial for policy learning. Pre-selected methods should meet basic evaluation standards (rigour, credibility, reliability, robustness, validity, and transparency).
- **Consider budget, time and data constraints:** Early thinking on evaluation design can facilitate better budget planning, save resources and improve data quality and quantity. If this is not the case, the evaluator may face severe budget, time and data constraints, which may act as disincentives to conduct rigorous evaluations. In general, there are several options for carrying out sound evaluation under budget, time and data constraints.
- **Selection of adequate methods:** An appropriate bundle of methods can minimise potential bias and systematic errors in evaluations. It is therefore recommended to carry out additional surveys in order to collect additional data on characteristics and performance of programme beneficiaries and non-beneficiaries instead of taking the risk of overstressing the interpretation of programme results based on simple yet biased evaluation techniques.

Expected outcome: Selected evaluation methods

What should be considered in the selection of suitable evaluation methods?

In the process of selecting a suitable method, the following issues are crucial⁵⁹:

- Users of evaluation have an interest that the results are reliable, scientifically sound, robust, and valid. When using these criteria, it is possible to classify methods and evaluation techniques with respect to their appropriateness to evaluate individual RD measures. Various methods meet the criteria of soundness, robustness, and validity at different levels. Each method has specific data requirements, strengths, and limitations. There is no best method with respect to the quality of all criteria used in evaluation⁶⁰. Restriction on access of individual data may decrease the level of transparency but model specification, estimation techniques, and test statistics should be transparent even in such cases.
- Methods that can provide quantitative results are generally preferred; however, other methods may still be useful when there are restrictions in the use of quantitative methods (e.g. data gaps, etc.)
- The majority of quantitative evaluation techniques are also not directly substitutable with each other. For example, a counterfactual approach that can be applied to the evaluation of programme results at micro-level cannot be replaced by another quantitative technique, e.g. Input-Output method which is applicable for the assessment of specific programme effects at regional- or macro-level.
- The results of different methods are expressed in different scales: e.g. counterfactual econometric and quantitative methods (e.g. Propensity-Score Matching⁶¹, Input-Output or programming methods) provide results on cardinal scales, the other (e.g. qualitative) on ordinal scales or on nominal scales. When results have to be expressed in numbers (cardinal scale), the scope of methods is limited because qualitative, theory-based and descriptive approaches allow ordinal statements at best.
- One of the biggest challenges in evaluation is to identify and test causal relations between the policy intervention and the outcomes. Only a small set of methods (typically econometric counterfactual models) are suited to provide exact results in this respect. Key in testing causal relation is the development of a counterfactual in order to see the effect of the programme, which cannot be directly observed. In general, there are two alternatives: results can be based on statistical evidence of randomised controlled trials or based on adequate counterfactual econometric assessments (e.g. Propensity-Score Matching). In case causal effects cannot be identified by observations, assumptions on causality need to be made possibly through the application of qualitative methods. The latter approach captures causal relations conceptually, but does not provide robust information on the causal relationship.
- Results based on counterfactual approaches (Propensity-Score Matching) should be tested by other methods in order to improve the validity of the results. The combination of methods contributes to the validity of results. For example, empirical evaluation studies reveal that the qualitative method can set the context and contribute to the development of hypotheses, which can subsequently be tested with

⁵⁹ See: Investment Support under RD policy, final report 2014.

⁶⁰ In scientific literature not all types of approaches are equally well accepted and the robustness of results may be questioned if results are based on a small sample size or a model specification that has not undergone peer reviews or if results were obtained in a non-transparent manner.

⁶¹ http://betterevaluation.org/evaluation-options/propensity_scores

quantitative methods. The results of quantitative methods can be validated and causal relationships explained more in-depth by applying qualitative methodologies.

- Indirect effects such as leverage, substitution, displacement effect and deadweight (windfall profit) are very important to consider when specific measures are evaluated. Any method can be used to take account of leverage and deadweight effect, but only counterfactual econometric methods (e.g. Propensity-Score Matching) can be used to quantify their size.
- Qualitative methods can be very helpful in developing hypotheses that can be tested or further explored with quantitative methods. For example, qualitative methods should be utilized to explore the range of possible RDP unintended and indirect effects. The magnitude, scale and importance of the above effects should however be verified by using quantitative approaches.
- Qualitative and quantitative methods are complementary and should not be regarded as substitutes. Quantitative methods are also complementary in many respects because some important results (e.g. deadweight, leverage or multiplier) can only be estimated using specific method/techniques.
- Given the above, it is advisable that the evaluation team chooses an evaluation design not based on a single method but on a bundle of evaluation methods and techniques which should be internally consistent and complementary to each other. Moreover, all selected methods should meet the basic evaluation quality criteria, i.e. rigour, reliability, robustness, transparency, validity, and practicability. The evaluator should try to select a strong evaluation design (consisting of a bundle of evaluation techniques and methods), bearing in mind time, money and practicability constraints (including data availability)⁶².

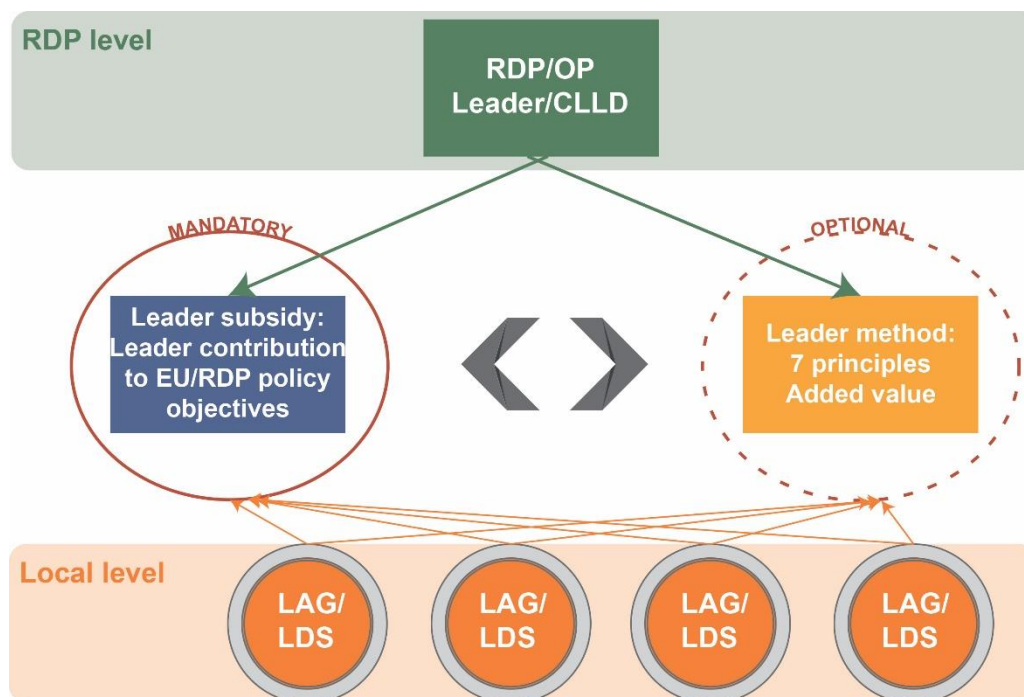
Specificities of LEADER

LEADER has to be seen as a multidimensional evaluation topic. It is evaluated at the RDP level as the measure programmed under FA 6B, but contributing to a range of RDP focus areas and priorities through CLLD strategy (see the left upper box of figure 11). However, LEADER is not only the subsidy delivered via the CLLD strategy, it is also the method which helps to create and develop partnerships (by integrating various sectors and areas of rural life), to foster participation of local people in development (bottom-up), innovation, cooperation and networking. Furthermore, LEADER can bring additional value, such as increased social capital, enhanced utilisation of local resources, improved local governance, etc. (see the right upper box of figure 11). LEADER is also part of CLLD where several ESI Funds work together towards achieving EU level objectives (see the top box of figure 11). All these aspects and others may become evaluation topics at national level.

LEADER should be also evaluated at local/LAG level (see the bottom part of figure 11). Here again the evaluation is multidimensional, covering the evaluation of CLLD strategy, assessment of the application of the 7 principles and LEADER added value for rural areas.

⁶² For more information see Guidelines for ex post evaluation of 2007-2013 RDPs, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html.

Figure 11. LEADER as an evaluation topic.



Source: European Evaluation Helpdesk for Rural Development, 2015

The complexity of LEADER as an evaluation topic requires a variety of quantitative and qualitative evaluation methods and combinations thereof. To capture the effects of the LEADER measure (subsidy) quantitative approaches will be used. In this respect LEADER operations and their contributions to the range of focus areas are collected via the LAG operations database and assessed together with other RDP operations by means of result indicators. Secondary contributions of LEADER are estimated via using sampling techniques.

Qualitative methods are used to triangulate quantitative findings. In case of other evaluation topics, such as the delivery of the LEADER method or the LEADER added value, mostly qualitative approach will be applied. It is important to recognise that the counterfactual is also a suitable tool to see programme effects caused by LEADER and can be used in quantitative, qualitative and mixed approaches. In countries where only parts of the territory are covered by LAGs, control groups could be selected within the same region with similar characteristics as the territory covered by LAGs. In "LEADER mainstream" countries (where the whole territory is covered by LAGs), LAGs with similar characteristics and different types of operations could be compared when assessing RDP LEADER effects.

Specificities of TA and NRN

Specific interventions such as TA and NRN often require specific considerations in the selection of both the evaluation approach and methods. Although mixed approaches are

proposed, in the majority of cases qualitative methods will be applied (focus groups, interviews, case studies, etc.), accompanied by the assessment of monitoring data. More information is provided in this respect in the NRN evaluation guidelines.

Further reading

Capturing the success of your RDP: guidelines for ex post evaluation of 2007-2013 RDPs, Chapter 4, Part II, enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

6.2 Establish the evidence for evaluation

6.2.1 Review data requirements for the chosen methods

High quality and timely available data is essential in order to apply suitable methods to accomplish all required evaluation tasks and to obtain robust evaluation findings. While different methods have different data needs, the evaluator also has to develop strategies to overcome data gaps.

Recommended working steps

- **Check the availability of data for all common, additional and programme-specific indicators and of financial data:** Identify data gaps (quality and availability). In case the data for certain common indicators are missing and cannot be obtained in a cost effective way, proxy indicators should be used.
- **Check the availability of data to calculate target values of indicators:** Examine if the data to calculate target values is in place. Some expected values, e.g. expected leverage of investments, should be set after specific research work has been carried out.
- **Decide which data is needed to calculate the net values of result (in 2017) and impact (in 2019) indicators and check their availability:** In this step, the data to conduct the counterfactual, calculate net values of indicators in line with selected methods, and additional data needed to answer the EQ, should be identified. This contains also the data for control groups, relevant sector data or any type of data which is needed for applying selected evaluation methods. Existing sources/databases of required data should be identified and assessed for their suitability in RDP evaluation.
- **Make the necessary arrangements for ensuring data availability:** Existing data might not be in the format required for the RDP evaluation. Therefore, it is necessary to make arrangements to obtain data for evaluation from existing databases, e.g. contract data providers, agree on data sets in specific format, apply legal procedure if necessary, etc.
- **Decide on additional data to be primarily collected:** In case existing databases do not satisfy the evaluation needs, data shall be collected primarily. Data for beneficiaries can be collected via the existing monitoring system. For control groups surveys can be carried out selecting the samples in line with the chosen evaluation method.
- **Propose adjustments of the monitoring system in order to make it better fit for evaluation:** A monitoring system, which collects data on beneficiaries also for the purpose of evaluation, is a cost effective tool to obtain data even for common and programme-specific result and impact indicators. As such, it can be used as source for counterfactual analysis providing data in suitable format and time.

Expected outcome: inventory of existing data sources, data gap analysis, arrangements for data provision and access.

Evaluation tasks and data needs

Data requirements for evaluation are linked to evaluation tasks and evaluation needs, which are expressed in the form of EQ answered by means of indicators. As such, data requirements are rooted in the intervention logic of the programme and its objectives, which address the “why” and “how” of the policy. The “why” describes the policy need and, to some extent, sets the quantified targets of the policy/measure. The “how” addresses the way by which instruments and resources will be used to achieve targets.

The following table presents the evaluation tasks to be reported in the AIR 2017⁶³ and respective data requirements.

Table 3. Evaluation task and data needs in 2017

Evaluation tasks	Types of data needed Relevant Data Needs	Time to collect	Legal Base
Assessment of the information and progress towards achieving the objectives of the programme: reporting and quantification of programme achievements, in particular through the assessment of the complementary result indicators and relevant evaluation questions	Data to calculate gross and, whenever possible, also net values of result indicators (complementary result indicators, target indicators expressed in ratios, additional indicators, and programme-specific indicators) Additional data & information needed to answer focus area-related EQ and EQ related to other RDP aspects	Since the programme start	Commission Implementing Regulation (EU) No 808/2014, Annex VII, point 7

⁶³ Annex VII, point 7 to Commission Implementing Regulation (EU) No 808/2014

Types of data in RDP evaluation

The following table provides an overview of the different types of data used in the evaluation of RDPs.

Table 4. Types of data used in rural development evaluation

Data	Sources of available data sets	Provider	Availability	Necessary arrangements
Primary data on beneficiaries	Operations database/monitoring	Paying Agency	Yes without additional costs	None
	Surveys	Evaluators	Yes with additional costs	Evaluator prepares and conducts the survey
	Income statements	Tax office	Yes, with additional costs	Agreement with the tax office
Primary data on non-beneficiaries	FADN (ensuring anonymity)	Member States responsible body	Yes	None
	National/regional statistics (ensuring anonymity)	Statistical office	Yes, no additional costs	Request for data sets
	Surveys	Evaluators	Yes, with additional costs	Evaluator prepares and conducts the survey
Secondary data on beneficiaries and non-beneficiaries	FADN	EU	Yes	None
	National/regional statistics	Statistical office	Yes, no additional costs	Request for data sets
	Surveys	Evaluators	Yes, with additional costs	Evaluator prepares and conducts the survey
Other data	Research reports	Research institutes	Yes, with additional costs	Contract with the research institute
	Other reports	NGO	Yes, with additional costs	Contract with the NGO

Evaluation methods and data needs

The evaluation approach and methods chosen have consequences on data needs, as the data availability has implications on the selection of methods. The following table provides an overview of data requirements in relation to selected methods.

Table 5. Methods and related data needs (example)

Method	Data needs	Existing data sources/providers (links if possible)	Arrangements to ensure access (if needed)	Data gaps /collection by evaluator
Quantitative				
Micro Methods				
Counterfactual surveys	The micro-level data may consist of bookkeeping data, survey data or both. Collected data should clearly identify programme beneficiaries and the level of support they received from individual RDP measures (e.g. measure X, etc.). Ideally, micro-level data panel should comprise no less than 150 beneficiary enterprises, farms or holdings for each of the analysed measure (or group of measures) and 2-3 times more for non-beneficiaries.	Data at micro-level (from beneficiaries and non-beneficiaries) should be collected on the basis of secondary data and/or own surveys. The FADN database combined with an anonymous data from the Paying Agency can be used for this purpose. http://ec.europa.eu/agriculture/ricaprod/	Arrangements carried out with Liaison Agencies: http://ec.europa.eu/agriculture/ricaprod/liaisonagency_en.cfm	Data on support from other programmes from the relevant authorities or the single Paying Agency.
Tracer surveys	Tracer studies address businesses, workers and trainees and attempt to trace through time the changing status of the survey's subject and record the time span of staying in a specific status.	Primary data collection	No access rules applied	All data to be collected by evaluator.
Macro Methods				
Input-Output (I-O) and related techniques	I-O data needs include a regional I-O model and RDP expenditure distinguished into demand for the product of model economic sectors and data on the change of productive capacity attributed to RDP projects. In the case of a Social Analysis Matrix, data needs include a regional (rural) I-O model, data specific to disaggregation of economic activities, production and households, and inter-institutional and factor-institution flows.	National I-O tables from Eurostat. http://ec.europa.eu/eurostat/web/esa-supply-use-input-tables/data/database	No access rules applied	National account, Consumer Expenditure Surveys and FADN data to complement the I-O and produce the SAM

Method	Data needs	Existing data sources/providers (links if possible)	Arrangements to ensure access (if needed)	Data gaps /collection by evaluator
CGE models	For Recursive Dynamic CGE model, data needs are more specific to the construction of the model rather than the RDP model input.	Does not apply, unless ready to use CGEs are used like the GTAP of Purdue University at: https://www.gtap.agecon.purdue.edu/	No access rules applied	I-O Table, National accounts, Consumer Expenditure Survey, Labour Force Survey and FADN data
Mostly Qualitative				
Focus groups and elite interviews	Focus groups and elite interviews are used when the evaluator aims at recording informed opinions by key stakeholders either through common participation and interaction (focus group) or individually (elite interview).	Does not apply	No access rules applied	All data to be collected by evaluator.
Mixed				
Quality of Life and of Local Governance, and Satisfaction Surveys	Surveys targeting quality of life in rural areas mainly refer to the provision, access and use of services and utilities. Local governance studies and life satisfaction surveys may also support the quality of life evaluation.	There are existing roll on surveys at a global as well as European scale. The problem is that they rarely distinguish between rural and urban areas and, of course, it is difficult to isolate net effects.	Eurostat SILK rules apply for anonymized data but the country-level samples are relatively small. The OECD does not provide access to micro-data.	All data to be collected by evaluator.

Addressing data gaps

Data gaps may be caused by various reasons: no systematic data collection, data only available for limited levels (e.g. national, not regional), missing data-sets on certain context indicators (e.g. environmental indicators).

There are different ways to address data gaps. Typical approaches include additional research and surveys, use of additional databases (e.g. national, regional statistics) or a more intensive use of qualitative methods as an alternative to hard data. However, the most effective and proactive approach to deal with data gaps is to construct baseline databases and establish the necessary data infrastructure from early stages of programme implementation/development. Unless defined otherwise, data are collected primarily by evaluators.

The requirements for evaluation in the current programming period underline the need for Member States to put in place the necessary procedures for data collection and management for evaluations⁶⁴. Managing Authorities are well advised to incorporate any data needs related to programme-specific elements in their monitoring and evaluation systems as early as possible.

Costs

To collect additional data is costly. Therefore, it is important that stakeholders (mainly managing authorities) know well all the existing data sources at EU, national and regional level and are able to provide arrangements to bring them to the required format. The need for additionally collected data by evaluators affects the evaluation budget.

Role of stakeholders in ensuring data availability

Several stakeholders should be involved in ensuring the availability of data for RDP evaluation. The communication among them should be ensured mainly by the Managing Authority or by the evaluation steering group, if established. The following table offers an overview of how these stakeholders can support data availability and quality.

Table 6. Evaluation stakeholders and their role in ensuring data availability and quality⁶⁵

Stakeholder	Description of role in the data management and collection
Managing Authority	<ul style="list-style-type: none"> Draw up a detailed evaluation plan with the monitoring and evaluation system Ensure a secure electronic system and a monitoring system for all common and programme-specific indicators Establish appropriate data infrastructure, including the necessary resources for collecting, storing and updating data Establish clear and transparent data collection and management procedures Screen all available data sources, communicate with providers, and make arrangements that data are available in format and quality needed for the RDP evaluation Ensure also communication among data providers Establish the processes for ensuring timely implementation of evaluations and reporting on them Design a good concept note and the terms of reference taking into account the available budget for evaluation, the evaluation scope and the preferred evaluation approach
Paying Agency	<ul style="list-style-type: none"> Establish clear and transparent procedures for the recording of monitoring information on applications, supported projects, payments and control Operations database Set up simple, transparent and fast procedures for the Managing Authority's and the evaluator's access to data Establish a common data system to facilitate access to data or an interface to facilitate the transfer and handling of data between the Paying Agency and the Managing Authority

⁶⁴ Article 54(2) of Regulation (EU) No 1303/2013 and Article 76(2) of Regulation (EU) No 1305/2013

⁶⁵ Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDPs, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

Stakeholder	Description of role in the data management and collection
	Ensure link to other databases, if required
Evaluation Steering Group (if established, otherwise Managing Authority)	<p>Facilitate and coordinate stakeholders' consultation in evaluation</p> <p>Establish a process for checking and ensuring the relevance of monitoring and evaluation activities to programme needs</p> <p>Ensure a Steering Group composition of experts in evaluation and rural development and experts who can provide advice on data availability, information and relevant contacts to evaluators</p> <p>Potentially involve beneficiaries in the Steering Group to facilitate access to micro-data and contacts of evaluators with beneficiaries</p>
Working Groups (if established)	<p>Ensure that working groups are composed of thematic and specialist experts so that they can advise evaluators on specific sectoral data availability (e.g. environmental issues such as water protection or nature conservation) and other data issues (e.g. on LEADER delivery)</p> <p>Establish thematic working groups according to the needs of each evaluation in order to enhance its quality</p>
LAGs	<p>Establish processes for bringing the LAG local knowledge and contacts to the service of the evaluation</p> <p>LAG level operations database</p> <p>Establish processes for feeding in self-assessments and the outcomes of CLLD Strategy into the evaluation</p> <p>Encourage the participation of LAG members in evaluation steering groups</p>
NRNs	<p>Establish an effective and transparent system for disseminating evaluation results</p> <p>Facilitate contacts and provide advice to evaluators on alternative data sources and other data requirements</p> <p>Establish mechanisms to transfer knowledge on monitoring and evaluation from one country or region to another, including the provision of capacity building to evaluators</p> <p>Develop a system for the establishment of regional proxies when only national data is available for indicators</p>
Regional governments and agencies	<p>Establish a process of communication with the Managing Authority and Paying Agency to provide help in addressing any data requirements</p> <p>Adapt their data collection systems to the monitoring and data collection requirements of RDPs and feed into or build an interface with the Managing Authority's systems</p>
Data providers	<p>Ensure early involvement of data providers (national statistical office, relevant ministries, research institutes, etc.) in monitoring and evaluation from the planning stage</p> <p>Provide data of relevance to RDPs and the results of research on relevant topics to evaluators</p> <p>Provide expert knowledge and even collect specific monitoring data for the Managing Authority (possibly on a contractual basis)</p> <p>Encourage the participation of data providers in evaluation steering groups and/or the Monitoring Committee</p>
Evaluators	<p>Ensure capacities and expertise appropriate to the evaluation type and topic</p> <p>Set up evaluation teams with capacities to implement the necessary evaluation methods (quantitative, qualitative)</p> <p>Establish a continuous interface with the Managing Authority for the timely collection of data, the identification of data gaps and solutions to bridge them</p> <p>Evaluators often have considerable experience with data collection and access to different data sources. Such experience can be brought into the evaluation teams to address any difficulties to access data.</p>

Specificities of LEADER

Data on operations implemented through CLLD strategy are collected via the LAG level operations database for output indicators and for target indicators for the focus areas to which operations have secondary contributions. The challenge will be to collect data which relate to complementary result indicators or additional indicators used to answer focus area-related EQ where CLLD strategy's operations are contributing. In this case, sampling will need to be applied and the data will be collected through surveys on CLLD strategy beneficiaries at LAG level.

In case of CLLD strategy specific intervention logic, the LAG should develop feasible indicators and ensure the data collection itself.

Specificities of TA

It is expected that mainly qualitative methods will be used in the evaluation of TA actions. This will require collecting mainly information from RDP stakeholders using qualitative tools, such as interviews, surveys, focus groups, etc. Data on accomplished actions divided by type and information needed in the assessment may be collected by programme authorities and used by the evaluator in addition to qualitative information.

Specificities of NRN

In case of NRN, the data on activities implemented via the action plan will be used from the monitoring system. The evaluator will have to use a sample in order to obtain data for NRN-specific result indicators. In addition, it is expected that qualitative information will be collected and analysed using qualitative methods. More information will be provided in this respect in the NRN evaluation guidelines.

6.2.2 Manage and collect data for evaluation

The fundamental condition of effective data management is the establishment of functioning and solid databases, which allow collecting data of high quality and timely availability. For each evaluation exercise it is important to create a database, which starts with the baseline data concerning all common and programme-specific indicators (context, output, results, including the complementary result indicators, targets, and financial data) on both RDP beneficiaries and their control groups. The baseline database needs to be linked and harmonised with the monitoring system (operations database), which collects both monitoring and evaluation data on beneficiaries.

Recommended working steps:

- **Allocate clear data collection responsibilities in order to clarify who is responsible to collect data for various types of indicators** (Context, Output, Result-Target, Complementary Result, and Impact):
 - collecting primary disaggregated data at application level, or through the payment request after project finalisation
 - transmitting or uploading disaggregated data at pre-defined time and frequency
 - inspecting data for abnormal and missing values (quality control)
 - aggregating data according to spatial units and the Member State level
 - storing data
- **Allocate clear data retrieval responsibilities:** Many indicators (e.g. impact indicators) demand data from Eurostat supported by national statistics. Other evaluation exercises may require anonymized data from European-wide Eurostat surveys, e.g. the FADN in order to establish counterfactuals or a simple sampling frame. At this stage, the Managing Authority should consult sources from Eurostat and national statistics:
 - retrieving data aggregated or disaggregated at pre-defined time and frequency;
 - inspecting data for abnormal and missing values and communicating with the relevant statistical authorities;
 - storing data and produce a metafile with the information concerning data.
- **Decide on the necessary primary surveys to be conducted by the evaluator:** At this stage, the evaluator should have a clear idea of which data are not collected via databases but are necessary to be collected. There will be

secondary or unintended effects, which may be judged as non-important and for which a decision not to collect data is made. Field surveys (of any kind) are expensive and time consuming so, make sure that the data under consideration cannot be retrieved from statistical sources and you cannot utilize a secondary analysis of recently collected data. At this stage you should decide on:

- The scope and objectives of the survey
 - The type of the survey (qualitative-quantitative) taking into account the evaluation methodology chosen
 - Statistical data which may serve as a sampling frame (e.g. FADN)
 - Opportunities to merge surveys and minimize the time and cost
- **Plan field surveys:** Field surveys should be planned as early as possible because many of the required baseline data may be retrieved at the time of application.
 - **Plan qualitative exercises:** Most qualitative survey exercises are based on key actors and informed agents. During the time of the programme's public consultation or the implementation of its various measures, the relevant authorities may have spotted key persons to be included in focus groups, experts to be consulted via in-depth interviews or simply members of key stakeholders that have followed the evolution of the programme and thus can express an informed opinion. All this contact information should be kept in a database of "potential" participants to qualitative studies.
 - **Monitor the database:** The database will require the cooperation among local and regional authorities, various agencies and the central Managing Authority/Paying Agency. The flow of data that are regularly collected or retrieved and the data from surveys should be monitored with milestones starting from the time an AIR should be delivered and going backwards. Close monitoring will flag delays, inconsistencies and lack of clarity in responsibilities very early in the data collection process. For each indicator and for each planned survey there must be an electronic system that will inform and alert the evaluators of any delays or abnormalities. Thus there will be enough time to take a corrective action either as concerns the flow of data or their quality.
 - **Create and maintain a dynamic database:** Database management and maintenance including security of personal data is of the outmost importance. The evaluators should have an access to this database limited with the competence of the given RDP and liaise with database managers.

Expected outcome: databases and data management for evaluation

Databases as the ground for proper data management

The evaluator should examine if a specific evaluation task can be accomplished by making use of the existing data sources, or if additional data needs to be collected. Additional data might be retrieved from other existing databases, may be collected primarily for the purposes of the evaluation, or may be obtained in the combination of both. For example, the values of impact indicators are collected from EU, national and in some cases regional databases. The dynamic link with these databases can facilitate the data availability.

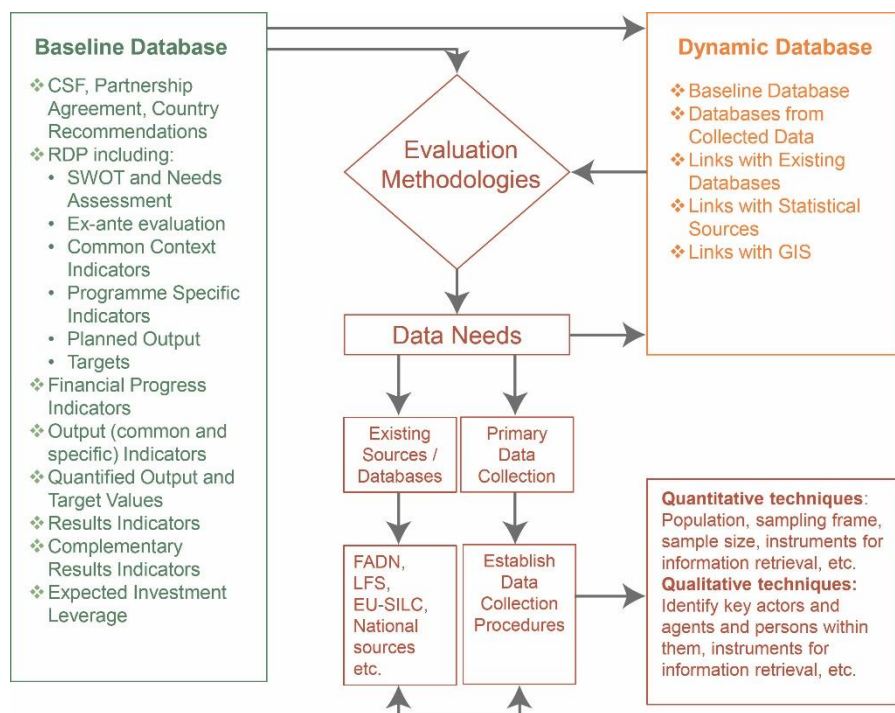
Finally, the response to certain EQ and definitely a response to the question of RDP's effects may require targeted evaluation research based on counterfactuals (net effects) or macroeconomic modelling (economy wide effects). This calls for a well-planned and tightly monitored system of data collection, data retrieval and survey implementation sets a framework of a dynamic and evolving data management system.

The baseline database should be solid, but at the same time allow for adjustments which may become necessary due to changing data needs across the programming period. For example, programme-specific indicators could be modified or newly added at later stages of programme implementation. This

also means to identify and collect more types of data through existing databases and to have a well-functioning and flexible data management system.

Data primarily collected for the evaluation or accessed from existing databases, together with the baseline database, form a dynamic database which will feed the evaluation process with appropriate data. In the dynamic database, data will be collected across the time respecting a panel of indicators. This database will allow also adding data in order to facilitate evaluation of RDP indirect and unintended effects and provide links with other databases, such as Geographic Identification Systems (GIS), national/regional statistics, etc. to respond to more complex data queries. The following figure provides a decision-making flow for data requirements respecting the purpose of the evaluation.

Figure 12. Decision-making process for data requirements and data management.



Source: European Evaluation Helpdesk for rural development, 2015

Further reading

Commission Implementing Regulation (EU) No 808/2014, <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32014R0808>

FADN Database (Eurostat),

http://ec.europa.eu/agriculture/ricaprod/database/database_en.cfm

National I/O Tables Database (Eurostat), <http://ec.europa.eu/eurostat/web/esa-supply-use-input-tables>

GTAP Database, Purdue University, <https://www.gtap.agecon.purdue.edu/>

EU SILC Database on subjective well-being (Eurostat), http://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_in_Europe_-_facts_and_views_-_overall_life_satisfaction

OECD "How is Life" Database, <http://www.oecd.org/statistics/how-s-life-23089679.htm>

Gallup-Healthways Global Well-Being Index, [gallup-healthways global well-being index 2015](http://www.gallup.com)

7 CONDUCTING THE EVALUATION

While the preparation and structuring phase of the evaluation are primarily led by the Managing Authority (sometimes with the help of an evaluator) the conducting phase is mainly carried out by the externally contracted evaluator.

The conducting phase includes observing, analysing and judging (whereas structuring has been already explained in the previous chapter). The evaluation steps can be further broken down into sequences, separated by different interim deliverables (Interim Report, Draft Final Report, etc.). The ultimate aim is that the evaluation is carried out based on good quality data, appropriate observations, an effective analysis of the policy results and impacts (through sound answers to the EQ), and accurate judgements, conclusions and recommendations.

7.1 Observing

In this phase, the evaluators must proceed with the collection of relevant information and data. The critical point will be the definition of tools and approaches for gathering the information to bridge any identified data gaps. The data management system should provide the evaluators with the necessary data. Evaluators are responsible to link monitoring data, data from other sources⁶⁶ and primarily collected data, employing various tools and techniques.

Recommended working steps:

- **Create the tools needed for the quantitative and qualitative analysis:** interview guides, questionnaires, queries for extractions from databases, requests for maps, guidelines for case studies, and any other data collection instrument that the contractor deems appropriate
- **Collect data and qualitative information needed for answering EQ:** databases, studies, people to be interviewed, appropriate case study areas, etc.
- **Describe the process of programme implementation,** composition of programmes, priorities and target levels, budget
- **Summary of the outputs:** Evaluators present the aggregated and quality-checked values for each of the common and programme-specific indicators.

Expected outcome: Data and information for evaluation

Create the tools needed for the quantitative and qualitative analysis

During the observing phase, the evaluators use the data files to calculate values for common and programme-specific indicators in line with the chosen evaluation methods. This process, however, can only happen if data on beneficiaries exists. In other words, the observing phase cannot be completed fully without sufficient RDP uptake!

As the final objective is to answer all EQ, evaluators should avoid to collect unnecessary data. Generally, the choice of data should be guided by the relevance of the variables for the assessment of achievements and the costs resulting from the planned investigation. For this reason, it is necessary to coordinate planning of data-gathering along with data-processing.

Many Member States have recently established more coordinated procedures for data management for the new programmes. In these procedures, both managing authorities and evaluation advisors have worked together on data management schemes. The workability of these databases should be appraised before real data are integrated. Glitches should be addressed as a matter of urgency.

⁶⁶ E.g. Farm Bird Index (FBI), Farm Accountancy Data Network (FADN).

A separate document contains templates for answering CEQ no 1 – 21. These templates provide among others also the guidance with respect to types and volume of data to be collected for assessment of indicators and answering EQ.

Quantitative data

In the observing phase data are collected for the following groups of indicators:

Context indicators (macro-data for RDP level)	<ul style="list-style-type: none"> Member States should collect data to quantify the context indicators (including impact indicators, which are part of the set, proxies agreed with the Commission and programme-specific context indicators that have been used in the description of the programme area). Data collected for context indicators are macro-level data (RDP, regional or national levels) and are used in the assessment of programme impacts (this includes netting out the respective impact indicators). Data for context indicators are collected from EU level databases⁶⁷ or national/regional databases, on an annual basis, independently of the programme uptake.
Input and output indicators	<ul style="list-style-type: none"> Micro-data collected at the unit level (farm, community, business) for beneficiaries. It is collected via the monitoring system (operations database). Input indicators are linked to the RDP financial allocation and payments for individual measures, activities and operations. Output indicators are linked to the basic description of the RDP implementation (number of beneficiaries, number of supported operations, etc.). Also used to calculate indicator values for some target indicators. Application forms, payment requests, standardised monitoring tables and other formats are used to collect these data and include them in the monitoring system. The structure of these forms and the decision on which data shall be collected (monitoring and evaluation) was part of the evaluation preparation and structuring.
Result indicators (micro-data at unit level)	<ul style="list-style-type: none"> Data for result indicators are micro-data, which are collected at the unit level (farm, community, business) for beneficiaries and non-beneficiaries. This data, if properly standardized, can be aggregated to get gross values. The operations database is used to collect data for result-target indicators on beneficiaries, and to identify the populations for assessment of secondary contributions and complementary result indicators. However, there is always the need for additional data to be collected by the evaluator: data on non-beneficiaries and data that cover the period after the implementation of the project, and data to calculate values for complementary result indicators. In line with requirements of specific evaluation methods, the data at micro-level should be collected on the basis of secondary data (if available) and of surveys. The earlier the collection of data starts in the RDP implementation, the better for ensuring data timeliness. This includes also secondary contributions and synergies. For practical reasons, it is important to automatize this system as much as possible, and ensure the electronic submission of the above mentioned forms. In optimal case all of this information is within a dynamic database, described in the chapter 6.2.
Impact indicators (micro- and macro-data at RDP, regional, national levels)	<ul style="list-style-type: none"> Collection of data for impact indicators depends on the approach selected to net out the impacts. In case the evaluators net out impacts using the micro-level data on units/farms – for both beneficiaries and control groups are needed for micro-level analysis, which are aggregated/extrapolated at RDP level. For beneficiaries, data can also be collected via operations database, or as for non-beneficiaries based from existing statistics (EU – e.g. FADN, national, regional). In case the evaluators will apply netting out impacts based on macro-data collected at administrative unit level, they will use mainly statistical data. In this case, the NUTS 3-5 level might be the lowest macro-level where effects of the RD support can be estimated (based on the situation in data availability in Member States)⁶⁸
Data for output and result-target indicators on beneficiaries are collected on an ongoing basis via the monitoring systems, along with the application forms and payment requests and used on an annual basis by	

⁶⁷ Annex IV to Commission Implementing Regulation No 808/2014; Working document. Context indicators fiches, Context indicator table and proposed list of common context indicators.

⁶⁸ Guidelines for the ex-post evaluation of 2007-2013 RDPs, Part II, page 132.

managing authorities in the AIR. Data for complementary result indicators, secondary contributions and non-beneficiaries are usually collected during the evaluation process (FADN, national/regional statistics, etc.)

Qualitative data and information

Apart from quantitative data, also qualitative information shall be collected by the evaluator during the observing phase, using interviews, focus groups, case studies, etc. For this purpose, tools to collect qualitative data and information are developed by the evaluator. As in case of quantitative data collection, it is important to:

- ensure the utility/relevance of the collected data to answer the EQ (check judgement criteria); and
- ensure that qualitative information collected on beneficiaries can be paired with the information collected on non-beneficiaries to ensure the counterfactual and netting out of the programme effects.

It should be noted that qualitative data collection aims to provide the evaluation with empirical information around the vision of certain entities and leaders of opinion, about the objectives, [potential] results and effects of the programme.

Description of the process of programme implementation

Description of the process of programme implementation, composition of programmes, priorities and target levels, budget, financial and physical execution, major modifications to the strategy, structure and resource allocation. Evaluators shall ensure that the relevant information in this respect is collected.

For this step it is considered essential to conduct interviews already in the initial phase of the evaluation, including representatives of the Managing Authority and the respective technical teams.

Summary of the outputs

Evaluators present the aggregated and quality-checked values for each of the common and programme-specific indicators and properly substantiated views on the progress observed or deviations occurring during the evaluation period. The aggregated values must be in harmony with what is foreseen in the templates for CEQ (to be published in separate document) and the WD Complementary Result Indicator fiches for Pillar II

7.2 Analysing

The key task of this phase is to process and synthesize all available data and information in a systematic way, with the view of assessing the (net) effects of the RDP interventions and consider them at both beneficiary and RDP territory levels. This means to analyse the collected quantitative and qualitative evidence in order to assess the effects, taking into consideration the baseline (CCI) and the trends over the programming period.

Recommended working steps:

- Introductory qualitative analysis
- Formulation of testable hypotheses regarding potential programme effects
- Testing and verification of hypothesis
- Calculation of result indicator values at micro-level (common, additional and programme-specific) and validation of values of target indicators
- Identifying programme net-results
- Considering the additional contribution of RDP operations to focus areas in the calculation of result indicators

Expected outcome: calculated values of indicators (gross, net)

For 2017, the result indicators will be in the form of gross and/or net values depending on the indicator form and data availability. As minimum requirement for reporting on evaluation in 2017, the complementary result indicator R2 "Change in agricultural output/AWU" should be presented in net values. For the remaining complementary result indicators, the minimum requirement is to calculate gross values, even though net values could be presented as a good practice (however, net values will have to be given for all the complementary result indicators in the AIR 2019). Guidance on the calculation of net/gross values is provided in templates for answering CEQ, published in a separated document.

The netting out of indicators implies that there is sufficient data collected on beneficiaries and non-beneficiaries in the system⁶⁹. Only under this condition it will be possible to compare control groups and net out the results observed and answer the EQ foreseen for 2017 (focus area-related) and beyond (all). This will influence also the assessment of achievements and enable collecting sufficient evidence to answer all EQ in the same way.

Net assessment of programme results including disentangling of programme effects from other intervening factors⁷⁰

The net assessment of RDP results is an integrative task. Transverse and various types of indirect effects are most likely to occur, and the evaluation design has to take this into account. Furthermore, numerous exogenous factors may influence a given result indicator observed at the micro level. The task to separate the programme-borne effects from other intervening factors is specifically critical concerning environmental results, where there is still a lot of uncertainty about methods to determine the proper scale of appraisal (micro or regional/local) and to derive appropriate counterfactual⁷¹.

Given the above, there are three key methodological challenges:

- The requirement to assess, wherever possible, programme results against their counterfactual, i.e. calculating the changes that would have occurred without the specific programme intervention;
- The requirement to estimate the net effects of the programme, by netting out deadweight, leverage, substitution effects, etc.;
- The requirement to construct a data and information base which allows for the unbiased computation of the effects as stipulated above.

The evaluators' task is to establish a logical series of steps that connect the data derived from the database and other sources to the chosen analytical models in a way to respond to all effects expected in the evaluation process:

- direct and indirect effects: expected and unexpected (positive and negative);
- secondary effects: expected and unexpected (positive and negative);
- synergies and transverse effects⁷².

These steps will be framed by the type and combination of measures and the availability of relevant context, input, output, and result data/information and their ability to appraise the programme results and impacts at later stages of RDP implementation. In some cases, quantitative data are available to estimate the RDP results. In other cases, missing quantitative data must be replaced with qualitative evidence. Moreover, the quantitative data should be triangulated with qualitative research, as it is proposed in these guidelines.

⁶⁹ The question is if the beneficiaries with completed projects are representative of the total.

⁷⁰ Capturing the success of your RDP: Guidelines for the ex post evaluation of 2007-2013 RDPs, PART II, Chapter 4, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

⁷¹ Due to these difficulties, the assessment of environmental effects was in many empirical studies restricted to the measurement aspect only. Indeed, the assessment of environmental results/impacts is still in its infancy and a wide range of possible methods (applicable in the field of socio-economic results/impacts) still wait for their empirical testing.

⁷² Also see chapter 5.1 of PART II in this document

An example of steps applicable in the assessment of RDP results is presented below:

Step 1: Introductory qualitative analysis

Introductory qualitative analysis (e.g. via application of theory of change) could be applied at the beginning of an evaluation process in order to reflect on various possible expected or unexpected, positive or negative effects of a RDP. If so, after linking it with the programme intervention logic, the evaluator may identify the magnitude and the scope of “observed” gross programme effects which at a later stage should be subject to more rigorous quantitative verification using causal inference approaches. The main aim of this introductory qualitative analysis is the prioritization of the different potential programme effects by indicating and selecting those which, in view of the RD stakeholders and policy makers, are expected to make the biggest “change” in comparison with a situation “without” the programme (in AIR 2017 to be reported at a micro-level only). Clearly, at the first stage of evaluation, the preliminary qualitative assessments are essential because they can provide invaluable insiders’ perspectives on a programme performance especially regarding its unintended results or direct and secondary effects. Stage 1 is also consistent with the approach recommended in order to develop PSEQ (see Chapter 5.2.3).

Step 2: Formulation of testable hypotheses regarding potential programme effects

At this stage CEQ and judgement criteria linked to the assessment of the expected results at a micro-level should be complemented with PSEQ and respective judgement criteria and result indicators, and testable hypotheses derived from Step 1.

Step 3: Testing and verification of hypotheses

Under this step, the above-mentioned hypotheses have to be tested and/or verified using methodological approaches based on causal inference. At this stage it is important to use previously collected result indicators to enable a robust analysis of the most important expected or unexpected, positive or negative and secondary programme effects grouped under economic, social and environmental domains. Such testing can only be carried out under rigorous evaluation methodologies inter alia using credible counterfactuals.

The main components of Step 3 are as follows:

- Defining the outcome variables (in economic, social and environmental domains). The analysis can be conducted with as many outcome variables as there are data for. The analysis can be extended not only to cover direct effects, but also secondary effects, and positive, negative, expected and unexpected effects.
- Defining the time dimension. For example, by comparing an average 2012-2013 (i.e. reflecting situation prior to the current programme) with 2016 (i.e. situation to be reported in AIR 2017).
- Applying suitable methodologies for finding credible control groups (a preference should be given to the application of both quantitative and qualitative methods). A suitable control group should be found for each type of effects analysed.
- Computing an average outcome effect for the group of programme beneficiaries.
- Computing an average outcome effect for a comparable control group.
- Calculating the expected or unexpected, positive or negative effect of the programme (for each outcome variable separately). Average Treatment Indicators (see definition below) can be applied to assess both programme direct as well as secondary effects – positive and negative.

Step 4: Calculation of specific programme indirect effects at micro-level⁷³

Indirect effects are usually a derived causal consequence of programme direct effects at the micro-level (e.g. deadweight loss, leverage effects, or substitution effects) and at the macro or regional level (e.g. multiplier effect, displacement effect, etc.). Both types of indirect effects can be intended or unintended.

Due to the focus of the AIR 2017, the analysis of the programme results covering the period 2014-2016 requires the calculation of programme indirect effects occurring at a micro-level only.

In general, a well-conceived evaluation design should be based on some verifiable hypothesis about whether programme indirect effects are present and why they exist. A typical feature of these types of effects is that they can make an overall economic, environmental and social effect substantially larger than direct effects alone. Yet, estimation and especially quantification of indirect effects may not be a trivial task (see the [Guidelines for the ex-post evaluation of 2007-2013 RDPs](#)). Methodologies recommended for estimating deadweight, leverage, and substitution effects are described and illustrated in several other publications.

Abundant micro-economic data and advanced evaluation methodologies are crucial elements in a robust quantitative analysis of programme indirect effects. Additionally, it is recommended to complement quantitative estimates with qualitative information, e.g. by carrying out additional surveys focusing on specific issues (e.g. environmental), or interviewing programme non-beneficiaries that feel affected by the RDP implementation in order to find out a correct reference for a quantitative analysis.

Step 5: Identifying programme net-results

Programme results should be expressed in “net” terms, which means after subtracting the effects that cannot be attributed to the intervention, and by taking into account indirect effects (deadweight loss, leverage, substitution, etc.). The evaluator should specify in detail what elements have been accounted for when calculating programme “net” effects.

Evaluation of indirect effects is complex and requires advanced methods that can capture them. Most important indirect effects are: multiplier effect, substitution effect and displacement effect (also see the Chapter 5 of PART II and tool for qualitative appraisal of RDP operation's primary and secondary contributions in PART III).

Considering secondary contributions of operations in calculating result indicators

When calculating result indicators (complementary result indicators or target indicators) secondary contributions should be considered. The definition of secondary contributions can be found in chapter 5.1 of PART II of these guidelines.

Secondary contributions to the value of each result indicator are estimated by evaluators using samples of beneficiaries who have flagged in the RDP operations database additional contributions of operations to focus areas other than those under which they have been programmed. Only significant additional contributions should be considered. For example, the beneficiaries of a given operation under measure 4, which is programmed under FA 2A, flag in the application form/payment request an additional contribution to FA 5B. The evaluator shall use the sample of those beneficiaries for the survey, assessing to what extent the operations implemented have contributed to the competitiveness of farms and to the energy efficiency. Survey results should be aggregated and extrapolated and used to estimate the additional contribution at the level of FA when calculating the respective result indicator.

⁷³ Here, we focus on reporting in AIR 2017, i.e. on effects occurring at micro-level only. In the real world, secondary effects occur also at regional- and macro-level and point out how positive or negative effects originated from a given programme are transmitted through and/or across a given territory.

7.3 Judging

The task of the judging phase is to interpret the evaluation findings, formulate answers to EQ on the basis of judgement criteria and indicators. The conclusions and recommendations relate to the effects of single focus areas, as well as the programme as a whole.

Recommended working steps:

- Answer the EQ to be reported in the AIR 2017,
- Judge on expected results and identify the internal and external factors which contribute to the success or failure of the programme objectives,
- Draft conclusions as well as recommendations.

Expected outcome: Answers to EQ, conclusions and recommendations

Answers to EQ must be based on sound evidence and accompanied by a critical discussion of such evidence. In this sense, evaluators shall discuss and interpret the qualitative information and data values for common and programme-specific indicators obtained from the analysis. Moreover, the evaluators need to consider the context (e.g. socio-economic situation, capacity of beneficiaries to co-finance projects). For example, where net values of indicators are very low or negative, a proper explanation shall be provided. Or, if certain focus areas or a part of the programme has not delivered the expected results and impacts, an analysis of the reasons for this unexpected effect is necessary. In this respect, the evaluator will have to:

- Answer the EQ to be reported in the AIR 2017, taking into account the context. In the answer to these EQ the effectiveness and efficiency principles (obtained results and resources spent) should always be present.
- While answering, the evaluator has to carefully judge to which extent the programme contributes to achieving the expected results and identify the internal and external factors which contribute to the success or failure of the programme objectives.
- Draft conclusions and recommendations appropriately substantiated by the findings.

In case the findings have limitations in their validity (e.g. in case of insufficient RDP uptake, or use of naïve evaluation methods), the resulting judgement should be critically reflected.

During the judging phase, the evaluator also draws the conclusions and recommendations to improve the strategy (design and implementation of RDP). The conclusions and recommendation should be based on the collected evidence and its robust interpretation and should be rooted in the answers to EQ.

The answers to EQ, together with the set of conclusions and recommendations, are the core messages to be integrated in the AIR 2017.

Further reading

Capturing the success of your RDP: Guidelines for ex post evaluation of 2007-2013 RDPs, PART II, Chapters 2, 3, 4 and 5, http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html
Handbook of the Common monitoring and evaluation framework 2007-2013 and Annexes, Annex 1, Guidance note B – Evaluation guidelines, http://ec.europa.eu/agriculture/rurdev/eval/index_en.htm

PART III - ANNEXES

ANNEX 1 – PROPOSED SFC TEMPLATE FOR POINT 7 OF THE AIR SUBMITTED IN 2017

The AIRs are submitted by Member States to EC via the electronic system for exchange of information, using the SFC template. In 2017 the AIR will also contain in point 7 the information resulting from evaluation activities, including the quantification of programme achievements through the assessment of the result indicators (in particular complementary result indicators) and the answers to relevant evaluation questions⁷⁴. For this reason, the SFC template of AIR 2017 will include in point 7 tables prepared separately for:

- d) each of the focus area-related common evaluation questions (CEQ), number 1 – 18,
- e) common evaluation questions related to other aspects of RDP
 - TA and NRN, CEQ number 20 and 21
 - Programme synergies, CEQ number 19
- f) programme-specific evaluation questions (PSEQ) linked to the assessment of:
 - programme specific focus areas (to be multiplied in accordance to number of specific focus areas included in the RDP)
 - RDP specific topics (e.g. assessment of delivery mechanism)

The information on evaluation collected via tables of point 7 of the SFC template summarises key issues important for Member States and the EU⁷⁵. The SFC template of point 7:

- translates and operationalizes CMES requirements in a feasible format.
- helps to avoid formal mistakes (e.g. wrong use of indicators; inconsistency of intervention logic with evaluation elements; data-sources, mix up of gross and net values etc.).
- facilitates judgement on the validity and reliability of the assessment of RDP results and the robustness and rigor of the answer to evaluation question
- encourages stakeholders to provide synthetic core-information emerging from more extensive evaluation reports, which can be used in user-friendly formats for policy makers and a wider audience.
- helps to collect and synthesize large amount of information at the EU level in a transparent way.
- provides a solid basis of information for EU-level synthesis and for CAP-level reporting on the achievements and results acts of rural development policy.

Methodological fiches how to fill the tables of the SFC template for point 7 are provided in a separate annex.

The table below shows an empty example of point 7 of the SFC template for answering common evaluation question number 4. This example has been prepared by the European Evaluation Helpdesk for Rural Development in consultation with the DG Agri services and its structure and content has been discussed with Member States during the 8th and 9th meeting of the Expert Group on Monitoring and Evaluating the CAP and during the 28th and 29th meetings of Rural Development Committee.

Table 7. SFC template for point 7 on Common Evaluation Questions no.4 (empty template)

<p>COMMON EVALUATION QUESTION No 4: “To what extent have RDP interventions contributed to improving the economic performance, restructuring and modernization of supported farms in particular through increasing their market participation and agricultural diversification?”</p>
<p>1. Intervention logic of the FA 2A</p> <p>Primarily programmed measures/sub-measures:</p>

⁷⁴ Annex VII, Commission Implementing Regulation (EU) No 808/2014

⁷⁵ This information does not replace the full evaluation report, which can be still developed in Member States but is not attached to the SFC template.

Measures programmed under the other FAs which show secondary contributions to FA 2A⁷⁶: <i>[A maximum of 1,000 characters = approx. ¼ page – Mandatory]</i>		
2. Link between judgment criteria, common and additional⁷⁷ result indicators used to answer the CEQ <i>[Mandatory]</i>		
Judgment criteria⁷⁸ <i>[Max. 255 characters]</i>	Common result indicators⁷⁹ <i>[Max. 255 characters]</i>	Additional result indicators⁸⁰ <i>[Max. 255 characters]</i>
3. Methods applied <ul style="list-style-type: none"> • Quantitative methods: <ul style="list-style-type: none"> i. Reasons for using the method ii. Description of methods to calculate ratios, gross or net (if applicable) values of common and additional result indicators, or other indicators used (output, common context indicators)⁸¹ iii. Challenges in the application of described methods and solutions encountered • Qualitative methods: <ul style="list-style-type: none"> iv. Reasons for using the method⁸² v. Description of methods used⁸³ vi. Challenges in the application of described methods and solutions encountered <i>[A maximum of 7,000 characters = approx. 2 pages – Mandatory]</i>		
4. Quantitative values of indicators⁸⁴ and data sources <i>[Mandatory]</i>		

⁷⁶ This also covers those measures/sub/measures which have shown the secondary effects during the evaluation not only those who have been planned to contribute additionally to focus areas others than programmed during the programme design

⁷⁷ Additional indicators are used if the common indicators are not sufficient to answer the CEQ, and if the success is specified with judgment criteria which are not measured with the common indicators.

⁷⁸ List judgment criteria used to specify the success of the intervention within the FA. The judgment criteria proposed by the WD: "Common Evaluation Questions for rural development programmes 2014-2020" can be used. Stakeholders in MS can propose their own judgment criteria to specify success in line with the specific intervention logic of the FA. In the system for answering the CEQ 1 – 21 one judgment criterion is linked to one result indicator (common or additional).

⁷⁹ List common result indicators used in answering the CEQ. Ensure that the indicators are listed in a consistent way with the judgment criteria and placed in the same line.

⁸⁰ List additional indicators used in answering the CEQ. Ensure that the judgment criteria are listed in a consistent way with the additional indicators and placed in the same line.

⁸¹ In case the same method was used for several indicators, the list of these indicators should be provided.

⁸² Explain the reasons why qualitative methods have been used to assess the RDP results linked to FA 2A - introductory qualitative analysis, triangulation of quantitative findings, no data available (e.g. no RDP uptake under the FA 2A), etc.

⁸³ In case the same method was used for several indicators, the list of these indicators should be provided.

⁸⁴ Values are based on the quantitative methods described above.

Indicator		Absolute value ⁸⁵	Ratio value ⁸⁶	Calculated gross value ⁸⁷	Calculated net value ⁸⁸	Data and information sources ⁸⁹
Common output indicators ⁹⁰						
Common result indicators	R1 – Percentage of agricultural holdings with RDP support for investments in restructuring or modernisation					
	R2 – Change in agricultural output on supported farms ⁹¹					
	R2 – AWU					
Additional result indicators ⁹²						
Common Context indicators ⁹³						
5. Qualitative findings⁹⁴ <i>[A maximum of 1,750 characters = approx. ½ page – Non-mandatory]</i>						
6. Problems encountered influencing the validity and reliability of evaluation findings⁹⁵ <i>[A maximum of 1,750 characters = approx. ½ page – Non-mandatory]</i>						

⁸⁵ Value is aggregated from Pillar II operations database for output indicators, and/or from statistics databases for common context indicators or additional indicators, if they are used to answer the common evaluation questions.

⁸⁶ This column is filled for the following result indicators: R1, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R16, R17, R20, R23 and R25.

⁸⁷ The gross value of following result indicators is inserted here: R2, R13, R14, R15, R18, R19, R21 and R22. The gross value of used common context indicators is inserted here as well, if relevant. The gross value of indicator is inserted also in case the net value has been inserted in the table.

⁸⁸ The net value of following result indicators (in case it was possible to calculate it) is inserted here: R2, R13, R14, R15, R18, R19, R21 and R22. The net value of used common context indicators is inserted here as well, if relevant.

⁸⁹ All data and information sources used for calculation of indicators should be listed here, e.g. Pillar II operations database, EU/national/regional statistics, GIS, etc.

⁹⁰ The common output indicators can be also used, especially if the value of output indicator is necessary to calculate the result indicator or in case it provides important information for answering the evaluation question, e.g. number of holdings/operations supported (O3 and O4), physical areas supported (O6), number of livestock units supported (O8), etc. The selection of output indicators for answering the evaluation question is done in MS.

⁹¹ Values of common complementary result indicators are collected separately for numerator and denominator

⁹² Additional indicators are optional and developed in MS to answer common evaluation questions in case the common indicators are considered insufficient for this purpose.

⁹³ The common context indicators can be also used to answer common evaluation questions if relevant (e.g. CCI 14, CCI 17, CCI 26).

⁹⁴ Findings from the qualitative assessment are summarised here.

⁹⁵ Explain problems faced by the evaluator during the assessment, e.g. data and information availability, timing and coordination issues, etc. which might have threatened the reliability and validity of evaluation findings.

7. Answer to evaluation question⁹⁶ <i>[A maximum of 10,500 characters = approx. 3 pages – Mandatory]</i>	
8. Conclusions and recommendations⁹⁷ <i>[Mandatory]</i>	
Conclusion <i>[Approx. 1,000 characters]</i>	Recommendation <i>[Approx. 1,000 characters]</i>
C.1	R.1
C.2	R.2
C.3	R.3

⁹⁶ On the basis of the evidence resulting from the analysis presented under points 4 and 5.

⁹⁷ On the basis of the information collected under points 4 and 5 and the answer given under point 7.

Table 8. SFC template for point 7 on Common Evaluation Questions no.4 (filled example)

The table below shows a filled example of point 7 of the SFC template for answering common evaluation question number 4.

COMMON EVALUATION QUESTION No 4: “To what extent have RDP interventions contributed to improving the economic performance, restructuring and modernization of supported farms in particular through increasing their market participation and agricultural diversification?”

1. Intervention logic of the FA 2A

Primarily programmed measures/sub-measures:

M4.1, M4.3, M6.3, M1.1, M2.1, M16.2

Tab.1: The level of uptake per primarily programmed measures (till end 2015)

Measure	2014	2015	2016	Total no. of operations
M 4.1	45	160	90	295
M 4.3	-	25	25	50
M 6.3	-	-	68	68
M 1.1	-	10	15	25
M 2.1	-	250	300	480
M 16.2	-	-	0	0
Total	45	445	498	918

Measures programmed under the other FAs which show secondary contributions to FA 2A⁹⁸:

FA 3A: M4.2, M9

FA P4: M9, M10, M11, M13

FA 5B: M4.1

FA 5C: M4.1

Tab. 2: The level of uptake per measures with secondary contributions to FA 2A (till end 2015)

Measure	2014	2015	2016	Total no. of operations
M 4.1 (5B)	-	35	40	75
M 4.1 (5C)	-	26	25	51
M 4.2 (3A)	35	105	68	208
M 9 (3A)	-	14	25	39
M 10 (P4)	240	280	300	300
M 11 (P4)	45	96	180	180
M 13 (P4)	350	360	375	375
Total	665	916	1 013	1 228

Only finalised projects/operations were considered in above tables. The level of uptake and amounts paid under operations are important to consider the use of advanced evaluation methods for the most important measures/sub-measures.

[A maximum of 1,000 characters = approx. ¼ page – Mandatory]

⁹⁸ This also covers those measures /sub-measures which have shown secondary effects during the evaluation not only those who have been planned to contribute additionally to focus areas other than those programmed during the programme design.

2. Link between judgment criteria, common and additional⁹⁹ result indicators used to answer the CEQ

[Mandatory]

Judgment criteria ¹⁰⁰ [Max. 255 characters]	Common result indicators ¹⁰¹ [Max. 255 characters]	Additional result indicators ¹⁰² [Max. 255 characters]
Due to RDP support Farms have been modernised and restructured (WD CEQ for RDPs 2014-2020)	R1 - % of agricultural holdings with RDP support for investments in restructuring or modernisation	
Agricultural output per AWU of supported agricultural holdings has increased due to implementation of above RD measures (WD CEQ for RDPs 2014-2020)	R2 – net Change in agricultural output on supported farms/AWU (Complementary result indicator)	
Farms' economic performance has improved due to implementation of above RD measures (RDP specific)		Net change of Gross Farm Income (RDP specific)
Farms' market participation has increased due to implementation of above RD measures (RDP specific)		Net change of Sales/total output (RDP specific)
Farms' agricultural diversification has increased due to implementation of above RD measures (RDP specific)		Net change of % of Agricultural income out of total income (RDP specific)

3. Methods applied

(a) Quantitative methods: Counterfactual analysis, combined PSM and DID

i. Reasons for using the method

The main reason was its flexibility and possibility to draw on existing secondary data and using them to verify causality of the identified primary and secondary contributions of supported operations. The counterfactual analysis using PSM+DID allows netting out the measure effects and calculate the extent of the key indirect effects of interventions at the level of direct beneficiaries (deadweight loss and leverage effects), which have been often proved in linkage with the investment support.

⁹⁹ Additional indicators are used if the common indicators are not sufficient to answer the CEQ, and if the success is specified with judgment criteria which are not measured with the common indicators.

¹⁰⁰ List judgment criteria used to specify the success of the intervention within the FA. The judgment criteria proposed by the WD: "Common Evaluation Questions for rural development programmes 2014-2020" can be used. Stakeholders in MS can propose their own judgment criteria to specify success in line with the specific intervention logic of the FA. In the system for answering the CEQ 1 – 21 one judgment criterion is linked to one result indicator (common or additional).

¹⁰¹ List common result indicators used in answering the CEQ. Ensure that the indicators are listed in a consistent way with the judgment criteria and placed in the same line.

¹⁰² List additional indicators used in answering the CEQ. Ensure that the judgment criteria are listed in a consistent way with the additional indicators and placed in the same line.

ii. Description of methods to calculate ratios, gross or net (if applicable) values of common and additional result indicators, or other indicators used (output, common context indicators)¹⁰³

In case of indicator R1, the value is taken from PA operational database (indicator O4 cumulative, data reported when the operation is completed) and ratio (%) is calculated from the total number of agricultural holdings in base year for the RDP area (Eurostat: FSS).

The calculation of indicator R2 and additional indicators was based on the operations of listed measures/sub-measures which had a reasonable uptake – M2.1, M4.1, M4.3, M6.3. Rest of the measures which had limited or no uptake till 31.12.2016 have not been taken into consideration. The operations of sub-measure M4.1 programmed under the FA 5B and FA 5C, as well as the measures programmed under 3A and P4, with the secondary contributions to the FA 2A have also been taken into consideration in calculation of R2.

Assessment was conducted in 2 stages:

Stage 1: Estimation of primary contributions of RDP measures directly attributable to the FA 2A (see Tab. 1, except M1.1 and M2.1).

Both, sample of farms supported in a given period (2014-2016) under measures listed in Tab.1 (participants) and samples of farms which didn't received support from respective measures (non-participants), have been searched from the National Database/FADN for values on agriculture output, labour inputs - AWU and other above result indicators. In order to increase the validity of the selection process some additional farm characteristics were used to achieve the best possible similarity of both samples. PSM was applied to identify a suitable "control group" from the sample of non-participants.

Average values of R2 indicator were computed for the group of participants and "control group" prior to the support (year 2013) and after support (year 2016). Calculation of Average Treatment Effects on Treated (ATT), using R2 indicator and net effect on R2 (by combining calculated ATTs with DID method). Finally, the aggregated value of the net indicator at a programme area was calculated by multiplying average micro-results computed at a farm level by a number of supported farms (extrapolation).

Stage 2: Estimation of secondary contribution of those measures whose main objective is linked to another FA (see Tab. 2, e.g. farm infrastructure investment, renewable energy, energy efficiency) but which are also expected to have an effect on farm restructuring and competitiveness and therefore on farm labour productivity (R2).

The steps in preparing group samples for statistical matching and methods of calculations are the same as described under Stage 1. Samples of beneficiaries implementing the operations under M4.1 programmed under the FA 5B, FA 5C, M4.2 and M9 under FA 3A have also been taken into consideration.

Stage 3: Estimation of indirect effects

Deadweight loss and leverage effects were calculated under M 4.1. only, by using data for identical comparable farms (e.g. similar size, location, specialization, agriculture output, employment etc.) and applying a proxy indicator - value of investments per farm for calculations of ATT (= average treatment on treated) combined with DID method to obtain figures for Deadweight loss. Same steps, but applying a proxy indicator - money transfers from farm to other business-related spending (e.g. property or land purchase, new buildings etc.), were repeated for calculation of Leverage effect, measured among direct beneficiaries (micro-level approach).

iii. Challenges in the application of described methods and solutions encountered

The key challenge was the preparation of non-participant group sample with the highest possible similarity with the group of participants, in order to establish a credible uncontaminated control group (elimination of selection bias). It is demanding for data classification and filtering due to fact, that the supported and control groups cannot be differentially exposed to other interventions.

Evaluator was unable to perform the calculation of R2 indicator values for the P4 area-based measures (M9, M10, M11, M13), where the appropriate matching of group samples would require additional data on environmental characteristics and parameters of location (altitude, landscape profile, soil structure, soil quality, soil erosion, use of fertilizers and pesticides, etc.) which were not available. In case of M13 there is an extra problem of identification of sample of non-participants,

¹⁰³ In case the same method was used for several indicators, the list of these indicators should be provided.

due to fact that great majority of farms located in NCA belong to participants and to compare them with other farms at lowlands would give unreal results. To encounter a problem the MA/PA will be instructed to expand the LPIS by appropriate GIS module with data layers to be fed from autonomous systems of several state institutions.

(b) Qualitative methods:

i. Reasons for using the method¹⁰⁴

Our choice was to complement quasi-experimental evaluation with the qualitative information collected via farm survey and stakeholder focus group discussions and use them to support findings from quantitative analysis, reasonably explain what worked better, why and how and also provide a useful bottom-up understanding of causal patterns related to intervention. At the final stage qualitative data allowed for more comprehensive and still user friendly formulation of findings to answer the CEQ, as well as conclusions and recommendations.

ii. Description of methods used¹⁰⁵

Qualitative techniques were used at initial stage of designing evaluation, for introductory qualitative analysis to develop a hypothesis how interventions logic worked and to clarify research questions that needed quantitative analysis. Qualified stakeholders and informants of MA, PA were interviewed for these reasons.

Qualitative data were used to assess the contribution of M1.1 and M2.1 to FA 2A, due to missing quantitative data providing a tangible link between knowledge transfer and change of the agriculture output needed for calculation of R2. Data were obtained during 3 sessions of specialised focus group of 9 selected participants from various direct and indirect participants of the training (M1) and advisory services (M2), representing the clients and also providers of knowledge. Compared to qualitative surveys, the focus group method proved to be a core element of obtaining qualitative information (also for other FA), securing ongoing dialog, learning process for stakeholders, stimulating their activation and chances to bring about new ideas and innovations.

Qualitative research was used to verify theory of change assumed under the intervention logics of several measures and to explain why intervention works for some farms or areas, but less for the others.

After computing of ATT and DID for R2, qualitative data helped to improve the triangulation of quantitative findings and assured more consistent formulation of achieved results.

iii. Challenges in the application of described methods and solutions encountered

Challenges in using the Focus group method lied with the limited choices of the right stakeholder representatives (personal bias, precondition of good theoretical and practical knowledge of RDP intervention logics and use of indicators), preparation of group sessions (questions, activities). In surveys, the main challenge was the elaboration of well-designed questionnaires.

[A maximum of 7,000 characters = approx. 2 pages – Mandatory]

4. Quantitative values of indicators¹⁰⁶ and data sources

[Mandatory]

¹⁰⁴ Explain the reasons why qualitative methods have been used to assess the RDP results linked to FA 2A - introductory qualitative analysis, triangulation of quantitative findings, no data available (e.g. no RDP uptake under the FA 2A), etc.

¹⁰⁵ In case the same method was used for several indicators, the list of these indicators should be provided.

¹⁰⁶ Values are based on the quantitative methods described above.

Indicator		Absolute value ¹⁰⁷	Ratio value ¹⁰⁸	Calculated gross value ¹⁰⁹	Calculated net value ¹¹⁰	Data information sources ¹¹¹ and
Common output indicators ¹¹²	O3 – No. of operations supported (primary/secondary measures)	918 / 1 228				Pillar II operations database
	O4 – no. of holdings supported (M4, 6, 11, 13)	1 350				Pillar II operations database
Common result indicators	R1 – percentage of agricultural holdings with rdp support for investments in restructuring or modernisation		14,7			Pillar II operations database
	R2 – change in agricultural output on supported farms ¹¹³			€ 56,000	€ 35,000	National database of farms/FADN
	R2 / AWU			€ 7,200	€ 6,000	National database of farms/FADN
Additional result indicators ¹¹⁴	Gross Farm Income			€ 126,700	€ 97,240	National database of farms/FADN
	Sales/total output		42	€ 100,000 / 237,951		National database of farms/FADN
	Percentage of agricultural income out of total income		86			National database of farms/FADN

¹⁰⁷ Value is aggregated from Pillar II operations database for output indicators, and/or from statistics databases for common context indicators or additional indicators, if they are used to answer the common evaluation questions.

¹⁰⁸ This column is filled for the following result indicators: R1, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R16, R17, R20, R23 and R25.

¹⁰⁹ The gross value of following result indicators is inserted here: R2, R13, R14, R15, R18, R19, R21 and R22. The gross value of used common context indicators is inserted here as well, if relevant. The gross value of indicator is inserted also in case the net value has been inserted in the table.

¹¹⁰ The net value of following result indicators (in case it was possible to calculate it) is inserted here: R2, R13, R14, R15, R18, R19, R21 and R22. The net value of used common context indicators is inserted here as well, if relevant.

¹¹¹ All data and information sources used for calculation of indicators should be listed here, e.g. Pillar II operations database, EU/national/regional statistics, GIS, etc.

¹¹² The common output indicators can be also used, especially if the value of output indicator is necessary to calculate the result indicator or in case it provides important information for answering the evaluation question, e.g. number of holdings/operations supported (O3 and O4), physical areas supported (O6), number of livestock units supported (O8), etc. The selection of output indicators for answering the evaluation question is done in MS.

¹¹³ Values of common complementary result indicators are collected separately for numerator and denominator.

¹¹⁴ Additional indicators are optional and developed in MS to answer common evaluation questions in case the common indicators are considered insufficient for this purpose.

Common Context indicators 115	CCI 17 - Total number of agricultural holdings	9,205				EU database of CCI
<p>5. Qualitative findings¹¹⁶</p> <p>According to the farm survey 75% of beneficiaries of measures/sub-measures under FA 2A, about 80% of beneficiaries of measures/sub-measures under FA 5B and 50 % of beneficiaries under FA 5C would have invested anyway, even without the RDP support. The similar result (70 %) in case of M4.1 was calculated by using earlier described quantitative methods (PSM+DID). Rate of deadweight will be re-calculated and verified in AIR 2019.</p> <p>From the surveys and Focus group interviews there is obvious an increasing number of farms acquiring modern buildings, energy-efficient new technology and technical equipment, installing ICT in order to increase efficiency by reducing production costs and minimizing losses. However, the prioritizing of support towards operations focussed on new products or improving the product quality and innovations should have been more unambiguous.</p> <p>It also highlights improvement in working and animal welfare conditions, and improvement of the effectiveness of traditional farming systems as being important in parallel with social and environmental benefits. At the same time a negative trend was observed, when almost 2 years after the investment support approximately a third of supported dairy farms did not increase neither agriculture output nor gross farm income and total sales. It is obviously caused by other external factors outside the RDP (e.g. price volatility, overcapacity, etc.).</p> <p>Focus group on knowledge transfer provided information on limited value added of about 50% of training events (M1.1) and 35% advisory activities (M2.1) for the participating stakeholders. Proposed to strengthen the MA capacities in order to perform rigorous training needs assessment among major stakeholder groups and obligatory quality assessment in terms of the knowledge content and modes of its delivery, but also some screening of participants in terms of their adequacy to avoid hunting for participant-days among providers.</p> <p><i>[A maximum of 1,750 characters = approx. ½ page – Non-mandatory]</i></p> <p>6. Problems encountered influencing the validity and reliability of evaluation findings¹¹⁷</p> <p>Findings of the quantitative analysis are robust and reliable since majority of data comes from National agriculture database (not surveyed farmers) where they regularly send data from their annual accounts. Used data on RDP beneficiaries are from the Pillar II operation database.</p> <p>For CCI 17 the data from national statistics of 2015 has been used, since the data for 2016 did not exist at the time of evaluation. It means that R1 indicator could be biased by that.</p> <p>As mentioned under challenges, Evaluator made a great effort to support credibility of findings from the surveys and Focus group interviews, by limiting the one-way or subjective opinions of the surveyed farmers or the representatives of stakeholders.</p> <p><i>[A maximum of 1,750 characters = approx. ½ page – Non-mandatory]</i></p> <p>7 Answer to evaluation question¹¹⁸</p> <p>In the “programme area” there are 7,250 holdings in the agricultural sector eligible for the RDP support. Out of these, 1,350 have been supported for investments under the FA 2A until 31.12.2016. Total number of operations under primarily programmed measures is 918 (Tab.1), in addition there were 1,228 supported operations under measures providing secondary contributions to FA 2A,</p>						

¹¹⁵ The common context indicators can be also used to answer common evaluation questions if relevant (e. g. CCI 14, CCI 17, CCI 26).

¹¹⁶ Findings from the qualitative assessment are summarised here.

¹¹⁷ Explain problems faced by the evaluator during the assessment, e.g. data and information availability, timing and coordination issues, etc. which might have threatened the reliability and validity of evaluation findings.

¹¹⁸ On the basis of the evidence resulting from the analysis presented under points 4 and 5.

namely FA 3A with 247 supported operations, FA 5B 75 supported operations, FA 5C with 51 supported operations and FAs of P4 with 855 supported operations.

This means altogether 14,7 % of eligible agricultural holdings have received support for investments in agriculture, which has enhanced the modernisation of supported holdings, mainly in the area of dairy cattle and poultry production and in the fruit and vegetables sectors. The majority of invested funds of M4 (about 70%) went to construction and remaining (30 %) to new machinery and equipment. These improvements lead to better competitiveness and therefore higher gross output and gross value added at the micro and macro level. At the same time there were investment operations creating overcapacities instead of structural adjustments and diversification (e.g. medium and larger dairy farms). The investment support was not clearly enough focused on development of new products or improving the product quality and innovations, these are the elements facilitating the shift towards agriculture diversification.

Majority of investment operations under M4 brought also environmental benefits of higher energy efficiency, reduced GHG production, more accurate application of pesticides/fertilizer, improved animal welfare and other social benefits through improved employees working conditions.

Economic performance of all above supported farms expressed in *agricultural output/AWU* (R1) has increased in gross values from € 29,000/AWU in 2013 to € 35,000/AWU in 2016, thus the change is € 6,000/AWU. Secondary contributions of operations implemented under the FA 5B and 5C (M4.1) to this value was estimated at 15 %.

The value of additional indicator - *gross farm income* has increased in average too, from € 81,300 in 2013 to € 97,240 in 2016. The share of secondary contributions of operations implemented under M4.1 of FA 5B and 5C to the change in value of indicator was 7% for M4.1 under the FA 5B, mainly due to energy savings.

Similar situation was in *sales on total output* of all supported farms, which has increased from 40 % in 2013 to 42% and has reached the share of € 100,000 of € 237,951 in 2016). Secondary contributions to the value of indicator under the FA 5B and 5C have been identified at 9% for M4.1. This shows the improved position on markets of those farms supported by M4.1 under FA 2A, 5B and 5C.

The % of *agriculture income on the total income* stays quite stable and high, and shows the low willingness of supported holdings to diversify their economic activities, this especially implies to dairy farms.

As for the indirect effects of the support, we have to highlight quite high deadweight effects: (75%) with primary beneficiaries under FA 2A, 80% with secondary beneficiaries under FA 3A and 5B, 50% with secondary beneficiaries under FA 5C.

Leverage effect measured by money transfers from farm to other business-related spending (e.g. property or land purchase, new buildings) was estimated at 13% with the primary beneficiaries under FA 2A.

[A maximum of 10,500 characters = approx. 3 pages – Mandatory]

8. Conclusions and recommendations¹¹⁹

[Mandatory]

Conclusion [Approx. 1,000 characters]	Recommendation [Approx. 1,000 characters]
C.1 The support has increased the economic performance of farms (increased agriculture outputs per AWU by € 6,000), as well as their modernisation and market participation, however at the cost of quite heavy deadweight effect.	R.1 By canalizing the given investment support for more financially reliant, e.g. smaller agricultural holdings the deadweight effect would decrease and overall benefits delivered would increase. The investment support should also be more accessible to new entrants into agriculture especially the young farmers, their investment needs were underestimated and outshout by existing holdings.

¹¹⁹ On the basis of the information collected under points 4 and 5 and the answer given under point 7.

<p>C.2 The support was accessible to quite a limited number of farms (14,7%) and has had only a limited influence on the overall sector and largely desired agricultural diversification.</p>	<p>R.2 Better targeting of the support toward the smaller farms or more limited rate of support for large farms could favour the diversification. To adjust the project selection criteria under investment measures so that the agricultural diversification will be supported to a larger extent (e.g. favour holdings which transparently outline their plans for diversification of agricultural production). The capacity adjustment effects of investment support should be of higher importance.</p>
<p>C.3 After an investment support farms in average became more productive, which is one goal of RDP. More productive firms usually produce more output which is available on the market and used as input for other sectors and induces value added. On contrary, several beneficiaries were identified with decreased sales and outputs after investment support.</p>	<p>R.3 New product development, innovations and capacity adjustment effect of potential beneficiaries should be more of interest while targeting investment support. They need to be more claimable and reflected in the project selection criteria.</p>
<p>C.4 Support contributed to the negative culture in many agricultural holdings of postponing or even cancelling own investments, while waiting for possible support measures to come, irrespective of the genuine needs, optimal timing of new products (e.g. linked to recent market developments).</p>	<p>R.4 Better design of measures/programme to avoid wasteful interventions, especially those increasing farm property without increasing its overall efficiency, with a heavy deadweight effects or without a clear market potential. To increase effectiveness of allocating funds towards the genuinely targeted groups (e.g. back checking of supported holdings).</p>

ANNEX 2 - ELEMENTS OF THE CMES AND THEIR USE IN EVALUATION

CMES element (Art.14 of 808/2014, point 1)	Short description	Use in the RDP assessment
Intervention logic	RDP intervention logic is composed of common EU and RDP specific policy objectives, which relate to RD priorities, focus areas and measures/activities, inputs (financial allocations) and expected outputs, results and impacts ¹²⁰ . The composition of the intervention logic should be coherent and relevant , reflecting the analysis of the territory (SWOT) and needs assessment.	<p>The RDP specific intervention logic is basis for setting the evaluation framework (common and programme specific evaluation questions and indicators) and for the assessment of the effectiveness, efficiency, and the RDP's results, impacts and relevance.</p> <p>The intervention logic is also the ground for application of theory of change as one the evaluation approaches.</p>
Set of common context, result and output indicators	<p>Common indicators¹²¹ for rural development are tools, set up at the EU level to describe and analyse the programme's territory or to measure rural policy effects. They relate to policy objectives. The common indicators set contains context (and impact), output and result indicators.</p> <p>Performance framework (PF) indicators¹²² relate to performance reserve, which is set up for each RDP and can be allocated only if milestones set up for PF indicators have been achieved. Performance framework (PF) indicators proposed by legal framework are output level indicators.</p>	<p>Common contexts indicators are applied to describe and analyse the programme's territory. Their values are used also in the SWOT analysis and as justification in the needs assessment.</p> <p>Programme authorities might decide to develop and use programme specific context indicators, if the common ones are not sufficient to describe the RDP territory.</p> <p>Common context indicators also contain the RD related impact indicators, which are used to measure the RDP effects at the level of programme.</p> <p>Common result indicators are used to measure RDP results within the group of programme's beneficiaries.</p> <p>Common output indicators are used to measure direct policy outputs at the measure level.</p> <p>Programme authorities might decide to develop and use programme specific output, result and impact indicators, if the common set does not capture all programme effects.</p>

¹²⁰ Working Paper/UPDATED Guidelines for strategic programming for the period 2014-2020

¹²¹ Working document: Target indicators fiches for Pillar II (Priority 1 – 6), Working document: Complementary result indicators fiches for Pillar II, Working document: Context indicators fiches, Working document: Impact indicators fiches and Working document: Defining proxy indicators for rural development programmes

¹²² Guidance fiche Performance framework review and reserve in 2014 – 2020, http://ec.europa.eu/fisheries/reform/emff/doc/09-performance-framework_en.pdf

CMES element (Art.14 of 808/2014, point 1)	Short description	Use in the RDP assessment
		As for the performance framework (PF), MS can use the pre-defined PF indicators, or decide to define the own set.
Common evaluation questions	Common evaluation questions ¹²³ (CEQ) are an important element of the common monitoring and evaluation system for rural development. They are formulated by the EC and they relate to policy objectives at the focus area level, horizontal policy objectives level (EU and CAP) and to other RDP aspects, such as TA or NRN. CEQ are further specified with the proposed judgment criteria and linked to common indicators.	CEQ define the focus of evaluations in relation to policy objectives and help to demonstrate the progress, impact, achievements, effectiveness, efficiency, and relevance of rural development policy. CEQs are answered with the help of indicators, which should be consistent with the CEQ' judgement criteria and allow for evidence based answers.
Data collection, storage and transmission	Data ¹²⁴ are the core of evaluation. There are primary data collected at the single holding level and secondary data providing the aggregated information. On RDP beneficiaries shall provide MA and evaluators with all information important to conduct monitoring and evaluation. ¹²⁵ Data on RDP's beneficiaries are collected via monitoring system. Data shall be collected, stored and transmitted electronically ¹²⁶ .	Data are used in evaluation to calculate indicators' values, collect evidence to answer the evaluation questions and provide robust judgments. Data for evaluation are collected electronically on beneficiaries (for each selected and completed operation via monitoring system) and non-beneficiaries (various statistics, other databases, and sampling) to construct counterfactual, if possible.
Regular reporting on M&E activities	MS shall report on M&E in Annual implementation reports (AIR) ¹²⁷ each year, starting in 2016 until 2024.	The regular AIR reports on M&E activities in the chapter 2, which also relates to other chapters, mainly 1 and 3. In the enhanced AIR submitted in 2017 and 2019, also chapters 6, 7, 8, 9 and 10 relate to reporting on M&E activities. Apart of reporting on M&E activities in AIR it is advisable to produce shorter versions of reports to inform and discuss with various audiences the evaluation activities and in 2017 and 2019 also the evaluation results.
Evaluation Plan	Evaluation plan (EP) is the document, which helps to prepare the programme's evaluation for the entire programming period. Since it is part of the RDP, any changes must be the subject of formal	The EP outlines the objectives of the evaluation, sets up the framework for management and governance of the evaluation, including stakeholders' involvement. EP also proposes the evaluation topics and activities, in particular

¹²³ Working paper: Common evaluation questions for rural development programmes 2014-2020, <https://enrd.ec.europa.eu/en/evaluation-helpdesks-publications>

¹²⁴ Working document: Rural development monitoring – implementation report tables, Working document: Rural development programming and target setting, Data item list for Pillar II, Operations database

¹²⁵ Regulation (EU) No 1305/2013, Art 71

¹²⁶ Regulation (EU) No 1305/2013, Art 70

¹²⁷ Commission implementing regulation No 808/2014, Annex VII

CMES element (Art.14 of 808/2014, point 1)	Short description	Use in the RDP assessment
	modification procedures. ¹²⁸ The implementation of the EP is regularly reported in the annual implementation reports. ¹²⁹	for data management and collection. It suggests the target groups, means for dissemination, communication and follow up of evaluation results, allocation of resources and timing conducting the evaluation tasks.
Ex ante and ex post evaluation	<p>Ex ante evaluation (EAE)¹³⁰ accompanies the RDP design since the beginning.</p> <p>Ex post evaluation (EPE)¹³¹ is conducted after accomplishment of all RDP operations and provides the learning opportunity on the development and implementation of rural policy interventions.</p>	<p>EAE aims to improve the RDP at the time of its development, since very early stages, starting with the assessment of territorial analysis and needs assessment. EAE should also provide the feedback on consistency of the RDP' intervention logic and its relevance in addressing identified needs.</p> <p>EPE shall examine the effectiveness, efficiency and impacts of the RDP and its contribution to EU2020 Strategy as well as to CAP policy objectives.</p>
Support to enable all actors responsible for M&E to fulfil their obligations	Support to stakeholders , active in the evaluation of rural policy is provided via legal and implementing acts, which outline evaluation tasks, and various working and guidance documents.	To complete the evaluation tasks, stakeholders may use various working and guidance documents and information.

¹²⁸ Regulation (EU) No 1305/2013, Art. 11

¹²⁹ Commission implementing regulation No 808/2014, Annex VII, point 2

¹³⁰ Regulation (EU) No 1303/2013, Art. 55 and Regulation (EU) No 1305/2013, Art. 77

¹³¹ Regulation (EU) No 1303/2013, Art. 57 and Regulation (EU) No 1305/2013, Art. 78

ANNEX 3 - OVERVIEW OF REPORTING REQUIREMENTS ON RDP IMPLEMENTATION AND EVALUATION IN THE PROGRAMMING PERIOD 2014-2020

Reporting requirement (No. in Annex VII of 808/2014 ¹³²)		Additional legal Reference	Reporting requirements		
			Standard AIR (from 2016)	Enhanced AIR submitted in 2017	Enhanced AIR submitted in 2019
1	Key information on the implementation of the programme and its priorities:				
a)	Financial data: Financial implementation data for each measure and FA, a statement on expenditure incurred and declared in the declaration of expenditure.	1303/2013 ¹³³ , Art. 50.2, 1305/2013 ¹³⁴ , Art. 75.2	✓	✓	✓
b)	Common and programme-specific indicators and quantified target values: Information on RDP implementation as measured by common and programme specific indicators, including progress achieved in relation to the targets set for each FA and on realised output compared to planned output as set out in Indicator Plan	1303/2013, Art. 50.2, and 54.2 1305/2013, Art.69 808/2014, Art. 14.1b), Annex IV	✓	✓	✓
2	Progress in implementing the evaluation plan:	1303/2013, Art. 56.1 808/2014, Art.14.1f), Art.1			
a)	Description of any modifications made to the Evaluation Plan		✓	✓	✓
b)	Description of evaluation activities undertaken during the year	1305/2014, Art. 75.2	✓	✓	✓
c)	Description of activities undertaken in relation to the provision and management of data	1305/2014, Art. 70	✓	✓	✓
d)	List of completed evaluations, incl. references to where they have been published on-line	1303/2013 Art. 50.2, 1305/2014, Art. 76.3	✓	✓	✓
e)	A summary of completed evaluations, focussing on evaluation findings	1303/2013 Art. 50.2,	✓	✓	✓
f)	Description of communication activities to publicise evaluation findings	1303/2013 Art. 50.9	✓	✓	✓
g)	Description of follow-up given to evaluation results	1303/2013 Art. 56.3	✓	✓	✓
3	Issues which affect the performance (quality and effectiveness of RDP implementation) of the programme and the measures taken		✓	✓	✓
4	Steps taken to implement technical assistance (including the establishment of the NRN) and programme publicity requirements	1305/2014, Art. 54.1 808/2014, Art. 13,			

¹³² Commission Implementing Regulation (EU) No 808/2014

¹³³ Regulation (EU) No 1303/2013

¹³⁴ Regulation (EU) No 1305/2013

Reporting requirement (No. in Annex VII of 808/2014 ¹³²)		Additional legal Reference	Reporting requirements		
			Standard AIR (from 2016)	Enhanced AIR submitted in 2017	Enhanced AIR submitted in 2019
a)	In case of coverage under the technical assistance of the establishment and functioning of NRN, the report shall describe actions taken and state of play as regards the establishment of the NRN and the implementation of this action plan	1305/2014, Art. 54.3	✓	✓	✓
b)	Steps taken to ensure that the programme is publicised	1305/2014, Art. 8.1m) iii),	✓	✓	✓
5	Actions taken to fulfil ex ante conditionalities (where relevant), description by priority/focus area/measure.	1303/2013 Art. 19.1, Art 50.2	✓	✓	
6	Description of implementation of sub-programmes as measured by common and specific indicators including on the progress achieved in relation to targets set in the indicator plan of the sub-programme	1305/2013 Art.7.1, 75.3		✓	✓
7	Assessment of the information and progress towards achieving objectives of the programme:				
	Reporting and quantification of programme achievements, in particular through assessment of the complementary result indicators and relevant evaluation questions	1303/2013, Art. 50.2 808/2014, Art. 14,		✓	✓
	Reporting on the progress towards the objectives of the programme and its contribution to achieving the Union strategy for smart, sustainable and inclusive growth through, inter alia, assessment of the programme's net contribution to changes in CAP impact indicator values, and relevant evaluation questions	1303/2013, Art. 54.1 808/2014, Art. 14,			✓
8	Implementation of actions to take into account principles set out in Articles 5, 7 and 8 of Regulation (EU) No 1303/2013				
a)	Promotion of equality between men and women and non-discrimination,	1303/2013 Art.7		✓	✓
b)	Sustainable development	1303/2013 Art.8		✓	✓
c)	Role of partners	1303/2013 Art.5		✓	✓
9	Progress made in ensuring an integrated approach to support the territorial development of rural areas, including through local development strategies	1303/2013 Art. 32-36 1305/2014, Art.42-44			✓
10	Report on implementation of financial instruments (as annex to AIR)				
	including for each financial instrument the information contained in Article 46.2 points a) to g) and i) of Reg. 1303/2013	1303/2013 Art. 46.2, points a) – g) and i)	✓	✓	✓

Reporting requirement (No. in Annex VII of 808/2014 ¹³²)		Additional legal Reference	Reporting requirements		
			Standard AIR (from 2016)	Enhanced AIR submitted in 2017	Enhanced AIR submitted in 2019
	Including the information contained in Article 46.2 points h) and j) of Reg. 1303/2013	1303/2013 Art. 46.2, points h) and j)		✓	✓

ANNEX 4 – WORKING STEPS IN SETTING UP THE SYSTEM TO ANSWER THE EVALUATION QUESTIONS IN 2017

Major steps	Steps	Responsibilities	Relevance for Terms of Reference	
			When the step is part of ToR	When the step is part of ToR
Preparing the evaluation	Revisit the RDP intervention logic	Evaluation experts within the Ministry of Agriculture (e.g. Evaluation Unit) and/or evaluators.	Revisiting the RDP intervention logic is not part of the terms of reference if the evaluation experts within the Ministry of Agriculture have the capacity to review the relevance and coherence of the intervention logic, e.g. with the help of a research institute, and, in case of NRN, the experts and/or the NSU have the capacity to formulate the NRN intervention logic.	This step is part of the terms of reference if the “in house” evaluation experts/NSUs do not have internal capacity to conduct this step.
	Link intervention logic to evaluation elements	Managing authorities are responsible for defining all programme-specific elements mentioned in the RDP, its Evaluation plan or in other internal evaluation planning documents. In case of CEQ and common indicators, the EU legal acts, EC working documents and guidelines must be taken into consideration when using the common evaluation elements. Evaluation experts/evaluators can define the PSEQ and programme-specific indicators and discuss them with the managing authorities. New programme-specific evaluation elements are designed when	The definition of evaluation elements (EQ, judgment criteria and indicators) is not part of the terms of reference if programme authorities or “in house evaluation experts” fully understand all terms used in the working documents mentioned in “Further reading”, in the RDP’s evaluation plan, internal evaluation planning documents or other relevant documents.	This step is part of the terms of reference if the evaluation elements have to be completed and defined after the intervention logic appraisal (additional programme specific elements).

Major steps	Steps	Responsibilities	Relevance for Terms of Reference	
			When the step is part of ToR	When the step is part of ToR
		those presented in the RDP are not sufficient to capture all programme effects or assess all evaluation topics.		
	Check consistency of evaluation questions and indicators with RDP intervention logic	Evaluation experts within the Ministry of Agriculture (e.g. Evaluation Unit) and/or Evaluators	Checking consistency between evaluation elements and RDP intervention logic is not part of the terms of reference if the evaluation experts within the Ministry of Agriculture have the necessary capacity to assess the consistency of the intervention logic with CEQ and PSEQ and indicators and if all PSEQ and indicators are already developed and defined,	This step is part of the terms of reference in case the programme-specific evaluation elements are not properly developed or defined, or if the “in house” evaluation experts do not have the necessary capacity to conduct this step and define programme specific elements if needed.
	Develop RDP specific evaluation elements	Managing authorities with the support of “in house” evaluation experts may define PSEQ and indicators in the RDP design stage and include them in the Evaluation Plan and Indicator Plan. They may also define programme-specific evaluation elements at later stages during the programme implementation, at the occasion of a newly identified evaluation need or topic, which cannot be covered by common or existing PSEQ. Evaluators may define other PSEQ and indicators, in case of: <ul style="list-style-type: none"> Inconsistency between the RDP intervention logic and common and programme-specific elements, or 		This step is always part of the terms of reference

Major steps	Steps	Responsibilities	Relevance for Terms of Reference	
			When the step is part of ToR	When the step is part of ToR
		<ul style="list-style-type: none"> ○ Inconsistency between the RDP intervention logic and the territorial and SWOT analysis, and needs assessment, or ○ Failure by common and programme-specific evaluation elements to fully capture RDP-specific effects, or ○ Identified significant unintended effects of intervention logic, which are not covered by existing PSEQ and indicators. 		
Structuring the evaluation	Set up a consistent evaluation approach	<p>Managing Authority (e.g. in collaboration with the evaluation experts within the Ministry) will propose the evaluation approach for the assessment of the RDP. If programme authorities have relevant capacity, they can also propose evaluation methods.</p> <p>Evaluators elaborate further on the evaluation approach and evaluation methods used to accomplish the evaluation tasks.</p>		This step is always part of the terms of reference: evaluators must elaborate on the proposed evaluation approach, select and apply the most suitable evaluation methods.
	Select evaluation methods and their combination	Managing authorities with the help of evaluation experts within the Ministry of Agriculture may express their preference for certain evaluation methods, which		This step is always part of the terms of reference

Major steps	Steps	Responsibilities	Relevance for Terms of Reference	
			When the step is part of ToR	When the step is part of ToR
		they may specify in the terms of reference. Evaluators select the evaluation methods		
	Establish the evidence for evaluation	Responsibilities for establishing the evidence for evaluation are described in detail in Chapter 6.2.		This step is part of the terms of reference and reflects the responsibilities of evaluators in bridging the gaps in existing data sources and/or ensuring additional data & information collection.
	Manage and collect data for evaluation	Responsibilities for establishing the evidence for evaluation are described in detail in Chapter 6.2.		This step is part of the terms of reference and reflects the responsibilities of evaluators in bridging the gaps in existing data sources and/or ensuring additional data & information collection.
Conducting the evaluation	Observing, analyzing, judging	Evaluators		<p>This task is always part of the terms of reference, among others:</p> <ul style="list-style-type: none"> ○ Prepare the evaluation system for 2017 and beyond (e.g. resolve data gaps by modelling, extrapolations or other means), ○ Focus the work on quantification of common result indicators (ideally net values), including the programme-specific results, ○ Develop and use the additional result indicators if needed to answer CEQ, ○ Use specific methods, tools and techniques for quantifying and assessing the effectiveness and

Major steps	Steps	Responsibilities	Relevance for Terms of Reference	
			When the step is part of ToR	When the step is part of ToR
				<p>efficiency of the RDP, and for answering the EQ,</p> <ul style="list-style-type: none"> ○ Validate values of target indicators ○ Reflect on the need to update the CCI, ○ Provide answers to evaluation questions ○ Draft conclusions and recommendations which are strictly based on evidence of the qualitative and quantitative assessment.

ANNEX 5 - CHECK-LIST FOR ASSESSING THE QUALITY OF THE EVALUATION REPORT

The following example of a check-list may be used by the Managing Authority to assess the quality of its evaluation report.

Title of the evaluation:				
Department / unit responsible:				
Evaluator / contractor:				
Assessment carried out by: (name organisations/units involved in the assessment)				
Date of quality assessment:				
RELEVANCE				
Does the evaluation respond to information needs of the commissioning body and fit the Terms of Reference?				
Poor	Satisfactory	Good	Very Good	Excellent
Arguments for scoring:				
SCOPE				
Is the rationale of the programme and its set of outputs, results and impacts examined fully, including both intended and unexpected effects of policy interventions?				
Poor	Satisfactory	Good	Very Good	Excellent
Arguments for scoring:				
APPROPRIATENESS OF EVALUATION DESIGN				
Is the methodological approach adequate for obtaining the results needed to answer the evaluation questions?				
Poor	Satisfactory	Good	Very Good	Excellent
Arguments for scoring:				
RELIABILITY OF DATA				
Are the data used in the evaluation adequate for the purpose and have their reliability been ascertained? Have data weaknesses and limitations been explained?				
Poor	Satisfactory	Good	Very Good	Excellent
Arguments for scoring:				
SOUNDNESS OF ANALYSIS				
Are qualitative and quantitative data appropriately and systematically analysed to answer evaluation questions and cover other information needs in a valid manner? Are cause and effect links between the intervention and its results explained? Are external factors correctly taken into consideration?				
Poor	Satisfactory	Good	Very Good	Excellent
Arguments for scoring:				
CREDIBILITY OF FINDINGS				
Do findings follow logically from and are justified by the data / information, analysis and interpretations based on pre-established criteria? Are findings based on carefully explained assumptions and rationale?				
Poor	Satisfactory	Good	Very Good	Excellent
Arguments for scoring:				

VALIDITY OF CONCLUSIONS

Are conclusions fully based on findings and non-biased? Are conclusions clear, clustered and prioritised?

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

USEFULNESS OF RECOMMENDATIONS

Are areas needing improvements identified in coherence with the conclusions? Are the suggested options realistic, impartial and sufficiently detailed to be operationally applicable?

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

CLARITY OF REPORT

Is the report well structured, balanced and written in an understandable manner?

Is the report easy to read and has a short but comprehensive summary? Does the report contain graphs and tables?

Poor	Satisfactory	Good	Very Good	Excellent

Arguments for scoring:

OVERALL ASSESSMENT OF THE FINAL REPORT

Overall, the quality of the report is:

Poor	Satisfactory	Good	Very Good	Excellent

ANNEX 6 - THE ROLE OF RURAL DEVELOPMENT STAKEHOLDERS IN THE DISSEMINATION AND COMMUNICATION OF EVALUATION RESULTS

WHO?	WHAT?	TO WHOM?	WHY?
MA	Report on evaluation in AIR submitted in 2017	EC	Inform the Commission, fulfil legal requirements
	Summary of major evaluation findings, conclusions recommendations	Policy makers	Inform policy makers, signal RDP modification needs
	Responses to recommendations, required action points	Policy makers	Ensure follow-up of recommendations, plan changes
	Organising a final conference / workshop	MA, PA, SG, stakeholders, evaluator	Summarising main findings, discussing actions to be taken, learning from the evaluation
	Press releases, articles in newsletters, round tables in TV, Radio,	Media, general public, researchers, stakeholders	Increase transparency of policy and knowledge about RDP results
	Publish report and citizen's summary on the website,	General public	Access to information on RDP results
Evaluator	Citizen's summary	MA	Concise summary of main results
	Presentation	MC, SG	Informing the MC & SG, allowing questions
	Presentation (if requested by the MA)	Other stakeholders	Informing stakeholders (e.g. farmers' organisations, environmental organisations)
	Presentation in a final conference / workshop	MA, PA, SG, stakeholders,	Summarising main findings, discussing action points,
NRN	Article in newsletter on findings, conclusions and recommendations	Stakeholders	Increase knowledge about the report and RDP results
	Post about evaluation report on website	General public, stakeholders	Increase knowledge about the report and RDP results
LAGs	Article in newsletter (from a Leader point of view)	Stakeholders	Increase knowledge about the report and RDP results

ANNEX 7 - MINIMUM REQUIREMENTS AND GOOD PRACTICE IN REPORTING ON EVALUATION IN THE AIR

Evaluation tasks and reporting requirements on evaluation in 2017 are described in Part I. Programme authorities need to inform on the progress on the implementation of the evaluation plan in each AIR. In the AIRs submitted in 2017 and 2019 additionally also the information on the assessment of programme results and impacts and answers to relevant evaluation questions need to be reported.

Moreover, programme authorities in Member States may decide to go beyond the minimum reporting requirements and provide more information on evaluation in various AIR sections, as shown below:

Requirements for reporting in relation to evaluation tasks	Optional elements to be reported as good practice - examples
Report on any issues affecting the performance of the programme and measures taken. ¹³⁵ (each AIR)	Report on the assessment of RDP delivery mechanisms, as they can influence the programme's effectiveness, efficiency, results, and impacts: e.g. assessment of RDP targeting, application of financial instruments in particular measures, project applications, selections, payment procedures, the role of RDP communication to beneficiaries etc.
Inform about activities undertaken in relation to and progress in implementing the Evaluation plan. ¹³⁶ (each AIR)	Report on activities, which were conducted beyond the minimum requirements, e.g. those planned and implemented in line with the internal planning documents for evaluation (e.g. steering group for evaluation, collaboration with evaluation experts, arrangements for data for evaluation beyond compulsory monitoring, etc.
Inform about financial commitments and expenditures by measure ¹³⁷ , (each AIR)	Provide information on the breakdown of financial commitments and expenditures in a way which helps later to conduct various assessments, for example with respect to the additional contribution (secondary effects) provide information on expected, unexpected, real contributions to flagged FA expressed in financial commitments and expenditures.
Report on the financial data, common, and programme specific indicators and quantified target values ¹³⁸ , where appropriate. (each AIR)	Provide information on the approach when and how programme specific indicators have been developed and how the values for all indicators have been calculated.
Inform on the quantification and assessment of programme achievements through common and programme specific result indicators ¹³⁹ , where appropriate. (AIR 2017 and 2019)	In case of low uptake with respect to financial commitments, conduct and report on studies, which assess the interest of potential beneficiaries to apply for support from the RDP measures and estimate results/achievements of RDP objectives.
Report on contribution of the financial instruments to the achievement of indicators of the priority of measure concerned ¹⁴⁰ (each AIR and more in AIR 2017 and 2019)	Explain why FI are used in the implementation of particular measures in supporting particular beneficiaries and how this might influence the effectiveness and efficiency of the programme. In case of sufficient uptake report on the role of targeting of the RDP support via financial instruments and its effects on the values of indicators.

¹³⁵ Regulation (EU) No 1303/2013, Art. 50 and Commission implementing regulation No 808/2014, Annex VII, point 2 e) and point 3

¹³⁶ Regulation (EU) No 1305/2013, Art. 75 and Commission implementing regulation No 808/2014, Annex VII, point 2

¹³⁷ Regulation (EU) No 1305/2013, Art. 75 and Commission implementing regulation No 808/2014, Annex VII, point 2

¹³⁸ Commission implementing regulation, Annex VII, point 1

¹³⁹ Regulation (EU) No 1303/2013, Art. 50 and 57 (ex post), Commission implementing regulation No 808/2014, Art. 14.1 b) and Annex IV and Annex VII, point 1 and point 7

¹⁴⁰ Regulation (EU) No 1303/2013, Art. 46

Requirements for reporting in relation to evaluation tasks	Optional elements to be reported as good practice - examples
Assess the potential additional contribution of operations to one or more focus areas (secondary effects) ¹⁴¹ (AIR 2017 and 2019)	In case of sufficient RDP uptake report on actual additional contributions of measures to focus areas, under which they have not been programmed, based on collected information via payment requests and/or surveys and compare it with those expected during the programme design.
Set out the synthesis of evaluations of the programme, available during the previous financial year ¹⁴² , (each AIR)	Report on evaluation or scientific studies, which have not been conducted in relation to RDP and ordered by the MA, but provide important findings with respect to the evaluation topics envisioned in the Evaluation plan or with respect to RDP objectives. For example, studies conducted by research & academia, various NGO or government agencies.
Assess the information and progress made towards achieving the objectives of the programme ¹⁴³ , and answer relevant (focus area related) evaluation questions ¹⁴⁴ (AIR 2017 and 2019)	Provide the information how it is ensured that the answers to evaluation questions are based on the robust quantitative and qualitative evidence, e.g. in the format of table from evaluation question to the collection of data and information.
Assess actions taken to ensure that objectives and implementation of EAFRD is in line with principles set out in Articles 6,7 and 8 of Regulation (EU) No 1303/2013 ¹⁴⁵ (AIR 2017 and 2019)	Explain the approach to assess the gender equality/non-discrimination, the sustainable development, and the role of partners in the RDP implementation, including necessary arrangements with respect to data collection and the findings of the assessments.
Inform on progress made in the implementation of the sub-programmes (including common and programme specific indicators) and the progress achieved in relation to targets set in the indicator plan ¹⁴⁶ (AIR 2017 and 2019)	In case of sub-programmes are part of the RDP, explain how the values of indicators have been calculated.

¹⁴¹ Commission implementing regulation No 808/2014, Art.14.4

¹⁴² Regulation (EU) No 1303/2013, Art. 50 and Common implementing regulation No 808/2014, Annex VII, point 2 d) and e)

¹⁴³ Regulation (EU) No 1303/2013, Art.50

¹⁴⁴ Regulation (EU) No 1303/2013, Art. 50 and Commission implementing regulation No 808/2014, Annex VII, point 7

¹⁴⁵ Common implementing regulation No 808/2014, Annex VII, point 8

¹⁴⁶ Common implementing regulation No 808/2014, Annex VII, point 6

ANNEX 8 - TIMING FOR PREPARING AND IMPLEMENTING THE SYSTEM TO ANSWER THE CEQ, AND ANSWERING CEQ

There are three different types of common evaluation questions:

1. CEQ related to RD focus area (1-18)
2. CEQ related to other aspects of RDP: operation performance/programme synergies, TA and NRN (19-201)
3. CEQ related to EU level objectives.

Although in the AIR submitted in 2017 only CEQ related to RD focus areas and other RDP aspects have to be answered, it is important to prepare the system for answering all evaluation questions prior the programme implementation's start. The table below illustrates at what point in time the three different types of CEQs need to be prepared, structured, observed, analysed and judged.

Table 9. Timing for preparing and implementing the system to answer the CEQ, and answering CEQ

Evaluation process		Years									
		2014-2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Preparing evaluation		CEQ-FA	CEQ-FA		CEQ-FA					CEQ-FA	
		CEQ-OA	CEQ-OA		CEQ-OA					CEQ-OA	
		CEQ-EU	CEQ-EU		CEQ-EU					CEQ-EU	
Structuring evaluation		CEQ-FA	CEQ-FA		CEQ-FA					CEQ-FA	
		CEQ-OA	CEQ-OA		CEQ-OA					CEQ-OA	
		CEQ-EU	CEQ-EU		CEQ-EU					CEQ-EU	
Conducting evaluation	Observing		CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA
			CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA
			CEQ-EU	CEQ-EU	CEQ-EU	CEQ-EU	CEQ-EU	CEQ-EU	CEQ-EU	CEQ-EU	CEQ-EU
	Analysing		CEQ-FA	CEQ-FA	CEQ-FA	CEQ-FA				CEQ-FA	CEQ-FA
			CEQ-OA	CEQ-OA	CEQ-OA	CEQ-OA				CEQ-OA	CEQ-OA
					CEQ-EU	CEQ-EU				CEQ-EU	CEQ-EU
	Judging			CEQ-FA		CEQ-FA					CEQ-FA
				CEQ-OA		CEQ-OA					CEQ-OA
						CEQ-EU					CEQ-EU
Reporting evaluation				CEQ-FA		CEQ-FA					CEQ-FA
				CEQ-OA		CEQ-OA					CEQ-OA
						CEQ-EU					CEQ-EU
		CEQ related to FA: CEQ-FA									
		CEQ related to other aspects: CEQ-OA									
		CEQ related to EU objectives: CEQ-EU									

ANNEX 9 – CRITERIA FOR SELECTION OF EVALUATION APPROACH

Criteria	Description
Validity of results	Users of evaluations want results that are reliable, scientifically sound, robust and valid. The validity of results can only be checked with a transparent research process. Attention must be paid to restrictions to data access which may impair a high level of transparency. In order to improve the validity of results, authorities commissioning evaluation studies should put emphasis on as much transparency as possible.
Scale of measurement	Different methods produce results at different scales: quantitative methods (e.g. econometric, input-output, computational models) provide results in numbers (cardinal scales), while qualitative methods produce results on ordinal or nominal scales. When results need to be expressed in cardinal scales, the choice of methods is limited because qualitative and participatory methods and theory-based and descriptive approaches allow ordinal statements at best. Quantitative methods would therefore be strongly preferred in this case.
Ability to analyse the counterfactual	A good evaluation should always develop a counterfactual. The counterfactual situation is the conceived and observed scenario that is used to compare firms/farms/territory/etc. with and without programme support. Although most methods can be used to analyse counterfactual scenarios, the challenge is to make judgements about an effect of the programme which by definition cannot be directly observed in most cases. Quantitative methods are more systematic in building counterfactuals; however, some qualitative methods can also include a counterfactual in their design.
Ability to identify and test causal relations	One of the biggest challenges in evaluation is to identify and test causal relations between the policy interventions and the outcomes. Only a small set of methods (typically econometric models) are suited to provide such results. If such results are not available, assumptions about causal effects need to be made, possibly through the application of qualitative methods.
Ability to link output and result with impact indicators	A good evaluation, even if carried out at an early stage of programme implementation, where impacts are not yet discernible, should be designed with a long-term perspective. This implies establishing mechanisms for linking outputs and results to impacts which at a later stage have to be assessed in relation to the intervention logic of the programme.
Ability to consider unintended effects	Several unintended effects are important in order to evaluate programme support. These include leverage, substitution, displacement effect and deadweight (windfall profit). Although most methods can be used to take account of leverage and deadweight effect, only econometric methods can be used to quantify their size.
Time horizon of the intervention	Many interventions, especially investments, have a period of operation that spans over many years (e.g. the planting of an orchard). The methods used to analyse the effects of such interventions are different from approaches that analyse interventions with short-term payoffs (e.g. an automatic milking system). It therefore has to be recognised that the effects of some interventions will be measurable only with a considerable delay that may go well beyond the programme period.

ANNEX 10 - TOOLS FOR QUALITATIVE APPRAISAL OF RDP OPERATIONS ´ PRIMARY AND SECONDARY CONTRIBUTIONS TO FOCUS AREAS AND SYNERGIES BETWEEN MEASURES, BETWEEN FAS AND BETWEEN RD PRIORITIES

The RDP intervention logic is complex and contains various horizontal, vertical and diagonal linkages between operations, measures, focus areas and rural development priorities. These linkages are:

- **Primary contributions** of operations implemented under measures/sub-measures¹⁴⁷ to focus areas under which they are programmed (vertical)
- **Secondary contributions** of operations implemented under measures/sub-measures¹⁴⁸ to focus areas under which they are **not** programmed (diagonal)
- **Transverse effects** between focus areas and between rural development priorities (horizontal)
 - Positive transverse effects, also called synergies
 - Negative/antagonistic effects

These linkages should not be mixed up with the programme indirect effects, which go beyond the objectives or address various issues. RDP indirect effects are explained in PART II of guidelines, in the chapter 5.1 Revisiting the RDP intervention logic.

Table 1 below is proposed as one possible tool to indicate primary and secondary contributions of RDP operations to RDP focus areas (vertical and diagonal linkages). The table is an example which needs to be adjusted to the specific RDP intervention logic with a view to show only priorities, focus areas and measures/sub-measures which are present in the respective RDP.

Primary contributions of operations are shown at the left side of the table where measures/sub-measures are linked to FAs under which they are programmed. Secondary contributions of operations are highlighted with orange boxes.

The table can be filled by the MA to appraise where the potential secondary contributions of operations to other FA can be expected during the evaluation. The table also can be used by the evaluators as an indicative qualitative assessment tool before setting up the samples of beneficiaries to assess the secondary contributions. Based on the information obtained from the table and the information from application forms/payment request, the evaluators can select samples of beneficiaries (e.g. according to sectors: dairy, grain etc.) with operations generating additional contributions to FA other than those under which they have been programmed and use the survey to find out their contributions to the value of result indicators).

The table is available also in an editable version.

Table 2 shows a classical pair wise matrix, allowing to assess synergies and negative transverse effects (horizontal linkages) between focus areas and between priorities in qualitative way, this table also can serve as an input to answer common evaluation question number 19: "To what extent have the synergies among priorities and focus areas enhanced the effectiveness of the RDP?"

Stakeholders in Member States can fill boxes of this table using two different colours. In the case of the example below, the green colour symbolises synergies and the red colour symbolises negative transverse effects. Unfilled boxes indicate neutral transverse effects.

Stakeholders in MS should adjust the table in accordance with the RDP intervention logic. The table example is also available in an editable version.

¹⁴⁷ Commission implementing regulation (EU) no 808/2014, Annex I, part 5

¹⁴⁸ Commission implementing regulation (EU) no 808/2014, Annex I, part 5

Table 10. Primary and secondary contributions of operations within measures/sub-measures to rural development focus areas – example

RDP													
RDP	PRIORITIES			PRIORITY 2		PRIORITY 3	PRIORITY 4		PRIORITY 5		PRIORITY 6		
		FA	Measures	FA 2A	FA 2B	FA 3A	FA 4A	FA 4B	FA 5A	FA 5B	FA 6A	FA 6B	
	PRIORITY 1	FA 2A	M4.1										
			M4.3										
			M6.2										
		FA 2B	M2.1										
			M6.1										
			FA 3A	M3.1									
	M4.2												
	M6.4												
	PRIORITY 4	FA 4A	M10.1										
			M11.1										
		FA 4B	M4.1										
			M11.1										
	PRIORITY 5	FA 5A	M4.1										
			M4.3										
			M8.1										
		FA 5B	M4.1										
			M4.3										
			PRIORITY 6	FA 6A	M6.4								
	M7.2												
	M7.4												
	FA 6B	M19.1											
		M19.2											
		M19.3											
M19.4													

Table 11. Synergies and negative transverse effects between focus areas and between priorities

RDP																											
RDP	PRIORITIES			PRIORITY 2				PRIORITY 3			PRIORITY 4				PRIORITY 5				PRIORITY 6								
		FA		FA 2A			FA 2B		FA 3A			FA 4A		FA 4B		FA 5A			FA 5B		FA 6A			FA 6B			
			Measures	M4.1	M4.3	M6.2	M 2.1	M6.1	M3.1	M4.2	M6.4	M10.1	M11.1	M4.1	M11.2	M4.1	M4.3	M8.1	M4.1	M4.3	M6.4	M7.2	M7.4	M19.1	M19.2	M19.3	M19.4
	PRIORITY 1	FA 2A	M4.1																								
			M4.3																								
			M6.2																								
			FA 2B	M2.1																							
	M6.1																										
	PRIORITY	FA 3A	M3.1																								
			M4.2																								
			M6.4																								
	PRIORITY 4	FA 4A	M10.1																								
			M11.1																								
			FA 4B	M4.1																							
				M11.1																							
	PRIORITY 5	FA 5A	M4.1																								
			M4.3																								
			M8.1																								
			FA 5B	M4.1																							
		M4.3																									
	PRIORITY 6	FA 6A	M6.4																								
			M7.2																								
			M7.4																								
		FA 6B	M19.1																								
			M19.2																								
			M19.3																								
			M19.4																								

European Evaluation Helpdesk


Boulevard Saint-Michel 77-79

B - 1040 BRUSSELS

T: +32 2 737 51 30

Email: info@ruralevaluation.eu

<http://enrd.ec.europa.eu>

A thick, horizontal bar with a textured, orange color, resembling a brushstroke or a distressed paint finish, located at the bottom of the page.